Estimating the Gains from New Rail Transit Investment: A Machine Learning Tree Approach *

Seungwoo Chin†
University of Southern California

Matthew E. Kahn ‡
University of Southern California

Hyungsik Roger Moon §
University of Southern California and Yonsei University

September 27, 2018

*Kahn and Moon thank the USC Lusk Center for Real Estate for financial support.

†Department of Economics, University of Southern California. email: chinseun@usc.edu

‡Department of Economics, University of Southern California. email: kahnme@usc.edu

§Department of Economics, Dornsife INET, Lusk Center of Real Estate, University of Southern California and the School of Economics, Yonsei University. email: moonr@usc.edu

Abstract

Urban rail transit investments are expensive and irreversible. Since people differ with respect to their demand for trips, their value of time, and the types of real estate they live in, such projects are likely to offer heterogeneous benefits to residents of a city. Defining the opening of a major new subway in Seoul as a treatment for apartments close to the new rail stations, we contrast hedonic estimates based on multivariate hedonic methods with a machine learning approach. This ML approach yields new estimates of these heterogeneous effects. While a majority of the "treated" apartment types appreciate in value, other types decline in value. We cross-validate our estimates by studying what types of new housing units developers build in the treated areas close to the new train lines.

JEL Classification Codes: O18, R21, C10

Keywords: real estate pricing, transit investment, Seoul, Supervised Machine Learning

1 Introduction

Major urban rail transportation projects increase travel speeds in cities and thus facilitate shorter commutes, labor market matching and consumer shopping and leisure opportunities. Housing located close to new transport nodes often increases in value as the demand to live close to fast public transit increases local demand to live there (McMillen and McDonald (2004), Blake (2016), Pivo and Fisher (2011)). Real estate developers will seek to build new housing units close to these new stations. In recent decades, Asia's major cities have made major investments in new subways (Gonzalez-Navarro and Turner (2016)). Cities ranging from Beijing, to Shanghai to Singapore have invested billions in new subways. In this paper, we study how the real estate market in Seoul has been affected by the construction of a major new subway. The line number 9 (hereafter LINE9) subway connects the Southern part of the city with the Gangnam District. This is one of the richest parts of the city.

The opening of new transit stations may increase nearby property values by both lowering commute times between the origin and destination and by potentially stimulating new "consumer city" amenities near the new transit stations (Fejarang (1993), Baum-Snow and Kahn (2000)), and Hess and Almeida (2007)). On the other hand, Bowes and Ihlanfeldt (2001) suggests that public transit access can have negative local effects if it increases nearby criminal activity. Based on a meta-analysis of 57 studies, Debrezion et al. (2007) concludes that every 250 meters closer to a station is associated with a 2.4% increase in property values. Cervero and Duncan (2002) finds that the effects vary with the specific transit technology.[1] Duncan (2008) shows that based on data from San Diego that the transit premium for multifamily housing is three times larger than that

[1] Based on data from San Jose, California, they find that the capitalization effect of access to a regional commuter rail system is much larger than having access to a city-wide light rail system.

for single-family housing. These studies highlight the importance of allowing for treatment heterogeneity across several dimensions of apartment and community attributes. Recently, Li (2016) finds that consumers value transit accessibility differently, depending on whether or not they live in congested area. It also documents consumers who are wealthy and own fewer cars show higher willingness to pay for rail transit.

Our methodological approach builds on past hedonic studies that use panel estimation strategies to recover estimates of the causal effects of new transit access (Kahn (2007), Billings (2011), Zheng and Kahn (2013), Gibbons and Machin (2005)). A distinguishing feature of our study is using machine learning to pare down the possible non-linearities in the hedonic pricing function. The typical hedonic analysis includes a large vector of housing attribute control variables. It would be computationally cumbersome to include a full set of interactions between each of these control variables and the treatment indicator.

Our solution to this challenge is to use the regression tree approach from machine learning (ML) (e.g., Breiman et al. (1984) and Friedman et al. (2001)). Past ML research has focused on predicting outcome variables using high-dimensional explanatory variables. A recent literature has used ML methods to estimate causal effects (Zeileis et al. (2008), Beygelzimer and Langford (2009), Su et al. (2009), Foster et al. (2010), Dudík et al. (2011), Imai et al. (2013), Athey and Imbens (2015), and Taddy et al. (2016)). Building on Athey and Imbens (2015), we apply ML methods in using a difference-in-difference approach to estimate conditional average treatment effects. This estimation strategy imposes few computational burdens. In our tree approach, we create dummy variables indicating whether the treatment has occurred or not and whether the housing unit is in the treatment area (i.e. close to the new transit stations). The ML algorithm splits the sample on these attributes as well as on the physical attributes of the housing unit. This approach allows us to test how housing price appreciation differs for treated units versus control units while allowing these effects to vary by housing unit and community attributes.

Based on our ML approach, we document that there is considerable variation in the conditional average treatment effect (CATE) across the apartment types. Some apartments types experience greater price appreciation than others. For example, one "winner" from the treatment is an apartment in the upper 25% of the apartment size distribution featuring 3 rooms, 2 baths that is less than five years old and is located within one kilometer of old transit in the Seocho county. We discuss the methodological advantages of our approach and also report estimates of the hedonic capitalization effect using a log-linear hedonic model (Redfearn (2009)).

As a validation test of our estimates, we study whether developers of new apartments are building units with the features that our ML estimates yield the highest marginal revenue. We document a positive correlation between our estimates of the real estate price appreciation gains from train network proximity and the specific type of new housing built by a developer. These findings support our claim that we have recovered key non-linearities of the true underlying pricing gradient and how they change over time.

2 Seoul's New Subway Construction and Project Financing

Seoul's first subway line was built in 1974. Over the last four decades, the subway system expanded to cover five lines. Since 2000, Seoul's government has built three additional subway lines(lines 6, 7 and 8). The last subway expansion plan is the introduction of LINE9. LINE9 was first designed in 1997. The detailed blueprint was released in 2000, and the ground-breaking construction ceremony took place in 2002. It began its service on July 24th 2009. As of 2014, 39% of trips in Seoul use subways or railways. Figures 1 and 2 display the early subway network and the LINE9. Figure 3 provides a chronology of the construction of this subway.

The total cost of this project was US$818 million[2]. 46.7% percent of the total costs were subsidized by the Seoul metropolitan government, and the METRO9, a private company, covered the rest. The METRO9 operates LINE9 for three decades without paying any rental fees while the Seoul metropolitan government owns it. The Seoul metropolitan government guarantees a minimum profit level for the first fifteen years of the project.[3] In 2005, the Korea Transport Institute predicted that 243,196 riders per a day in 2014 would use LINE9. However, the actual ridership has been 384,423 riders a day. This prediction stands in contrast to the U.S literature that argues that transit agencies routinely over-state the ridership of a new subway before it is built (Kain (1990)). Such strategic predictions increase the likelihood that the project is funded.

2.1 The Demand for Housing Close to Transit

Seoul's residents rely on public transit. In 2014, cars accounted for 22.8% and buses accounted for 27%, while subway and light rail take 39%. The share of trips by taxi is 6.8%. Standard network logic suggests that the value of subway access increases in the set of potential destinations one can reach in a short time. A fast train that connects to a desirable city sub-center should lead to gentrification along its nodes, and transit improvements (McMillen and McDonald (1998), Glaeser and Kahn (2001) and Baum-Snow et al. (2005)). If such a train is fast enough then it could reduce the demand to live very close to the destination because people can decentralize while still having access to the destination area. Glaeser et al. (2008) documents that poor people live close to slow

[2]The construction costs are based on an exchange rate of 1100won/ US$1.

[3]The Seoul metropolitan government promised this private company 90 percent of the expected profits for the first five years, 80 percent for the next five years, and 70 percent for the last five years.

public transit, while rich people are attracted to fast public transit in centralized cities such as Boston and New York City.

LINE9 significantly reduces travel times within Seoul. To document this fact, we calculate the travel time between each apartment unit to twenty major destinations before and after it is built, based on the average train's speed of 50 kilometers per hour and a walking speed of 4 kilometers per hour. Table 1 presents the one way reduction in travel time (measured in hours) to 20 major destinations in Seoul.[4] For example, across our sample, the average apartment resident experienced a reduction in travel time to Gangnam by roughly 5 minutes each way. Those living within a 1 kilometer radius of new transit enjoyed a 14.4 minute reduction in one way travel time to Gangnam. This is a 35% reduction.

While the new train reduces commute times, we do not believe that its effects are large enough to cause important general equilibrium shifts in the entire Seoul housing market. Starting with the work of Sieg et al. (2004) there has been a growing recognition that local public goods improvements can have general equilibrium effects on a given city. They studied how Clean Air Act regulations sharply reduced pollution in major sections of Los Angeles and this caused a reshuffling of the population such that richer people moved to previously poorer polluted areas of the city. A hedonic researcher who ignores this migration effect would likely over-state the role of clean air improvements as the sole cause of real estate price appreciation. In our setting, we believe that such GE effects are a second order concern. As we discuss in the next section, the new train's treatment area is only a small physical area of Seoul.

[4] The estimated travel time is based on the assumption that travelers walk to the closest subway station and take a subway.

2.2 The Supply of Housing Close to the New Transit Stations

When LINE9 was completed, there were 322 residential apartment complexes within a kilometer of the new transit stations. Since LINE9 opened the owners of these properties began enjoying a capitalization effect that we will estimate below. As we will document in section 5.1, we do not find evidence of a capitalization effect caused by the announcement of the LINE9 construction.

A key assumption in our difference-in-difference approach is identifying the treatment date. The plan for LINE9 was announced in the year 2000 but it was only completed years later. In section 5.1, we test and reject the hypothesis that the subway construction plant had an ex-ante capitalization effect.

Once the new subway opens, nearby land becomes more valuable in these "treated" areas. Thus, real estate developers have incentives to upgrade existing structures and to build new structures. But, Seoul features stringent construction regulation. It takes an average of 33.3 months to build a new apartment complex (Jeon et al, 2010). Redeveloping existing housing entails overcoming many regulatory obstacles. Each urban housing redevelopment project proposal in Seoul undergoes a nine stage process that includes a strict safety investigation. For the typical redevelopment project completed between the years 2000 and 2015, it took an average of 8.7 years to complete the reconstruction process. Seoul's regulations also require developers to supply a certain proportion of small apartment types. A U.S literature has studied how regulations limits housing supply (see Glaeser et al. (2005)).

While developers face many restrictions in building, they will have a greater incentive to do so if the marginal revenue from building an apartment is higher. The total revenue a developer collects from producing apartments of certain type in a given location is the price per unit multiplied by the units sold. Our ML estimates provide an estimate of the former. If each developer is a price taker, then facing the non-linear hedonic pricing function (their revenue curve) they will have an incentive to supply new housing that offers greater revenue. Below, we will use data on the new housing supply by developers combined with our CATE estimates to study this.

3 The Empirical Approach

3.1 The OLS Model

Following Kahn (2007), Billings (2011), Zheng and Kahn (2013) and Gibbons and Machin (2005), we begin by estimating average treatment effects of the LINE9 on apartment prices using a difference-in-difference estimator. For the apartment type i in district j at time t, its price is expressed as follows:

$$Log(Price_{ijt}) = \beta_1 Line9_i + \beta_2 Line9_i \times AFTER_t + \beta_3 X_{ijt} + \mu_j + \lambda_t + \varepsilon_{ijt}, \tag{1}$$

where $Line9_i$ is distance between the apartment type i and the closest LINE9 station and $AFTER_t$ takes one if time period is after the LINE9 opened, and 0 otherwise. X_{ijt} is a set of the apartment characteristics except proximity to the LINE9, and μ_j and λ_t are the district fixed effect and the quarter fixed effect, respectively. ε_{ijt} is the error term.

To specify the treatment area we split the area with G groups, $\mathcal{G} = \{1, ..., G\}$, based on the distance between the LINE9 station and the apartment. For apartment i in district j at time t, its price is expressed as follows:

$$Log(Price_{ijt}) = \sum_{g=1}^{G} \alpha_g \mathbb{I}\{Group_i = g\} + \sum_{g=1}^{G} \beta_g \mathbb{I}\{Group_i = g\} \times AFTER_t$$
$$+ \gamma X_{ijt} + \mu_j + \lambda_t + \varepsilon_{ijt}, \quad (2)$$

where $Group_i$ is the group dummy of apartment i that takes a value in $\mathcal{G} = \{1, ..., G\}$. In the empirical analysis in Section 5, we consider three groups ($G = 3$). The first group includes the apartments within 1km of transit stations. The second group includes between 1km and 2km from a transit station, and the third group includes all other apartments. For both econometric specifications, we allow serial correlation in ε_{ijt} within the district of "Dong".

For OLS to yield consistent estimates of the average treatment effects, the unobserved error term $(\epsilon_{ij1}, ..., \epsilon_{ijT})$ must be uncorrelated with the "treatment" variables even after we control for the observed apartment characteristics and the two sets of fixed effects. If β_2 measures the "average" price effect associated with the treatment, the creation of such new subway stations may have several effects on apartments nearby. First, it reduces travel times to destinations. Second, once the new stations open, this may trigger the opening of new restaurants and stores close to the new stations. In this case, the LINE9 causes both a reduction of commute times, say, to Gangnam and an improvement in the quality and quantity of local restaurants and stores. We will return to this point below.

We recognize that home prices reflect future expectations of local amenity changes. If home buyers anticipate that the new train would raise future rents, then they may bid more aggressively for houses before the LINE9 opens. In section 5.1 below, we will study trends in real estate prices in the treatment areas and the control areas before the actual opening of the line. We find that there is little evidence of an anticipation effect in the treated areas.[5]

3.2 The Machine Learning Approach

Our Machine Learning approach allows us to disaggregate the average treatment effect associated with new transit access along a high dimensional set of observed attributes. For those interested in the economic incidence of public policies, this ML approach provides more precise estimates of which incumbent apartment owners are the biggest winners from the city's public goods investment.

We follow the general supervised machine learning approach to grow our regression tree. The detailed algorithm is described in the appendix. We define apartments within one kilometer of a LINE9 station as the treatment group, whereas those more than one kilometer away from it are considered to be in the control group. We have chosen this definition based on the OLS results presented in table 3. This definition is consistent with the earlier literature (Bartholomew and Ewing (2011)). Another issue to address is a double-treatment effect. As the average of direct distance between the LINE9 stations is 1.1km, some apartments have two new stations within 1km. Using our methods, we cannot address this double treatment effect. In our analysis, we treats these data points as in the treatment group but do not attempt to estimate a separate treatment effect for this subset.

According to Athey and Imbens (2015), our approach represents a single tree model because

[5]Some papers provide empirical evidence that housing prices rise in advance of when new transit lines open (McDonald and Osuji (1995), Knaap et al. (2001), McMillen and McDonald (2004)). However, Gibbons and Machin (2008) argue that the impact of transport improvements are heavily dependent on the economic context. If housing represents a consumption good rather than as an investment asset, then the anticipated capitalization effect can be small.

the treatment dummy, the time dummy and all covariates are included in the single tree. We can extend our approach to the two tree model or the four tree model, based on how splitting variables are included when the tree grows. If the post-treatment effects and the pre-treatment effects are estimated separately from two different trees with sub-sample of $T = 1$ and $T = 0$, respectively, then it is referred to as the two tree model. In implementing our machine learning approach, the main assumption we are making is that both the treated and controls would follow a similar trajectory in the absence of the intervention. We discuss pre-trends below to address the issue.

We recognize that the LINE9's geographic placement was not randomly determined. Thus, we are conducting a conditional analysis. Given the new transit stations that were built how has it affected real estate pricing? This approach is relevant for an ex-post evaluation of what apartment types have gained from this public investment. Our approach cannot be used to predict what will be the future impact of a new Seoul subway built in another location.

4 Data

In this section, we describe the data and the summary statistics.

4.1 Apartment Data

We use apartment price data provided by the Ministry of Strategy and Finance of South Korea. This covers more than 90% of entire apartments in South Korea since 2000 and contains a rich set of apartment characteristics, including size, the number of rooms and bath and parking spaces. Since our goal is to investigate the effects of LINE9 on apartment prices, we restrict our sample to the locations where LINE9 passes. In figures 1 and 2, the light green area represents our districts of interest. These data restrictions result in our sample that includes 1,102 apartment complexes and 4,161 apartment types. Market prices are surveyed based on an apartment type rather than at the apartment unit. This means that the price data represents the average price of all apartment types that share the same characteristics within the same complex. For example, 84 square meter apartment units with two beds and a bath within the same complex are considered to be the same product and thus have an identical price in our data.[6] Our data has a panel structure such that the price for each apartment type has been surveyed on a weekly basis. We use the quarterly average price for our analysis. As shown in Table 1, the average apartment has three beds and 1.7 baths, and it is 8.2 years old.

4.2 Geographic Data

Geographical data are obtained from the Seoul Metropolitan Government. It provides administrative borders, locations of bus stops and hospital, and all subway systems including LINE9. Figure 2 shows the locations of LINE9 subway stations. Using ArcGIS, we measure the distance between center of each apartment complex and the closest LINE9 station. This is a key variable in our analysis. Our control group consists of apartments more than one kilometer away from the new transit.[7]

As shown in table 2, the mean distance between an apartment and the closest LINE9 station is 1.92 kilometers, and each apartment has other subway station, excluding LINE9, within 0.7

[6] We are interested in the value of a certain apartment type, not individual apartments. An apartment complex has a limited number of apartment types. Under a certain apartment type, there are many homogeneous units. For example, an apartment complex has 600 units, but they can be categorized into four different types, meaning that each type has 150 units on average. We cannot observe transaction prices of individual units but we observe an appraisal of the type. The appraisal process uses each property's arms-length transaction sale price and then an averaging takes place. Though our price data has some measurement error due to this process, this is classical error.

[7] We also construct one kilometer buffers for each apartment to count the number of bus stops and hospitals within one kilometer. These are utilized as regression controls.

kilometers on average. Our sample of apartments consists of those in districts where the LINE9 passes through.

Roughly 42.4% of Seoul residents lived in apartments in 2014. Apartments in Seoul are organized into complexes. The "complex" is composed of several apartment buildings. In our sample, each "complex" features an average of 5.5 apartment buildings. Each apartment building contains many apartment units where households reside in. All "apartment units" can be classified into a few number of "apartment type" that share the same apartment characteristics e.g. size, the number of rooms and baths to name but a few. Within the same "apartment type", it is reasonable to assume that apartment units are homogenous. Table 2 shows that each complex has 435 "apartment units" in our sample, meaning that each "apartment building" contains more than 80 "apartment units'".

The Seoul housing stock is quite young. The average age of a Seoul apartment in our data is eight years. It is worth noting that the supply of apartments in Seoul has been expanding since the mid 1970s. The city's growing population and rapid economic development since the 1970s catalyzed the need for high-density residential structures. The development plan for Gangnam caused a massive apartment supply increase in the 1980s. Our data shows that the oldest apartment building is 46 years old, but a large fraction of apartments were built between 1997 and 2007.

5 Results

5.1 The Pre-Treatment Trend

In conducting a difference in difference study it is important to demonstrate that the pre-trends for the treatment and control groups are not statistically different. In Figure 4, we define the treatment group as the set of apartments within 1 kilometer of LINE9 transit station and the control group is the set of those apartments located more than 2 kilometers away. In figure 5, the treatment group we further refine this set to represent the apartments located within 1/2 of a kilometer of the closest LIEN9 station while the controls are the same as in figure 4. These figures are based on apartments located in districts where LINE9 passes through. Both figure 4 and figure 5 show that the pre-trends are parallel, which implies that both the treated and the controls would have followed the similar path in the absence of the intervention. Figure 5 shows that the gap between the treated and controls has been decreasing significantly after LINE9 opened.

Each bar in figure 6 represents the coefficient for interaction between the dummy of within 1 kilometer and the estimated year dummy along with a 90 percent confidence interval.[8] This shows when LINE9 started affecting apartment prices, and the new transit station effects becomes statistically significant at 10% after 2009 when LINE9 opened. This implies that the treated and the untreated had experienced a similar path prior to LINE9, meaning that they would have followed the same trajectory in the absence of LINE9 even though detailed blueprint was announced at the early period. Our pre-trend analysis indicates that the difference in difference approach is suitable for estimating the impact of LINE9 on apartment prices.

5.2 OLS Results

Our first set of results builds on earlier work studying the consequences of Seoul's investment in transit infrastructure (see Kim et al. (2005), Cervero and Kang (2011), Bae et al. (2003), Agostini

[8] We estimate $Log(Price_{ijt}) = \beta_1 Within1km_i + \sum_{Y=2000}^{2015} \beta_Y Within1km_i \times YEAR_Y + \beta_3 X_{ijt} + \mu_j + \lambda_t + \varepsilon_{ijt}$ where $Within1km_i$ takes one if an apartment type i has the new station within 1km, and zero otherwise. $YEAR_Y$ is a year dummy and X_{ijt} is a set of apartment type i's characteristics in district j at time t. μ_j and λ_t are a district fixed effect and a quarter fixed effect, respectively.

and Palmucci (2008) and Ahlfeldt (2013)).

Table 3 reports results from a linear hedonic pricing regression. Controlling for structure attributes, the double difference approach indicates that an extra kilometer reduction in the distance from the LINE9 station is associated with a 1.7% increase in the home's price. We further explore these results by including distance to transit dummies. All else equal, properties within 1 kilometer of the transit experience a 4% price appreciation compared to those more than two kilometers away from transit (see column 3). Column 4 documents that there is heterogeneity in the treatment effects across apartments' size and age. The larger an apartment is, the smaller the price premium is. Old apartments benefited less from the LINE9 than new ones did.

To test whether or not an anticipation effect confound our estimates, we present rent regressions using the same parametric specification. These results are reported in the appendix. If there is an anticipation effect or if speculative demand played a major role in determining the LINE9's price appreciation, then housing prices would have experienced a larger increase than rents. Equilibrium rents are less subject to an anticipation effect. However, the rent results are qualitatively either quite similar to or larger than our results from the house price regressions. This suggests that the anticipation effect is limited.

5.3 The Machine Learning Results

The ML approach yields 142 estimates of the treatment effect.[9] In Figure 7, we present a histogram of these estimates. With standard errors calculated by the bootstrap, all estimates are reported in

[9] We present the regression tree constructed by our data in appendix. First, the entire sample is split into two subsamples based on its size. Then one sample is divided by its size again, while the other sample is split by whether or not it is located in Gangnam. Our algorithm keeps splitting the samples based on apartment attributes until it reaches the terminal condition described above.

appendix. Table 4 shows the top-tier or the low-tier leafs. For example, 17.8% of the apartment types in the top 10% leafs are smaller than the 25% quartile, while 21.3% are larger than the 75% quartile. Based on our estimates, we find that 89 are positive and 53 are negative. One way to interpret the results is that the demand for a certain apartment type is low. One explanation for the negative treatment effects is suburbanization. Negative leafs are associated with many apartments in Gangnam (CBD) which suggests that some residents are willing to pay less for this desirable location when they can more easily access it from other locations. The new fast subway encourages center city residents to decentralize (Baum-Snow (2007)). Negative externalities caused by, in part, increased noise (Diao et al. (2016)) or higher crime (Bowes and Ihlanfeldt (2001)) might be another mechanism. As shown in table 5, most apartment types featuring a low CATE (the Bottom 10% or Bottom 25%) have other subway transit stations within one kilometer. This finding suggests that there are diminishing marginal returns to access to the quantity of transit lines. A transit hubbing hypothesis would posit that such multi-nodal points would be even more valuable.

Comparing column 4 in table 3 with table 4, we document the differences in the estimated treatment effects between the OLS model and the ML model. Table 3 implies that small apartment benefit more from LINE9, but table 4 shows that the low CATE groups are characterized by small apartments. In addition, the linear hedonic model's estimates of the treatment effect do not statistically vary across distances to other existing subway station. Our ML model shows most apartments in the low CATE group have another transit station within one kilometer. The OLS model indicates new apartments experience greater price premium, but a distinct linear association between apartment age and treatment effect is not observed in the ML model.

5.4 Developer Responses to the Shifting Real Estate Price Gradient

If we have recovered the true underlying gradient, then the non-linear pricing function sketches out the developer's revenue function for producing different new housing units. Assuming smooth cost functions with respect to apartment size, if developers can earn a large marginal revenue for bundling certain features then they have a profit incentive to build these in. To be specific, developer i supplies apartment type j to maximize the following profit function.

$$\begin{aligned} \max_{j \in J} \Pi_{ij} &= \pi_{ij} + \epsilon_{ij} \\ \pi_{ij} &= R_{ij} - C_{ij}(L, P, B) \\ B &= B(size) \end{aligned} \tag{3}$$

where R_{ij} represents revenue of apartment type j and C_{ij} is a cost function of constructing apartment type j. L indicates required amounts of land, and P is costs related to attaining permits. B denotes building costs that hinge mainly on size, and ϵ_{ij} is a random component. The probability that developer i supplies apartment type j is

$$Pr\{Y_i = j\} = Pr\{max(\Pi_{i1}, \ldots, \Pi_{iJ}) = \Pi_{ij}\} \tag{4}$$

where Y_i indicates an apartment type chosen by developer i. If ϵ_{ij} is independent and identically distributed with Gumbel (type 1 extreme value) distributions[10], then the probability that type j is

[10]Its cumulative density function is $F(\epsilon_{ij}) = exp(-exp(-\epsilon_{ij}))$.

chosen by developer i is as follows (McFadden (1984)).

$$Pr\{Y_i = j\} = \frac{exp(\pi_{ij})}{\sum_{j=1}^{J} exp(\pi_{ij})} \tag{5}$$

We do not have any cost of construction data, but our ML estimates provide information on the shape of the revenue function. We test whether new construction is positively correlates with our estimates of the revenue function. Potential buyers are more willing to pay for more attractive apartment types. Our CATE estimates provide a proxy for the revenue a developer will receive from selling a given type of apartment. This suggests that the flow of new construction's attributes should be positively correlated with our CATE estimates.

We study this by constructing three histograms in figure 8. These histograms are based on properties built at three points in time; before 2002, between 2002 and 2009, and after 2009. These three stages can be thought of as the before, middle period and after the construction of LINE9. The histograms display the share of all housing units as a function of each of their respective CATEs. We find that units built before 2002 (when developers at that time were unaware of what the future treatment effects of LINE9 would be) build housing units that are symmetrically distributed around zero. In the post-period, the new units built feature positive CATE estimates. We interpret this as evidence that developers are focusing their efforts on constructing what the market signals is scarce and valued.

In addition, we estimate $CATE_i = \beta_0 + \beta_1 Age_i + \beta_2 Age_i^2 + \beta_3 X_i + \varepsilon_i$. Age_i represents how old an apartment is in years, and X_i is a vector of apartment characteristics. In this analysis, apartments within one kilometer from the LINE9 transit stations are only included. Table 5 shows that the CATE and apartment's age are negatively associated, and that the relationship is not sensitive to model specification. This implies that real estate developers respond to the heterogeneous positive demand shock and began providing the apartments offering the greatest return given the non-linear hedonic pricing gradient.

5.5 Testing Two Explanations for the Price Appreciation Effects

In this section, we explore two potential reasons for why transit access is associated with rising real estate prices. One explanation is reduced travel time to popular destinations and the other is that new "consumer city" retail and restaurants co-agglomerate near the new train stations.[11] We test each of these by augmenting our linear hedonic regression to include additional explanatory variables and then we test if the capitalization of transit effect changes.

In table 6, we report results where we return to the parametric hedonic specification reported in equation (1). Across the eight regressions reported in table 6, we include different combinations of extra control variables to test if the treatment effect shrinks as we control for these variables. We use the distance between each apartment and the LINE9 station in column (1) to (4), and we use the dummy variable that indicates whether an apartment locates within a kilometer from the LINE9 station. We add travel times to twenty key destinations in column (2) and (6). We include a measure of the count of new restaurants and the count of retail establishments co-agglomerated near the new transit lines in columns (3) and (7).[12] All of the controls are included in columns (4) and (8). The first four columns show that the treatment effect does not shrink much as the travel time vector and the restaurant and retail store count variables are included separately. The

[11]Transportation innovation causes, in some cases, local amenity development that generates additional price premiums (Tu and Eppli (1999), Dronyk-Trosper (2017), Billings (2015). We cannot test whether this happens in Seoul due to data limitation.

[12]We use the count of restaurants and retail establishments and the number of employees in those industries at the "Dong" level. "Dong" is the smallest administrative level. The data is drawn from the Seoul metropolitan government

results reported in column (4) indicate that the treatment effect shrinks around 20% and becomes no longer statistically different from zero at 10%. This implies that the travel time saving and the creation of a local "consumer city" are the main reasons for the capitalization effect. Columns (5) to (8) provide additional evidence on the role that reductions in travel times play in determining the treatment effect. Apartments within a kilometer of the LINE9 station experience a price premium of 3.36%. Controlling for travel times, the treatment effect is not statistically different at the 10% significance level. The treatment effect is statistically significant when we control for the count of retail and restaurants. The findings suggest that the price appreciation is mainly caused by travel time savings rather than by the "consumer city" effect.

We also take our panel CATE estimates from the ML procedure and we compare these leaf specific estimates to those obtained when we conduct a "long difference" ML estimation. In this second case, we only keep the data for the first year and the last year of our sample and we rerun the ML estimator. In Figure 9, we graph the relationship between the long run CATE and the short run CATE. The slope is 0.42. This suggests that the CATE effects shrinks over time. The first possible explanation is that the local "consumer city" effect is small as time passes. Another explanation is that there is a general equilibrium effect as developers build new desirable housing units (as revealed by the CATE responses by developers). As the developers engage in this activity, increase in supply lowers the equilibrium prices.

5.6 Estimating the Value of Time

We measure the implied value of time for different Seoul residents under the assumption that all of the observed capitalization effect is due to time savings. To study this, we first regress the rent for apartment i in district j at time t on each travel time to twenty major destinations presented in table 1 with apartment type fixed effect as follows.

$$Rent_{ijt} = \sum_{g=1}^{20} \beta_g Hours_{igt} + X_{jt} + \mu_i + \lambda_t + \varepsilon_{ijt} \tag{6}$$

where $Hours_{igt}$ represents travel time between apartment i and destination g at time t, and X_{jt} includes counts of restaurants and retail shops and the number of employees in those industries in district j at time t. μ_i is apartment type fixed effect and λ_t is quarter fixed effect.

The main reason we use the rent data is to rule out any speculative demand that may affect the property price, and focusing on instant benefits. Though an apartment is not located in the vicinity of the transit station, travel time from the apartment to each destination changed because riders might use a faster route after LINE9 opened. This reveals the correlation between an hour reduction to each destination and rent. We find that tenants are likely to pay US$ 1,454,545 more as one travels to Kangnam(CBD) an hour earlier (Table 7, A). Note again that rent is not monthly payments but two-year deposit in advance unlike the U.S. and many countries. Tenants put down deposits at the first day of contract and retrieve them when they move out. No other fee is paid to home owners for renting. This suggests that willingness to pay to be an hour closer to Kangnam is not the amount of deposit per-se, but foregone interest that tenants would earn if they lived in their own apartment. Assuming an interest rate of 2%, the opportunity cost for two years is US$ 29,090(Table 7, B), which means tenants that sacrifice US$ 39 every day (Table 7, C).

We compare our estimated value of saving an hour in commute time to Gangnam to the taxi fare in Seoul to cross-validate our estimates. Based on the current taxi fares, riders pay US$ 2,73 for first 2km, even though they travel less than 2km. Beyond 2km, riders pay US$ 0.09 for every 142m. With an average speed of 35.4km in Seoul (The Korea Transport Institute, 2011), the estimated taxi fare to travel for an hour is US$ 24.11, which implies that riders pay US$ 48.22 for a round-trip. Moreover, traveling speed becomes slower during rush hours due to traffic congestion. For commuters to travel 35.4km during rush hours, it costs $29.13 in the morning and $29.84 in the evening, respectively. This back-of-the-envelope calculation implies that a commuter would pay $58.97 for roundtrip, should she travel only by a taxi during rush hour. She would save $19.97 everyday as a net effect by moving to the region where she is able to travel to Gangnam an hour faster.

6 Conclusion

Over the years 2000 to 2009, US$818 million were spent to build a new subway in Seoul, South Korea. Such place based investments offer the opportunity to explore how a city's urban form and real estate pricing are affected by such an investment. This paper has used ML methods to contribute to the urban transit infrastructure effects literature.

Our paper implements a difference in difference empirical design. We find that the introduction of the train is associated with apartment price appreciation for certain leafs but actually lowered apartment price growth in other neighborhoods. We posit that the fast train is most likely to reduce prices for apartments in the destination area of Gangnam because people can now decentralize and still access this location by using the fast train. The notable feature of our study is our ability to document significant heterogeneity on observable dimensions. The payoff for urban research from ML methods is the ability to search across a large number of dimensions of heterogeneity at low cost. Such conditional average treatment effects disaggregate the overall average treatment effect that has been the typical object of interest in earlier real estate studies. By estimating the CATEs our work has new implications for estimating the economic incidence of public transit improvement projects.

References

Agostini, C. A. and Palmucci, G. A. (2008). The anticipated capitalisation effect of a new metro line on housing prices. *Fiscal studies*, 29(2):233–256.

Ahlfeldt, G. M. (2013). If we build it, will they pay? predicting property price effects of transport innovations. *Environment and Planning A*, 45(8):1977–1994.

Athey, S. and Imbens, G. W. (2015). Machine learning methods for estimating heterogeneous causal effects. *stat*, 1050:5.

Bae, C.-H. C., Jun, M.-J., and Park, H. (2003). The impact of seoul's subway line 5 on residential property values. *Transport policy*, 10(2):85–94.

Bartholomew, K. and Ewing, R. (2011). Hedonic price effects of pedestrian-and transit-oriented development. *Journal of Planning Literature*, 26(1):18–34.

Baum-Snow, N. (2007). Did highways cause suburbanization? *The Quarterly Journal of Economics*, 122(2):775–805.

Baum-Snow, N. and Kahn, M. E. (2000). The effects of new public projects to expand urban rail transit. *Journal of Public Economics*, 77(2):241–263.

Baum-Snow, N., Kahn, M. E., and Voith, R. (2005). Effects of urban rail transit expansions: Evidence from sixteen cities, 1970-2000 [with comment]. *Brookings-Wharton papers on urban affairs*, pages 147–206.

Beygelzimer, A. and Langford, J. (2009). The offset tree for learning with partial labels. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 129–138. ACM.

Billings, S. B. (2011). Estimating the value of a new transit option. *Regional Science and Urban Economics*, 41(6):525–536.

Billings, S. B. (2015). Hedonic amenity valuation and housing renovations. *Real Estate Economics*, 43(3):652–682.

Blake, T. (2016). Commuting costs and geographic sorting in the housing market. *Real Estate Economics*.

Bowes, D. R. and Ihlanfeldt, K. R. (2001). Identifying the impacts of rail transit stations on residential property values. *Journal of Urban Economics*, 50(1):1–25.

Breiman, L., Friedman, J., Stone, C. J., and Olshen, R. A. (1984). *Classification and regression trees*. CRC press.

Cervero, R. and Duncan, M. (2002). Transit's value-added effects: light and commuter rail services and commercial land values. *Transportation Research Record: Journal of the Transportation Research Board*, (1805):8–15.

Cervero, R. and Kang, C. D. (2011). Bus rapid transit impacts on land uses and land values in seoul, korea. *Transport Policy*, 18(1):102–116.

Debrezion, G., Pels, E., and Rietveld, P. (2007). The impact of railway stations on residential and commercial property value: a meta-analysis. *The Journal of Real Estate Finance and Economics*, 35(2):161–180.

Diao, M., Qin, Y., and Sing, T. F. (2016). Negative externalities of rail noise and housing values: Evidence from the cessation of railway operations in singapore. *Real Estate Economics*, 44(4):878–917.

Dronyk-Trosper, T. (2017). Searching for goldilocks: The distance-based capitalization effects of local public services. *Real Estate Economics*, 45(3):650–678.

Dudík, M., Langford, J., and Li, L. (2011). Doubly robust policy evaluation and learning. *arXiv preprint arXiv:1103.4601*.

Duncan, M. (2008). Comparing rail transit capitalization benefits for single-family and condominium units in san diego, california. *Transportation Research Record: Journal of the Transportation Research Board*, (2067):120–130.

Fejarang, R. A. (1993). Impact on property values: A study of the los angeles metro rail. In *PUBLIC TRANSPORT PLANNING AND OPERATIONS. PROCEEDINGS OF SEMINAR H HELD AT THE EUROPEAN TRANSPORT, HIGHWAYS AND PLANNING 21ST SUMMER ANNUAL MEETING (SEPTEMBER 13-17, 1993), UMIST. VOLUME P370*.

Foster, J., Taylor, J., and Ruberg, S. (2010). Subgroup identification from randomized clinical data. *Statistics in Medicine*, 30:2867–2880.

Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, volume 1. Springer series in statistics Springer, Berlin.

Gibbons, S. and Machin, S. (2005). Valuing rail access using transport innovations. *Journal of urban Economics*, 57(1):148–169.

Gibbons, S. and Machin, S. (2008). Valuing school quality, better transport, and lower crime: evidence from house prices. *oxford review of Economic Policy*, 24(1):99–119.

Glaeser, E. L., Gyourko, J., and Saks, R. (2005). Why is manhattan so expensive? regulation and the rise in housing prices. *The Journal of Law and Economics*, 48(2):331–369.

Glaeser, E. L. and Kahn, M. E. (2001). Decentralized employment and the transformation of the american city. Technical report, National Bureau of Economic Research.

Glaeser, E. L., Kahn, M. E., and Rappaport, J. (2008). Why do the poor live in cities? the role of public transportation. *Journal of urban Economics*, 63(1):1–24.

Gonzalez-Navarro, M. and Turner, M. (2016). Subways and urban growth: evidence from earth. Technical report, London School of Economics and Political Science, LSE Library.

Hess, D. B. and Almeida, T. M. (2007). Impact of proximity to light rail rapid transit on station-area property values in buffalo, new york. *Urban studies*, 44(5-6):1041–1068.

Imai, K., Ratkovic, M., et al. (2013). Estimating treatment effect heterogeneity in randomized program evaluation. *The Annals of Applied Statistics*, 7(1):443–470.

Kahn, M. E. (2007). Gentrification trends in new transit-oriented communities: Evidence from 14 cities that expanded and built rail transit systems. *Real Estate Economics*, 35(2):155–182.

Kain, J. F. (1990). Deception in dallas: Strategic misrepresentation in rail transit promotion and evaluation. *Journal of the American Planning Association*, 56(2):184–196.

Kim, J., Zhang, M., et al. (2005). Determining transit's impact on seoul commercial land values: An application of spatial econometrics. *International Real Estate Review*, 8(1):1–26.

Knaap, G. J., Ding, C., and Hopkins, L. D. (2001). Do plans matter? the effects of light rail plans on land values in station areas. *Journal of Planning Education and Research*, 21(1):32–39.

Li, T. (2016). The value of access to rail transit in a congested city: Evidence from housing prices in beijing. *Real Estate Economics*.

McDonald, J. F. and Osuji, C. I. (1995). The effect of anticipated transportation improvement on residential land values. *Regional science and urban economics*, 25(3):261–278.

McFadden, D. L. (1984). Econometric analysis of qualitative response models. *Handbook of econometrics*, 2:1395–1457.

McMillen, D. P. and McDonald, J. (2004). Reaction of house prices to a new rapid transit line: Chicago's midway line, 1983–1999. *Real Estate Economics*, 32(3):463–486.

McMillen, D. P. and McDonald, J. F. (1998). Suburban subcenters and employment density in metropolitan chicago. *Journal of Urban Economics*, 43(2):157–180.

Pivo, G. and Fisher, J. D. (2011). The walkability premium in commercial real estate investments. *Real Estate Economics*, 39(2):185–219.

Redfearn, C. L. (2009). How informative are average effects? hedonic regression and amenity capitalization in complex urban housing markets. *Regional Science and Urban Economics*, 39(3):297–306.

Sieg, H., Smith, V. K., Banzhaf, H. S., and Walsh, R. (2004). Estimating the general equilibrium benefits of large changes in spatially delineated public goods. *International Economic Review*, 45(4):1047–1077.

Su, X., Tsai, C.-L., Wang, H., Nickerson, D. M., and Li, B. (2009). Subgroup analysis via recursive partitioning. *Journal of Machine Learning Research*, 10(Feb):141–158.

Taddy, M., Gardner, M., Chen, L., and Draper, D. (2016). A nonparametric bayesian analysis of heterogenous treatment effects in digital experimentation. *Journal of Business & Economic Statistics*, 34(4):661–672.

Tu, C. C. and Eppli, M. J. (1999). Valuing new urbanism: The case of kentlands. *Real Estate Economics*, 27(3):425–451.

Zeileis, A., Hothorn, T., and Hornik, K. (2008). Model-based recursive partitioning. *Journal of Computational and Graphical Statistics*, 17(2):492–514.

Zheng, S. and Kahn, M. E. (2013). Does government investment in local public goods spur gentrification? evidence from beijing. *Real Estate Economics*, 41(1):1–28.

Figure 1: Map of Seoul

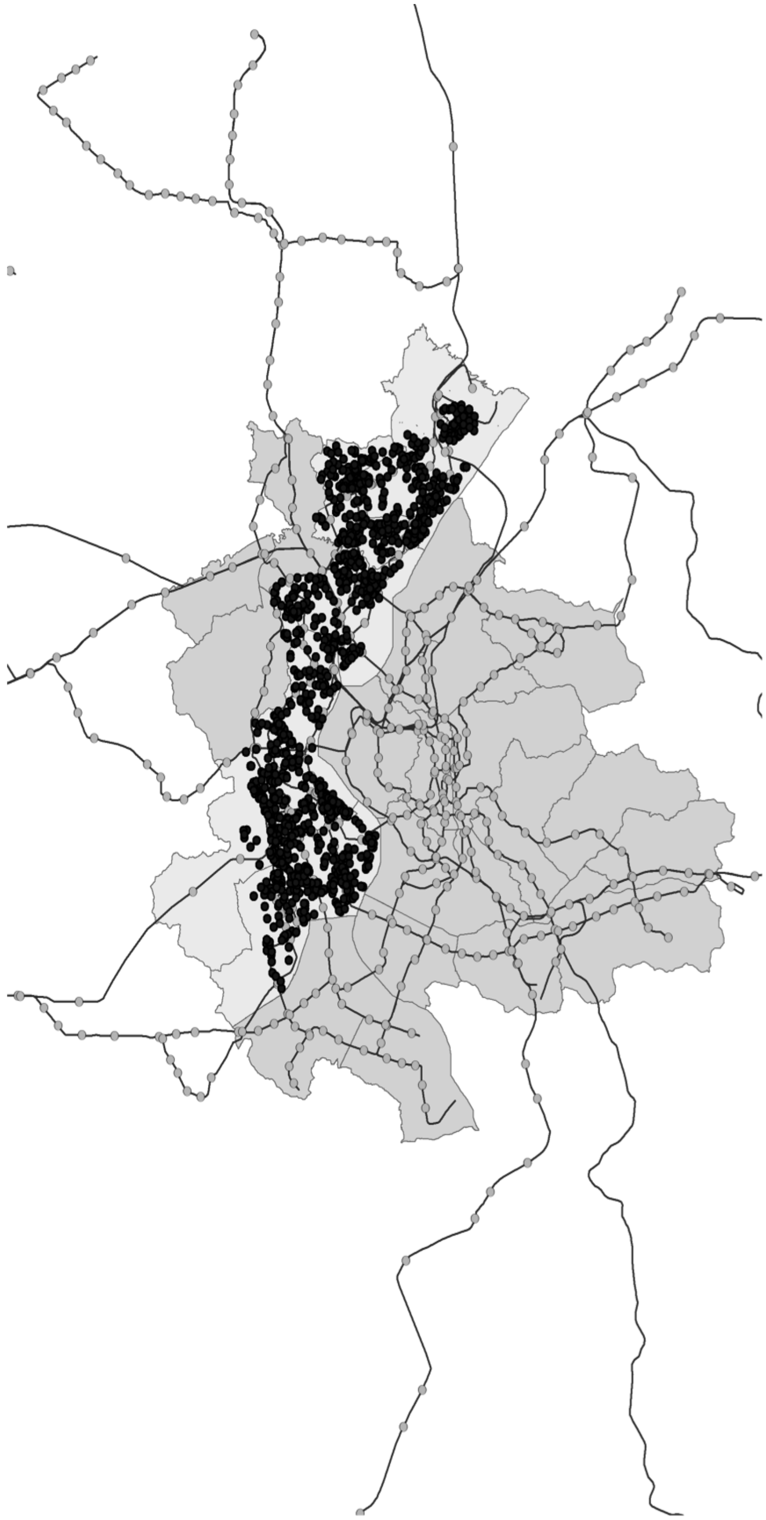


Figure 2: LINE9

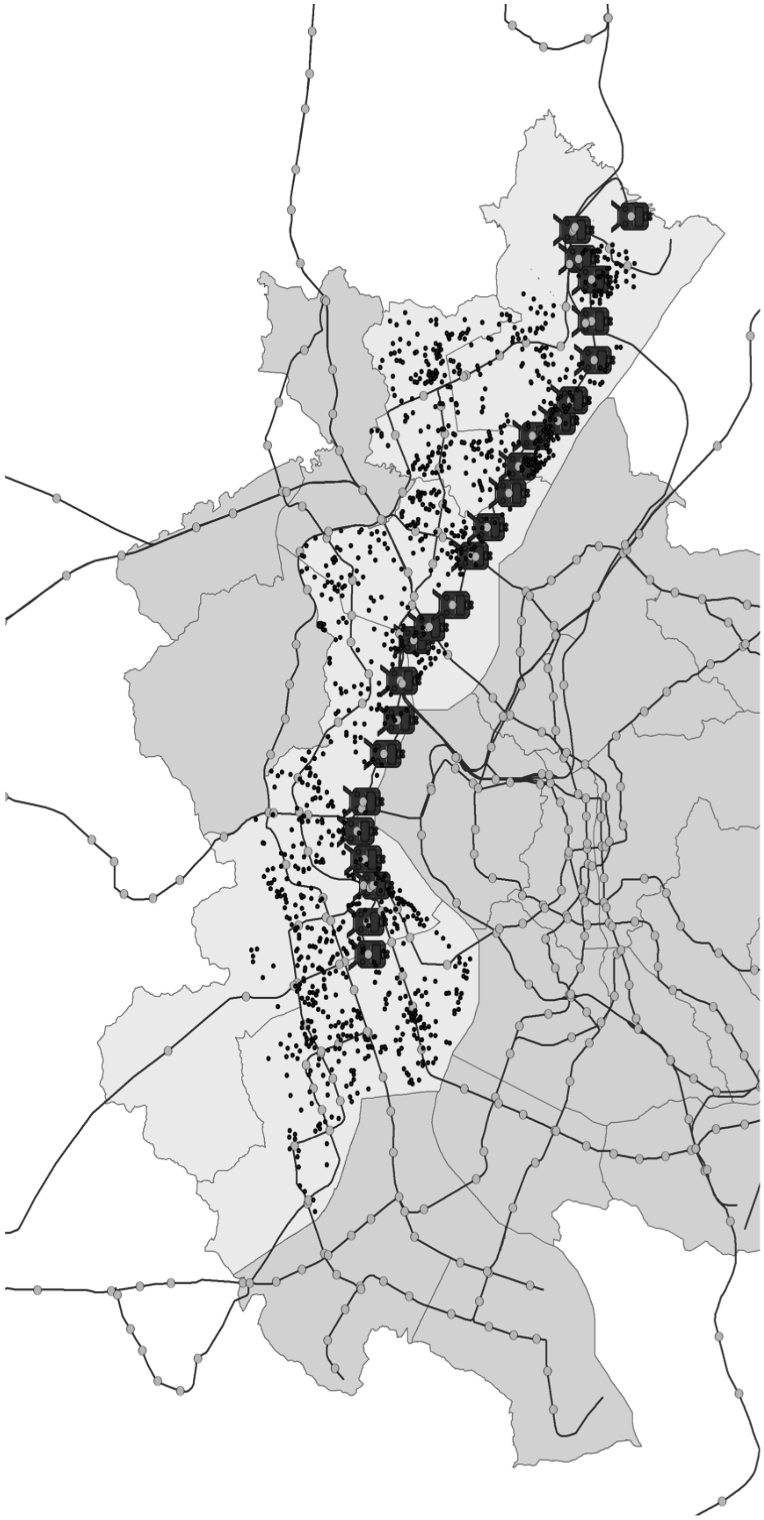


Figure 3: Time line

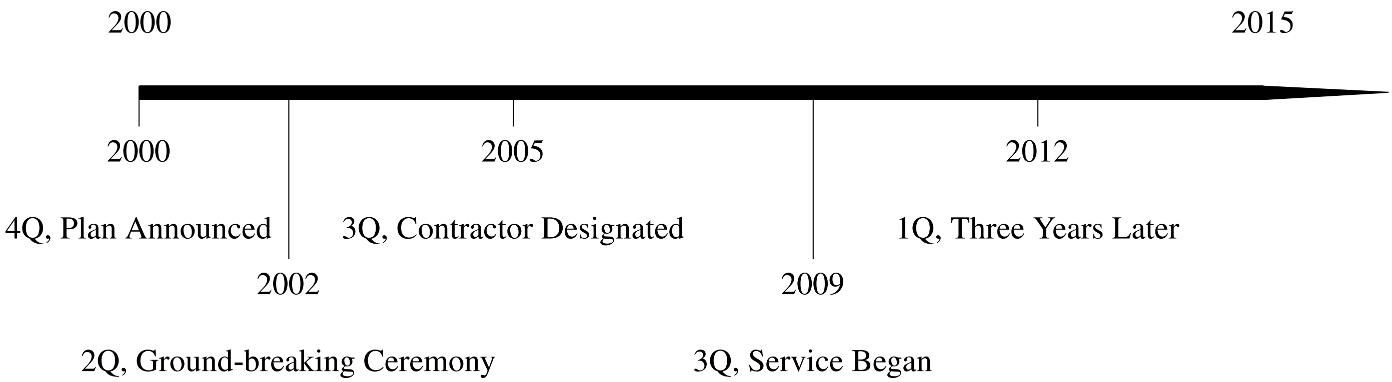


Table 1: Travel Time to Major Destination in Seoul

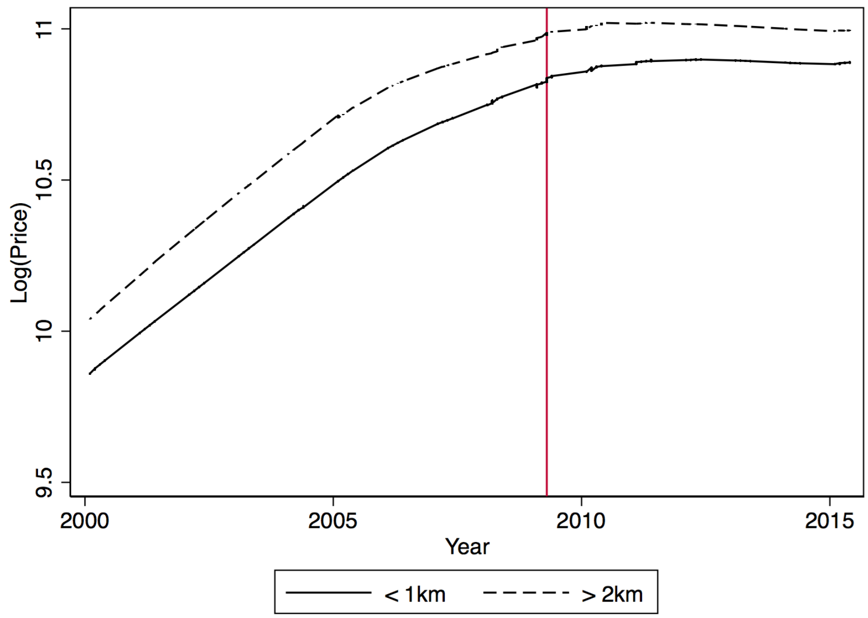
Destinations	Category	Whole Sample Traveling Time before LINE9 (Hours)	Whole Sample Traveling Time after LINE9 (Hours)	Whole Sample Difference (Hours)	Within 1km Traveling Time before LINE9 (Hours)	Within 1km Traveling Time after LINE9 (Hours)	Within 1km Difference (Hours)
Kangnam	CBD	0.51	0.42	0.09	0.67	0.43	0.24
Yeouido	Business	0.46	0.36	0.1	0.5	0.29	0.21
Myungdong	Business	0.53	0.47	0.06	0.56	0.46	0.1
Hongik University	Entertainment	0.46	0.43	0.03	0.5	0.44	0.07
Express Bus Terminal	Bus Terminal	0.51	0.36	0.19	0.65	0.38	0.27
Shinchon	Entertainment	0.48	0.43	0.05	0.51	0.44	0.08
Sadang	Entertainment	0.67	0.46	0.21	0.71	0.42	0.29
Gimpo Airport	Airport	0.91	0.46	0.45	0.93	0.36	0.57
Incheon Airport	Airport	1.66	1.61	0.05	1.67	1.59	0.08
Seoul Zoo	Entertainment	0.77	0.57	0.21	0.81	0.53	0.28
Gwanghwamun	Business	0.52	0.46	0.06	0.56	0.46	0.1
Kungook University	Entertainment	0.58	0.54	0.03	0.67	0.56	0.11
Nambu Bus Terminal	Bus Terminal	0.51	0.43	0.08	0.67	0.44	0.23
Kangbyun	Bus Terminal	0.59	0.54	0.05	0.7	0.56	0.14
Jongro	Old CBD	0.54	0.48	0.05	0.58	0.48	0.1
Apgujeong	Entertainment	0.53	0.47	0.06	0.65	0.48	0.17
Yeoungdeungpo	Train Station	0.42	0.4	0.02	0.47	0.41	0.07
Seoul Train Station	Train Station	0.49	0.43	0.06	0.53	0.42	0.11
Dongdaemun	Entertainment	0.49	0.47	0.03	0.57	0.46	0.09
Korea University	Entertainment	0.62	0.59	0.03	0.71	0.61	0.1

Notes: The estimated traveling time is based on the assumption that travelers walks to the closest subway station and take a subway, with walking speed of 4km/hour and subway running at 50km/hour.

Table 2: Summary Statistics

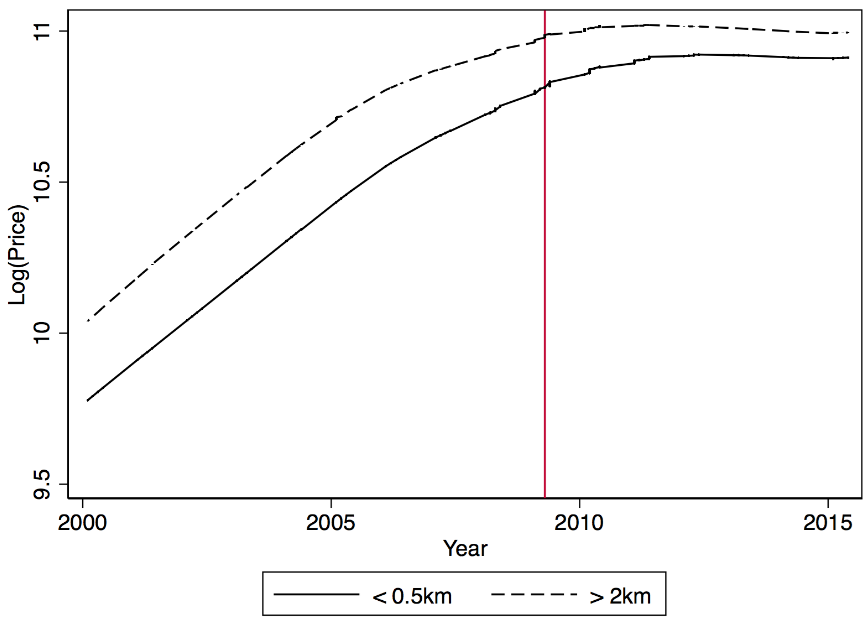
Variable	Mean	Std. Dev.	N
Subway			
Distance to line9 (km)	1.915	1.244	265600
Distance to closest other line (km)	0.702	0.560	265600
Apt Characteristics			
Area (m^2)	94.98	39.5	265600
Room	3.13	0.96	265344
Bath	1.74	0.51	260381
Age, in Years	8.21	9.95	265600
Parking spaces within Complex	546.05	777.99	265600
Number of Apartment units within Complex	435.59	558.62	265600
Number of Apartment buildings within Complex	5.6	9.06	265600
Bus stops within 1km	62.86	19.66	265600
Hospitals within 1km	2.45	1.77	265600

Figure 4: The Pre-Treatment Trend: 1km vs 2km (Lowess Graph)



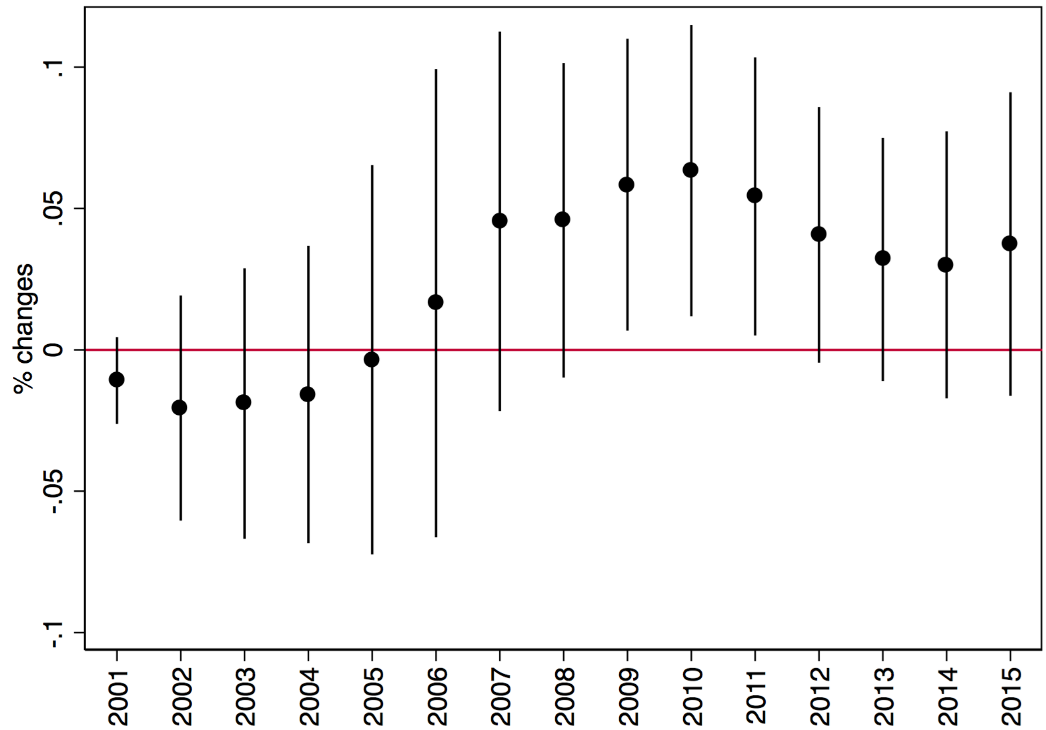
Notes: The vertical line represents when LINE9 opened.

Figure 5: The Pre-Treatment Trend: 0.5km vs 2km (Lowess Graph)



Notes: The vertical line represents when LINE9 opened.

Figure 6: Treatment Effect Estimates Over Time



Notes: Each circle indicates the coefficient of interaction between a "within 1km dummy" and each calendar year. Each bar represents a 90 percent confident interval.

Table 3: OLS Estimates of the Value of Rail Access

VARIABLES	(1) Log(Price)	(2) Log(Price)	(3) Log(Price)	(4) Log(Price)
Distance (km)	0.0339*			
	(0.0193)			
Distance (km) $\times$ AFTER	-0.0174**			
	(0.0078)			
Log(Distance, km)		0.0138		
		(0.0181)		
Log(Distance, km) $\times$ AFTER		-0.0246**		
		(0.0103)		
Within 1km			-0.0785**	-0.0641**
			(0.0367)	(0.0254)
Between 1 $\sim$ 2km			-0.0253	
			(0.0400)	
Within 1km $\times$ AFTER			0.0392*	
			(0.0205)	
Between 1 $\sim$ 2km $\times$ AFTER			0.0224	
			(0.0214)	
Within 1km $\times$ AFTER				0.4073***
				(0.0970)
Within 1km $\times$ AFTER $\times$ Size (m^2) (1)				-0.0017**
				(0.0006)
Within 1km $\times$ AFTER $\times$ Other Line (km) (2)				0.0092
				(0.0101)
Within 1km $\times$ AFTER $\times$ Room (3)				-0.0178
				(0.0285)
Within 1km $\times$ AFTER $\times$ Bath (4)				-0.0269
				(0.0233)
Within 1km $\times$ AFTER $\times$ Age (5)				-0.0117**
				(0.0051)
Within 1km $\times$ AFTER $\times$ Age^2 (6)				0.0001
				(0.0001)
Joint F-value (1) $\sim$ (6)				11.19
(P-value)				(0.000)
Observations	201,985	201,985	201,985	201,985
R-squared	0.9079	0.9076	0.9078	0.9102

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are reported in parentheses. The standard errors are clustered at the district("Dong") level. Controls include the apartment's size, the number of rooms and baths, parking spaces, the apartment's age, age squared, distance to other closest station, the number of bus stops within one kilometer, the number of hospitals within one kilometer, a dummy for whether the complex is built by major construction companies, the number of households within the complex and the number of apartment building within the complex. District fixed effects and quarter fixed effects are included.

Figure 7: The CATE Distribution

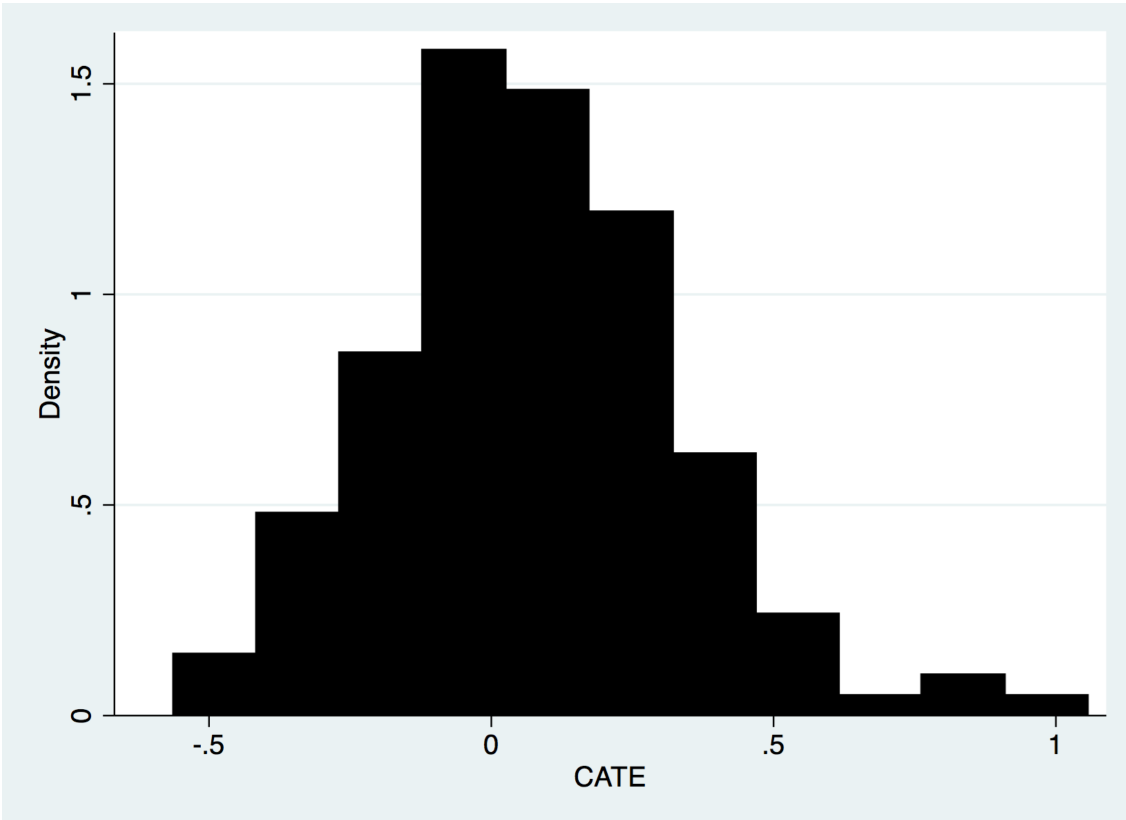
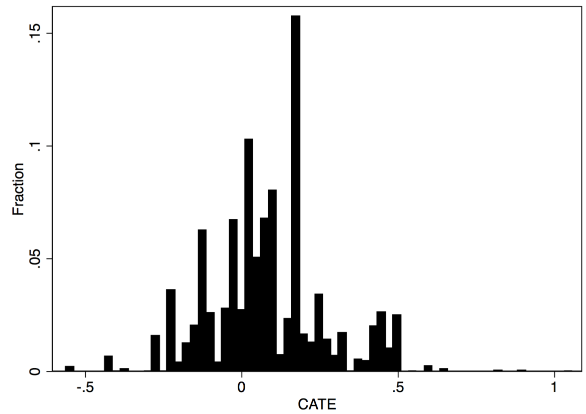


Table 4: Apartment Characteristics in Each CATE Group

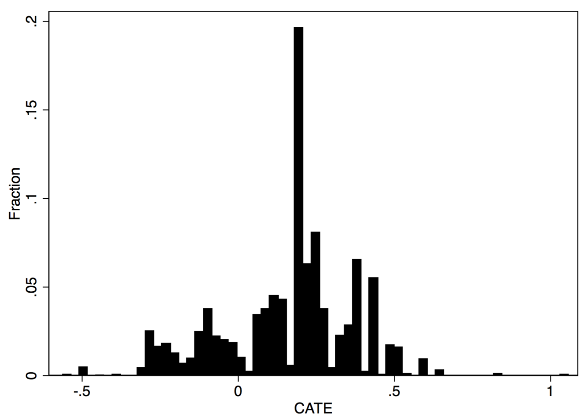
Variables	**Top 10%** (CATE)	**Top 25%** (CATE)	**Bottom 10%** (CATE)	**Bottom 25%** (CATE)
Apartment Characteristics				
Size (m^2)				
$0 \sim 66.24$ (=1)	0.18	0.15	0.34	0.35
$66.24 \sim 84.88$ (=1)	0.32	0.22	0.14	0.21
$84.88 \sim 114.93$ (=1)	0.29	0.35	0.26	0.24
$114.93 \sim 301.45$ (=1)	0.21	0.29	0.26	0.2
Room (Number)	3.13	3.18	3.29	3.23
Bath (Number)	1.8	1.82	1.67	1.70
Apartment's Age				
Less than 5 yrs (=1)	0.27	0.27	0.39	0.27
Between $5 \sim 10$ yrs (=1)	0.66	0.61	0.21	0.28
More than 10 yrs (=1)	0.07	0.12	0.41	0.45
Another Nearby Station (=1) (Within 1km)	0.555	0.76	0.89	0.98
District				
Yeongdeungpo	0.07	0.09	0.12	0.16
Dongjak	0	0.06	0.23	0.18
Gangnam	0.09	0.16	0.37	0.28
Gangseo	0.47	0.3	0.14	0.17
Seocho	0.29	0.34	0.05	0.1
Yangcheon	0.09	0.05	0.11	0.12
The Average of CATE			0.08	
Number of types with positive impacts			89	
Number of types with negative impacts			53	
Observations			201,530	

Figure 8: The Empirical Distribution of New Construction as a Function of the CATE

(a) Before 2002



(b) Between 2002 and 2009



(c) After 2009

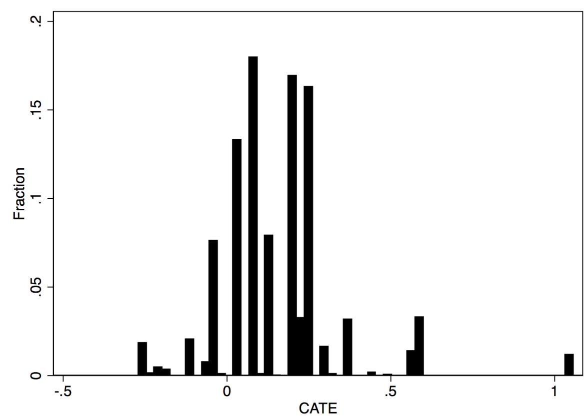


Table 5: New Apartment Yields High CATE

VARIABLES	(1) CATE	(2) CATE	(3) CATE	(4) CATE
Age	-0.0169**	-0.0140**	-0.0119*	-0.0080
	(0.0073)	(0.0065)	(0.0062)	(0.0062)
Age^2	0.0004**	0.0003**	0.0003**	0.0002
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Other Controls	No	No	Yes	Yes
District Fixed Effect	No	Yes	No	Yes
Observations	878	878	878	878

Notes: *** p<0.01, ** p<0.05, * p<0.1. Robust standard errors are reported in parentheses. The standard errors are clustered at the district("Dong") level. Controls include the apartment's size, the number of rooms and baths, parking spaces, the apartment's age, age squared, distance to other closest station, the number of bus stops within one kilometer, the number of hospitals within one kilometer, a dummy for whether the complex is built by major construction companies, the number of households within the complex and the number of apartment building within the complex. District fixed effects and quarter fixed effects are included.

Table 6: Rail Transit Capitalization: The Role of Travel Time Savings and the Local Consumer City

VARIABLES	(1) Log(Price)	(2) Log(Price)	(3) Log(Price)	(4) Log(Price)	(5) Log(Price)	(6) Log(Price)	(7) Log(Price)	(8) Log(Price)
Distance (km)	0.0256	0.0181	0.0251	0.0182				
	(0.0189)	(0.0185)	(0.0187)	(0.0185)				
Distance (km) × AFTER	-0.0172**	-0.0156*	-0.0153*	-0.0138				
	(0.0078)	(0.0090)	(0.0077)	(0.0087)				
Within 1km					-0.0478*	-0.0328	-0.0447*	-0.0314
					(0.0244)	(0.0253)	(0.0244)	(0.0254)
Within 1km × AFTER					0.0336*	0.0248	0.0284*	0.0225
					(0.0173)	(0.0256)	(0.0166)	(0.0256)
Travel Times	No	Yes	No	Yes	No	Yes	No	Yes
Retails and Restaurants	No	No	Yes	Yes	No	No	Yes	Yes
Observations	185,745	185,745	180,569	180,569	185,745	185,745	180,569	180,569

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Robust standard errors are reported in parentheses. The standard errors are clustered at the district("Dong") level. Controls include the apartment's size, the number of rooms and baths, parking spaces, the apartment's age, age squared, distance to other closest station, the number of bus stops within one kilometer, the number of hospitals within one kilometer, a dummy for whether the complex is built by major construction companies, the number of households within the complex and the number of apartment building within the complex. District fixed effects and quarter fixed effects are included.

Figure 9: Long Difference Result

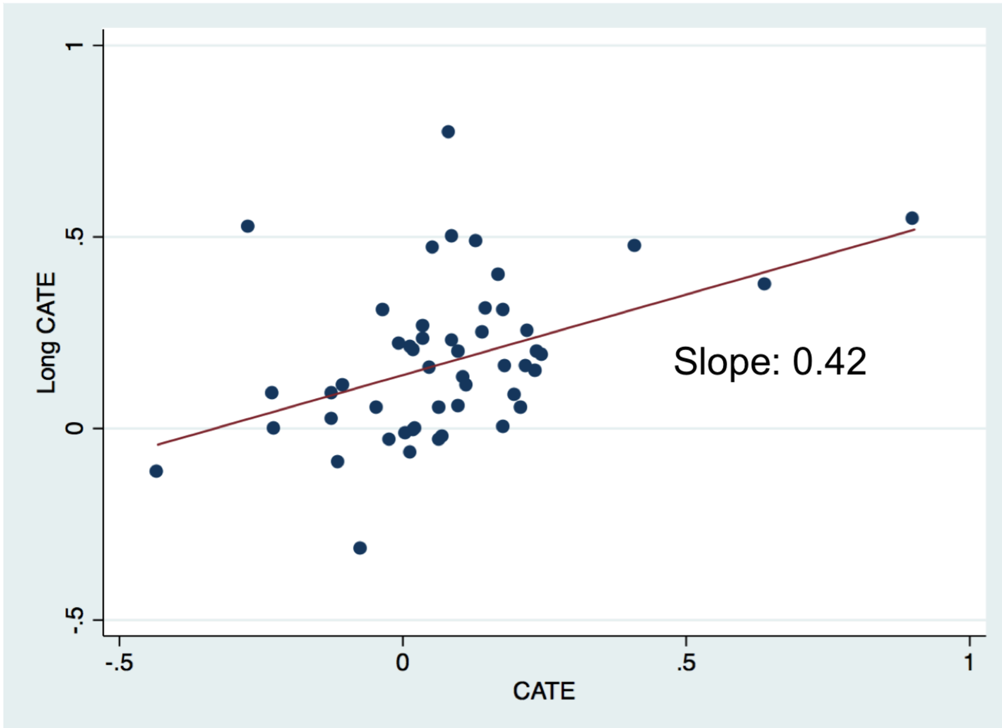


Table 7: Estimates of Value of Time

Estimated Value of an Hour	
Correlation b/w An hour reduction in travel time to CBD and 2-year rent deposit	US\$ 1,454,545 (A)
Interests for two years (Interest rate 2 %)	US\$ 29,090 (B = A $\times$ 0.02)
Daily opportunity costs (2 year = 730 days)	US\$ 39 (C = B / 730)
Estimate Taxi Fare for an Hour	
Basic Fare (First 2km)	US\$ 2.73
Approximate Extra Fare	US\$0.09 per 142m
Average Estimated Driving Distance in an Hour (With an average speed of 35.4km/hour, 2011)	35.4km
Average Estimated Taxi Fare to go 35.4km	US\$ 24.11
Traveling Speed in Morning Rush Hours* (8-9 AM)	29.3km/hour
Traveling Speed in Evening Rush Hours* (6-8 PM)	28.6km/hour
Estimated Taxi Fare to go 35.4km in Morning Rush Hours*	US\$ 29.13 (= 24.11 $\times \frac{35.4}{29.3}$)
Estimated Taxi Fare to go 35.4km in Evening Rush Hours*	US\$ 29.84 (= 24.11 $\times \frac{35.4}{28.6}$)

* Taxi fare in South Korea is determined by distance as well as traveling time. Due to traffic congestion, we expect taxi fare to be higher during rush hours if a commuter travels the same distance. All numbers are based on back-of-the-envelop calculation.