Measuring the spatial effect of public transportation: An empirical study on rental price and metro stations in Shanghai

by

Jiachen Huang

An honors thesis submitted in partial fulfillment

of the requirements for the degree of

Business and Economics Honors Program

NYU Shanghai

May 2019

Professor Marti G. Subrahmanyam Professor Yiqing Lu Professor Shuang Zhang	Professor Enric Junqué de Fortuny
Faculty Advisers	Thesis Adviser

Measuring the spatial effect of public transportation: An empirical study on rental price and metro stations in Shanghai

Jiachen Huang

May 17, 2019

Abstract

Studies on how to determine rental price have become increasingly important with the development of modern cities. The accessibility to public transportation sites, such as metro stations, is a fundamental factor that influences rental price of housing units. In the past few decades, the effect of nearby metro stations on rental price was usually estimated via accessibility measures. On the basis of research by Taisuke Sadayuki,[1] this paper modifies the theoretical model that relates metro stations proximity to rental prices and then tests this model using data from Shanghai. My model introduces metro passenger traffic data as an improved measure of the quantitative characteristics of the station. Estimation results reveals the spatial effect of the metro system on housing rents, which decreases with the distance and increases with the quantitative score, and proves the feasibility of applying our proposed model to Shanghai. The modification to the model shows expected improvement by making parameter estimations maintain stable as adding more stations into the model.

1 Introduction

Determining what factors influence rental prices has always been a considerable problem for real estate studies and urban economics. In the house rental market, housing unit's geographical relationship with nearby public transportation sites, such as metro stations, is a fundamental factor in determining rent. Previous studies generally attempted to build a model for estimating the spatial effects of multiple sites meeting the following assumptions. (A1) The shorter the distance to nearby transportation sites, the higher influence from the site on the housing unit. (A2) The characteristics of each site may

1. Taisuke Sadayuki, "Measuring the spatial effect of multiple sites: An application to housing rent and public transportation in Tokyo, Japan," *Regional Science and Urban Economics* 70 (2018): 155–173.

lead to different impacts on rents. (A3) As the "rankings" near the site increase, the site has an increasing impact on housing rents.[2] The key part of the model is measuring a housing unit's accessibility to surrounding transportation sites. While previous studies established several patterns of "accessibility measures," this study develops a highly accurate estimation based on a weighted gravity proximity measure and performs a thorough application study on the influence of metro stations on housing rents in Shanghai.

Several proximity measures were used in previous research. The most straightforward measure calculates the distance from the closest site to the housing unit.[3] An alternative is to use the number of sites located in a given range around the housing unit.[4] Another approach widely used in early times is a binary-type measure, specifically, whether the site is included in a given circle centered at the housing unit.[5] These three measures have not only failed to satisfy the general assumptions, each of those measures is restricted under specific criteria and may generate a biased estimate if any requirements do not hold. For example, one obvious problem of the first proximity measure is that it automatically ignores the effect from the other sites except the closest one. Two methods were proposed to address the impact of multiple sites: (1) run a regression on the closest, second closest, and so on; and (2) use the sum of the distances between sites to the housing unit.[6] However, these two remedies brought with them new issues. Adding multiple sites to the model can result in serious multicollinearity problems, preventing scholars from deriving useful interpretations of the spatial effects. On the contrary, simply using the sum of distances requires a selection of buffers, which is usually determined by the researcher in any way. Researchers like McMillen and McDonald tried to avoid the spatial heterogeneity problems by limiting the use of housing samples that are very close to locations rather than playing with variables of multiple sites.[7]

2. **Sadayuki**; **157.**.

3. Gabriel Ahlfeldt, "If Alonso Was Right: Modeling Accessibility And Explaining The Residential Land Gradient," *Journal of Regional Science* 51, no. 2 (2010): 318–338.

4. Xian F. Bak and Geoffrey J.d. Hewings, "Measuring foreclosure impact mitigation: Evidence from the Neighborhood Stabilization Program in Chicago," *Regional Science and Urban Economics* 63 (2017): 38–56.

5. Ben Hoen et al., "The Impact of Wind Power Projects on Residential Property Values in the United States: A Multi-Site Hedonic Analysis," 2009,

6. John Campbell, Stefano Giglio, and Parag Pathak, "Forced Sales and House Prices," 2009,

7. Daniel P. Mcmillen and John Mcdonald, "Reaction of House Prices to a New Rapid Transit Line:

Some researchers made modifications to the model and adopted a new proximity measure that addresses those three assumptions. Those amendments included a redefinition of the traditional proximity measure to transfer the model into a spatial analysis model, which allows estimation of the point-to-point effect. This amendment uses an "accessibility measure" based on the gravity function that increase with the attractiveness of the destination and decrease with distance.[8] Then, an improved proximity measure was introduced with an additional parameter addressing the order of proximity. Sadayuki's work is the most recent study on the topic of rental price and metro stations using this kind of accessibility measure. Compared with previous studies focused on zone-to-zone effects, the focus of the proposed model in Sadayuki's study is more local and attempts to derive the spatial effect of multiple metro stations. The accessibility measure allows for a more flexible implementation when constructing the functional form compared with proximity measures developed in early times. Second, the number of parameters remains unchanged in the measure when the site number shifts. Therefore, this measure can well solve the multicollinearity problem and render useful implications about the spatial effect.

In the past two decades, efforts to implement the accessibility measure into the hedonic model have increased, particularly in fields like real estate study and transportation.[9] Although recent studies have proposed a proximity measure that fulfills the three assumptions and can be applied in general cases, the implementations of previous models still missed important quantitative characteristics in the case of the metro system. To our knowledge, most studies that used hedonic analyses to examine the housing market only considered the effect of the closest station, such as Nakagawa[10] and Diewert.[11] In Sadayuki (2018), up to nine closest stations were included in the model and the order of proximity was considered to estimate the determination of housing rent in Tokyo,

Chicagos Midway Line, 1983-1999," *Real Estate Economics* 32, no. 3 (2004): 463–486.

8. Sadayuki, "Measuring the spatial effect of multiple sites: An application to housing rent and public transportation in Tokyo, Japan."

9. John R. Ottensmann and Greg Lindsey, "A Use-Based Measure of Accessibility to Linear Features to Predict Urban Trail Use," *Journal of Transport and Land Use* 1, no. 1 (2008).

10. Masayuki Nakagawa, Makoto Saito, and Hisaki Yamaga, "Earthquake risk and housing rents: Evidence from the Tokyo Metropolitan Area," *Regional Science and Urban Economics* 37, no. 1 (2007): 87–99.

11. W. Erwin Diewert and Chihiro Shimizu, "Hedonic regression models for Tokyo condominium sales," *Regional Science and Urban Economics* 60 (2016): 300–315.

Japan.[12] However, the quantitative characteristics chosen to describe a metro station were still quite arbitrary. Most studies only used the number of lines and ignored the notion that different stations may have different importance according to their traffic situation. By considering the importance of a station in the whole metro system, our model allows us to have a productive and more practical explanation for the spatial effects of surrounding metro stations with an improved evaluation of sites.

In the following section, we introduce the hedonic model and traditional accessibility measures to give a brief explanation about the basis of our research. Then, we describe our proposed modification to the accessibility measure and the functional forms of exponential and linear types in the methodology section. In the application section, our dataset is displayed and explained, including the house and transportation data, before we proceed to the model estimation. Estimation results from our proposed model and accessibility measure are discussed in the results section, which suggests our application can well solve the research problem with several metro stations included in the model. It has been shown that distance, quantitative characteristics, and proximity order all influence the impact of a station to nearby housing.

Considering that no research has explicitly studied the spatial effect of the metro system on Shanghai's rental prices, the findings of this empirical study suggest that the proposed accessibility measure has good explanation power with Shanghai's data. This study also indicates that proximity order and quantitative characteristics of stations are critically influential factors on rents and are thus worthy consideration when new transportation sites are assessed. The conclusion section discusses our results and offers several potential improvements in the future, such as using the new methodology to identify the role of different stations in the metro system.

12. Sadayuki, "Measuring the spatial effect of multiple sites: An application to housing rent and public transportation in Tokyo, Japan."

2 Hedonic model and the proposed accessibility measure

2.1 Traditional accessibility measures

In previous studies on transportation sites, point-to-point spatial analysis was realized by implementing an accessiblity measure developed from conventional ones used to study zone-to-zone effects. To define our question, i refers to the ith housing unit and the jth closest site around housing i is indicated by j. We supposed that K types in total of different sites exist in our model with distinct spatial effects, and each site $s_{i(j)}$ falls into one of those K categories. Generally, the traditional gravity-based accessibility measure used in hedonic model is as the follow.

$$G_i^J = \sum_{j=1}^{J} \left(\sum_{k=1}^{K} D^k(s_{i(j)}) f^k(d_{i(j)}, q_{i(j)}) \right) + c_{(j)} \tag{1}$$

The gravity-based function G_i^J, which represents a well-used accessibility measure, has functional forms of $d_{i(j)}$, $q_{i(j)}$, and $D^k(s_{i(j)})$, in which $j = \{1; 2; ...; J\}$. $d_{i(j)}$ is the distance between housing unit i to the station with index j, $s_{i(j)}$. As for $q_{i(j)}$, it is the variable standing for the quantitative characteristics of station $s_{i(j)}$. Different types of sites exist, and these types are represented by a dummy variable $D^k(s_{i(j)})$, which takes a value of 1 if $s_{i(j)}$ belongs to type $k \in \{1; ...; K\}$. The model can include J closest sites according to different settings, with $c_{(j)}$ representing the intercept in the measure. The two most commonly used specifications for $f^{k(.)}$, which is the accessibility measure, namely, exponential type and linear type.

$$f^k(.) = \tau^k q_{i(j)} e^{\alpha^k d_{i(j)}} \tag{2}$$

In the exponential-type model defined as function (2), $\tau^{(k)}$ and $\alpha^{(k)}$ are parameters associated with q and d, respectively. According to our prediction that as the distance increases the spatial effect declines, $\alpha^{(k)}$ is expected to have a negative sign. As for the metro station's quantitative characteristic, its associating parameter $\tau^{(k)}$ should be positive and the characteristic q itself should be non-negative regardless of type-k.

$$f^k(.) = \tau^k q_{i(j)} + \alpha^k d_{i(j)} + \omega^k \tag{3}$$

In the linear-type model, $\tau^{(k)}$, $\alpha^{(k)}$, and $\omega^{(k)}$ are three parameters. Using the linear-type specification has a huge benefit because the estimation results of $\tau^{(k)}$ and $^{(k)}$, can be directly interpreted as two marginal effects: quantitative characteristics and distance. Moreover, we can release the assumption that those two effects should have a negative correlation to give our model improved flexibility and allow us to obtain results when the spatial effect is not very significant (i.e. when parameters are small).

2.2 Proposed proximity measures

This study adopts an improved proximity measure developed by previous research,which adds a proximity order into our model. The basic idea is simple: the first closest station has a higher impact on the housing rent than the second closest one and so on. We consider a straightforward implementation that multiplies $f^k(.)$ by a new parameter $g^k(j)$ to weigh the spatial effect differently according to the proximity order of a station.

$$G_i^J = \sum_{j=1}^{J} \left(\sum_{k=1}^{K} D^k(s_{i(j)}) g^k(j) f^k(d_{i(j)}, q_{i(j)}) \right) + c_{(j)} \tag{4}$$

In this study, we use the specification $g^k(j) = j^{\theta^k}$ as the weighting function. For the special case where all metro stations have an identical impact, θ^k is 0 and (5) decreases to the original functional form (1). We can see where the improvement of the proximity measure is making a difference. The discounting impact is considered and a wrong specification of the impact of the sites with different proximity orders is avoided. However, an inevitable identification problem occurs with this proximity measure in the computation, during which the estimation can sometimes fail to converge when θ^k or τ^k is relatively small.

Next, we introduce the explicit models with their functional form and use the models to test the spatial effect of Shanghai's metro stations in the Application section. Weighted proximity measure is directly applied and the linear-type measure is likewise implemented. One major difference between the measure in this study and the traditional measures in previous literature is that we only consider the first J closest sites and study how people tried these sites differently according to their orders. By contrast, many previous studies simply included all sites within a certain range when computing the accessibility. In this specific point-to-point effect study on housing and public transportation, using a certain number of sites rather than all surrounding sites to construct our proximity measure is appropriate and aligns with common sense.

3 Application

3.1. Estimation models

The research question is explored using a hedonic regression analysis of a large sample of housing rent and metro system data. Hedonic regression is often used in real estate and property studies to investigate the impact of several factors that affect housing value. The advantage of using this model is that it breaks down the factors into constituent characteristics, which allow us to estimate the impact of each category of characteristics. In this study, the model of housing rent is reduced to two constituent parts, housing attributes and public transportation accessibility. Generally, hedonic models are estimated using regression analysis, and MLE is used for most parts of this project.

The hedonic function of rental price prediction:

$$Rent_i = G_j^J \left\{ j, d_{i(j)}, q_{i(j)}, D_{i(j)}^k \right\} + X_i\beta + e_i \tag{5}$$

$Rent_i$ is the value of the housing rent with index i. G_i^j is the gravity-based proximity measure, including the J closest stations around the unit i. X_i is the vector of housing attributes, and β is the vector of coefficients associated with X_i. The proximity measure is a function of $d_{i(j)}$, the distance from the station to the housing unit, and $q_{i(j)}$, the quantitative parameter of station j. D is a dummy variable indicating the type of station. Here, the dummy variable $D^1_{(j)} = 1$ if a station $s_{i(j)}$ is the closest one to access a certain metro line, and $D^1_{(j)} = 0$ otherwise.

Now, we define the explicit functional forms of the two weighted accessibility measures.

$$G_i^J = \sum_{j=1}^{J} \left[D^1_{i(j)} j^{\theta^1} \tau^1 q_{i(j)} e^{\alpha^1 d_{i(j)}} + D^0_{i(j)} (j-1)^{\theta^0} \tau^0 q_{i(j)} e^{\alpha^0 d_{i(j)}} \right] + c_{(j)} \tag{6}$$

$$\begin{aligned} G_i^J = & \sum_{j=1}^{J} (D^1_{i(j)} j^{\theta^1} (\tau^1 q_{i(j)} + \alpha^1 d_{i(j)}) + D^0_{i(j)} (j-1)^{\theta^0} (\tau^0 q_{i(j)} + \alpha^0 d_{i(j)}) \\ & + D^0_{i(j)} (j-1)^{\theta^2} \omega^0) + c_{(j)} \end{aligned} \tag{7}$$

Function (7) defines Model A of the exponential type, and the order weighting is realized by adding a term j^{θ}. θ^0 and θ^1 are expected to be negative if a diminishing spatial effect of metro stations with the proximity order exists. On the contrary, if the spatial effect has no correlation with the proximity order of stations, then $\theta^0 = \theta^1 = 0$. Especially, the term $(j-1)$ is constructed to represent the effect of the second closest station and so on, while the coefficient τ^1 and α^1 indicate the case when $D^1_{(j)} = 1$ and represent the effects of the closest station. Model B in Function (8) is a linear-type extension of traditional measures with newly introduced weighting parameters. Here, $omega^0$ stands for the intercept, which is just the setting of this measure. We are not going to discuss the explicit interpretation of *omega* in the following applications. The hedonic functions with two proximity measures are estimated using the maximum likelihood method (MLE).

One major improvement of our implementation is the choice of quantitative characteristics. Rather than simply using the number of lines at station s_j as the quantitative parameter $q_{i(j)}$, we derive a more accurate quantitative measure by using metro card swipe data from the Shanghai metro system. By constructing this variable $q_{i(j)}$, our model can highly reflect the importance of each station depending on its relative traffic volumes along the metro line.

$$q_j = \left(\frac{ln(traffic_j) - \overline{ln(traffic_j)}}{max(ln(traffic_j)) - min(ln(traffic_j))} + 1 \right) \times number\_of\_lines_j^{1/2} \tag{8}$$

3.2 Data

Table 1: Definition of Variables

Variable	Definition
Rent	Housing unit rental price per month in RMB
$d_{(j)}$	Euclidian distance in meters to the $s_{(j)}$
$q_{(j)}$	Quantatitive score at $s_{(j)}$
$D^1_{(j)}$	Station type dummy, =1 if has a new line
$D^0_{(j)}$	Station type dummy, =1 if no a new line
Area	Floor space in square meters
Bedrooms	Number of bedrooms
Kitchens	Number of kitchens
Restrooms	Number of restrooms
Level	Floor level
Age	Age of the housing
N	Housing direction, =1 if the direction is north
S	Housing direction, =1 if the direction is south
E	Housing direction, =1 if the direction is east
W	Housing direction, =1 if the direction is west

Two major sources of data are used. Housing rental data in Shanghai City are from Lianjia, a housing rental and real estate platform.[13] Specifically, we apply a web crawler in October 2018 to obtain the data from the Lianjia Shanghai website, thus ensuring its timeliness and reliability. A total of 11,649 sample housing units from 4,559 communities

13. *Lianjian Wang*, `https://sh.lianjia.com/`.

(residential quarter) are obtained after the preliminary data processing, such as dropping the missing values and removing the top and bottom one-percentiles. The housing rental data include monthly rents and housing attributes, such as name of the community, age of the unit, floor level, number of living rooms and bedrooms, number of total stories of the building, and direction of windows. Table 1 gives detailed definitions of the variables we need, while Table 2 shows several basic statistics of those variables. The average monthly rent in the sample is approximately 7,747 RMB after the top and bottom values are removed. The average size of the housing unit is 92 square meters, and the average age in year of all housing units is 17.97.

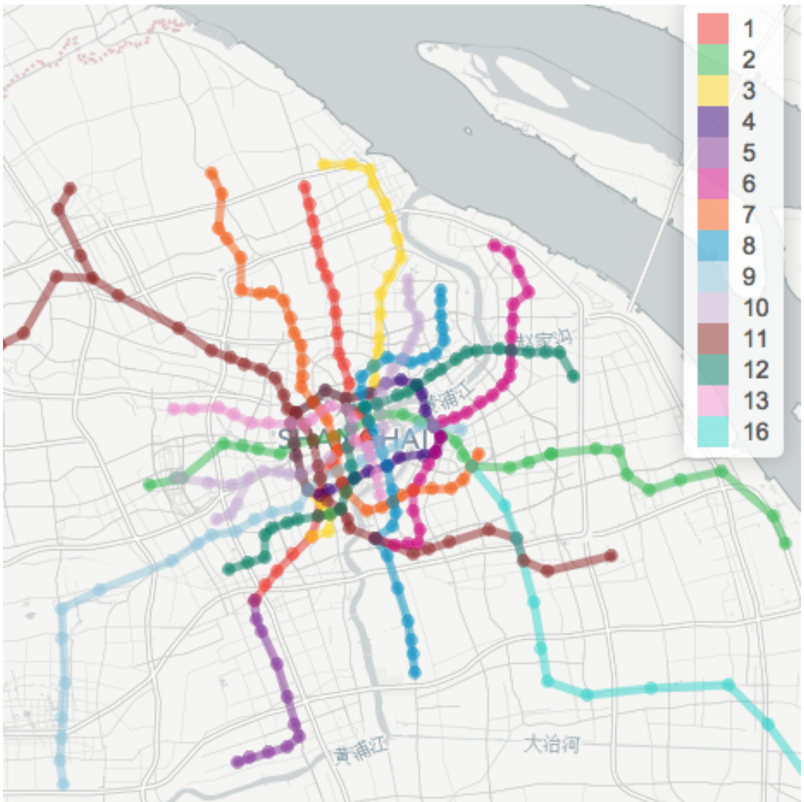

Figure 1: Metro lines and stations in Shanghai

The metro data are accessed from the database of the Shanghai Metro Data Team. The metro line data is in ArcGIS format and we convert it into our desired dataset by using the “geopanda” package offered in Python. Our metro line data consist of 16 lines. Figure 1 shows the lines and stations in Shanghai City. This dataset includes the name of every station, the line number available at that station, and the geometrical coordinates (longitude and latitude). As of 2018, Shanghai’s metro system had 323 stations, of which

243 stations only have one line, 62 with two, 16 with three, and 2 with four. From this processed dataset, we compute the Euclidian distance from a housing unit to its surrounding metro stations using a haversine method. The first $n = 5$ closest stations are then identified by comparing the distances, and we will test the conditions from when $n = 1$ to when $n = 5$.

Table 2: Basic statistics

Variable	Mean	S.D.	Min	Max
Rent	7747	5032	2230	25900
Variables				
Age	17.97	9.56	3	108
Size	92.98	53.64	6	619
Distance $d_{(j)}$				
$d_{(1)}$	1220	1656	29	29470
$d_{(2)}$	1754	1689	203	30208
$d_{(3)}$	2252	1822	388	30706
$d_{(4)}$	2746	2104	656	31403
$d_{(5)}$	3150	2333	865	31246
Quantitative Score $Q_{(j)}$				
$Q_{(1)}$	1.32	0.65		
$Q_{(2)}$	1.29	0.61		
$Q_{(1)}$	1.32	0.64		
$Q_{(1)}$	1.30	0.61		
$Q_{(1)}$	1.32	0.61		

Table 2 shows the metro data, including basic statistics on housing attributes and variables derived from metro station data. The average value of the closest distance variable d_1 is 1,220 meters, and the respective average values of the second and third closest stations are 1,754 and 2,252 meters. Approximately 20% of the second closest stations have a "new line", which means they are the closest station lead to a new line other than that in station 1, and such a proportion remains similar as we go to the third and fourth closest stations. This feature is considerably different from the research conducted in Tokyo, and one possible explanation is the difference in the metro system density in two cities.

4 Results

Model A is the first model to be estimated, and it contains the traditional measures and weighting parameters that address the proximity ranking order. Table 5 shows the results of parameters estimated using the MLE. Control variables in the housing attribute vector X are currently skiped here.

Table 3: Parameter estimates of Model A1

	Models with stations from 1 to 5				
	(J=1)	(J=2)	(J=3)	(J=4)	(J=5)
τ^1	2.20***	2.21***	2.12***	2.04***	1.95***
	(0.072)	(0.068)	(0.064)	(0.060)	(0.057)
α^1	-0.94***	-0.85***	-0.70***	-0.58***	-0.50***
	(0.053)	(0.045)	(0.036)	(0.030)	(0.024)
τ^0		-1.01	-1.31	-2.75	-2.96
α^0		-4.71	-7.02***	-3.82*	-1.23
			(0.667)	(2.282)	
θ^1		-1.18***	-1.82***	-1.30***	-1.11***
		(0.094)	(0.059)	(0.054)	(0.036)
θ^0			-1.54	-0.64	-0.47
Log-Likelihood	-27,459	-27,309	-27,221	-27,117	-27,003
AIC	54,935	54,643	54,466	54,259	54,031

The dependent variable here is set to be the monthly rent in 1000 RMB.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Starting with the results from Model A, the first step is to examine the accessibility measure parameters τ^1 and α^1. Both parameters show the expected signs and are statistically significant, which aligns with our prediction that the rents of neighboring housing can be positively influenced by a station leading to a new metro line. In addition, finding that a housing unit is expensive if it is close to a station or that the station has a high quantitative score is natural. Parameters τ^0 and α^0 show a low statistical significance, implying that when evaluating housing rent, a station has almost no influence if there exists a closer station leading to a particular line.

Estimation results shown in Table 5 correspond with traditional measures. Although the assumption of the traditional accessibility measure is that all close stations have equal importance to a residential housing unit regardless of the proximity order, this model can reveal the effect difference with orders when n is two or larger. According to the result, θ^1 has a negative sign and is statistically significant, implying that the model containing proximity order performs well using data from Shanghai. As we can observe, θ^1 remains negative in $n = 3$ to $n = 5$, which meets our previous prediction. In Model A, the log likelihood and AIC have the same trend of improvement when we include more stations into the model and maintain a good level, same as in previous studies. However, the estimation is noticeably not as stable as the result from Tokyo, especially when more than three stations are added. This result may be attributed to the built-in structural difference between the metro systems of two cities and their diverse logic behind the real estate market.

Figure 2 (Model A)

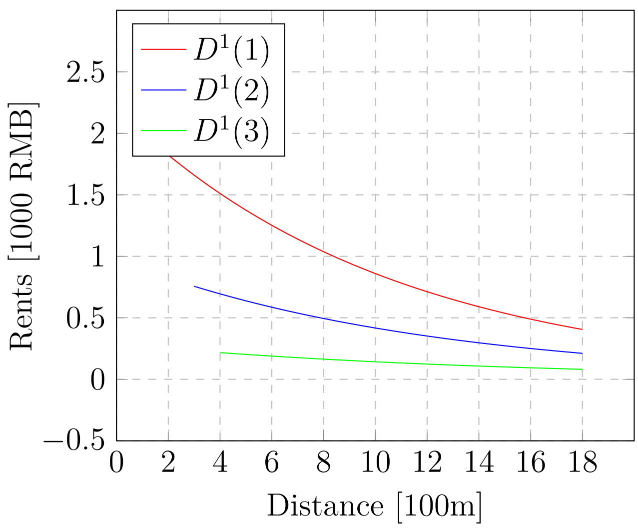

Figure 2 is a straightforward display of the spatial effect of a metro station on housing rents and how the effect shifts according to the distance of a housing unit to the station. For the sake of computational convenience, the quantitative characteristics q at the station is set to 1. Although we draw all station orders in one graph, the levels of their effect cannot be directly compared across orders because the term c is not identifiable in this hedonic function implementation. As displayed in Figure 2, a drop of 1021 yuan in the monthly rent can be expected when the d_1 increases from 400 meters to 1,600 meters. When the distance to the second closest "new-line" station (i.e. d_2) changes by 1200 meters, the rent falls by approximately 283 yuan. The change of stations without a new line is ambiguous, which agrees with common sense that people do not care about additional stations if an existing station already leads to the same line. The functional form of Model A already restricts two major parameters d and q to be negatively correlated. This assumption should hold to ensure a reasonable interpretation of the result.

Estimation results for Model B are described in Table 4. Our desired parameters α and τ all show the expected signs with high significance, same as the case in Model A. For the case with the closest station, τ^1's value 0.72 means the housing rent will increase by 720 RMB when the quantitative score of the station goes up by 1. As for the parameter α^1, the value of -0.65 means the rent will decreases by 650 yuan as the distance between the housing unit and the first closest station increases by 1000 meters. The negative value of θ^1 proves our assumption that distance and quantitative characteristics have diminishing marginal effects. One simplified interpretation is that the spatial effect of the first closest station is roughly as six times as that of the second closest station. As for stations leading to no new metro line, α^0 and τ^0 are relatively unstable and insignificant in several cases. Therefore, we may continue the assumption that residents only care about the closest station leading to a particular metro line they desire. Finally, θ^0 is not statistically significant, and so we just skip ω for now.

Table 4: Parameter estimates of Model B1

	Models with stations from 1 to 5				
	(J=1)	(J=2)	(J=3)	(J=4)	(J=5)
τ^1	0.72***	0.72***	0.81***	0.89***	0.72***
	(0.048)	(0.048)	(0.048)	(0.048)	(0.048)
α^1	-0.65***	-0.65***	-0.66***	-0.63***	-0.65***
	(0.019)	(0.019)	(0.019)	(0.019)	(0.019)
τ^0		-1.54	-0.89	-0.65	-0.68
α^0		-1.91	-1.09***	-1.08**	-1.23
			(0.234)	(.342)	
ω^0		-4.83 ***	-3.21***	-1.90***	-1.51***
			(0.005)	(0.192)	(0.206)
θ^1		-1.02***	-0.73***	-1.36***	-1.25***
		(0.106)	(0.062)	(0.149)	(0.089)
θ^0			-1.73***	-2.09	-1.93
			(0.003)		
θ^2			-4.91 ***	-0.58 ***	-1.05
			(0.101)	(0.106)	
Log-Likelihood	-27,335	-27,335	-27,235	-27,115	-27,335
AIC	54,687	54,699	54,498	54,258	54,698

The dependent variable here is set to be the monthly rent in 1000 RMB.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Figure 3 (Model B)

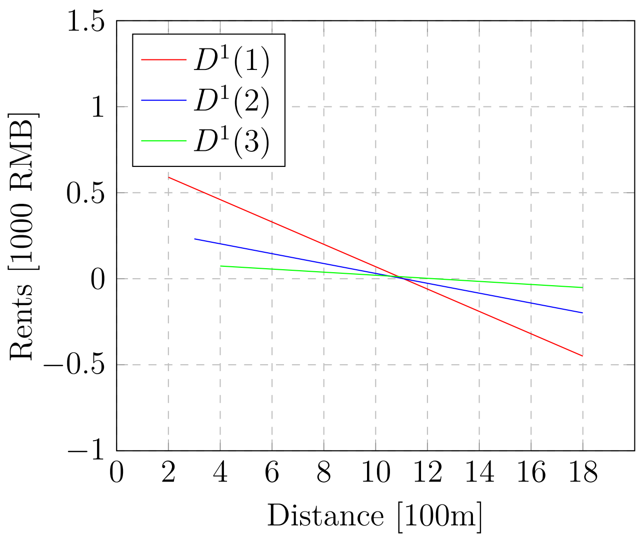

Figure 3 displays of the spatial effect of a metro station estimated using model B. For the sake of computational convenience, the quantitative score q is also set to 1. The levels of their effect cannot be directly compared across orders because the term c is not identifiable in this hedonic function implementation.In Figure 3, a drop of 780 yuan in the monthly rent can be expected as the distance to the closest station increases from 400 meters to 1,600 meters. When it turns to the second closest station, the decreasing value in rent is approximately 220 yuan. Same as the result in model A, the change of stations without a new line is ambiguous.

5 Conclusion

This study's objective is to estimate the spatial effect of Shanghai's metro system using the improved accessibility measure developed from previous studies. Results indicate that the weighted proximity measure keeps its explanation power when implemented in the context of Shanghai and shows a diminishing spatial effect according to the proximity order. With regard to the updated quantitative characteristics, constructing an adjusted indicator variable achieves the initial goal of including the passenger traffic volume while avoiding the endogeneity problem. This new indicator variable evaluates the traffic and the importance of a station among all stations in the metro system.

Although this study has met most of its objectives, some of its estimation results are less stable than those in previous studies that used hedonic models. Two possible explanations exist for this result. One is that the model is not entirely specified when it is implemented with data from Shanghai. Another is the built-in complexity within Shanghai's real estate market that limits the accuracy of the spatial effect estimation of any kind. For example, like many cities in China, Shanghai's real estate market is highly dependent on policies that can lead to misspecification with the hedonic model, such as the purchasing quota or the school quota that goes with a house. Another challenge comes from the functional form of the hedonic model, which makes MLE estimation very difficult to get because it is not convex in some cases. We expect to improve the MLE estimate in future studies.

The results also have practical interpretations and significance besides the improvement in proximity measures. First, our findings suggest that the proximity order is vital in assessing a new housing or transportation project because residents usually care more about distance than which line they are close to. The detailed estimation of parameters in the two models offers a clear reference when studying how metro stations explicitly influence housing rents in the neighborhood. Academic researchers, real estate developers, and public transportation planners can obtain insight from this result to assist their future projects. Particularly, interested researchers can implement this model by using historical data or simply waiting for one more year until a new metro line is built to check the fixed effects of this model and address currently unobserved variables. Considering there are very limited studies using the hedonic model to study Shanghai's real estate and transportation, this paper offers a novel approach for local researches. Besides the application on the metro system, the use of accessibility measure we used in the paper can be extended to study different kinds spatial effects, such as the effect of hospitals, malls, and numerous other public amenities, as long as the constructed proximity measure has appropriate quantitative and qualitative characteristics.

Furthermore, this research topic can be elaborated in the future by working on the measures of distance and quantitative characteristics. Some researchers, such as Derrible and Mishra, already use network indices to develop another specification in transportation studies.[14] Other future works can come up with improved identification of different stations by using advanced techniques, such as neural networks.[15] A research group from China has published a study that identifies hub stations of bus networks in Xiamen.[16] A more complicated algorithm will also help improve the estimation of the weighting parameter because, currently, our model only considers the factor of new lines but ignores the metro transferring concern. Research techniques from numerous disciplines can be

14. Sybil Derrible, "Network Centrality of Metro Systems," *PLoS ONE* 7, no. 7 (2012).

15. Felipe Jimenez et al., "Bus line classification using neural networks," *Transportation Research Part D: Transport and Environment* 30 (2014): 32–37.

16. Hui Zhang, Chengxiang Zhuge, and Xiaohua Yu, "Identifying hub stations and important lines of bus networks: A case study in Xiamen, China," *Physica A: Statistical Mechanics and its Applications* 502 (2018): 394–402.

applied to solve this spatial effect problem, and we believe future works on specifying d, q, and the weighting parameter will generate an improved estimate.

To summarize, this study is a worthwhile examination of the proposed model and accessibility measure in Shanghai. We suggest that future studies on the geospatial structure of Shanghai consider this proposed methodology. Although this study still has limitations, considering the significant results, this study is relatively meaningful in practice, particularly for transportation and real estate studies. It would be interesting for future studies to focus on improving station identification and quantitative score construction using advanced methodologies, so that the spatial effects from different geographical sites will be better revealed.

References

Ahlfeldt, Gabriel. "If Alonso Was Right: Modeling Accessibility And Explaining The Residential Land Gradient." *Journal of Regional Science* 51, no. 2 (2010): 318–338.

Bak, Xian F., and Geoffrey J.d. Hewings. "Measuring foreclosure impact mitigation: Evidence from the Neighborhood Stabilization Program in Chicago." *Regional Science and Urban Economics* 63 (2017): 38–56.

Campbell, John, Stefano Giglio, and Parag Pathak. "Forced Sales and House Prices," 2009.

Derrible, Sybil. "Network Centrality of Metro Systems." *PLoS ONE* 7, no. 7 (2012).

Diewert, W. Erwin, and Chihiro Shimizu. "Hedonic regression models for Tokyo condominium sales." *Regional Science and Urban Economics* 60 (2016): 300–315.

Hoen, Ben, Ryan Wiser, Peter Cappers, Mark Thayer, and Gautam Sethi. "The Impact of Wind Power Projects on Residential Property Values in the United States: A Multi-Site Hedonic Analysis," 2009.

Jimenez, Felipe, Francisco Serradilla, Alfonso Roman, and Jose Eugenio Naranjo. "Bus line classification using neural networks." *Transportation Research Part D: Transport and Environment* 30 (2014): 32–37.

Mcmillen, Daniel P., and John Mcdonald. "Reaction of House Prices to a New Rapid Transit Line: Chicagos Midway Line, 1983-1999." *Real Estate Economics* 32, no. 3 (2004): 463–486.

Nakagawa, Masayuki, Makoto Saito, and Hisaki Yamaga. "Earthquake risk and housing rents: Evidence from the Tokyo Metropolitan Area." *Regional Science and Urban Economics* 37, no. 1 (2007): 87–99.

Ottensmann, John R., and Greg Lindsey. “A Use-Based Measure of Accessibility to Linear Features to Predict Urban Trail Use.” *Journal of Transport and Land Use* 1, no. 1 (2008).

Sadayuki, Taisuke. “Measuring the spatial effect of multiple sites: An application to housing rent and public transportation in Tokyo, Japan.” *Regional Science and Urban Economics* 70 (2018): 155–173.

Lianjian Wang. `https://sh.lianjia.com/`.

Zhang, Hui, Chengxiang Zhuge, and Xiaohua Yu. “Identifying hub stations and important lines of bus networks: A case study in Xiamen, China.” *Physica A: Statistical Mechanics and its Applications* 502 (2018): 394–402.