Initial results

Sih-Yu Wei

April 20, 2023

1 Statement of question

How accessibility to transportation affects the prices of houses is a subject that has been widely investigated in real estate and economics research. In this study, I propose to pursue this topic again, and this time using the evidence of the latest transportation construction plan in Taiwan, namely, the proposal for a new station of Taiwan High-Speed Rail (THSR). The THSR connects the northern part of Taiwan, Taipei, all the way down to the southern part, Kaohsiung, spanning a total distance of 350 km. In 2019, the government announced the Railway will be extended from Taipei to the eastern county, Yilan. In this research study, I want to measure how house prices change because of the construction plan.

A public construction plan has different stages of publicizing. Following Yiu and S. Wong (2005), I plan to exploit the differences in the stages. According to official records, as early as 2017, the evaluation of the construction had started, but no formal proposal was raised. It was not until 2019 did the plan become known publicly when the then Minister of Transportation announced the proposal for a new THSR station in Yilan. Then in February 2021, 4 sites in Yilan were chosen as the candidates for the THSR station. I call the period after the announcement Phase 1. After 10 months, in December 2021, the Government announced the final decision on the location among the four. The period after this announcement is referred to as Phase 2.

My hypothesis is that the price of the houses would reflect people's expectations of the accessibility to the THSR station in the future. After the announcement of the proposal, despite some speculation, the station's location was not clear. So one would expect the overall house prices in Yilan to rise moderately. However, in Phase 1, I expect to see a more obvious increase around the 4 proposed locations. As for in the Phase 2, the areas around the winning location would have the most obvious growth among the four sites in house prices.

More precisely, I plan to apply the hedonic model to measure the premium of expected accessibility to the THSR station. I would like to use a difference-in-differences design in the hedonic model framework, and conduct with 2 parts: pre-announcement stage *v.* Phase 1, and stage 1 *v.* Phase 2. The hedonic model for the price of a house evaluates the effect of expected accessibility at the time of transaction. And I employ different kinds of expected accessibility, including a dummy variable, distance to the nearest station, and bins of distance.

For the data sources, I will use the data set published by the Taiwan Department of Land Administration.This data set contains details of all real estate transactions each season, including the exact house address, price, usage, area, date of transaction, etc. Starting in 2012, the law requires each transaction of real estate to be recorded, creating a complete and accurate source of house price data.

2 Literature review

The classical theory of travel cost and house prices can be traced back to von Thünen's model of agricultural land use, who posits that the rent of a property is inversely related to its distance to the market. Stemmed from von Thünen's agriculture-focused model, Alonso (1964), Mills (1967) and Muth (1969) further developed the monocentric urban land use model, called the "bid rent theory", focusing on how house prices change as the distance from the central business district (CBD) increases, where the bid rent refers to the land user's willingness to pay for the accessibility to the CBD.

This research would serve as a piece of evidence for the theory above by showing the decrease in travel costs to the CBD (in this case, Taipei) will lead to an increase in housing prices. There have been several related studies done examining the effect of improvement in transportation. For example, Debrezion, Pels, and Rietveld (2011) use the hedonic model on the Dutch housing price data from 1985-2001 to conclude houses that are close to the station are 25% more expensive than houses that are at a 15 km or longer distance; Levkovich, Rouwendal, and Marwijk (2016) explores the effect of highway development on housing price, also in the Netherlands, with repeat sales and DID method, and finds a positive effect of proximity to the highway on housing price; Mohammad et al. (2013) conducts meta-analysis on 23 empirical studies, mostly in the US, and finds that the distance to railway stations has a positive relationship with the house prices and that the change in the purchase price is actually similar across the studies.

As opposed to the extensive evidence found in Europe and the US, the studies focused on East Asia provide relatively weaker evidence, some due to a lack of good quality data (*e.g.* Andersson, Shyr, and Fu 2010; Hu 2010), and some due to imprecise measurement (*e.g.* Geng, H. Bao, and Liang 2015). However, the investigation of the question in East Asia remains important as we are currently observing a large number of ongoing and planned constructions of transportation infrastructure in East Asian countries, and the effect on real estate prices may be different from the western countries because of differences such as house owning culture.

Another strand of literature related to this research is the effect of expected accessibility to transportation. The existing literature studying the dynamics of price uses the timing of the completion or the opening of the transportation as the shock of the transportation (*e.g.* Levkovich, Rouwendal, and Marwijk 2016). However, this would likely be an underestimate of the change since the price of the housing would be likely to increase right after the announcement of the project and well before the opening of the service due to the expected increase in accessibility. There are a few articles focusing on the expected increase in accessibility, such as Yiu and S. Wong (2005) and H. X. Bao, Larsson, and V. Wong (2021) investigating the effect of expected tunnels on housing prices, and Cengiz, İnce, and Çelik (2022) find that the average increase in house prices before construction is greater than the increase after the construction had started.

3 Contribution

Most of the studies measuring the effect of transportation are conducted using cross-sectional data. However, only using the cross-sectional data would suffer from endogeneity problems, such as omitted variable bias. It is possible that transport infrastructure is chosen to be built in a neighborhood with certain traits that are not included in the hedonic model. This study, however, proposes a semi-DID design. Not only does the design preserve the benefits of the hedonic model, but it also incorporates the spirit of a control group. Using the "losing" locations for the THSR station as control, we can ensure that the treated and controlled locations share important characteristics in common, for example, proximity to the township center and train stations.

Additionally, this paper would use the data set made available due to the enforcement in 2012 of a law in Taiwan requiring the registration of the actual selling price in real estate transactions. The law requires people to disclose the complete address of the property and the actual selling price. The data set is thus comprehensive and accurate. Indeed, it contains detailed features of a property such as the number of rooms in the house, so the specification can be richer when constructing the hedonic model. This would solve the problem mentioned above of poor-quality data, and give a more precise measure of the effect. By estimating the effect of expected accessibility instead of actual change in accessibility, this research would increase our understanding of the dynamic of housing prices in terms of changes in expectations. In fact, the proposed method in this research will explore the price change in 4 different stages of announcements, so that the change can be observed even more closely.

Furthermore, the study would add to our knowledge of the housing market in East Asian countries. The housing cultures are certainly different across regions. For example, in China and Taiwan, there is more owner-occupied housing than in the US. This piece of knowledge is important as we see transportation construction growing in East Asia.

Finally, the result of the study will benefit policy-making in Taiwan. Firstly, there are plans to extend the south part of THSR, and the experience this time in Yilan can serve as a lesson. Secondly, it is beneficial for future urban planning. The planned TSHR extension will make the connection between Taipei and Yilan tighter, so understanding the change in house prices in Yilan will help better urban policy for topics such as commuting.

4 Data

The dataset used in this analysis is Actual Price Registration Dataset (APRD). The dataset covers almost all sales transactions of the Taiwanese real estate market for a period of 5 years from 2018 to 2022. These transactions are recorded by the Department of Land Administration, Taiwan Ministry of the Interior. In 2011, the Actual Price Registration Act was passed in the congress. Thus, starting in August 2012, the transaction parties are required to register the actual price of sales transaction of real estates within 30 days after the transaction. This greatly improved the precision of the price information and the coverage of the transactions.

Table 1: Sample size compared to total observations

Year	In Sample	Not In Sample	Total
2018	1,446	4,828	6,274
2019	1,694	5,231	6,925
2020	2,104	5,883	7,987
2021	2,327	6,470	8,797
2022	1,422	4,605	6,027
Total	8,993	27,017	36,010

Table 1 summarizes the sample size for each year, and compares the sample in use with the full observations in the dataset. After filtering out the transactions that happened before 2018, the full transaction data for Yilan county contains a total of 36 010 real estate transactions, with each year ranging from around 6 000 to 9 000 observations.

For the purpose of this research, I restrict the data to only residential houses. Furthermore, to limit the size of computation, I only include the single-family houses, which means that condominium buildings are excluded in the analysis. The remaining sample is cut to a total of 8 993 observations, with an average of around 1 400-2 400 observations each year. Notice that the numbers of transaction in 2022, both in sample and out, are not on the increasing trend. This is likely because transactions that happened in 2022, especially the ones in December, have not yet been registered by the end of the year (recall the 30-day registration rule). Since I only use date to see if the house was sold pre- or post- announcement, the incompleteness of the 2022 data should not cause a problem.

Table 2 lists the variables provided in the dataset. According to Malpezzi (2002), hedonic models for residential houses usually include the following aspects: structural, neighborhood, locational, and contract characteristics. This dataset incorporates information related to the addresses, the price of the real estates, and the structural characteristics. Whereas the structural characteristics and contract characteristics can be directly applied from the dataset, the address requires further processing to exploit the locational effect. I plan on geocoding the houses to compute the proximity to the train station, which would serve as the main independent variable of which the effect we would like to understand.

5 Research Design

5.1 Model specification

The method that I plan on using is the hedonic model. The development of hedonic model is often traced back to Lancaster (1966), where he proposes to measure the utility of the non-market goods by viewing them as characteristics of a market good. Using this model, we can single out the consumer's utility of a certain non-market good if we control all other characteristics equal. Housing is the most common example because people receive the utility of the house not through the building *per se*, but through the service it provides, for example, the shelter from the rain, its accessibility to the park, the rooms in the house that can stack groceries, etc. (Malpezzi 2002) In this research, I want to measure the expected value of the accessibility to a THSR station, using the housing value as the market good where our interest attaches.

Within the hedonic model framework, I will use a difference-in-differences design to specify the effect. I plan on evaluating the expected accessibility to the THSR station in 2 parts. In the first part, after the announcement of the candidate locations, I compare the differences in the house prices within a certain distance to the the candidate locations and those that are not. In the second part, I use the candidate locations as the control group, and observe the difference in house prices before and after the announcement of the winning location. The difference in the effects between the two stages can be used to explain the people's expected utility with different level of probability.

For the model specification, I follow H. X. Bao, Larsson, and V. Wong (2021) and Debrezion, Pels, and Rietveld (2011) to adopt a semi-logarithm functional form. According to Malpezzi (2002), this semi-log model has a few important benefits, including that we can simply interpret the coefficients as percentage change in the house prices given a unit change in the corresponding characteristic. This means using the form can help us compare the same effect across different countries.

The general model used in the research is the following:

$$\begin{aligned} ln(P_{i,t}) = \alpha + \beta Post_t \times EAT_{i,t} \\ + \gamma EAT_i + \delta Post_t + \zeta' \mathbf{X}_{i,t} + \eta_i + \epsilon_{i,t}, \end{aligned} \tag{1}$$

where $ln(P_{i,t})$ is the logarithm of the price of house i transacted in month t; α is the constant; $\mathbf{X}_{i,t}$ is a vector of structural and locational characteristics of house i (I will come back to this); η_i signifies the village/town fixed effect. Our main coefficient in interest is β, which is the effect of the expected accessibility to the THSR; and ϵ is the error term.

I use different measurement for the expected accessibility in the two stages because they have different features. For the first part, I would like to use 2 different metrics: (a) the Euclidean distance between house i and its nearest candidate, and (b) a dummy variable indicating whether house i is within 2 km (or other number) to the nearest candidate location. In case (a), since the larger distance implies less accessibility, it has to be noted that we expect β to have a negative sign; and in case (b), I expect β to have a positive sign since the accessibility should have a positive impact on the price of the house.

For the second part, I want to use a slightly different approach: I treat the "losing locations" as the control group. Thus, the metrics I use here are different: (a) using the proposed station locations as centers and splitting the areas around them into 5 zones for increasing distance away from the centers, (b) using a dummy variable indicating whether a house is within 2 km to the "winning" location, and (c) the continuous Euclidean distance between house i and the winning location. More specifically, for part (a), the model will become the following:

$$
\begin{aligned}
ln(P_{i,t}) = \alpha &+ \sum_{k=1}^{5} \beta_k Post_t \times Treat_i \times Zone_{k,i} \\
&+ \sum_{k=1}^{5} \theta_k Zone_{k,i} \\
&+ \gamma Treat_i + \delta Post_t + \zeta' \mathbf{X}_{i,t} + \epsilon_{i,t},
\end{aligned} \tag{2}
$$

where $Treat_i$ is a dummy variable indicating whether house i is located within proximity to the finalized station location; $Zone_{k,i}$ is a dummy variable which equals 1 if house i is located in zone k to the nearest candidate location. This specification allows us to compare the effects on different zones through $\beta_1, \beta_2, ..., \beta_5$. The variables $Zone_{k,i}$ can control for the effect of locating in different zones. To avoid confusion, this model omits the interaction term of any duo combination of $Post$, $Treat$, and $Zone$. Moreover, the model drops the village/town fixed effect since most of the houses with respect to the same station location are in the same village/town, and so it is not useful to keep the fixed effect considering we already have a control variable for the treatment group.

For part (b) and (c), I am using a similar model as model (1), but the data are restricted to the houses located within the specified distance to a candidate location. Thus, in case (b), $EAT_{i,t} = 1$ only when house i is located near the "winning" location, and $EAT_{i,t} = 0$ when house i is located near a "losing location". In case (c), $EAT_{i,t}$ is measured using the distance between house i and the winning location.

Finally, I want to mention $\mathbf{X}_{i,t}$, the vector of characteristics of house i. In this vector, I would include all the structural variables listed in Table **??**, namely number of stories, main building material, built date, total floor area, number of bedrooms, number of loving ooms, number of bathrooms, and the dummy variable of containing partition. Additionally, the zoning variable under the location category would also be included. I am also planning on including the distance of the house to other important facilities, e.g., schools, hospitals, and the entrance and exit of the highways.

I will use the google maps API to geocode the location of each house by their address, and calculate the distance with QGIS, an open-source geographical information system.

5.2 Method limitation

The model attempts to capture the effect of increase in price due to the expected accessibility to the THSR station. However, the models in this research have the following limitations:

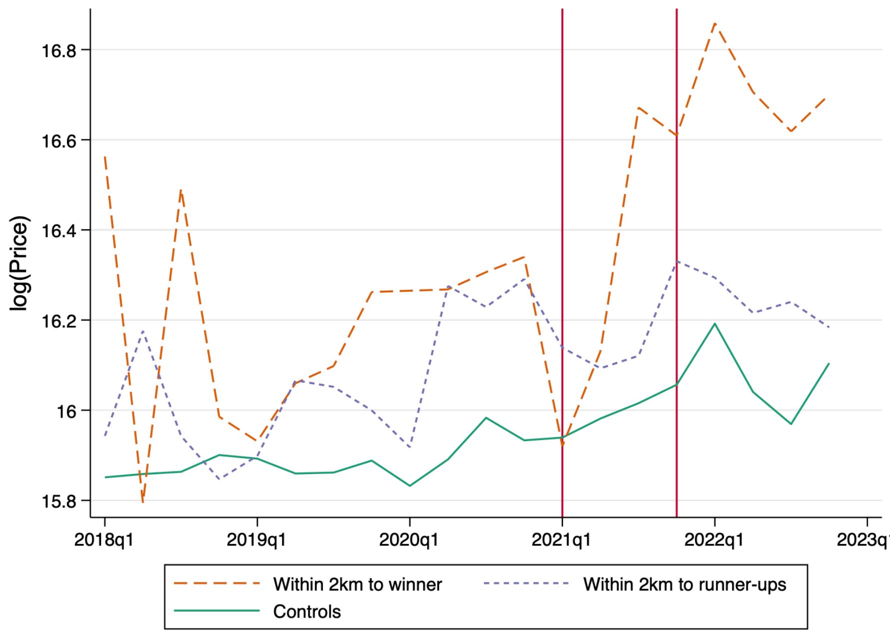
1. *Capture of only partial expected accessibility effect:* The government announced the new construction plan in Yilan in 2019. People already expected the accessibility in Yilan by the time, and the candidate-in-future (e.g. area around train station) were likely already expected to have higher probability to have the station. Since the models in this research can only capture the effect after the announcement of the candidate locations, a part of expected effect would be ignored. As a result, the estimate of this research would serve as a lower bound of the total expected effect.
2. *Different probability across the candidates*: In Model (1), I assume the probability of winning for each candidate is the same, whereas in real life people had different beliefs in which location were more likely to win. The estimate should reflect the average probability across the four candidates perceived by people in the period.
3. *Change of characteristics of houses on market:* The design of the model compares the difference before and after the announcements, assuming the house transacted in the two periods are similar. However, it can happen that once the announcement came out, the unobserved characteristics on-market houses located in the designated area changed. The direction of this bias is ambiguous, and it can bring concern to the accuracy of the estimates.
4. *Change of house types on market:* This is similar to the previous point. Since I only include transaction of single-family houses in the analysis, I ignore other types of residential houses. It could be the case that people have greater preference for apartments after the announcement because they are generally closer to the station. If this were true, then this research would be an underestimate of the effect.

6 Results

Table 2 summarizes the key variables used in the regression analysis. The distance is sourced from Google distance matrix API, which calculates the distance between two locations by car route. This largely improves on the measure of ”accessibility” compared to other traditional measures such as Euclidean distance.

Figure 1 presents the time trend of houses for different groups. Following the announcement of candidate locations in February 2021, we observe a rise in the price of all three groups, with the

Figure 1: Log prices of houses within 2 km



largest increase in the winner location. After the announce of the winner in December 2021, we observe a small spike in 2022 Q1 for the winner group.

6.1 Phase 1

Table 3 presents the regression results for Phase 1. Panel A uses distance to the nearest candidate stations as measurement of expected accessibility. In model (1), the effect is interpreted as a 1.2 percent decrease in price as the house is located 1 km further from the nearest candidate THSR stations. The effect is significant for model (1) and (2), and the scales of the effect are similar (1.2 and 1.0 percent). However, once the structural controls are added in the model, the effect almost completely vanishes. In columns (3) and (4), there is an insignificant 0.1 percent point decrease for 1 km further away from the train station. Similar story happens when I use the dummy of whether the house is located within 1 km to the nearest candidate locations. Without the structural controls, we observe significant 24.4 and 26.8 percent increase in price if the house is located within the distance. However, the effect is much smaller with the structural controls. As I replace the 1 km dummy with the 2 km dummy in Panel C, the effect of expected accessibility is small without structural controls, and the sign is even negative in model (3) and (4). It should also be noticed that the standard errors are similar across models within the same metric of expected accessibility, which implies the controls do not make the effect noisier.

The result in Table 3 suggests that the increase in house prices near the candidate locations can be explained by the difference in the structural characteristics between the houses. One can suspect the houses transacted after the announcement are newer, or have more bathrooms, since these are the variables that are significant in the regression analysis, but without further analysis, it is hard to make sure which specific structural characteristics are the main source of the difference.

6.2 Phase 2

Table 4 presents the effect of the expected accessibility to winner THSR station. The models with the full sample reports a significant effect, with and without structural controls. Compared to the effect in Panel A of Table 3, the effect is considerably larger in all 4 models, reaching 2.4 percent increase in house price for every kilometer closer to the winner location in model (2). Controlling for structural characteristics decreases the effect to 1.4 percent, but that is still much larger than the Phase 1 effect. Once I zoom in and restrict the sample to the houses that are located within 2 km to the candidate locations, the pattern in Phase 1 reappears. Although the effect is even stronger in columns (5) and (6), but it also has large variation, and most of that vanishes if controlled for structural characteristics. The result for the restricted sample, however, does not offer clear insight, since the expected accessibility to THSR of the houses near runner-up locations is distorted by other ways of transportation. All of the candidate locations have a train station nearby, so people are more likely to choose traveling from one location to another by train, which makes the distance measured by car ride a bad measure of accessibility.

I further examine the effect by looking at the discrete metric, which is whether the house is located within 2 km to the winner location. The sample used in this specification is the one with houses that are located within 2 km to all the candidate locations. This time I also find mixed results. While in columns (1) and (2), the increase of 24.5 and 29.1 percent in prices for houses located within the winner location are large in magnitude, the standard error are also large (0.209 and 0.197), resulting in difficulties understanding the effect. Controlling for structural characteristics decreases the standard errors, but the signs of the coefficients become the opposite, suggesting that the house prices near the winner location actually increase less compared houses close to other locations.

Overall for Phase 2, we can conclude that there is consistently more increase in the value for houses near the winner location if we compare those to all other houses in Yilan county. However, as we change the control group to the houses located near the runner-up locations, then results become inconclusive. It should be noticed that the precision of the estimation is poor because the large standard error. If we still want to try to interpret the results, I can see two main directions. First, we can speculate that the houses located near the runner-up locations experience spill-over effect from the THSR station, even when they are not exactly within a short distance to the winner. Since the candidate locations also serve as train stations, the accessibility to THSR also increases largely for their nearby houses. Thus, the spillover can blur the difference between the winner and the runner-ups. The other explanation, based on Figure 1, is the likelihood of the winner being selected in Phase 1 is already larger than other locations, and so the increase happened in Phase 1 more than in Phase 2.

References

Alonso, William (1964). *Location and Land Use: Toward a General Theory of Land Rent*. Cambridge: MA: Harvard University Press.

Andersson, David Emanuel, Oliver F. Shyr, and Johnson Fu (Jan. 2010). "Does high-speed rail accessibility influence residential property prices? Hedonic estimates from southern Taiwan". en. In: *Journal of Transport Geography* 18.1, pp. 166–174. ISSN: 09666923. DOI: 10.1016/j.jtrangeo.2008.10.012. URL: https://linkinghub.elsevier.com/retrieve/pii/S0966692308001257 (visited on 03/02/2023).

Bao, Helen XH, Johan P Larsson, and Vivien Wong (2021). "Light at the end of the tunnel:The impacts of expected major transport improvements on residential property prices". In: *Urban Studies* 58.14, pp. 2971–2990. DOI: 10.1177/0042098020967308. eprint: https://doi.org/10.1177/0042098020967308. URL: https://doi.org/10.1177/0042098020967308.

Cengiz, Elif Can, Enver Cenan İnce, and Hüseyin Murat Çelik (June 2022). "Impact of urban railway announcements on housing market: Preponderance of pre-construction and post-construction phase effect of urban railway projects". en. In: *Case Studies on Transport Policy* 10.2, pp. 1090–1101. ISSN: 2213624X. DOI: 10.1016/j.cstp.2022.03.021. URL: https://linkinghub.elsevier.com/retrieve/pii/S2213624X22000736 (visited on 03/02/2023).

Debrezion, Ghebreegziabiher, Eric Pels, and Piet Rietveld (Apr. 2011). "The Impact of Rail Transport on Real Estate Prices: An Empirical Analysis of the Dutch Housing Market". In: *Urban Studies* 48.5. Publisher: SAGE Publications Ltd, pp. 997–1015. ISSN: 0042-0980. DOI: 10.1177/0042098010371395. URL: https://doi.org/10.1177/0042098010371395 (visited on 03/12/2023).

Geng, Bin, Haijun Bao, and Ying Liang (Oct. 2015). "A study of the effect of a high-speed rail station on spatial variations in housing price based on the hedonic model". en. In: *Habitat International* 49, pp. 333–339. ISSN: 01973975. DOI: 10.1016/j.habitatint.2015.06.005. URL: https://linkinghub.elsevier.com/retrieve/pii/S0197397515001228 (visited on 03/02/2023).

Hu, Chich-Ping (2010). "An Impact Analysis of Housing Prices from the Taiwan High Speed Rail System Opening-an Example of Hsin Chu Station". In: *Journal of architecture and planning* 2.

Lancaster, Kelvin J. (1966). "A New Approach to Consumer Theory". In: *Journal of Political Economy* 74.2, pp. 132–157. ISSN: 00223808, 1537534X. URL: http://www.jstor.org/stable/1828835 (visited on 03/14/2023).

Levkovich, Or, Jan Rouwendal, and Ramona van Marwijk (Mar. 2016). "The effects of highway development on housing prices". en. In: *Transportation* 43.2, pp. 379–405. ISSN: 0049-4488, 1572-9435. DOI: 10.1007/s11116-015-9580-7. URL: http://link.springer.com/10.1007/s11116-015-9580-7 (visited on 03/02/2023).

Malpezzi, Stephen (Oct. 2002). "Hedonic Pricing Models: A Selective and Applied Review". en. In: *Housing Economics and Public Policy*. Ed. by Tony O'Sullivan and Kenneth Gibb. Oxford, UK: Blackwell Science Ltd, pp. 67–89. ISBN: 978-0-470-69068-0 978-0-632-06461-8. DOI: 10.1002/9780470690680.ch5. URL: https://onlinelibrary.wiley.com/doi/10.1002/9780470690680.ch5 (visited on 03/12/2023).

Mills, Edwin S. (1967). "An aggregative model of resource allocation in a metropolitan area". In: *American Economic Review Papers and Proceedings* 2.

Mohammad, Sara I. et al. (Apr. 2013). “A meta-analysis of the impact of rail projects on land and property values”. en. In: *Transportation Research Part A: Policy and Practice* 50, pp. 158–170. ISSN: 09658564. DOI: 10.1016/j.tra.2013.01.013. URL: https://linkinghub.elsevier.com/retrieve/pii/S0965856413000207 (visited on 03/01/2023).

Muth, Richard F. (1969). *Cities and Housing*. Chicago: University of Chicago Press.

Yiu, C.Y. and S.K. Wong (2005). “The Effects of Expected Transport Improvements on Housing Prices”. In: *Urban Studies* 42.1. Publisher: SAGE Publications Ltd, pp. 113–125. ISSN: 0042-0980. DOI: 10.1080/0042098042000309720. URL: https://doi.org/10.1080/0042098042000309720 (visited on 02/28/2023).

Table 2: Descriptive statistics

	Obs	Mean	Median	SD	Min	Max	Sum
Dependent variable							
log(Price)	8993	15.84	15.89	0.60	10.82	19.41	142406.5
Locational							
Distance to winner	8993	9.31	8.18	5.11	0.09	61.53	83728.3
Distance to runner-up 1	8993	12.97	13.11	7.02	0.12	62.20	116681.9
Distance to runner-up 2	8993	10.46	10.13	6.56	0.16	57.26	94098.79
Distance to runner-up 3	8993	9.18	8.67	5.90	0.18	67.08	82538.6
Within 1km to winner	8993	0.01	0	0.07	0	1	48
Within 1km to runner-up 1	8993	0.01	0	0.08	0	1	61
Within 1km to runner-up 2	8993	0.02	0	0.14	0	1	177
Within 1km to runner-up 3	8993	0.01	0	0.10	0	1	96
Zone 1	8993	0.01	0	0.08	0	1	57
Zone 2	8993	0.04	0	0.19	0	1	326
Zone 3	8993	0.08	0	0.27	0	1	695
Zone 4	8993	0.06	0	0.24	0	1	566
Zone 5	8993	0.09	0	0.28	0	1	798
Is urban	8993	0.48	0	0.50	0	1	4319
Structural							
Total floor area	8993	166.16	168	64.75	2	1066	1494319
Total lot area	8993	127.09	102	273.19	4	17375	1142914
% floor area as balcony	8993	4.86	5	4.50	0	29	43705.73
% floor area as auxiliary building	8993	1.26	0	3.75	0	60	11351.55
Age	8736	18.62	10	18.57	0	96	162656
# bedrooms	8993	3.68	4	1.41	0	18	33097
# living rooms	8993	1.92	2	0.73	0	8	17263
# bathrooms	8993	3.31	3	1.40	0	18	29762
# stories	8991	2.76	3	0.71	1	5	24808
Has compartment	8993	0.97	1	0.16	0	1	8766
Has manager	8993	0.02	0	0.15	0	1	197
Has hotspring	8993	0.00	0	0.05	0	1	19
Is leaking	8993	0.00	0	0.06	0	1	34
Including renovation fee	8993	0.00	0	0.05	0	1	23
Material	8230	0.04	0	0.19	0	1	308
Contractual							
Is family transaction	8993	0.04	0	0.19	0	1	343
Is presale house	8993	0.12	0	0.32	0	1	1048
House is not registered	8993	0.06	0	0.24	0	1	545

Table 3: Phase 1 effect on log price of expected accessibility to candidate THSR stations

	(1)	(2)	(3)	(4)
Panel A: Distance				
$Dist \times Post1$	-0.012***	-0.010***	-0.001	-0.001
	(0.004)	(0.004)	(0.003)	(0.003)
Panel B: Within 1km dummy				
$1km \times Post1$	0.244***	0.268***	0.023	0.044
	(0.065)	(0.063)	(0.049)	(0.049)
Panel C: Within 2km dummy				
$2km \times Post1$	0.062	0.104**	-0.025	-0.003
	(0.053)	(0.052)	(0.039)	(0.038)
Contractual controls	No	Yes	No	Yes
Structural controls	No	No	Yes	Yes
Observations	8761	8761	7759	7759

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

Table 4: Phase 2 effect on log price of distance to winner THSR station

	Full sample				Restricted sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Dist \times Post2$	-0.026***	-0.024***	-0.015***	-0.014***	-0.050*	-0.059**	-0.004	-0.006
	(0.007)	(0.007)	(0.005)	(0.005)	(0.030)	(0.028)	(0.018)	(0.017)
Observations	2171	2171	2100	2100	418	418	369	369
Contractual controls	No	Yes	No	Yes	No	Yes	No	Yes
Structural controls	No	No	Yes	Yes	No	No	Yes	Yes

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.

Table 5: Phase 2 effect on log price of within 2 km to the winner THSR station

	(1)	(2)	(3)	(4)
$Dist \times Post2$	0.245	0.291	-0.102	-0.069
	(0.209)	(0.197)	(0.126)	(0.119)
Observations	418	418	369	369
Contractual controls	No	Yes	No	Yes
Structural controls	No	No	Yes	Yes

* p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses.