Differences in the effects of distance to subway on rental prices. The case of 7 Russian cities

Mishalkin Ivan[a]

[a] *The Higher School of Economics*

Abstract

Subway in Russian cities differs significantly. Some cities have too many stations, some have lack of them. Some subway are convenient, some only for small part of the city. The main purpose of this article is to estimate how different by structure subway effect on the rental prices of the apartments.

Keywords: subway, Airbnb, rental prices.

Classification JEL: 38; I3; R21.

Contents

1. Introduction

In Russia there are 15 cities with the population over one million people. Such a big amount of citizens requires strong transportation facilities. One of the possible solutions, that humanity has invented is subway. However in Russia it is a special case. Only seven cities have subway: Moscow, Saint-Petersburg, Kazan, Nizhniy Novgorod, Novosibirsk, Samara, Ekaterinburg. One more city, Volgograd, has so called subway-tram. It is a tram line with stations underground. Moreover, the subway structure is quite different for this cities. The most developed underground is Moscow one, it has 244 stations, one underground ring and monorail above the ground, connected with subway stations. It includes 13 lines and covers almost the whole city. Little less developed is Saint-Petersburg underground, which includes 67 stations and 5 lines. It mainly covers the central part of the city and some distant regions to the North and to the South. The North-East part of the city is not covered, though there is a residential area. In Nizhniy Novgorod there are two lines. subway has the T-shape. There are significantly less stations-only 14. Nizhniy Novgorod is situated along the river Oka, where it flows into the Volga and subway is also built along the river. It also crosses the river, so citizens can get to the other bank by subway. Another line is situated along the river Volga. It seems to be rather convenient underground. In Novosibirsk there are two lines and 13 stations. Novosibirsk is situated on the both sides of the river Ob. subway allows to cross the river. Most of the stations are located on the right bank of the Ob and mainly cover the centre of the city. Samara subway has 10 stations and one line. As in the other cities, in Samara subway started being set up in Soviet period. Its main aim was to deliver people to the factories. After the collapse of the USSR the factories stopped or significantly shortened its production, so most of people have no more need to reach the the places the stations are situated in. This subway does not also lead to the historical centre of the city. In Kazan subway was projected to connect distant residential areas in the South and in the North. It goes through the centre and crosses the Kazanka river. For the citizens in the East part of the city subway is rather inconvenient. In Ekaterinburg subway has 1 line with 9 stations. The city is oblong from North to South and subway connects upper and lower parts of the city, crossing the centre. It seems to be very convenient.

We are going to use data from Airbnb.com, an online marketplace and hospitality service, that connects people looking to rent their homes with people who are looking for accommodations. It is a good source of data for several reasons. The first one is data structure. Using Airbnb api we get the same categories of data throughout the whole Russia. The second reason is the amount of observations. It is especially big in Saint-Petersburg and Moscow, however the other cities also have enough data for statistical tests. The next reason is policy of Airbnb. The service is built in a way that both hosts and guests have their rating. This minimizes the amount of fake apartments. Data also contains geocoordinates which we use to enrich the dataset. Due to popularity of this service and structure of the data this research is reproducible in any country.

2. Literature background

The effect of subway on housing prices is widely observed topic.Agostini and Palmucci (2008) estimate the effect of building a new line in Santiago. They use database of transactions and get as a result the increase in price from 4.2 to 7.9% after the announcement of construction and from 3.1 to 5.5% more after announcement of exact locations of the stations. The data contain several regressors on physical parameters of the housing and authors added distances to the nearest subway station, to the nearest hospital, to the nearest school and to the nearest green area. The other work [Forrest et al. (1996)] considers the relation between the availability of commuter rail services and housing prices. Authors also investigate whether modernisation of facilities can modify prices. authors use data from Nationwide Building Society of 892 observations, which include property characteristics, neighbourhood characteristics and property location. The distance is provided with 5 categories divided by distance to and type of the nearest subway station. The authors estimate hedonic regression with semilog specification. The results are following: closeness to subway gives the discount for house price from 2.1 to 8.1%. The coefficients are negative because of unpopularity of rail transport in Manchester, as citizens have other well developed public transport facilities and closeness to subway cause negative externalities such as noise. [Damm et al. (1980)] investigate influence of subway on Washington houses prices. The authors use three distinct datasets owner-occupied single-family dwellings, multi-family buildings and retail establishments. They estimate several hedonic regression on each type of data. Authors estimate elasticities and they are in range of 0.05 to

0.19 and for retail-established model the elasticity was equal to -0.68. The other article [Mulley et al. (2016)] compares difference between distance to bus stations and distance to train station. The authors estimated hedonic regression and got that each 100 meters farther from the busstop decreases the price by 0.13% and at the same time each 100 meters from the train station increases the price by 0.15%.

Unlike the researches above this one has the source of data which is interesting by itself. There are several researches on this service. Voytenko Palgan et al. (2016) investigate the "old phenomenon in a new disguise", specifically, sustainability framings held by operators and users of accommodation sharing platforms with the framings identified in the literature. In the other article [Kathan et al. (2016)] the influence of sharing economy on existing business. The special case of influence may be the employment in spheres which may be effected by sharing services. The effect on tourism industry employment is described in Fang et al. (2016). This small research is provided with simple polynomial regression. The authors say that number of listings has some impact on employment. Some authors also make researches from host [Liang et al. (2017)] and guest [Karlsson et al. (2017)] points of view. The most interesting for us may be the research on characteristics which influence the rental prices by [Wang and Nicolau (2017)]. The research is based on previous papers, which determined the regressors of the model and the specification. The authors use 25 variables from 5 categories: host attributes, site and property attributes, amenities and services, rental rules and online review ratings. Data was collected from 33 cities. The authors estimate quantile and ols regressions. The results of the paper are following: being a superhost increases the price of the housing by 8.73%. Similar results were obtained by both models. Next authors find out that number of listings obtained by the host effects the price by 0.06%. Also increases prices verified host identities. Distance from the centre has negative effect, what is not surprising. Also positive and significant effects have number of people accommodated(6.1%), amount of bathrooms(10.8%), real beds(15.5%), bedrooms(12.4%), wireless Internet(9.98%), free parking on premises(8.44%). Negative effect on price have instant booking(-6.44%), breakfast(-1.05%).

In this paper we will try to apply Airbnb data to Russian cities and estimate the differences, caused by different subway structure in Russian cities.

3. Data description

The data used in this research was taken from Airbnb service. This service is an online marketplace and hospitality service, enabling people to lease or rent short-term lodging including vacation rentals, apartment rentals, homestays, hostel beds, or hotel rooms. It has its API which is not public, but not forbidden to use. All methods are described at the http://airbnbapi.org web-site. We tried parser based on search method, which iterated through pages, but this returned only 300 unique values for each cities. So the possible solution was found at http://tomslee.net/airbnb-data-collection-methodology-and-accuracy. This researcher provides us with source code of the scraper on https://github.com/tomslee/airbnb-data-collection. It allows to use three methods: search by neighbourhood(special Airbnb marked area in some cities), by zipcodes and by so called bounding box. Neighbourhood didn't work for us, so we didn't checked this. Zipcodes were parsed from http://postindexapi.ru/. This method worked but some errors occurred. Moreover it was not better than "bounding box" method. The last requires to estimate square around the city. If there are any apartments in the square, the method divides it into four equal squares and searches further. The process continues while there are new apartments in the square, then goes to the next square or stops if it was the last one. "Bounding box" showed the greatest results, but it did not collect a lot of possible information about the apartments. So we took only apartments' ids from this parser and modified the first one, to collect all possible information. Later undocumented hint was found. Parser may request search data in certain price interval. At more frequent prices we took the interval of 50 rubles. From 300 to 1000 and from 2000 rubles the interval was set to 100 rubles. At big values the interval was set to 1000 rubles. The parser may be found at https://github.com/iamishalkin/airbnb/

Next the data was enriched with the distance to the nearest subway station(coordinates of subway-stations were parsed with HeadHunter API https://api.hh.ru/subway/. On the first step two nearest subway-stations were determined with straight distance. For that purpose we used https://overpass-turbo.eu/. Next we counted the distance to both of them by road, using the package sending requests to Google by its api, and chose the shortest one. We also calculated distance to the centres of the cities: The Kazan Kremlin for Kazan, The Red Square for Moscow, The Palace Square for Saint-Petersburg, The Lenin Square for Novosibirsk, The

Kuybyshev Square for Samara, The Bolshaya Pokrovka for Nizhniy Novgorod and The Square of 1905 in Ekaterinburg. This distance is calculated with haversine formula as we only would like to know, how far from city-centre the apartment is. We also include prices for three nearest neighbours to catch the region features which are complicated to estimate within including all possible location regressors. We divide this regressors by the person capacity, as it is obvious that larger apartments ceteris paribus are more expensive, so the division by person capacity should eliminate or at least decrease this effect.

We do not consider hostels, because this data is very noise and it may cause errors in estimation. Moreover, it seems, that hostels appear in the places which are interested to the travellers, so we are not able to compare this "usual" hostels with ones which are distant from places of interest or subway.

4. Model description

Inspired by Theebe (2004) we will take the semilog regression specification in order to make the research independent from the exchange rates what means reproducibility. Unfortunately, we didn't collected data about the host, like [Wang and Nicolau (2017)]. We assume, that guest chooses apartment only by its location on the map, characteristics and relying on the other users' reviews. We add 6 dummy variables indicating the city. The basis is Moscow. The regression includes constant. Prices of the nearest housings are also logarithmed.

The regressors we would like to include are following: presence of air conditioning, presence of breakfast presence of the elevator, essentials in the apartment, family and kid friendliness. Availability for guests with pets may be also important. Smoking_allowed is added because it may reflect the condition of the flat in some way. Suitable_for_events regressor may inderectly reflect the size of the apartment. Presence of TV may indirectly show the level of comfort provided in the apartments. Presence of wireless internet may also reveal some characteristics of the apartments.[Wang and Nicolau (2017)] showed that cancellation policy and availability to instant booking also matter, so we add these variables too.

All the variables, which are related with the size of the flat(amount of bathrooms, bedrooms, beds, person capacity) are included. Minimum nights to stay should also be included and the effect it may cause - increase the price of the apartment with more minimal nights the guest has to stay. We also include the amount of pictures made and the amount of reviews and rating from other users.

To estimate differences between cities cross effects on distance to subway(dist_walk) and distance to the centre(to_centre) are added. Distance to subway was counted in meters and then divided by 1000 to show kilometers. This is more convenient to interpret

After estimating model with spatial lag problems related to multicollinearity occurred, that is why distance to centre, suitable_for_events and cancellation_policy were excluded from the model.

As it was mentioned above we used model with spatial lag. Moran's I test rejected the null hypothesis of uncorrelated errors.1

1: Moran I output

```
Moran I statistic standard deviate = 15.892, p-value < 2.2e-16
alternative hypothesis: two.sided
sample estimates:
Observed Moran I      Expectation        Variance
    1.148956e-01    -1.377284e-03    5.353275e-05
```

The formula for spatial lag model looks like

$$Y = \rho * WY + X`\beta + e$$

Y is the dependent variable, X is the matrix of independent variables, β is the vector of regression parameters to be estimated from the data, ρ is the autoregressive coefficient, which tells us how strong the resemblance is, on average, between Y_i and it's neighbors. The matrix W is the spatial weight matrix, describing the spatial network structure of the observations. In this paper we use 3 nearest neighbours.

Another specification that may be suitable for our research is spatial error model. The problem is formulated in such a way:

$$Y = X`\beta + e$$

$$e = \lambda * We + v$$

The difference between to model is in the way we think of data. In the lag model it is assumed that the spatial dependence affects only the dependent variable. We can liken this to a diffusion scenario, where your neighbors have a diffusive effect on you.

The error model says that dependence affects the residuals only. We can liken this to the missing spatially dependent covariate situation, where, if only we could measure another really important spatially associated predictor, we could account for the spatial dependence. Unfortunately, we cannot, and we instead model dependence in our errors.

We build both models and choose the best one comparing Akaike information criterion.

The results are represented at the table 1. Both models perform almost equally. Spatial autoregressive model has slightly better AIC, so the effects are counted according to the spatial autoregressive model.[Table 2]

5. Results

As we estimated our regression with dummy variables, the impacts of distance to subway are compared to values for Moscow. In spatial lag models, interpretation of the regression effects is complicated. Each observation will have a direct effect of its predictors, but each observation will also have in indirect effect of the information of its neighbors. In tables 3, 4, 5 we express absolute effects for each city and percentage change in price if the apartment will be 1 km farther from subway-station. We will pay more attention to total effects than to the other. The most important subway seems to be in Moscow. In spite or, probably, because of the size of it one more kilometer from the station will about 1.5%.

The most indifferent prices to the location relative to the station are in Samara. As it was mentioned above subway-stations in Samara are very inconvenient, so the increasing in distance from subway station may reflect the shortening in distance to centre of the city or some popular among citizens places(recall that stations are located in the industrial regions).

The highest direct effect on prices is in Saint-Petersburg. By the total effect this city is on the second place. In Saint-Petersburg there is a good coverage of the centre of the city, however the other area develops rather rapidly and suffer from lack of fast public transport, independent from traffic jams. Especially the right bank of the Neva suffer from lack of stations.

The insignificant coefficient has Ekaterinburg city. This may be because the line is parallel and next to two, probably, main roads of the city. So the unpopularity may be explained with substitutes to underground transport.

Kazan Has also insignificant coefficient. This may be because of well developed non-underground public transport and the fact that Kazan subway is useful for those who go from south to north and vice versa. It might occur that such apartment may be unpopular among Airbnb users, so the sample may be little biased.

Nizhniy Novgorod and Novosibirsk have nearly same effect of distance on price. The subway in these cities are also very similar(2 lines, 13 and 14 stations).

It seems that subway starts to play an important role in apartment pricing when it has certain amount of stations. This amount lies somewhere between 14 and 64.

What is also interesting to mention is the negative effect on price of the presence of wireless Internet, number of bathrooms and beds. Lower price on the flats with Internet may explained in such a manner that host's target is youth, who often want to save money and are ready to refuse from some conveniences but not from the Internet. The number of beds is not the same as the size. Person capacity, for example, has positive coefficient. So the more beds the apartments have the less convenient they may be. The number of bathrooms may have such sign because 1.5 bathrooms in Airbnb means united toilet and bath. So if the majority of bathrooms has decimal part than this may lead to negative coefficient.

6. Discussion and Limitations

The results of this research are not consistent with papers in literature review section. This may be because of the uniqueness of Russian realities. In many cities subway was built to lead to industrial parts. The small amount of stations is also a problem. The profit from building one more station in the short-run period the government may get only by supporting quite developed subway in Moscow and Saint Petersburg. So the development of subway in small cities seems to be unrealistic.

About the limitations: as we investigated special area, Russia, the results may significantly deffer from the results in well developed countries. The research may be suitable to compare subway development in countries that were included into the USSR after its collapse.

7. References

Agostini, C. A. and Palmucci, G. A. (2008). The Anticipated Capitalisation Effect of a New Metro Line on Housing Prices. *Fiscal Studies*, 29(2):233–256.

Damm, D., Lerman, S. R., Lerner-Lam, E., and Young, J. (1980). Response of Urban Real Estate Values in Anticipation of the Washington Metro. *Journal of Transport Economics and Policy*, 14(3):315–336.

Fang, B., Ye, Q., and Law, R. (2016). Effect of sharing economy on tourism industry employment.

Forrest, D., Glen, J., and Ward, R. (1996). The Impact of a Light Rail System on the Structure of House Prices: A Hedonic Longitudinal Study. *Journal of Transport Economics and Policy*, 30(1):15–29.

Karlsson, L., Kemperman, A., and Dolnicar, S. (2017). May I sleep in your bed? Getting permission to book. *Annals of Tourism Research*, 62:1–12.

Kathan, W., Matzler, K., and Veider, V. (2016). The sharing economy: Your business model's friend or foe? *Business Horizons*, 59(6):663–672.

Liang, S., Schuckert, M., Law, R., and Chen, C.-C. (2017). Be a "Superhost": The importance of badge systems for peer-to-peer rental accommodations. *Tourism Management*, 60:454–465.

Mulley, C., Ma, L., Clifton, G., Yen, B., and Burke, M. (2016). Residential property value impacts of proximity to transport infrastructure: An investigation of bus rapid transit and heavy rail networks in Brisbane, Australia. *Journal of Transport Geography*, 54:41–52.

Theebe, M. A. J. (2004). Planes, Trains, and Automobiles: The Impact of Traffic Noise on House Prices. *The Journal of Real Estate Finance and Economics*, 28(2):209–234.

Voytenko Palgan, Y., Zvolska, L., and Mont, O. (2016). Sustainability framings of accommodation sharing. *Environmental Innovation and Societal Transitions*.

Wang, D. and Nicolau, J. L. (2017). Price determinants of sharing economy based accommodation rental: A study of listings from 33 cities on Airbnb.com. *International Journal of Hospitality Management*, 62:120–131.

8. Appendix

	Dependent variable:	
	price	
	spatial autoregressive	*spatial error*
	(1)	(2)
Air_conditioningTrue	0.182***	0.183***
	(0.010)	(0.011)
BreakfastTrue	−0.022	−0.021
	(0.014)	(0.014)
Elevator_in_buildingTrue	0.011	0.015
	(0.009)	(0.010)
EssentialsTrue	−0.074***	−0.076***
	(0.014)	(0.014)
Family_kid_friendlyTrue	0.075***	0.075***
	(0.011)	(0.011)
Pets_allowedTrue	−0.048***	−0.047***
	(0.011)	(0.011)
Smoking_allowedTrue	−0.070***	−0.068***
	(0.012)	(0.012)
TVTrue	0.157***	0.159***
	(0.012)	(0.012)
Wireless_InternetTrue	−0.062***	−0.065***
	(0.017)	(0.017)
bathrooms	−0.033**	−0.036***
	(0.014)	(0.014)
beds	−0.051***	−0.050***

	(0.006)	(0.006)
bedrooms	0.105***	0.104***
	(0.008)	(0.008)
instant_bookableTrue	0.009	0.008
	(0.009)	(0.009)
min_nights	−0.012***	−0.012***
	(0.003)	(0.003)
person_capacity	0.100***	0.101***
	(0.004)	(0.004)
ekb	−0.133***	−0.146***
	(0.019)	(0.022)
dist_walk	−0.013***	−0.016***
	(0.002)	(0.002)
kazan	−0.317***	−0.355***
	(0.022)	(0.027)
nn	−0.256***	−0.288***
	(0.025)	(0.030)
nvsbrsk	−0.225***	−0.260***
	(0.023)	(0.027)
smr	−0.280***	−0.320***
	(0.030)	(0.036)
spb	−0.117***	−0.131***
	(0.012)	(0.015)
ekb:dist_walk	0.002	0.002
	(0.003)	(0.004)
dist_walk:kazan	0.009*	0.011*
	(0.005)	(0.006)
dist_walk:nn	0.009***	0.011***
	(0.003)	(0.004)
dist_walk:nvsbrsk	0.010***	0.012***

	(0.003)	(0.003)
dist_walk:smr	0.012***	0.015***
	(0.002)	(0.002)
dist_walk:spb	−0.0005	−0.002
	(0.004)	(0.004)
Constant	5.970***	7.337***
	(0.084)	(0.025)
Observations	10,733	10,733
Log Likelihood	−6,326.660	−6,349.377
σ^2	0.189	0.190
Akaike Inf. Crit.	12,715.320	12,760.750
Wald Test (df = 1)	281.461***	242.536***
LR Test (df = 1)	272.664***	227.229***
Note:		*p<0.1; **p<0.05; ***p<0.01

Table 2: Impact

	Direct	Indirect	Total	Significance
Air_conditioningTrue	0.1837399076	0.0376498758	0.2213897833	***
BreakfastTrue	-0.0221312985	-0.0045348920	-0.0266661905	
Elevator_in_buildingTrue	0.0113491611	0.0023255400	0.0136747011	
EssentialsTrue	-0.0744608506	-0.0152576640	-0.0897185146	***
Family_kid_friendlyTrue	0.0751803002	0.0154050854	0.0905853855	***
Pets_allowedTrue	-0.0488580111	-0.0100114236	-0.0588694346	***
Smoking_allowedTrue	-0.0707617623	-0.0144996892	-0.0852614516	***
TVTrue	0.1587667381	0.0325326601	0.1912993982	***
Wireless_InternetTrue	-0.0625957992	-0.0128264137	-0.0754222129	***
bathrooms	-0.0333928695	-0.0068424841	-0.0402353536	***
beds	-0.0513660472	-0.0105253416	-0.0618913888	***
bedrooms	0.1054134362	0.0216001131	0.1270135493	***
instant_bookableTrue	0.0086897243	0.0017805987	0.0104703230	
min_nights	-0.0118941198	-0.0024372067	-0.0143313265	***
person_capacity	0.1004656328	0.0205862659	0.1210518987	***
ekb	-0.1342980145	-0.0275188098	-0.1618168243	***
dist_walk	-0.0128406361	-0.0026311560	-0.0154717921	***
kazan	-0.3192013228	-0.0654070762	-0.3846083990	***
nn	-0.2580505318	-0.0528767571	-0.3109272889	***
nvsbrsk	-0.2270265933	-0.0465196872	-0.2735462805	***
smr	-0.2820506558	-0.0577945874	-0.3398452432	***
spb	-0.1183108857	-0.0242429106	-0.1425537963	***
ekb:dist_walk	0.0019085599	0.0003910802	0.0022996401	
dist_walk:kazan	0.0090058873	0.0018453832	0.0108512705	*
dist_walk:nn	0.0087264864	0.0017881316	0.0105146180	***
dist_walk:nvsbrsk	0.0098316507	0.0020145892	0.0118462398	***
dist_walk:smr	0.0118287334	0.0024238085	0.0142525418	***

dist_walk:spb	-0.0005009717	-0.0001026534	-0.0006036251	***

Table 3: Absolute direct change

City	Absolute coefficient	$100 * (e^{\beta} - 1)$
Moscow	-0.0128406361	-1.275855
Ekaterinburg	-0.01093208	-1.087254
Kazan	-0.003834749	-0.3827406
Nizhniy Novgorod	-0.00411415	-0.4105698
Novosibirsk	-0.003008985	-0.3004463
Samara	-0.001011903	-0.1011391
Saint-Petersburg	-0.01334161	-1.3253

Table 4: Absolute indirect change

City	Absolute coefficient	$100 * (e^{\beta} - 1)$
Moscow	-0.0026311560	-0.2627698
Ekaterinburg	-0.002240076	0.03911567
Kazan	-0.0007857728	0.1847087
Nizhniy Novgorod	-0.0008430244	0.1789731
Novosibirsk	-0.0006165668	0.201662
Samara	-0.0002073475	0.2426748
Saint-Petersburg	-0.002733809	-0.01026481

Table 5: Absolute total change

City	Absolute coefficient	$100 * (e^{\beta} - 1)$
Moscow	-0.0154717921	-1.535272
Ekaterinburg	-0.01317215	0.2302286
Kazan	-0.004620522	1.091036
Nizhniy Novgorod	-0.004957174	1.057009
Novosibirsk	-0.003625552	1.191668
Samara	-0.00121925	1.435459
Saint-Petersburg	-0.01607542	-0.0603443