Impact of Expected Transport Accessibility on House Prices: A Dual-Phase Investigation of the Proposed Taiwan High-Speed Rail Station *

Sih-Yu Wei

May 11, 2023

1 Introduction

The effect of transportation accessibility on housing prices has been a subject of significant interest in real estate and economics research. In the late 18th century, Johann Heinrich von Thünen formulated the influential model named after him, which emphasized the exchange between land cost and transportation cost. In contemporary urban studies, the focus shifted to house prices as a substitute for land cost, and transportation accessibility emerged as a vital determinant of transportation cost. However, limited scholarly attention has been devoted to investigating the impact of ”expected accessibility to transportation,” an essential factor affecting real estate prices. Furthermore, previous studies have faced challenges in addressing the endogeneity inherent in transportation location analysis. This study addresses the question of how transportation affects housing costs by

*This paper serves as the culmination of my coursework in ECON 191, completed during the Spring semester of 2023 at UC Berkeley.

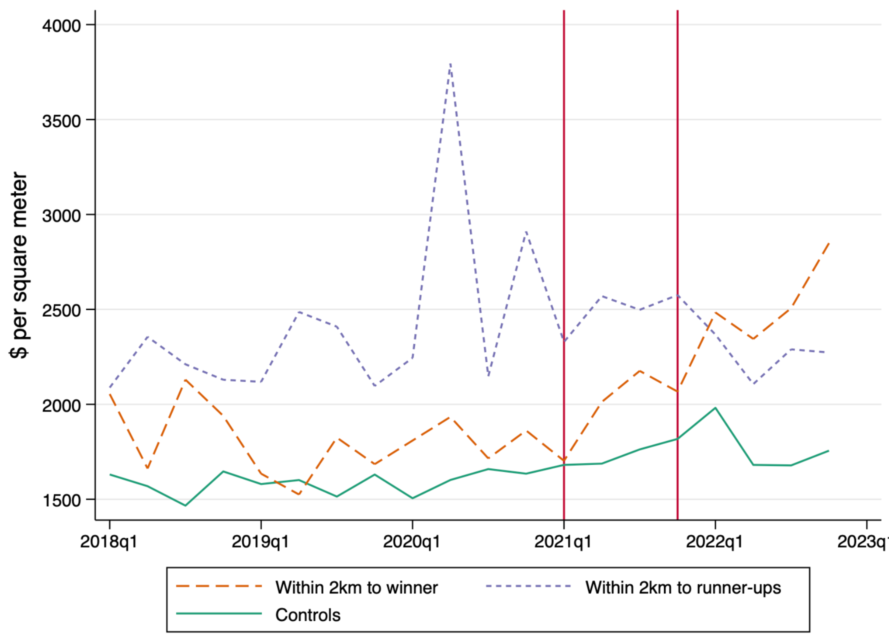

Figure 1: The change of single-family house prices in Yilan county

examining the evidence provided by the latest transportation construction plan in Taiwan.

This study utilizes the proposed construction of a new station for the Taiwan High-Speed Rail (THSR) as a case study. The THSR spans 350 km, connecting Taipei in the north to Kaohsiung in the south. In 2019, the government announced the extension of the railway from Taipei to Yilan, an eastern county in Taiwan. The announcement consisted of two stages: the initial announcement of four candidate sites for the station, followed by the official announcement of the chosen location. This unique setup enables a dual-phase analysis, capturing effects during different stages of the process. Moreover, the candidate design helps mitigate the issue of reverse causality, as the candidates share similar characteristics that also influence the prices of nearby houses.

Figure 1 illustrates the change of single-family house prices in Yilan between 2018 and 2022. According to official records, the evaluation of the construction had already commenced as early as 2017, albeit without a formal proposal. The plan gained public awareness in 2019 when the then Minister of Transportation unveiled the proposal for a new THSR station in Yilan. Subsequently, in February 2021, four sites in Yilan were identified as potential locations for the THSR station. For clarity, I refer to the duration following this announcement as Phase 1. Ten months later, in December 2021, the Government declared its final decision on the selected location among the four candidates, marking the initiation of Phase 2.

The hypothesis posits that house prices would reflect people's expectations of future accessibility to the THSR station. In Phase 1, following the announcement of the candidates, a more pronounced price increase is expected around the four proposed locations. In Phase 2, the area surrounding the winning location is anticipated to experience the most significant growth among the four sites in terms of housing prices.

Methodologically, I employ the hedonic model to assess the premium associated with anticipated accessibility to the THSR station. To achieve this, I utilize a difference-in-differences (DID) design within the framework of the hedonic model, conducted in two parts: the pre-announcement stage versus Phase 1, and Phase 1 versus Phase 2. In the first part, I compare the price changes for houses in close proximity to the candidate sites with those farther away. In the second part, I examine the price changes for houses near the selected winner site compared to other houses, including those near the other runner-up sites. The findings reveal a significant 1.1% increase in price for every 1 km closer to the candidate sites during Phase 1. However, this effect diminishes to 0.1% once house characteristics are taken into account. Moreover, different specifications yield mixed evidence.

In Phase 2, using the full sample, I observe a 2.5% increase in price for every 1 km closer to the selected winner site, and controlling for characteristics reduces the effect to 1.1%, still statistically significant at a 5% level. When examining only houses near the candidate sites, the effect magnifies to 4.8%. However, this is largely attributable to differences in house characteristics, as the effect reduces to a mere 1.8% as controlled for them. The estimation using alternative model specifications provides inconclusive results.

Given the significant impact of controlling for structural characteristics on the findings, I further investigate the determinants of price changes. I discover that the age and size of the buildings undergo significant changes following the announcement of the candidate sites. In fact, the announcement is associated with a 1.0% increase in building area and a 0.4-year reduction in age for every 1 km closer to the site, both statistically significant at a 1% level. As these structural characteristics are primarily determined during the construction of the houses, it can be inferred that houses in better conditions were made available for transactions after the announcements. The effect of expected accessibility to the train station appears to play a relatively minor role in the observed price increase.

The structure of this paper is as follows: In Part 2, I review relevant literature, including theories and recent empirical studies on the impact of accessibility on house prices, while highlighting the contribution of this paper. Part 3 introduces the data used in this research, namely the Actual Price Registration data, and provides the background information that facilitated its availability. In Part 4, I specify the hedonic models and the DID design employed for estimation, along with the limitations that arise. Part 5 presents the results. Part 6 discusses in detail about the heterogeneity of the effect, the changes in characteristics on houses, as well as some limitations left for future studies. Part 7 concludes the paper.

2 Literature review

The classical theory of travel cost and its impact on house prices can be traced back to Johann Heinrich von Thünen's model of agricultural land use. According to von Thünen, the rent of a property is inversely related to its distance from the market. Building upon this agricultural-focused model, subsequent researchers Alonso (1964), Mills (1967) and Muth (1969) further developed the monocentric urban land use model known as the ”bid rent theory.” This theory focuses on how house prices change as the distance from the central business district (CBD) increases, with the bid rent representing the willingness of land users to pay for accessibility to the CBD.

This research aims to provide empirical evidence supporting the aforementioned theory by demonstrating that a decrease in travel costs to the CBD (specifically in Taipei) leads to an increase in housing prices. Several related studies have been conducted examining the effect of transportation improvements. For instance, Debrezion, Pels, and Rietveld (2011) utilized a hedonic model on Dutch housing price data from 1985 to 2001 and found that houses in close proximity to the station were 25% more expensive than those located 15 km or more away. Levkovich, Rouwendal, and Marwijk (2016) explored the impact of highway development on housing prices in the Netherlands using repeat sales and difference-in-differences (DID) methods, revealing a positive effect of proximity to highways on housing prices. Mohammad et al. (2013) conducted a meta-analysis of 23 empirical studies, primarily in the US, and discovered a positive relationship between house prices and distance to railway stations, with consistent findings across the studies.

In contrast to the extensive evidence found in Europe and the US, studies focused on East Asia provide relatively weaker evidence due to factors such as a lack of high-quality data (*e.g.* Andersson, Shyr, and Fu 2010; Hu 2010) and imprecise measurement (*e.g.* Geng, H. Bao, and Liang 2015). However, investigating this question in East Asia remains important given the significant number of ongoing and planned transportation infrastructure projects in the region. The impact on real estate prices may differ from Western countries due to cultural differences such as attitudes towards homeownership.

Another relevant strand of literature explores the effect of expected accessibility to transportation. Existing studies often rely on the timing of transportation completion or opening as a shock, underestimating the price change since housing prices tend to increase after the announcement of a project due to expected improvements in accessibility (*e.g.* Levkovich, Rouwendal, and Marwijk 2016). Some articles focus on expected accessibility, such as Yiu and S. Wong (2005) and H. X. Bao, Larsson, and V. Wong (2021), who investigate the effect of expected tunnels on housing prices. Cengiz, İnce, and Çelik (2022) find that the average increase in house prices before construction begins exceeds the increase observed after construction starts.

Most studies examining the effect of transportation infrastructure use cross-sectional data, which can suffer from endogeneity problems, such as omitted variable bias. It is possible that transport infrastructures are built selectively in neighborhoods with certain characteristics not included in the hedonic model. To address this issue, this study proposes a semi-difference-in-differences (semi-DID) design. This design combines the benefits of the hedonic model with the spirit of a control group. By using the ”losing” locations for the Taiwan High-Speed Rail (THSR) station as controls, the study ensures that the treated and controlled locations share important characteristics, such as proximity to the township center and train stations.

Additionally, this study highlights the utilization of a unique dataset that became available following the enforcement of an act in Taiwan in 2012, which mandated the registration of actual selling prices in real estate transactions. This act necessitates the disclosure of the complete property address and the actual selling price, resulting in the creation of a comprehensive and accurate data set. The inclusion of detailed property features, such as the number of rooms in a house, further enhances the specification in the hedonic model. Consequently, this addresses the previously mentioned issue of poor-quality data and provides a more precise measure of the effect under investigation.

By estimating the effect of expected accessibility rather than the actual change in accessibility, this research seeks to enhance our understanding of the dynamics of housing prices in relation to changes in expectations. The proposed method in this study will examine price changes across two different stages of announcements, allowing for a closer observation of the changes.

Moreover, this study aims to contribute to our knowledge of the housing market in East Asian countries. Housing cultures differ significantly across regions, and understanding these variations is crucial, particularly in light of the growing transportation construction in East Asia. For instance, countries like China and Taiwan exhibit higher rates of owner-occupied housing compared to the United States. This insight becomes increasingly relevant as transportation infrastructure development continues in the region.

Ultimately, the findings of this research will have implications for policy-making in Taiwan. The lessons learned from the current study, particularly in Yilan, can inform future plans to extend the southern part of the THSR. Secondly, these findings will be valuable for future urban planning endeavors. As the planned THSR extension strengthens the connection between Taipei and Yilan, a better understanding of the changes in house prices in Yilan will aid in formulating effective urban policies, particularly in relation to commuting and transportation.

3 Data

The dataset used for this analysis is the Actual Price Registration Dataset (APRD), covering the period from the first quarter of 2018 to the first quarter of 2023. It encompasses a vast majority of sales transactions within the Taiwanese real estate market over a span of five 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 enacted, making it mandatory for transaction parties to register the actual price of real estate sales within 30 days of the transaction. This legislation significantly improved the accuracy and coverage of price information.

For the purpose of this research, the dataset has been restricted to include only residential houses. Furthermore, to streamline the computational process, only single-family houses have been included, while condominium buildings have been excluded from the analysis. The resulting sample consists of 9 377 observations in total, with an annual average of approximately 1 500-2 500 observations. The smaller number of observations in 2022 is likely due to transactions that occurred in that year, particularly in December, had not been registered by the end of the year (recall the 30-day registration rule). It should be noted that houses registered at a later time may possess different characteristics.

Table 1 provides an overview of the variables used in the analysis and presents a summary of

Table 1: Descriptive statistics, by group

	All	< 2 km to winner	< 2 km to runner-ups	Control
Dependent Variable				
log(Price)	15.819	16.334	15.906	15.791
	(0.602)	(0.515)	(0.667)	(0.583)
Structural				
Total floor area (m^2)	166.705	220.922	147.890	169.588
	(64.858)	(72.266)	(76.268)	(60.938)
Total lot area (m^2)	127.100	146.171	107.947	130.731
	(269.048)	(59.092)	(129.949)	(291.901)
% floor area as balcony	4.895	7.567	2.752	5.291
	(4.490)	(3.998)	(4.266)	(4.405)
% floor area as auxiliary	1.252	1.429	1.464	1.205
	(3.734)	(3.557)	(4.213)	(3.629)
Age	18.475	10.150	30.102	16.485
	(18.600)	(12.097)	(18.863)	(17.826)
# bedrooms	3.688	3.578	3.551	3.719
	(1.417)	(1.059)	(1.788)	(1.331)
# living rooms	1.922	2.095	1.829	1.938
	(0.727)	(0.501)	(0.867)	(0.696)
# bathrooms	3.320	3.687	2.968	3.386
	(1.403)	(1.103)	(1.567)	(1.360)
# stories	2.764	3.245	2.556	2.797
	(0.708)	(0.658)	(0.838)	(0.669)
Has compartment	0.965	1.000	0.934	0.970
Has manager	0.022	0.116	0.019	0.020
Has hotspring	0.002	0.000	0.000	0.003
Is leaking	0.004	0.014	0.009	0.002
Includes renovation fee	0.003	0.000	0.001	0.003
Contractual				
Is family transaction	0.040	0.020	0.060	0.036
Is presale house	0.121	0.048	0.037	0.140
House is not registered	0.059	0.041	0.087	0.053
Observations	9377	147	1594	7636

Note: Each cell reports the mean of the characteristic in certain group. Standard errors are displayed in parentheses.

the characteristics for each group. According to Malpezzi (2002), hedonic models for residential houses typically incorporate structural, neighborhood, locational, and contract characteristics. This dataset includes information concerning addresses, real estate prices, and structural characteristics. While the structural and contract characteristics can be directly extracted from the dataset, processing of the address information is required to explore the locational effects. I geocoded of the houses using the Google Distance API to estimate the approximate travel distance by car to the train stations, which enhances the measurement of accessibility compared to traditional methods such as Euclidean distance.

Table 1 also reveals that the control groups are not perfectly matched, as the characteristics of the houses differ noticeably across the three groups. One striking difference is observed in the total floor area, with houses near the winner being, on average, almost 50% larger than those near the runner-ups, and 30% larger than those near neither. Another significant disparity lies in the age of the houses. Houses near the winner have an average age of 10.15 years, while the average ages of houses near the runner-ups and the remaining houses are 30.10 and 16.49 years, respectively. These differences are expected to have a substantial impact on the subsequent analysis.

4 Research Design

In this study, I aim to examine the increased utility of accessibility to train services on housing prices using the hedonic model. The hedonic model, first proposed by Lancaster (1966), measures the utility of non-market goods by considering them as characteristics of a market good. By applying this model, we can isolate the consumer's utility derived from a specific non-market good while controlling for other characteristics. Housing is a common example of a market good because its utility is not solely derived from the physical structure itself, but also from the services it provides, such as shelter, proximity to parks, and storage capacity (Malpezzi 2002). Thus, in this research, I intend to estimate the expected value of accessibility to a THSR station, using housing prices as the market good of interest.

Within the framework of the hedonic model, I employ a difference-in-differences (DID) design to investigate the effect. I evaluate the expected accessibility to the THSR station in two stages. In Phase 1, I exploit the price difference caused by accessibility to the train after the announcement of candidate locations. In Phase 2, I use the candidate locations as a control group and observe the difference in house prices before and after the announcement of the winning location. By comparing the effects between these two stages, we can explain people's expected utility with varying levels of probability.

To specify the model, I adopt a semi-logarithmic functional form following the approach used by H. X. Bao, Larsson, and V. Wong (2021) and Debrezion, Pels, and Rietveld (2011). According to Malpezzi (2002), the semi-log model offers several advantages, including the ability to interpret coefficients as percentage changes in house prices given a unit change in the corresponding characteristic. This facilitates the comparison of effects across different dollar values. Additionally, the semi-log form ensures that the value gained is proportionate to the magnitude of each quality. For instance, the value gained from an additional bathroom in a one-bathroom house is expected to be higher than in a house that already has three bathrooms.

The general model used in the research is the following:

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

where $ln(P_{i,t})$ is the natural logarithm of the price of house i transacted in month t; α is the constant term; $\mathbf{X}_{i,t}$ is a vector of structural and contractual characteristics of house i. Our main coefficient in interest is β, which is the effect of the expected accessibility to the THSR; and ϵ is the error term.

Different measurements are used for the expected accessibility in the two stages due to their distinct features. In the first part, two metrics are adopted: (a) the driving distance between house i and its nearest candidate location, and (b) a dummy variable indicating whether house i is within a specified distance (e.g., 2 km) of the nearest candidate location. In case (a), it should be noted that a longer distance implies reduced accessibility, thus we expect β to have a negative sign. Conversely, in case (b), we anticipate a positive sign for β as increased accessibility is expected to positively impact the house price.

In the second part, I employ a slightly different approach, treating the losing locations (or ”runner-ups”) as the control group. Consequently, the metrics used in this part differ: (a) the continuous driving distance between house i and the winning location, (b) a dummy variable indicating whether a house is within a specified distance (e.g., 2 km) of the winning location, and (c) dividing the areas surrounding the proposed station locations into five zones based on increasing distance from the centers.

Specifically, for parts (a) and (b), a similar model to equation (1) is used, but the data is restricted to houses located within the specified distance of a candidate location. Thus, in case (a), $EAT_{i,t} = 1$ 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 (b), $EAT_{i,t}$ is measured as the distance between house i and the winning location. For part (c), the model is as follows:

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

where $NearWinner_i$ is a dummy variable indicating whether house i is located in proximity to the finalized station location, ; $Zone_{k,i}$ is a dummy variable that equals 1 if house i is situated in zone k relative to the nearest candidate location. The inclusion of these variables enables us to examine the effects across different zones, represented by the coefficients $\beta_1, \beta_2, ..., \beta_5$. By controlling for the impact of locating in different zones using the variables $Zone_{k,i}$, we can assess the variation in effects. To maintain clarity, the model does not include interaction terms between any pairs of variables, such as $Post$, $Treat$, and $Zone$.

Furthermore, the vector $\mathbf{X}_{i,t}$ encompasses a range of structural and contractual characteristics of house i. These variables, as outlined in Table 1, include but are not limited to total floor area, number of stories, age, number of bedrooms, number of living rooms, number of bathrooms, as well as dummy variables for family transactions and presale houses. These variables provide a comprehensive representation of the housing characteristics that may influence prices.

5 Result

5.1 Phase 1

Table 2 presents the regression results for Phase 1, examining the relationship between expected accessibility and house prices. In Panel A, the distance to the nearest candidate THSR stations is used as a measure of expected accessibility. Model (1) indicates a 1.1 percent decrease in price for every 1 km increase in distance from the nearest candidate station. This effect is significant for models (1) and (2), with similar effect sizes (1.1 and 0.8 percent). However, when structural controls are introduced in the model, the effect becomes almost negligible. In columns (3) and (4), the effect shows an insignificant 0.2 and 0.1 percent decrease for every 1 km increase in distance from the train station.

Table 2: Phase 1 effect on log price of expected accessibility to candidate THSR stations

	(1)	(2)	(3)	(4)
Panel A: Distance				
$Dist \times Post1$	-0.011***	-0.008**	-0.002	-0.001
	(0.004)	(0.004)	(0.003)	(0.003)
Panel B: Within 1km dummy				
$1km \times Post1$	0.204***	0.228***	-0.019	-0.002
	(0.063)	(0.061)	(0.043)	(0.042)
Panel C: Within 2km dummy				
$2km \times Post1$	0.032	0.073	0.011	0.026
	(0.051)	(0.049)	(0.033)	(0.033)
Contractual controls	No	Yes	No	Yes
Structural controls	No	No	Yes	Yes
Observations	9126	9126	8865	8865

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses. The observations considered for analysis are limited to single-family houses in Yilan county that were transacted between January 2018 and November 2021. $Post1 = 1$ if the house is transacted from February 2021 onwards. $Dist$ represents the driving distance from the house to the nearest candidate.

A similar pattern emerges when using a dummy variable to indicate whether a house is located within 1 km of the nearest candidate location. Without the inclusion of structural controls, the results show a significant 20.4 and 22.8 percent increase in price for houses within this proximity. However, when structural controls are added, the effect becomes much smaller. When the 1 km dummy is replaced with a 2 km dummy in Panel C, the effect of expected accessibility is further diminished without structural controls, and in model (3) and (4), the effect even becomes negative. It is worth noting that the standard errors remain similar across models within the same metric of expected accessibility, indicating that the inclusion of controls does not introduce more noise but rather reduces standard errors. [1]

[1] Distances longer than 2 km are not represented by dummies since they cover too large a portion of the sample.

Table 3: Phase 2 effect on log price of distance to winner THSR station

	Full sample				Restricted sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
$Dist \times Post2$	-0.025***	-0.024***	-0.012***	-0.011**	-0.048*	-0.058**	-0.014	-0.018
	(0.007)	(0.007)	(0.005)	(0.004)	(0.026)	(0.025)	(0.017)	(0.017)
Observations	2244	2244	2178	2178	433	433	385	385
Contractual controls	No	Yes	No	Yes	No	Yes	No	Yes
Structural controls	No	No	Yes	Yes	No	No	Yes	Yes

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses. The full sample includes all single-family houses in Yilan county made between December 2021 and December 2022. The restricted sample limits the sample to the houses that are located within 2km of driving distance to the nearest candidate locations. $Dist$ signifies the house's driving distance to the winner location. $Post2 = 1$ if the house is transacted from December 2021 onwards.

The results in Table 2 suggest that the increase in house prices near candidate locations can be attributed to differences in the structural characteristics of the houses. In fact, after the announcement, the transacted houses closer to the candidates tend to be larger and younger, and there exists heterogeneity among the candidates. Further discussion on this topic will be provided in the following section.

5.2 Phase 2

Table 3 provides insights into the effect of expected accessibility to the winner THSR station. In the models using the full sample, a significant effect is observed, both with and without the inclusion of structural controls. In comparison to the effect presented in Panel A of Table 2, the effect size is considerably larger across all four models. For instance, in model (2), the effect indicates a 2.4 percent increase in house price for every kilometer closer to the winner location. When controlling for structural characteristics, the effect diminishes to 1.1 percent, yet it remains larger than the Phase 1 effect.

However, when focusing on a narrower subset of houses located within 2 km of the candidate

Table 4: Phase 2 effect on log price of within 2 km to the winner THSR station

	(1)	(2)	(3)	(4)
$Within2km \times Post2$	0.273	0.325*	-0.054	-0.019
	(0.201)	(0.192)	(0.129)	(0.124)
Observations	433	433	385	385
Contractual controls	No	Yes	No	Yes
Structural controls	No	No	Yes	Yes

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses. The sample includes single-family houses that are located within 2km to the nearest candidate locations, and the transaction of which happen between December 2021 and December 2022. $Within2km$ signifies the house's driving distance to the winner location is less than 2km. $Post2 = 1$ if the house is transacted from December 2021 onwards.

locations, the pattern observed in Phase 1 reemerges. In columns (5) and (6), the effect becomes even stronger, reaching 4.8 percent and 5.8 percent, respectively. Nonetheless, it is important to note that these estimates exhibit large standard errors, and the majority of the effect decreases when controlling for structural characteristics.

To further investigate the effect, I examine the discrete metric of whether a house is located within 2 km of the winner location. [2] This specification focuses on the subsample of houses that are within 2 km of all candidate locations. However, the results obtained from this analysis yield mixed findings. In columns (1) and (2), a substantial increase of 27.3 and 32.5 percent, respectively, in prices for houses located within the winner location is observed, indicating a large effect. However, the presence of large standard errors (0.209 and 0.197) makes it challenging to interpret the effect accurately. When controlling for structural characteristics, the standard errors decrease, but interestingly, the signs of the coefficients become reversed. This suggests that the prices of houses near the winner location actually experience a smaller increase compared to houses near other locations when holding other characteristics constant.

Finally, the zone design is applied to delineate the neighborhood surrounding the candidate sites into five zones based on their distance to the nearest stations. The results are presented in Table 5.

[2] I refrain from using the 1 km dummy variable due to the limited number of observations in the treatment group.

Table 5: Phase 2 zone effect on log price of the houses

	Full sample				Restricted sample			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Zone 1 inter.	0.168	0.088	-0.257	-0.274	0.102	0.050	-0.325	-0.336
	(0.450)	(0.424)	(0.280)	(0.271)	(0.442)	(0.418)	(0.266)	(0.256)
Zone 2 inter.	0.508*	0.573**	-0.042	0.021	0.432	0.521*	-0.007	0.069
	(0.279)	(0.263)	(0.171)	(0.166)	(0.281)	(0.266)	(0.167)	(0.161)
Zone 3 inter.	0.226	0.232	-0.101	-0.077	0.157	0.174	-0.143	-0.116
	(0.304)	(0.287)	(0.184)	(0.178)	(0.303)	(0.287)	(0.176)	(0.169)
Zone 4 inter.	0.030	0.033	-0.188	-0.189	-0.037	-0.027	-0.189	-0.167
	(0.587)	(0.554)	(0.358)	(0.346)	(0.576)	(0.544)	(0.340)	(0.327)
Zone 5 inter.	0.140	0.236	0.123	0.166	0.067	0.212	0.037	0.122
	(0.252)	(0.238)	(0.152)	(0.147)	(0.254)	(0.241)	(0.149)	(0.144)
Observations	2244	2244	2178	2178	639	639	590	590
Contractual controls	No	Yes	No	Yes	No	Yes	No	Yes
Structural controls	No	No	Yes	Yes	No	No	Yes	Yes

Note: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors in parentheses. The full sample includes all single-family houses in Yilan county made between December 2021 and December 2022. The restricted sample limits the sample to the houses that are located within 2.5 km of driving distance to the nearest candidate locations. To define the proximity around each candidate location, the area is divided into zones. Each zone is defined as a disk with a radius of 500 meters, where Zone 1 corresponds to the closest proximity to the candidate location.

It is important to note the limited precision of the estimates indicated by the large standard errors in parentheses, which hampers the interpretability of the results. However, certain patterns can still be observed. Without the inclusion of structural controls, specifically in columns (1), (2), and (5), (6), most zones exhibit positive effects, with the exception of Zone 4 in (5) and (6). Notably, Zone 2 shows the largest effects across these columns, ranging from 43 percent to 57.3 percent. Conversely, when structural controls are added in columns (3), (4), and (7), (8), the effects tend to become predominantly negative. Due to the imprecise nature of these estimates, this model does not provide strong evidence for the observed effects.

Overall, for Phase 2, it can be concluded that houses near the winner location consistently experience a greater increase in value compared to all other houses in Yilan county. However, when the control group is changed to houses located near the runner-up locations, the results become inconclusive. It is important to acknowledge the limited precision of the estimates due to the large standard errors. If we still aim to interpret the results, two main directions can be considered. Firstly, it is possible to speculate that houses located near the runner-up locations might experience a spill-over effect from the THSR station, even if they are not in immediate proximity to the winner. Since the candidate locations also function as train stations, the accessibility to THSR is significantly enhanced for houses in their vicinity. Consequently, the spill-over effect can blur the distinction between the winner and the runner-up locations. Secondly, as discussed in more detail in the subsequent section, characteristic divergence after the announcement could offer an alternative explanation for the observed findings.

6 Discussion

The primary objective of the research design employed in this paper is to analyze the impact of the anticipated accessibility to the THSR station on the price increase. In the ensuing section, an in-depth examination of the discernible patterns in the findings is presented, along with the identification of certain design-related concerns. While some of these concerns are explored in this study, others are earmarked for future investigations.

6.1 Heterogeneity across candidates

As outlined in the data description, notable variations in the structural attributes are observed among different treatment and control groups, as displayed in Table 1. It is plausible that candidates may also differ in unaccounted characteristics, or individuals might possess diverse expectations regarding the selection of these stations as winners. Consequently, heterogeneity arises in the magnitude of value increase subsequent to the initial announcement. Therefore, this subsection focuses on

Table 6: Phase 1 effect on different candidates

	(1)	(2)	(3)	(4)
Winner: County Government				
$Dist \times Post1$	-0.0050*	-0.0031	-0.0010	-0.0002
	(0.0029)	(0.0028)	(0.0020)	(0.0019)
Runner-up 1: Sicheng				
$Dist \times Post1$	-0.0075***	-0.0066***	-0.0023*	-0.0021*
	(0.0020)	(0.0019)	(0.0012)	(0.0012)
Runner-up 2: Yilan City				
$Dist \times Post1$	-0.0078***	-0.0068***	-0.0027**	-0.0024*
	(0.0022)	(0.0021)	(0.0013)	(0.0013)
Runner-up 3: Luodong				
$Dist \times Post1$	0.0086***	0.0096***	0.0062***	0.0067***
	(0.0025)	(0.0024)	(0.0017)	(0.0017)
Contractual controls	No	Yes	No	Yes
Structural controls	No	No	Yes	Yes
Observations	9126	9126	8865	8865

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses. The observations considered for analysis are limited to single-family houses in Yilan county that were transacted between January 2018 and November 2021. $Post1 = 1$ if the house is transacted from February 2021 onwards. $Dist$ represents the driving distance from the house to the candidate specified in the italic title.

examining the different effects in Phase 1.

Table 6 treats the specified candidate as the only treatment group in Phase 1. Let's begin by noting an interesting observation regarding Runner-up 3, Luodong, which exhibits an opposite trend compared to other candidate locations. This suggests that houses located further away from this potential station experience higher prices following the announcement, a pattern that remains consistent irrespective of the inclusion of controls. This finding is not surprising considering that the other three candidates are concentrated in the northern region, while Luodong stands alone in the south. Analyzing the entire sample includes the spillover effects from other candidates. However, when focusing on a subset of the sample within a 2 km radius of Luodong, which should be less influenced by other candidates, consistently positive effects for further houses are still observed, although they are statistically insignificant. A plausible explanation for this phenomenon could be that people have lower expectations regarding the construction of a train station in this particular area.

Furthermore, Table 6 reveals additional insights, primarily highlighting the dilution of effects. All coefficients in this table are smaller than their counterparts in Table 2, which is to be expected as the effects are attenuated by houses located near other candidates. Moreover, differences can be observed across the Winner, Runner-up 1, and Runner-up 2 regarding the impact of controlling for structural characteristics on the coefficients in columns (3) and (4). Although the inclusion of controls diminishes the effect for Runner-up 1 and Runner-up 2, it results in a reversal of direction for the Winner. This discrepancy leads to the suspicion that the changes in structural characteristics are more substantial for houses in proximity to the Winner.

6.2 Structural characteristics

In the results section, it is evident that the impact of expected accessibility to the train diminishes when controlling for structural characteristics, observed in both Phase 1 and Phase 2. Consequently, this subsection is dedicated to comprehending the changes in these structural characteristics.

Figure 2 illustrates the coefficients associated with each structural characteristic employed in the regression models. For further analysis, floor area and age are specifically selected. The choice to focus on floor area stems from its high correlation with other characteristics, such as the number of rooms; age is of particular interest due to the substantial differences observed across treatment and control groups, as indicated in Table 1.

Using a similar DID model as described in Model 1, but with the dependent variables replaced by specific characteristics and without the inclusion of controls, Table 7 presents the results obtained

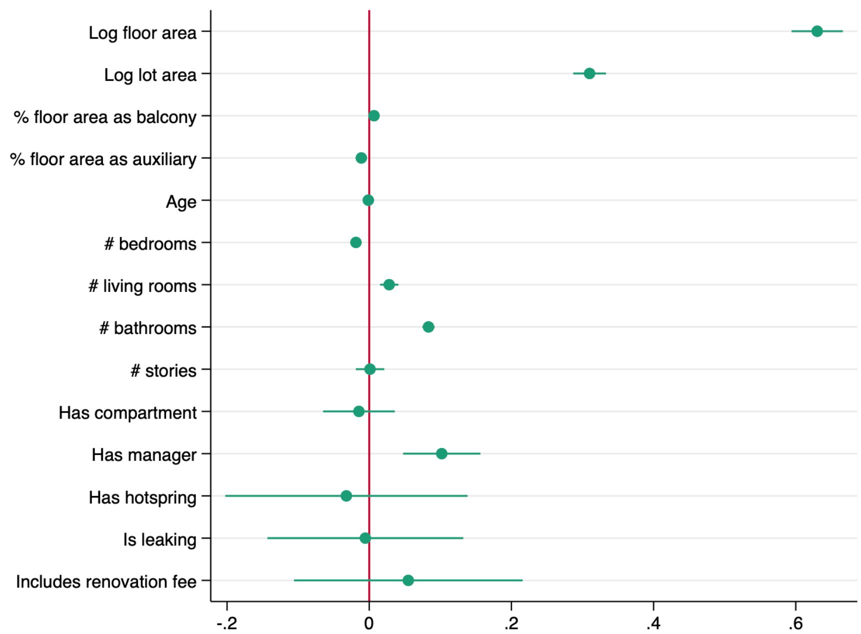

Figure 2: The coefficients for the structural characteristics

Note: The dependent variable is log(Price). The 9126 observations used in the regression include the houses transacted between January 2018 and December 2022.

in Phase 1. The analysis focuses on floor area and age of the houses. Column (1) examines the impact of the announcement by interacting the distance to the nearest candidate and the post-treatment dummy variable. Columns (2) to (5) focus on the distance to specific candidates. Regarding floor area, the results indicate that the announcement is associated with a significant 0.97% increase for every kilometer closer to the nearest candidate site. Upon closer examination, a similar significant effect, albeit smaller due to dilution, is observed for the Winner and Runner-ups 1 and 2. However, estimating the effect using the distance to Runner-up 3 reveals no discernible increase, and if anything, a slight decrease in floor area. Turning to the age of the houses, a ”younger” effect is observed in relation to the announcement. The estimation demonstrates the effect is that for every kilometer closer to the nearest candidate, the houses are approximately 0.42 years younger. Examining individual candidate sites also reveals significant effects in the same direction, with the exception of Runner-up 3. The pattern observed for the age of the houses is similar to that of the floor area.

Table 7: Phase 1 effect on house characteristics, by distance to certain candidate

	(1) Nearest	(2) Winner	(3) Runner-up 1	(4) Runner-up 2	(5) Runner-up 3
Dependent: Log floor area					
$Dist \times Post1$	-0.0097***	-0.0054***	-0.0051***	-0.0056***	0.0008
	(0.0023)	(0.0019)	(0.0014)	(0.0015)	(0.0016)
Dependent: Age					
$Dist \times Post1$	0.416***	0.205**	0.223***	0.247***	0.015
	(0.098)	(0.080)	(0.058)	(0.062)	(0.068)

Note: * p < 0.1, ** p < 0.05, *** p < 0.01. Standard errors in parentheses. The observations considered for analysis are limited to single-family houses in Yilan county that were transacted between January 2018 and November 2021. $Post1 = 1$ if the house is transacted from February 2021 onwards. $Dist$ represents the driving distance from the house to the candidate specified in the column title.

Based on the results presented in Table 7, it can be concluded that the announcement of the train station brings about changes in the characteristics of houses involved in transactions. This evidence is further supported by the observation that a higher proportion of new houses were sold in proximity to the candidate sites following the announcement. Specifically, transactions involving newly completed houses (with an age of 0) account for 24% of the total sample, and newer houses tend to have larger floor areas. In fact, the data used in this study indicates that a decrease of 1 year in age is associated with a 1.38% increase in floor area (with a standard error of 0.02%). Additionally, applying the DID design in probit form reveals that the announcement is associated with a 1.18% increase in the probability of a house transaction being classified as a new house for every 1 km closer to the candidate site.

The observations made in this subsection serve as a reminder of the limitations of the hedonic model. While the model considers a house's value as a combination of its provided characteristics, it relies on transaction prices to capture this value. However, it should be noted that prices are also influenced by market supply factors. For instance, as housing demand increases in a particular area, prices rise, leading construction companies to allocate more resources to that area, resulting in early construction completion and intensified marketing efforts. In this subsection, noticeable changes in the houses on the market are observed, some of which are likely influenced by changes in housing supply. Therefore, to accurately isolate the impact of expected transportation accessibility on individuals' valuation of houses, further research is necessary.

6.3 Other limitations

Due to time and resource constraints, this study was unable to address all the pertinent issues that arise during the research process. However, these unresolved matters hold significance and merit further investigation in future studies. Some of these notable issues include:

1. *Capture of only partial expected accessibility effect:* The research design of this study focuses on capturing the effect of the announcement of candidate locations on house prices and characteristics. However, it is important to recognize that the announcement itself represents only a portion of the overall expected accessibility effect. Prior to the announcement, people in Yilan were already anticipating increased accessibility, and areas designated as candidate locations were likely perceived to have a higher probability of being selected for the train station. Since the models employed in this research can only measure the effect following the announcement, a significant portion of the expected accessibility effect may not be fully accounted for. Consequently, the estimates provided in this study serve as a lower bound, representing only a partial capture of the total expected effect.

2. *Change of house types on market:* It is important to acknowledge that this study exclusively focuses on single-family houses in its analysis, omitting other types of residential properties. This narrow scope may overlook potential changes in housing preferences that could arise after the announcement. For instance, it is plausible that individuals may develop a greater preference for apartments, as they tend to be closer to the train station. If such a preference shift towards apartments were to occur, the estimates presented in this research would underestimate the overall effect. While the available dataset does have the potential to address

this question, constraints in terms of time and study duration prevented its inclusion in the current analysis.

3. *Unobserved characteristics*: According to Malpezzi (2002), it is recommended to include a comprehensive set of control variables encompassing structural, locational, neighborhood, and contractual characteristics in hedonic models. In this study, the controls primarily focused on structural and contractual characteristics, while other relevant factors such as proximity to hospitals, schools, and parks, were not fully accounted for. Incorporating these variables can provide valuable insights. Furthermore, the potential distortion of the Phase 2 effect by other modes of transportation should be explored, given the presence of train stations near all candidate locations. Exploring the impact of alternative transportation options on the observed effect would be a valuable avenue for future investigation. The available data for this study possess the potential for a complete examination of these aspects, presenting an opportunity in future research.

7 Conclusion

This study examines the impact of increased accessibility to transportation on the prices of single family houses by focusing on the newly announced THSR station in Taiwan. By utilizing the Actual Price Registration Dataset of Taiwan and employing DID design within the hedonic framework, this research provides valuable insights into the expected effect on real estate prices and the broader understanding of the value of transportation accessibility.

The analysis is conducted in two phases, where the first phase involves the announcement of four potential station candidates and the second one selection of the winning location. In Phase 1, following the announcement of the four candidate locations, the prices of single-family houses near these areas experienced greater increases compared to houses located further away. However, the magnitude of this effect diminishes when controlling for the structural characteristics of the houses. In Phase 2, the study identifies even larger effects of the anticipated increase in accessibility, although these effects are less consistent when holding the structural characteristics constant.

Moreover, this research finds heterogeneity in the effects of the initial announcements across the different station candidates. Furthermore, upon closer examination, it is observed that the announcements led to an increased prevalence of larger and younger houses in proximity to the stations. These findings suggest the presence of additional factors and complexities that warrant further investigation. Future research endeavors could incorporate neighborhood characteristics as control variables, focus exclusively on pre-owned houses, and compare different house types over time.

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 (2010). "Does high-speed rail accessibility influence residential property prices? Hedonic estimates from southern Taiwan". In: *Journal of Transport Geography* 18.1, pp. 166–174. DOI: 10.1016/j.jtrangeo.2008.10.012.

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.

Cengiz, Elif Can, Enver Cenan İnce, and Hüseyin Murat Çelik (2022). "Impact of urban railway announcements on housing market: Preponderance of pre-construction and post-construction phase effect of urban railway projects". In: *Case Studies on Transport Policy* 10.2, pp. 1090–1101. DOI: 10.1016/j.cstp.2022.03.021.

Debrezion, Ghebreegziabiher, Eric Pels, and Piet Rietveld (2011). "The Impact of Rail Transport on Real Estate Prices: An Empirical Analysis of the Dutch Housing Market". In: *Urban Studies* 48.5, pp. 997–1015. DOI: 10.1177/0042098010371395.

Geng, Bin, Haijun Bao, and Ying Liang (2015). "A study of the effect of a high-speed rail station on spatial variations in housing price based on the hedonic model". In: *Habitat International* 49, pp. 333–339. DOI: 10.1016/j.habitatint.2015.06.005.

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.

Levkovich, Or, Jan Rouwendal, and Ramona van Marwijk (2016). "The effects of highway development on housing prices". In: *Transportation* 43.2, pp. 379–405. DOI: 10.1007/s11116-015-9580-7.

Malpezzi, Stephen (2002). "Hedonic Pricing Models: A Selective and Applied Review". In: *Housing Economics and Public Policy*. Oxford, UK: Blackwell Science Ltd, pp. 67–89. ISBN: 978-0-470-69068-0 978-0-632-06461-8. DOI: 10.1002/9780470690680.ch5.

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. (2013). "A meta-analysis of the impact of rail projects on land and property values". In: *Transportation Research Part A: Policy and Practice* 50, pp. 158–170. DOI: 10.1016/j.tra.2013.01.013.

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, pp. 113–125. DOI: 10.1080/0042098042000309720.