Anticipation Effects and the Second Avenue Subway

SARAH SCHOENGOLD, CHRISTIAN MOSCARDI, HAO XI

Abstract— **The Second Avenue Subway opened on January 1, 2017 after a century of fits and starts. While rent prices from a more transient community have fluctuated leading up to the opening, our question focused on housing sales. In particular, were interested in what extent anticipated improvements around this long-running project increase property values before the substantive effect of new transit took hold. Our study looks uses a Difference-in-Difference model to look at housing prices before and after the opening of the Second Avenue Subway on the Upper East Side in New York City. The model reveals that although there was an increase in sales leading up to the opening, with particular turnover within 400ft of a new stop, the D-I-D coefficient is not significant.**

I. INTRODUCTION

A. OVERVIEW

The first phase of the Second Avenue Subway (SAS) line in New York City opened on January 1, 2017 after nearly a century of fits and starts. The first mention of the plan to extend the line North along Second Avenue was in 1919, a keystone feature in Daniel Turners Comprehensive Rapid Transit System plan [1]. The plan caused the housing stock prices to jump 50% according to one New York Times article [2]. Thus began the saga of development in which Upper East Side real estate prices waffled with the SASs stuttering progress after delays from the stock market crash, a World War, and another fiscal crisis in the 1970s.

Of course, it wasnt just the new 72nd, 86th, and 96th Street Q stops that influenced real estate hikes in one of New Yorks most well-known neighborhoods. Our question is, exactly how much did the SAS change the real estate prices compared to other exogenous factors. In particular, were interested in what extent anticipated improvements around this long-running project increase property values before the substantive effect of new transit took hold.

B. LITERATURE REVIEW

There has been much work demonstrating the positive relationship between proximity to transportation and housing prices, but the effects are widely different based on many factors including the density and existing commuter patterns of the city. When streetcar lines were replaced with a subway line in the late 1970s, several studies looked to the city of Toronto to explain this relationship. D.N. Deweess study from 1976 concluded that the rise in housing price increased perpendicular to the subway entrances, but not all the way along the line[3]. It also showed the housing price increases diminished mile (walking distance) from the station. Vladimir Bajics study on the same subway introduction showed that the savings from commuting costs were factored into the price of housing along the new line, and that the rent increases impacted the housing stock close to the new stations, but not along the line in its entirety [4]. These studies are interesting to keep in mind as a baseline for the relationship between new transit options and housing, though its impossible to generalize their findings in the case of the SAS.

Several studies investigate these effects using difference-in-difference models, which compare a control and treatment area both before and after the transportation change. One recent example looks at the introduction of Singapores Circle Line, a new underground subway system, and its effect on property values. The study highlights several techniques, but primarily uses a spatial difference-in-difference (D-I-D) model to study these impacts. In addition to aftereffects of pricing, it also captures anticipation effects in its modeling, and the fluctuation in prices during different stages of construction. This paper uses three different dates at which to compare the control and treatment group in order to capture the D-I-D coefficients change over time [5]. We use aspects of this papers methodology to zoom in on SAS anticipation effects at one year, six months, and at time of the opening.

A similar model is used for a specific case in in Montreal, where two new commuter rail stations were opened [6]. This model builds on top of the baseline econometric models that have been historically used, such as the Hedonic Price model[5]. This same article goes on to provide a specific case study applying this model to commuter rail patterns in Montreal - another North American city with well-established transit. It is important to note, however, that the Second Avenue line is not specifically a commuter rail (although based on Yorkvilles demographics we may expect the line to serve a similar function). One other interesting point of contrast is that the Second Avenue Subway has been a long time coming, whereas according to the authors of this paper the announcement / implementation of new commuter rail in Montreal was done on a timescale short enough to make anticipation effects negligible. Our curiosity is if the opposite is also true: perhaps a dramatically extended timescale also makes anticipation effects negligible.

Honing in further on studies that look at anticipation effects specifically, Damm's 1980 study looks at the anticipated effects of the the Washington DC metro system, and in the 90s Grass studied the effects of the systems development[7]. Damm studied the anticipated effects of development across a variety of real estate types, to the end of supporting some sort of value capture policy - in use in various municipalities, most notably Hong Kong, today [8]. Whats interesting about D.C.s system is that it was developed entirely in the 1960s/70s; before that, there was no subway network, and suddenly there was. Damm found that prices for retail real-estate jumped more immediately than residential real-estate prices, indicating elasticity in that market [7]. Grass looked at residential values and found a positive increase in prices after the fact [9].

Another case from London provided inspiration about how to distinguish between changes in prices caused by the new subway site, and changes with economic growth and decline [10]. Comparing the implied value of travel time savings (VTTS) Gibbons' study concluded that households value rail access reflected in about 1.5% on price and that these valuations are large compared to the valuations of other local amenities.

The case of SAS is unique because of the density of the city, the numerous options for other transit, and potentially the frustration of the community that had stopped holding its breath for an opening. The New York Times, among other major newspaper publications covered the subways saga extensively [11][12]. These articles include many personal narratives, interviewing people whose lives may be or have been affected by the opening. This journalism focuses primarily on anecdotes and fluctuating rent prices for a more transient community. We want to instead use these difference-in-difference models championed by similar housing studies to concretely define if the SAS had an impact on buyers, in particular.

II. Experiment design

In this study we focused on identifying the extent to which the SAS opening affects housing sales prices over time, leading up to and after the SAS opened. Using a novel difference-in-difference model, we explore the change in impact of the SAS coefficient leading up to the opening to define anticipation effects within this context. In addition to looking at how the impact changes over time, we also explore how the distance to subway station influences sales prices before and after the opening.

Our question is rooted in the subways long history. Renters signing leases a year at a time leading up to the opening may benefit greatly from the opening. But did buyers, after being yo-yo-ed around for nearly a century, factor in the prospective opening into what they were actually willing to pay, before the SAS opened its turnstiles?

Understanding the extent to which buyers value extensions of the line close to home can help policy makers think ahead as the city prepares for the next phases of the project. The next two phases extend into areas of Manhattan with cheaper housing stock, potentially making rent and sales price hikes more acute. Defining the patterns of real estate price increases may help small business anticipate and the hype cycle, and make more informed property decisions leading up to the opening.

figure[h]

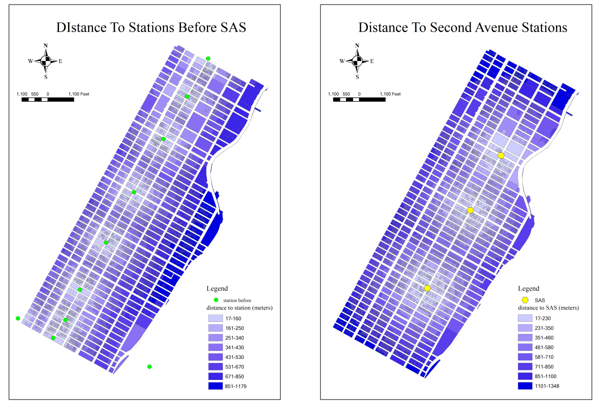

Fig. 1. distance to station before SAS and distance to SAS

III. DATA AND METHODOLOGY

A. THE DATA

The Upper East Side is an older, and typically wealthier neighborhood within New York City, but theres a stark contrast in housing prices between the East and West sides of the neighborhoods, in part due to transportation accessibility. Using community boards as general guidelines, we defined the Upper East Side as everything east of Central Park, between 59th and 110th streets. We first used American Community Survey data in our exploratory analysis phase, but because we could only compare mean rental data at the census tract level, it wasnt as useful for questions at the building level.

Instead, our base dataset for this study is the Property Land Use Tax Lot Output (PLUTO), managed by the NYC Department of City Planning. This dataset contains extensive geographic and land use data for New York City. Layered on top of PLUTO, we used subway shapefiles available from the Department of Transportation. Because PLUTO includes spatial data, it enabled us to calculate the distance of the building to the nearest subway stop. Unlike similar studies that explored the introduction of transportation systems in areas of diverse landscape, the Upper East Side is quite uniform on a grid block system. Although we calculated distance using Euclidean distance, an interesting extension may be to use the L1-norm taxicab distance (possibly on rotated X/Y coordinates), as an acknowledgement of the strict grid pattern on the Upper East Side.

Next, we used sales information from the NYC Department of Finance (DOF) Sales Record data. We filtered sales data by location to the Upper East Side, by type to residential, and in particular, by sales data that included a specific apartment number. With a little bit of extra commonsense cleaning (sales prices >$100,000, a legitimate year built column), this left us with 7014 sales records.

—

a) : Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

—

Various studies used various methods for subsampling the dataset. Diao et. al.[5] looked at the entire dataset of sales and used a fairly standard DID model with sale price as the output variable. We believe they used all sales without concern for repeat-sales. We also believe they had a far more homogeneous market than the one we see in Yorkville, as their base DID model (not controlling for any substantive housing features) managed an R^2 of 0.77.

Dube et. al. , however, used a repeated-sales approach, so only took records where there was a repeat sale - one before the opening of their transit line, and one after. They modeled . We chose to look only at repeat sales, but to model $\Delta log(price)$ (rather than Δ as Dube et. al.[6] do). We believe looking at repeat sales helps control for other factors that we may not be able to control for, as we are looking at the exact same units price development over time.

After filtering for repeat sales, we were left with 74 pairs - or 148 records.

B. Integrating housing characteristics features

Next, in order to control sales price by unit size, we used the Zillow API to acquire additional housing characteristic statistics at the apartment level. In particular, number of bedrooms proves to be an important factor to control for r=0.78.

Between resources merged from Zillow, PLUTO, and DOF sales data, we had a complete dataset to compare sales over time, for both control and treatment groups. We use log price in our actual DID modeling. This standard across similar papers [5]. This is because it is more Gaussian.

While we were able to extract enough pairs to find significant results, the Zillow API limited the amount of data we could extract. In future iterations of the project, wed be interested in looking at Upper East Side sales in its entirety. This would also enable us to capture the D-I-D coefficient in more frequency leading up to the opening. Further, wed be interested in looking at how rent prices fluctuate leading up to the opening. StreetEasy prevented us from scraping the data needed for this study, but its worth exploring other methods for data collection in the future.

IV. METHODOLOGY

As a reminder, chose the entire area of our study to be the Upper East Side, as defined above.

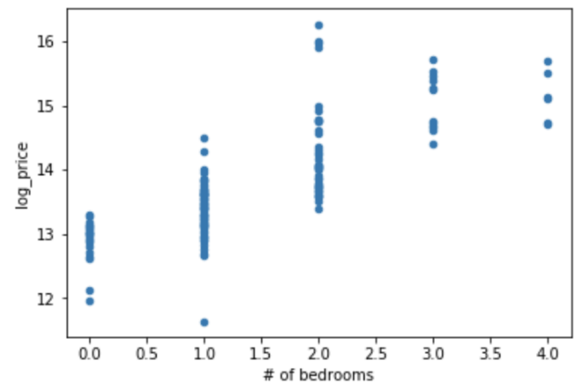

Fig. 2. log price vs. number of bedrooms

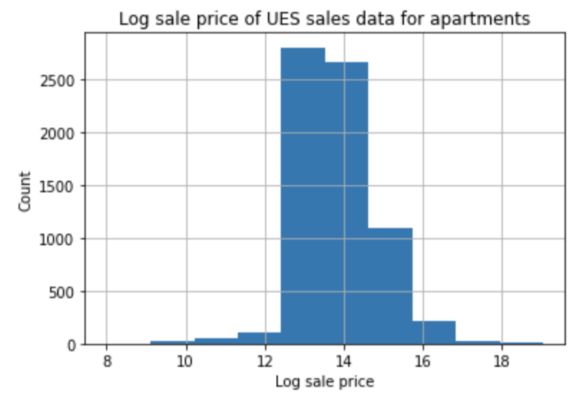

Fig. 3. histogram of log sale price of UES sales data for apartments

We used a Difference-in-Difference (D-I-D) model to determine the effect of the SAS opening on prices. To start, we tried two different controls. The first was distance from the SAS. Different studies have set different thresholds. We tried the method of Diao et. al. to set our threshold - we plotted median sales price before and after the opening, by distance (Fig. 1). In this figure, we can see median sales price before and after SAS opening, by distance to SAS. We can see that at about 200m the after-opening price change becomes small.

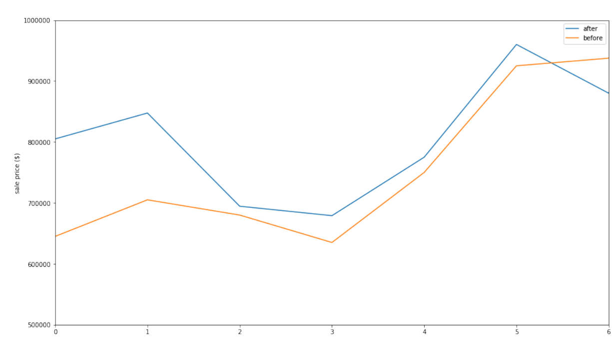

Fig. 4. sales price vs. distance

With this in mind, we chose <200m to a new SAS station

TABLE I
SALES PRICE BEFORE/AFTER SAS, DIVIDED BY 2ND AVE

	Sale Price(east of 2nd Ave),before	Sale Price(east of 2nd Ave),after	(west of 2nd Ave),before	Sale Price(west of 2nd Ave),after
count	2,381	516	3,329	788
mean	$859,057.09	$935,996.61	$2,440,821.97	$2,199,408.87
std	$759,238.12	$844,057.95	$5,505,758.62	$3,349,628.33
min	$12,734.00	$20,280.00	$3,000.00	$14,167.00
25%	$425,000.00	$481,237.50	$615,000.00	$664,250.00
50%	$625,000.00	$685,000.00	$1,275,000.00	$1,264,000.00
75%	$977,520.00	$1,056,250.00	$2,550,000.00	$2,550,000.00
max	$10,250,000.00	$8,600,000.00	$185,000,000.00	$55,000,000.00

TABLE II
SALES PRICE BEFORE/AFTER SAS, DIVIDED BY 200M RING BUFFER

	Sale Price (<200m),before	Sale Price(<200m),after	Sale Price($\geq$ 200m),before	Sale Price($\geq$ 200m),after
count	611	139	5,099	1,165
mean	$929,762.41	$1,075,570.33	$1,883,276.41	$1,773,909.15
std	$718,215.38	$708,685.25	$4,536,527.45	$2,866,578.57
min	$27,000.00	$330,000.00	$3,000.00	$14,167.00
25%	$464,459.00	$602,500.00	$505,250.00	$550,000.00
50%	$700,000.00	$845,000.00	$870,000.00	$890,000.00
75%	$1,229,500.00	$1,282,500.00	$1,875,000.00	$1,940,000.00
max	$5,250,000.00	$3,825,000.00	$185,000,000.00	$55,000,000.00

TABLE III
SALES PRICE BEFORE/AFTER SAS, DIVIDED BY 2ND AVE

	Sale Price(east of 2nd ave)before	Sale Price(east of 2nd ave)before	Sale Price(east of 2nd ave)before	Sale Price(east of 2nd ave)before
count	9	9	65	65
mean	$690,773.78	$803,999.89	$1,530,268.00	$1,599,683.85
std	$153,689.77	$160,345.96	$2,030,060.37	$1,804,032.37
min	$470,000.00	$499,999.00	$113,669.00	$186,000.00
25%	$554,000.00	$700,000.00	$445,000.00	$580,000.00
50%	$720,000.00	$845,000.00	$800,000.00	$890,000.00
75%	$800,000.00	$940,000.00	$1,550,000.00	$1,975,000.00
max	$873,964.00	$962,000.00	$11,600,000.00	$8,910,000.00

TABLE IV
SALES PRICE BEFORE/AFTER SAS, DIVIDED BY 200M RING BUFFER

	Sale Price (<200m),before	Sale Price(<200m),after	Sale Price($\geq$ 200m),before	Sale Price($\geq$ 200m),after
count	27	27	47	47
mean	741,272.04	$885,090.74	$1,822,766.79	$1,857,729.77
std	$574,289.29	$551,764.03	$2,288,349.45	$2,028,915.18
min	$113,669.00	$430,000.00	$155,000.00	$186,000.00
25%	$414,250.00	$548,000.00	$576,000.00	$762,500.00
50%	$554,000.00	$685,000.00	$860,000.00	$995,000.00
75%	$770,000.00	$921,975.00	$1,705,000.00	$2,306,000.00
max	$2,494,721.00	$2,487,500.00	$11,600,000.00	$8,910,000.00

as our treatment/control criteria. This makes intuitive sense in New York, as density is very high, and even 200m west of the SAS is actually closer to the Lexington Avenue line, and thus already has strong transit accessibility.

Additionally, given our location-specific understanding of Yorkvilles dynamics, we also picked another treatment / control threshold - east of 2nd avenue. We suspect that this area benefits most from the new subway accessibility, due to its distance from the SAS.

Difference-in-difference models also require a choice of event time - the time when a potentially meaningful event occurred. Our choice of time point was clear - the SAS opened on January 1, 2017.

Diao et. al. use the following as a base D-I-D model. This is standard in the literature.

$$log(price) = intercept + \alpha * T_{opening} + \beta * S_{treatment} + DID * (T_{opening} \cdot S_{treatment})$$

Upon trying this model, we received poor results (R^2 0.09 or less) - indicating our model probably does not accurately control for factors in our market. In fact, we are confused about how Diao et. al. managed an R^2 of 0.77 with this model. As mentioned above, we suspect market homogeneity is to blame.

The most immediate control we thought to use was number of bedrooms. As described above, we know this correlates to log price. Thus, we augmented our model to:

$log(price) = intercept + \alpha * T_{opening} + \beta * S_{treatment} + DID * (T_{opening} \cdot S_{treatment}) + \gamma * N_{bedrooms}$

In other-words, we controlled for the size of the apartment. This led to a significantly higher R^2 of 0.6. At this point, we felt more confident interpreting our Difference-In-Difference coefficient as meaningful.

V. RESULTS

First, we show results for our two respective choices of treatment variable.

TABLE V
RESULTS OF DID MODEL FOR SAS OPENING

Distance	200 Meters	East of @nd Ave
Treatment Coefficient	-0.0613	-0.3148
DID Coefficient	-0.0054	-0.1456
R-Squared	0.621	0.640
R-Squared	0.610	0.630

As we can see, east of Second Avenue proved a far more interesting treatment area. With a DID coefficient of 14%, we interpret this as the SAS opening providing a 14% average improvement to these property values. On the other hand, the <200m treatment group was much less interesting, with a DID coefficient of -0.5%. e did not have enough data to make our coefficient 95% confident, but hope to do so in future models by including a larger set of better-controlled data, and not just repeated sales.

TABLE VI
TEST OF "ANTICIPATION" EFFECTS OF SAS OPENING

Sample Period	2010-2017	2010-2017	2010-2017
Treatment Coefficient	-0.2812	-0.2989	-0.2996
Time Coefficient			
Time t = 12	0.0768		
Time t = 6		0.1056	0.0358
Time t = 0			0.0713
Interactive Variable			
DID Coef t = 12	0.0679		
DID Coef t = 6		0.1092	**-0.4275**
DID Coef t = 0			**0.5578**
Observations	146	146	146
R^2	0.633	0.638	0.641
Adjusted R^2	0.623	0.627	0.626
P-value DID Coef R^2	0.706	0.540	0.457

From here, we decided to stop looking at the <200m treatment group and focus on east of second avenue. In particular, we wanted to look at anticipation effects on sale prices. Was the market excited for the Second Avenue Subway opening? To examine this, we set up new Difference-in-Difference models at various time windows. In particular, we added a variable for 6 months out, and for 1 year out.

This table contains three separate models: one, looking at difference-in-difference from 1 year ahead of opening and beyond, the next, from 6 months ahead of opening and beyond. These two models demonstrated lower DID coefficients than the at-opening DID model (6.79% and 10.92% respectively) - indicating an absence of anticipation effects. Last, we tried a combined model with DID terms accounting for both 6 months ahead as well as after the opening. We find this model particularly interesting - the DID coefficient 6 months in advance was negative (-42.74%), while accounting for these effects led our after-opening DID coefficient to be more strongly positive at 50% (with a wide confidence band). We interpret this to mean that anticipation effects were, in fact, not particularly strong. However, upon seeing the subway finally open with their own eyes, markets responded favorably to the opening.

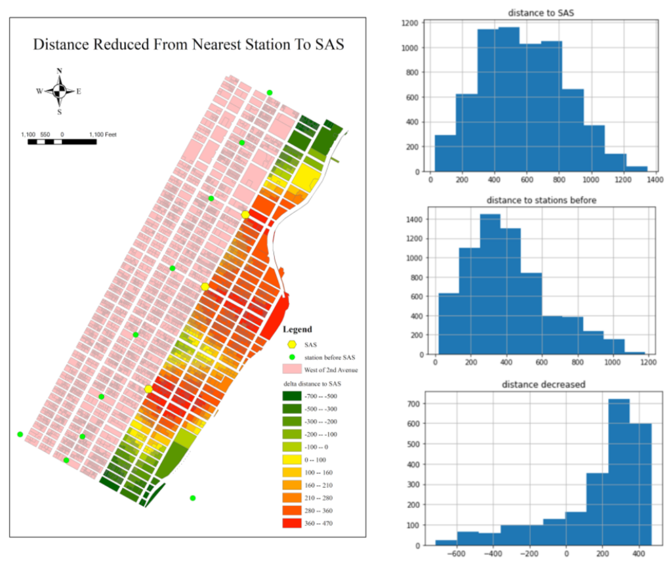

Fig. 5. distance reduced from nearest station

VI. CONCLUSIONS

Our model reveals that while there was an increase in sales leading up to the SAS opening, buyers were not paying more for real estate in expectancy of the opening. The anticipation effects of the SAS did not factor into to ultimate sales prices. However after it opened, the proximity to a new line was in fact bakes into sales prices.

It makes sense that buyers wouldnt factor in the SAS in advance of its opening, given the SAS tumultuous history perhaps this is an Ill believe it when I see it mentality in the market. Further, real estate buyers for the price point for Upper East Side apartments may be less concerned about the proximity of public transportation than a demographic that may rely more on the SAS for commuting purposes.

Interpreting these anticipation results using data instead of the popular qualitative reporting gives concrete evidence on which to plan for future phases of the SAS line. Small business owners and landowners in Harlem can use this information as a way to prepare for what will ultimately change their neighborhood. By measuring real estate values, we look at one piece of this puzzle, but many others remain to be explored.

A. FUTURE WORK

We hope to improve upon our model by including spatial autoregressive terms, to better account for spatial differences we have not yet included. In general, we hope to use Diao et. al.s paper as a template for ours, as we believe they have done a good job accounting for many factors so as to isolate the effect of the lines opening. By doing this, we allow for more accurate quantification of the lines opening impacts, which is helpful for assessing future projects such as Mayor De Blasios streetcar proposal, or perhaps bus line improvements.

References are important to the reader; therefore, each citation must be complete and correct. If at all possible, references should be commonly available publications.

REFERENCES

[1] Fichbein, R. (2016, December 21). A Brief History Of The Long-Awaited Second Avenue Subway. Gothamist. Retrieved November 20, 2017, from

[2] Anon. (1929). MANHATTAN SALES; Second Avenue Subway Project Causes 50% Rise in Prices. The New York Times.

[3] Dewees, D. (1976). The effect of a subway on residential property values in Toronto. Journal of Urban Economics, 3(4), 357-369. doi:10.1016/0094-1190(76)90035-8

[4] Bajic, Vladimir. The Effects of a New Subway Line on Housing Prices in Metropolitan Toronto. Urban Studies, vol. 20, no. 2, 1 May 1983, journals.sagepub.com/doi/abs/10.1080/00420988320080029

[5] Diao, M., Leonard, D., & Sing, T. F. (2017). Spatial-difference-in-differences models for impact of new mass rapid transit line on private housing values. Regional Science and Urban Economics, 67, 64-77.

[6] Jean Dub, Digo Legros, Marius Thriault, Franois Des Rosiers, A spatial Difference-in-Differences estimator to evaluate the effect of change in public mass transit systems on house prices, In Transportation Research Part B: Methodological, Volume 64, 2014, Pages 24-40, ISSN 0191-2615, https://doi.org/10.1016/j.trb.2014.02.007.

[7] Damm, D., Lerman, S., Lerner-Lam, E., & Young, J. (1980). Response of Urban Real Estate Values in Anticipation of the Washington Metro. Journal of Transport Economics and Policy,14(3), 315-336. Retrieved from http://www.jstor.org/stable/20052588

[8] Kembrey, M. (2015, December 18). Hong Kong metro system operators MTR spread 'value capture' message to Australia. The Sydney Morning Herald. Retrieved October 20, 2017, from http://www.smh.com.au/national/hong-kong-metro-system-operators-mtr-spread-value-capture-message-to-australia-20151215-glo0wq.html

[9] Grass, R. G. (1992). The Estimation of Residential Property Values Around Transit Station Sites in Washington, D.C. Journal Of Economics & Finance, 16(2), 139.

[10] Gibbons, S., & Machin, S. (2005). Valuing rail access using transport innovations. Journal of urban Economics, 57(1), 148-169.

[11] Fitzsimmons, E. (2017, February 01). Second Avenue Subway Relieves Crowding on Neighboring Lines. Retrieved December 01, 2017, from https://www.nytimes.com/2017/02/01/nyregion/second-avenue-subway-relieves-crowding-on-neighboring-lines.html

[12] Hughes, C. J. (2017, July 07). Second Avenue Subway Brings New Development. Retrieved December 01, 2017, from https://www.nytimes.com/2017/07/07/realestate/second-avenue-subway-brings-new-development.html