Cornell Undergraduate Economic Review

Spring 2023

- The Economic Effects of Subway Parallelism
- Tipping Behavior and Delays
- Impact of Implementing Weighted School Funding on High School Educational Attainment

Cornell Undergraduate Economic Review

Contents

Staff

CUER Thanks:

The Student Activities Funding Commission for their generous continued financial support

CUER is published each fall and spring.

Letter From the Editor:

My name is Andrés Aradillas-Fernández and I am a sophomore in the College of Arts and Sciences majoring in Economics & Mathematics. I was once again very impressed with the papers we ended up selecting in CUER. While they vary in subject, they are all unified in their insightful quantitative contributions to economics. I continue to be amazed by the work of my fellow undergraduates.

I have a tremendous amount of gratitude for all the members of *the Cornell Undergraduate Economic Review–Specifically, all the copy-editors and referees who have spent countless hours this semester reading, editing, analyzing, and discussing all submissions. Without such an incredible team of students, this journal would certainly be impossible to create. I would especially like to thank Chief Referee Michael Negrea and Chief Copy-Editor Jonathan Stolow for their help this semester.*

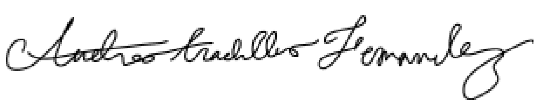
I am of the firm belief that the most noble pursuit of all is the pursuit of knowledge, and many of the papers submitted to the journal this edition embodied this to the fullest, making this semester's decisions one of the hardest yet. It is this pursuit that CUER seeks to praise with every publication, and I strongly believe it is a mission that ought to be continued for years to come.

On a personal note, I would like to thank my family back home in State College, Pennsylvania, for all of their love and support. I would especially like to thank my father, Prof. Andrés Aradillas-López for inspiring me to pursue economics.

All I ask is that when you read this, please become intellectually stimulated and truly engaged in the readings just as our team has. I would love to hear your comments or suggestions on the publication. Please feel free to reach out to me at our email.

Sincerely Yours,

Andrés Aradillas-Fernández



Editor-In-Chief
cueconomicreview@gmail.com
Cornell Undergraduate Economic Review

The Economic Effects of Subway Parallelism: A comparison study of common track subways of New York City and Shanghai City

Weiyi Zhou

New York University

Abstract

Through the comparison study and statistical analysis of the similarities and differences in the planning and practical usage of lands along the subway tracks of New York and Shanghai, relationships of the parallelism of subway system and surrounding area are revealed, which leads to several proposals of promoting the commercial function wielding in an effective subway system.

1.Introduction

New York City, one of the biggest metropolises, first developed its subway system in the early twentieth century. Consequently, New York is famous (or rather, infamous) for its relatively outdated underground shuttle system. Despite its antiquity, however, it still manages to maintain a unique yet effective mode of transport to connect and promote the surrounding commercial centers. Shanghai, one of the biggest international cities on a comparable stage of economic development to New York, equips a highly developed, modern, multi-functional subway network that constructs various ground commercial spaces. There exists an intriguing contrast between the developing context and condition of subway commercialization of these two big cities.

With more and more attention devoted to the study of network construction with high intensity, I deeply realize that seizing the opportunity to explore and expand rail transit to cover property development has significant and far-reaching influence on land-intensive use, urban functional layout and rail transit (Ye, 2021). My research mainly focuses on the economic correlation of track-sharing subways and surrounding areas and compose comparison study between two cities with various comparable economic aspects at the current stage, exemplified as New York City (NYC) and Shanghai City (SHC).

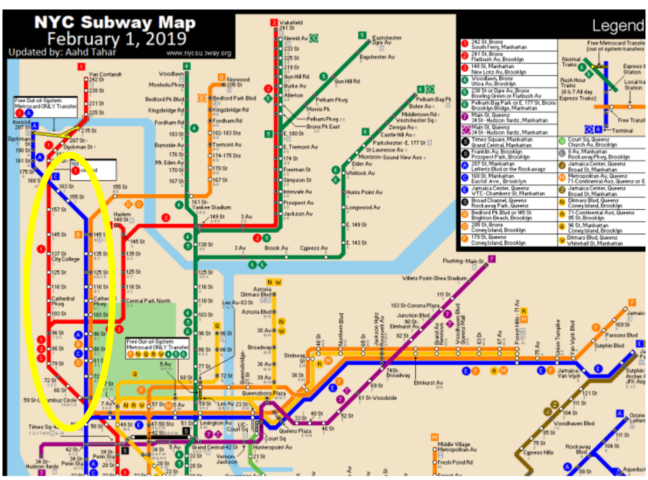
2. Background

New York's first official subway network was opened in Manhattan on October 27, 1904, while the very first subway line of Shanghai, Shanghai Metro Line 1, was completely built up on May 28, 1993, which is nearly 90 years later.

a) Current NYC MTA Line A (former Express A) and Line D (former Express C) (circled by yellow stroke in graph 1 and graph 2)

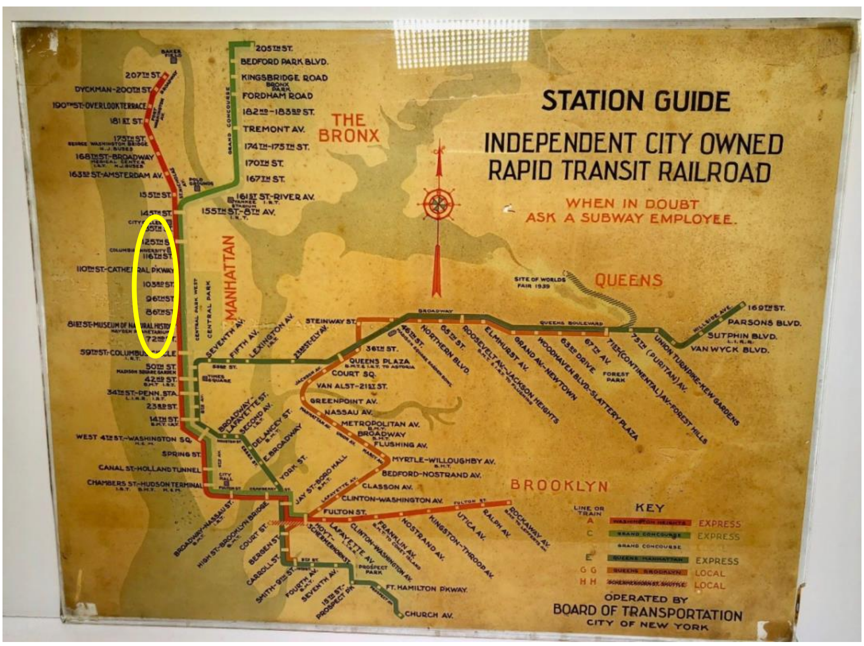
Initially, the ground transportation of New York started with electric cars and railroads. There was an interdependent relationship between the construction of ground transit and city expansion before the occurrence of the subway system. Around 1831, many businessmen requested that the New York State Legislation give the approval to the New York and Harlem Railroad Company and allow it to build a Manhattan rapid transit system (Fischer and Henderson, 311). Therefore, IRT (The Interborough Rapid Transit Company, purchased by the city in June 1940, along with the BMT—The Brooklyn-Manhattan Transit—and IND—The Independent Subway System) and BRT (the Brooklyn Rapid Transit) constructed the earliest BMT Broadway Line. On October 27th 1904, George McClellan, then New York's mayor, took the joystick and started the first ride of the New York City subway.

Figure 1: Part of NYC Subway Map from 2019 with Highlight of Common Track of Line 3 and 4



After that, around the 1920s, when the Independent was first conceived, there was severe overcrowding on the subways. Hence, on April 25th, 1921, the New York Transit Commission was formed to develop a plan to resolve the overcrowding and delays on the current lines, including “recapturing” existing lines from the privately owned rapid transit subway and elevated lines. (Feinman) On August 3, 1923, the New York City Board of Estimate approved the Washington Heights Line, an extension of the Broadway Line to Washington Heights, i.e., the birth of IND Eighth Avenue Line. On its opening day, the IND system had a relatively small subway car fleet of 300 cars, while the IRT had 2,281 subway and 1,694 elevated cars, and the BMT (Brooklyn-Manhattan Transit) had 2,472 cars”. On September 10, 1932, the Eighth Avenue Line opened from 207th Street to Chambers Street, inaugurating the IND. Most lines have four tracks with one local and one express track in each direction, except for the extreme north and south ends, where only the two express tracks continue. Internally, the Line A is composed of tracks A1, A3, A4, and A2 from west to east, running from approximately 800 at the south end to 1540 at the north end (measured in hecto feet).

Figure 2: Map of the IND System from 1939



Here, as the first Manhattan trunk line, the “Eighth Avenue Line” along with Cranberry Street Tunnel composite former Express A in 1939 IND system from 207th Street to Jay Street–Borough Hall. The Concourse Line here was the former Express C in 1939 IND system from 205th Street to 145th Street.

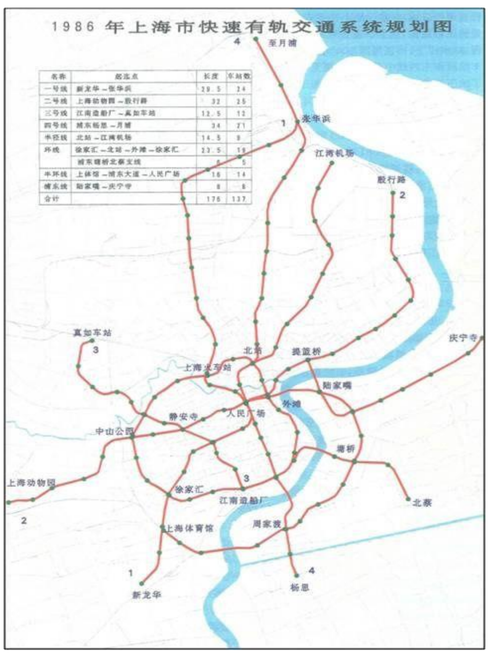
Therefore, I may conclude that at the very initial stage of railway design, common track of current Line A (former Express A) and Line D (former Express C) from 145 St to 59 St Columbus Circle were built on purpose to relief the overcrowding situation caused by private-owned subway lines, at a cost of $191 million (Feinman).

b) Shanghai Metro Line 3 and Line 4

The Shanghai Metro system is the world's second biggest metro system by route length after the Beijing Subway, totaling 676 kilometers (420 mi). It is also the second biggest by the number of stations with 413 stations on 16 lines after the Grand Central Terminal in New York Manhattan. But its first subway line Southern section of Line 1 (Shanghai South Railway Station – Xujiahui) did not enter trial operation until May 28, 1993.

Shanghai Metro Line 3 was modified from the former Hu'song (Shanghai-Wusong Wharf) Railway and Hu'hang (Shanghai-Hangzhou) Railway. The original purpose for Shanghai railway construction was to satisfy the need at war time, so commercial use was not included in its designated intention at first.

Figure 3: The Planning Map of Shanghai Rapid Railway Transportation 1986



In August 1986, Shanghai Commission of Science and Technology organized Shanghai Urban Planning and Design Research Institute and Shanghai Railway Institute (Corporated by Tong'ji University in April 2000) to conduct "The Feasibility Study of Developing Urban Railway to Relief Urban Traffic," which focused on the usage of Hu'hang Railway Inner Ring and Hu'song Railway. In May 1987, Shanghai Urban and Rural Planning and Environmental Protection Commission invited the Land Transport Bureau of the Ministry of Transport of France, the Municipal Planning Institute, and Shanghai Railway Institute to cooperate in a new round of "Preliminary Feasibility Study of Shanghai North-south Rail Transit Line," which was completed in September 1989.

In "A Study on the Early Completion of Metro Circle Line in Shanghai" published in 1997, it was proposed that the section of Caohejing to Baosteel section of Metro Line 3 should be constructed from 1996 to 2000. From 2000 to 2019, the southeast half ring of Subway Line 11 (namely the subway circle line) was constructed with an underground line, and the two ends of Hongqiao Road station transferred with rail transit Line 3. After 2020, the northwest half ring of Line 11 should be built to form a complete ring line.

Also in 1997, Mingzhu Line (current Line 3) began construction. Baoshan Road and Hongqiao Road were therefore preserved as a common interface. In 1999, the general plan for Shanghai Subway, designed by Systra (a multinational consulting and engineering firm) was put into Shanghai government's city general planning. In Systra's design, the former temporary common track became a permanent one, because common track cost lower budget than parallel track. This critical advantage would make line 4 (the original Mingzhu Line phase 2) be able to pass the approval of the government faster and therefore promote local economic development earlier. However, the government hasn't executed an adequate civil procedure law that supports the private companies entering the market of underground commercialization. With construction projects involving special fields, the government needs to invest in an enormous amount of funds as a stable asset. Once built, there is a constant need of management and maintenance, so retrieving the basic expenditure of construction for the government requires a very long term. Since most of the exploration is conducted by the government so far and the private sectors with more potential in financial investment have not given many opportunities of getting involved, the profitability of Shanghai City underground railway cannot be significantly boosted unless more market availability for private companies were approved. Thereafter, the common track of Line 3 and Line 4 was born due to the shortage of budget.

3. Hypothesis

According to the essay of David King, "Developing Densely: Estimating the Effect of Subway Growth on New York City Land Uses," the subway construction of New York did not lead but followed the residential and commercial growth of the surrounding areas. Yet, the order of development of the Shanghai subway system and the growth of the surrounding area seems to be the opposite. According to Xiying Sun and Junju's essay "The Effect of Shanghai Subway Construction to the Economic Growth of the Surrounding Area," after the construction of the subway, the population, industrial output, and housing prices of the area around the metro station have increased significantly.

Therefore, my theory is that such a difference in relationships between the construction of the subway and the economic development of surrounding areas leads to certain degrees of difference in the usage of surrounding lands, as well as affects the function mode of the subway system. As the New York subway was constructed a lot earlier than the Shanghai subway, the Shanghai subway produces a higher revenue due to its more various income sources, for example the bigger areas and facilities designed for commercial advertisements inside the station. Also, Shanghai metro takes the chronicle advantages of having many other cities' subway construction plans as references. Obviously, a later time sequence enables the city government to have more adequate cases to study from and have more time to design a comprehensive plan for the subway construction, based on the experience and possible problems run by the former subway networks.

4.Analysis of Data

I define y_i as the average ridership for station interval i, and define x_{ji} as the number of unit/buildings that serves different purpose, s.t. $j \in [1,7]$, where j=1 represents resident units, j=2 shops and supporting facilities like convenient stores and clinics and/or churches, j=3 restaurants, j=4 schools (without affiliated dormitory), j=5 parks and museums, j=6 office buildings, and j=7 shopping plazas correspondingly, and i represents the parameter of j for each station interval i. And β_i represents the corresponding coefficients of x_i.
The model I run is:

$$y_i = \alpha + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_7 x_{7i} + \varepsilon_i$$

a) Current NYC MTA Line A and Line D from 145 St Station to 72 St Station

In this study, I focus on ten stations on the common track of Line A and D, which are 145th St Station, 135th St Station, 125th St Station, 116th St Station, Cathedral Pkwy Station, 103rd St Station, 96th St Station, 86th St Station, 81st St Station, and 72nd St Station, with a layout shown in Figure 4.

Figure 4: Part of NYC Subway Map from 2019

According to Google Maps, from 145 St Station to 135 St Station, the subway is lined with residential buildings and supporting facilities and services (such as schools, clinics, restaurants, churches), convenient retail outlets, and private commercial offices. From 135th street station to 125th Street station, there are many residential buildings and supporting facilities and services along the subway. From 125th street station to 116th Street station, many residential buildings and supporting facilities and services are distributed along the subway. From 116th Street station to Cathedral Pkwy station, many residential buildings and supporting facilities and services are distributed along the way with a small number of office buildings. From Cathedral Pkwy station to 103rd Street station to 96th Street station to 86th Street station, numerous high-class housing and supporting facilities are distributed on both sides of the subway track. From 86th street station to 81st Street Station, there are many residential and supporting facilities. From 81st street to 72nd Street, upscale residences and amenities, museums, shops, and hotels are located along the way.

(Note: Due to the special circumstance that the east side of the subway from Cathedral Pkwy station to 72nd Street and the corresponding road on the ground are fully occupied by Central Park, most of the plots included in the statistics are solely West side of the road. This is not a typical urban design that can be easily generalized or compared. Therefore, parameter x_5 parks and museums will not be included in this part of analysis but will be included later in the expansion analysis of Shanghai subway common track interval.)

Figure 5: Average Land Usage Share Along NYC Common Track Line A&D

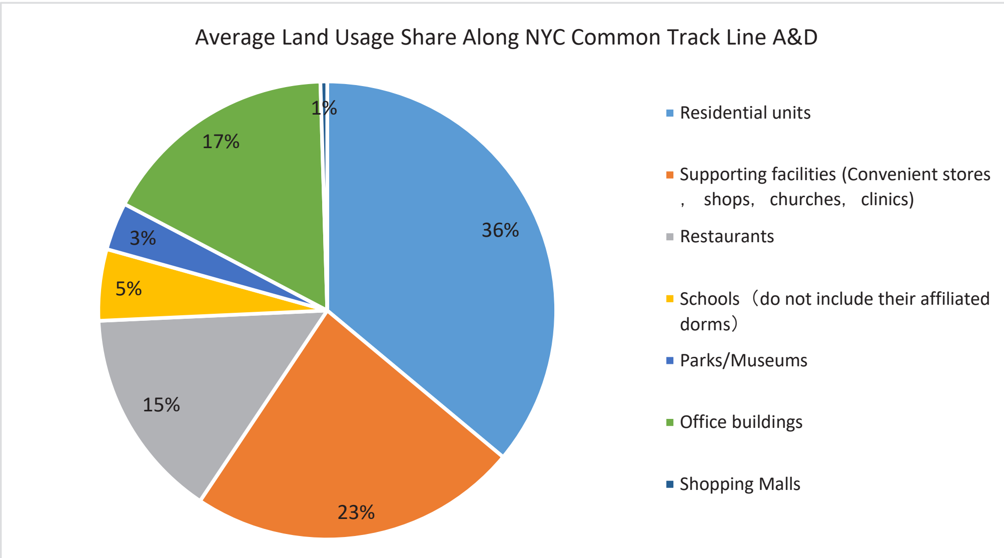


Table 1: Partial Data of Land Usage along NYC Subway Line A and D

	Residential units (number*21)	Supporting facilities	Restaurants	Schools	Office buildings	Shopping Malls	interval unit ridership
	x1	x2	x3	x4	x6	x7	y
145th St Station to 135th St Station	462	16	7	4	6	0	4604287
135th St Station to 125th St Station	672	14	5	3	1	1	5476259.5
125th St Station to 116th St Station	231	24	21	0	4	1	5748354.5
116th St Station to Cathedral Pkwy Station	315	12	21	1	9	0	2342367.5
Cathedral Pkwy	357	6	1	2	14	0	1947493.5

Station to 103th St Station							
103th St Station to 96th St Station	294	5	2	2	9	0	2268613.5
96th St Station to 86th St Station	336	11	1	6	15	0	3135750.5
86th St Station to 81th St Station	336	6	2	2	7	0	3820122
81th St Station to 72th St Station	147	3	2	1	5	0	3622324
Average	350.00	10.78	6.89	2.33	7.78	0.22	3662841.33

Note: parameter of parks and museums is excluded here.

For data in the column "Residential units" in Table 1, I adjust the initial collected number of residential buildings by multiplying it by 21, my evaluation of average households per building in a standard block. The sketch map of a standard block is shown as Figure 4. I derive this evaluation by first getting the average number of people who live on one side of buildings (μ_{side}) in a standard block 700 through dividing the average number of people living in one standard block (μ_{Block}) which is 1400 by 2, the number of rows of buildings in a standard block (N_{Row}).

$$\mu_{Side} = \frac{\mu_{Block}}{N_{Row}} = \frac{1400}{2} = 700$$

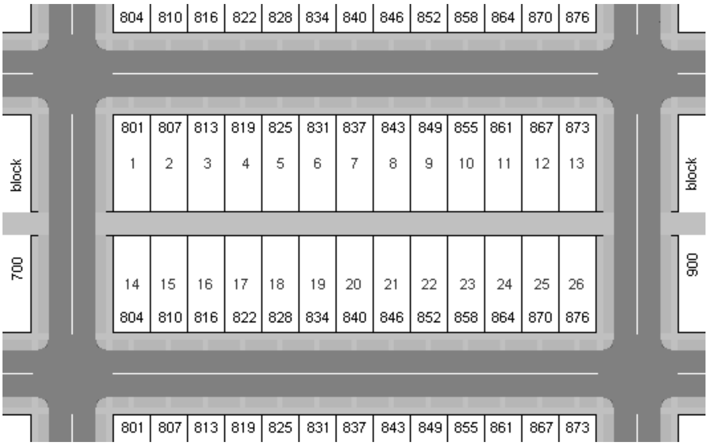
Then I divide the average number of people per side of a standard block by 2.45, the average people per household ($\mu_{\frac{Ppl}{HH}}$), to get the average household per side of a standard block ($\mu_{\frac{HH}{Side}}$), which is 285.71.

$$\mu_{\frac{HH}{Side}} = \frac{\mu_{Side}}{\mu_{\frac{Ppl}{HH}}} = \frac{700}{2.45} = 285.71$$

Since the average occupancy rate (r^{o}) in NYC is 95%, I multiply the average household per side of a standard block by 0.95 to have the adjusted average household per side of a standard block ($\mu'_{\frac{HH}{Side}}$) 271.43.

$$\mu'_{\frac{HH}{Side}} = \mu_{\frac{HH}{Side}} * r^{o} = 285.71 * 95\% = 271.43$$

Figure 6: Sketch Map of a Standard Block



Assume there are 13 residential buildings per side of a standard block ($N_{\frac{BLD}{Side}}$). I finally derive the average of households per building in a standard block ($\mu_{\frac{HH}{BLD}}$) to be 20.88, which is later rounded to 21, by dividing the adjusted average household per side of a standard block 271.43 by 13.

$$\mu_{\frac{HH}{BLD}} = \mu'_{\frac{HH}{Side}} \div N_{\frac{BLD}{Side}} = \frac{271.43}{13} = 20.88 \approx 21$$

According to the summary in Table 2, I found that residential units, restaurants, and office buildings as independent variables do not seem to develop positive coefficients with respect to the ridership of subway with common track as dependent variables, while supporting facilities like convenient stores, schools, and shopping malls do. However, the P-values of all independent variables except x6 (office building) do not reach the significant level of 0.05 to provide 95% confidence level in the result. Yet, on the other hand, the R Square of 0.96 shows that this multivariable linear model does a fair job describing the relationship between the ridership of common track subway and the land usage of surrounding areas. More specifically, the variance of its errors is 96% less than the variance of the dependent variable (ridership of common-track subway) and the standard deviation of its errors is approximately 80% less than the standard deviation of the dependent variable. Further analysis is conducted in part (c).

Table 2: Summary Output for Common Track Subway of NYC

Regression Statistics	
Multiple R	0.98037207
R Square	0.9611294
Adjusted R Square	0.88338819
Standard Error	447539.197
Observations	10

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	1.4857E+13	2.4762E+12	12.3631916	0.031985976
Residual	3	6.0087E+11	2.0029E+11		
Total	9	1.5458E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 90.0%*	*Upper 90.0%*
Intercept	4512104.88	704212.006	6.40731035	0.00770217	2270987.981	6753221.77	2854838.09	6169371.66
x1	-2228.7688	1624.83932	-1.3716857	0.26374455	-7399.732679	2942.19509	-6052.6062	1595.06864
x2	154030.807	69022.0488	2.23161743	0.11182837	-65628.15677	373689.772	-8403.1584	316464.773
x3	-75408.386	48456.5116	-1.5562075	0.21751703	-229618.6322	78801.8608	-189444.17	38627.397
x4	141378.259	201829.776	0.70048266	0.53406488	-500934.1662	783690.685	-333600.56	616357.075
x6	-211575.89	53062.7436	-3.9872777	0.02824287	-380445.2269	-42706.562	-336451.82	-86699.974
x7	477259.86	787300.698	0.60619768	0.58717728	-2028282.338	2982802.06	-1375544.8	2330064.53

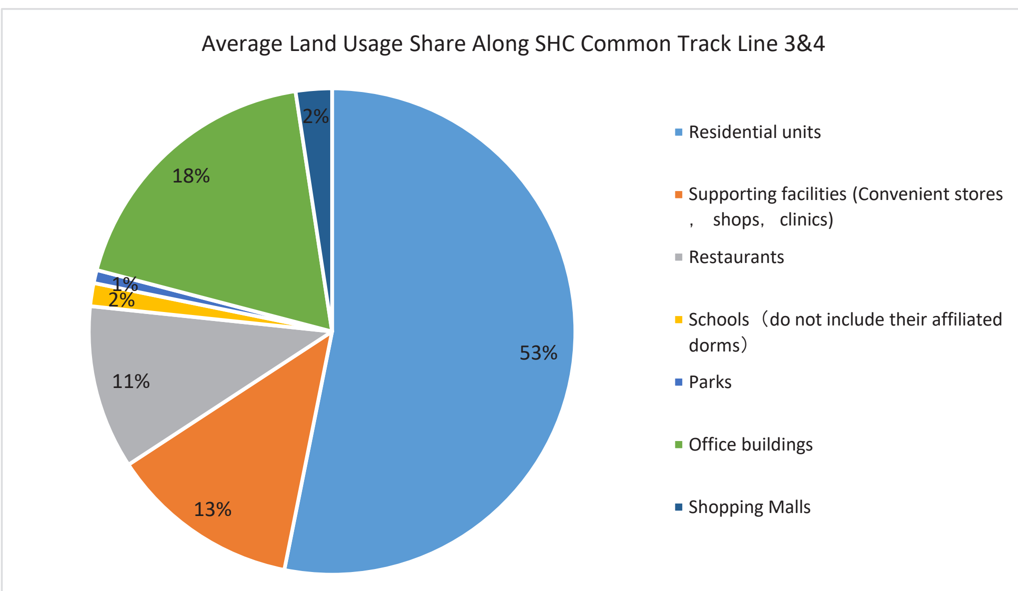
b) Current Shanghai Metro Line 3 and Line 4 from Shanghai Stadium Station to Baoshan Road Station

In this study, I focus on nine stations on the common track of Line 3 and 4, which are Hongqiao Road Station, Yan 'an West Road Station, Zhongshan Park Road Station, Jinshajiang Road Station, Caoyang Road Station, Zhenping Road Station, Zhongtan Road Station, Shanghai Railway Station, and Baoshan Road Station.

Image6: Partial Shanghai Subway Map from 2018

Taking Baidu Maps as a reference, from Hongqiao Road Station to Yanan Ist Road Station, there are residential buildings, convenient retail outlets, restaurants, and schools along the common track. From Yanan West Road Station to Zhongshan Park Road Station, more residential buildings and supporting facilities located closer to the former the station, and many large commercial centers, shopping plazas (such as Shanghai Raffles, Mengzhilong Shopping Park, Zhaofeng Square) and high-end residential areas surrounding the latter. From Zhongshan Park Road Station to Jinshajiang Road Station, similar is that the two subway stations are mostly near large commercial centers, shopping plazas, along the way many residential buildings. From Jinshajiang Road Station to Caoyang Road Station, office buildings, residential buildings and shopping buildings are distributed along the way. From Caoyang Road Station to Zhenping Road Station, there are many office buildings, houses and supporting facilities (such as restaurants, parks, and schools) along the way. From Zhenping Road Station to Zhongtan Road Station, there are many office buildings, residences and supporting facilities along the way. From Zhongtan Road Station to Shanghai Railway Station, there are many houses and supporting facilities along the way, and many hotels and catering near the latter. From Shanghai Railway Station to Baoshan Road Station, residential buildings and buildings are distributed along the way.

Figure 6: Average Land Usage Share Along SHC Common Track Line 3 & 4



Note: parameter of parks and/or museums is excluded here.

Table 3: Partial Data of Land Usage along Shanghai Subway Line 3 and 4

	Residential units (number*21)	Supporting facilities	Restaurants	Schools	Office buildings	Shopping Malls	interval unit ridership
	x1	x2	x3	x4	x6	x7	y
Hongqiao Road Station to Yanan Road Station	3780	11	16	1	12	0	4031673
Yanan Road Station to Zhongshan Park Station	864	6	5	1	10	3	5265825
Zhongshan Park Station to Jinshajiang Road Station	2538	11	8	2	11	1	5274609
Jinshajiang Road Station to Caoyang Road Station	756	2	1	0	5	1	5397036
Caoyang Road Station to Zhenping Road Station	1512	7	5	1	13	0	4793136
Zhenping Road Station to Zhongtan Road Station	648	8	5	1	22	1	3401421
Zhongtan Road Station to Shanghai Railway Station Station	1134	9	8	1	8	3	11082846
Shanghai Railway Station Station to Baoshan Road Station	2160	4	2	0	7	0	10996287
Average	1674.00	7.25	6.25	0.88	10.63	1.38	6280354.13

For data in the column "Residential units" in Table 3, I adjust the initial collected number of residential buildings. My methodology of evaluation can be elucidated as follows: I divide the number of residential buildings between each station (N_{BLD_i}) by the length (L_i, in kilometer) of common track of that interval, then I average them to get the evaluation of average residential building per kilometer ($\mu_{\frac{BLD}{km}}$) 21.

$$\mu_{\frac{BLD}{km}} = \frac{1}{8} * \sum_i \frac{N_{BLD_i}}{L_i} \approx 21$$

I also calculate the average population density of these areas (ρ_i) where the stations are located which is 3823 (people per square kilometer) and divide it by

Shanghai average people per household (μ_{HH}) 2.32 to get the average household density of the areas involved ($\overline{\rho_{HH}}$) 167.84 (household per square kilometer).

$$\overline{\rho_{HH}} = (\frac{1}{8} * \sum_i \rho_i) \div \mu_{HH} = \frac{3823}{2.32} = 167.84$$

Then I multiply the length of each common track between each station (L_i) by average household density ($\overline{\rho_{HH}}$) and average them out to get the average household per kilometer ($\mu_{\frac{HH}{km}}$) 2267.1.

$$\mu_{\frac{HH}{km}} = \frac{1}{8} * \sum_i (L_i * \overline{\rho_{HH}}) = 2267.1$$

By considering the average occupancy rate of residential buildings (r^O) which is 50%, I adjust the average household per kilometer ($\mu'_{\frac{HH}{km}}$) as 1133.55.

$$\mu'_{\frac{HH}{km}} = \mu_{\frac{HH}{km}} * r^O = 1133.55$$

Finally, I divide the adjusted average household per kilometer 1133.55 by my evaluated average residential building per kilometer ($\mu_{\frac{BLD}{km}}$) 21 to get the average household per buildings 53.98, which rounds to 54, my evaluation of average household per building for the neighborhoods along the subway span used in my calculation.

$$\mu_{\frac{HH}{BLD}} = \mu'_{\frac{HH}{km}} \div \mu_{\frac{BLD}{km}} = 53.98 \approx 54$$

Under this calculation procedure, I generate the summary output of Shanghai common track subway in Table 4, which has high R square value, corroborating the rationality of my evaluation model.

Table 4: Summary Output for Common Track Subway of SHC

Regression Statistics	
Multiple R	0.98731691
R Square	0.97479467
Adjusted R Square	0.89917869
Standard Error	895856.164
Observations	9

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	6	6.20765E+13	1.03461E+13	12.8913835	0.07372607
Residual	2	1.60512E+12	8.02558E+11		
Total	8	6.36816E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 90.0%*	*Upper 90.0%*
Intercept	2793756.13	1532497.834	1.823008208	0.2098751	-3800049.9	9387562.117	-1681115.4	7268627.71
x1	167.766345	487.439761	0.344178621	0.76353119	-1929.5177	2265.050363	-1255.5507	1591.08342
x2	2471882.5	477559.2064	5.17607549	0.03535723	417111.08	4526653.926	1077416.51	3866348.5
x3	-943392.32	223057.19	-4.229374169	0.05161461	-1903129.9	16345.3099	-1594716.1	-292068.54
x4	-6996421.3	1443644.475	-4.846360313	0.04003684	-13207922	-784920.4492	-11211842	-2781000.2
x6	-368061.57	73099.61466	-5.035068524	0.03725446	-682583.83	-53539.31231	-581511.39	-154611.75
x7	882424.559	348860.3653	2.529449163	0.12715908	-618600.44	2383449.562	-136242.68	1901091.8

c) Data Comparison Analysis

Note: x1 represents variable residential units (number*21), x2 represents variable supporting facilities, x3

Table 5: Comparative Analysis of OLS Regression coefficients of NYC and SHC

OLS (Ordinary Least Square) Regression Coefficient	NYC	Shanghai City	NYC/Shanghai City	NYC Average/Shanghai City Average
Intercept	4512104.88	2793756.13	1.62	-
x1	-2228.77	167.77	-13.28	0.21
x2	154030.81	2471882.50	0.06	1.49
x3	-75408.39	-943392.32	0.08	1.10
x4	141378.26	-6996421.30	-0.02	2.65
x5	-	-	-	3.12
x6	-211575.89	-368061.57	0.57	0.73
x7	477259.86	882424.56	0.54	0.16

represents variable restaurants, x4 represents variable schools, x5 represents variable parks and museums, x6 represents variable office buildings, x7 represents variable shopping malls.

From the ratio of average of each type of land use of NYC and Shanghai City, I notice that the number of residences, office buildings, and shopping malls are larger in Shanghai, while land usages as supporting facilities, restaurants, especially schools and parks and museums are seen more often in NYC. As shown in the Table 5, the coefficients for x_1 residential units (NYC:-2228.77, SHC:167.77) and x_4 schools (NYC:141378.26, SHC: -6996421.30) show an opposite direction in positive and negative values of coefficients for these two cities, suggesting that the common tracks of NYC MTA Line A and D and Shanghai Metro Line 3 and 4 develop very different relationships between subway intervals of common track subway and the economic activities of surrounding lands, especially in the residential and schooling usages, due to the differences in the causality of subway construction and surrounding lands.

While leaving one parameter x_5, parks and/or museums, out of regression in the comparison, I conclude that for New York City, the total effect of residences, schools, and park and/or museum to common track ridership is that people prefer not to take common track subway as transportation, while people in the similar situation in Shanghai City prefer taking common track subway.

Considering tourism and residence together, my explanation to the negative coefficient of NYC x_1(-2228.77) is that on the one hand, due to the compactness of NYC subway station design of common track Line A and D with museums, parks, and hotels, people may choose to walk over taking the subway at lower cost but same or shorter time. On the other hand, the ticket price of the NYC subway is relatively higher than the Shanghai subway. According to my estimation of relative ticket cost, by comparing the results from dividing two standard one-way subway tickets of NYC ($2.75*2=$5.5) and Shanghai common track subways (¥3*2=¥6) each by their 2020 individual daily incomes ($34,386/365=$94.208 and ¥171,884/365=¥470.915) to get the relative ticket cost in each city. Two standard one-way NYC subway tickets would cost approximately 5.838% of an average New Yorker's daily income, while the same subway service in Shanghai City would cost 1.274% of an average Shanghainese daily income. On top of that, the common track subway in NYC delays more often than Shanghai Subway, because of different auxiliary mechanisms of common track subway schedule and facilities. NYC subway depends on "train crews" who monitor and submit the signal of opening and closing the gates of each car on the train, but Shanghai City subway employs terminal attendants who monitor and assist the passengers to keep order on the platform. And NYC lines carry more passengers than their designs allow, and extra trains cannot be added. Furthermore, Asian railway systems take delay analysis to "levels hitherto unseen in U.S. agencies" (Reddy, Lu, and Wang) These wedges make common track subway in NYC less punctual. All the reasons combined could conclude that compared to Shanghai City, tourists and residents in NYC prefer less to take the subway with common track, so that residential buildings have negative correlation with the ridership of subway with common track. Yet, considering the high p-values of the model examined (NYC:0.76353119, SHC:0.76353119), the actual correlation remains to be further confirmed.

Differences in the coefficient of x4 schools and the ridership of common-track subway (NYC:141378.26, SHC: -6996421.30) can be easily explained in the context of different cultures. Generally, in China, it is quite popular among the parents to buy or rent a house near their child/children's school, especially for those with child/children in junior high and senior high school and for those who live in a developed city like Shanghai. Also affected by Chinese domestic historical policy of encouraging students to attend the schools that are closed to their home "学区房"(xue'qu fang), from 1986 the demand for such housing has only been increasing, which drives the price of *xue'qu fang* all the way up, disturbing the housing market. For the exact same reason, the government has recently announced to cancel the market of *xue'qu fang* in 2024. So, this relationship of school and the ridership of subway could be completely changed for long run in the future, but the remaining momentum of housing market and cultural convention of *xue'qu fang* as well as the delayed effect of new policy might not change the current relationship too much. Due to different education systems, Chinese students generally have a relatively heavier workload and longer class schedule compared to the students in the US. By living near the schools, parents can cut students' commuting time and save more time for the students to get rest at home. As a result, there's not so much need for public transportation like subways for students in Shanghai, many of whom can walk themselves to school instead. While in NYC, where traffic jams are heavy around morning and afternoon peak, public transportation like the subway is still a popular choice for students. Therefore, the opposite value of coefficient for factor x_4 can be understood.

Surprisingly, the intercepts (NYC:4512104.88, SHC:2793756.13) shows that the NYC subway interval carried approximately 1.62 times more annual ridership than the Shanghai City interval I studied, contrasting the fact that Shanghai City has 3.14 times more population. Yet, this can be explained once I take the extraordinarily high population density of NYC into consideration. NYC has 27,000 people per square mile, compared to 10,600 people per square mile in Shanghai City.

Table 6: Comparative Analysis of Common Track and Individual Track Design

	Pros	Cons
Common track	1.Costs lower budget and covers larger area.	1.Complex map may confuse passengers.
	2.Suburban lines may operate on different frequencies from urban lines, so it ensures the frequency of trains in urban areas.	2.Requires expensive construction to achieve railway connection.
	3.Taking up a smaller space.	3.Less punctuality and therefore increases subway traffic pressure and lower efficiency.
Individual track	1.If the actual ridership is large, it can effectively improve the frequency of trips and relieve the traffic pressure.	1.High construction cost.
	2.Clear layout rarely confuses passengers.	2.Taking up more space
	3.More punctual	3.If the ridership is not large enough, it could induce loss.

For factors x_2, shops and supporting facilities, and x_6 office buildings, my understanding toward the opposite result values is that many workers in both cultures choose to get off at the station where the shops and convenient stores at, because they can have a quick breakfast or lunch near their office buildings and then walk to office instead of taking the train again. According to one report of reputable Chinese domestic media en.people.cn (人民网) from 2019, it is revealed by a study of Lai Yang, an expert in the China General Chamber of Commerce, that an increasing number of white-collar workers buy ready-made meals in convenience stores and fewer cook at home. Besides, based on the data from the statistical website Statista, the Industry revenue of "convenience stores" in New York has been continuously rising from 2012 to 2024 (except for 2020 due to the strike of Covid-19 to the macroeconomy). Also, punctuality is another reason people might choose individual track subways over the common track ones. I also consider the difference in the magnitude and infer that because this kind of "getting off earlier" behavior is more popular in today's culture of China than the US, so that more people choose to get off at the station surrounded by shops in the common track subway interval Shanghai.

Differences in the magnitude of parameter x_3 (NYC: -75408.39, SHC:-943392.32), restaurants can also be explained in the context of different cultures. Here, I define "restaurant" as the catering service which operates individually outside the shopping mall. Since the majority of the consumers of the diner inside the shopping mall are young people, senior consumers who are more old-fashioned as well as wealthier who would drive and dine at the restaurants I monitored. Hence, the factor of restaurants x_3 has negative correlation with the ridership of common track subway. Similarly, variety in the magnitude of coefficients can be unveiled if I see it under the filter of culture. Senior Chinese are more likely to differentiate themselves from the other social groups by dining at individual restaurants to show their exquisite styles of life.

Finally, the coefficients of factor x_7(NYC:477259.86, SHC:882424.56) shopping malls are not hard to understand. Shopping malls attract a high volume of visitors and therefore ridership. I noticed that the coefficient of x_7 of Shanghai is almost 2 times that of NYC, which means that the business model of shopping malls in Shanghai is more efficient than that of NYC. All the shopping malls I studied in Shanghai City are built as interior pedestrian streets, where consumers enjoy a more composed and comprehensive shopping experience while being less disturbed by the changes in weather. But still, if taking the p-values and confidence level of the result into consideration, (p-value of NYC: 0.58717728, SHC:0.12715908) the solidity of this conclusion needs future study to verify.

(i) The categories of subway commercial spaces

"The commercial space of subway station is generally divided into three categories: aisle shops, underground commercial street and ground commercial center" (Liu, 8).

One type of these commercial spaces can be seen the most often in the subway of New York, which is aisle shops. Aisle shop refers to the shop built in the subway station public passage or near the subway ticket station. And usually, they are distributed in a dot shape. (Liu) In the case of New York City, aisle shops are presented as vending machines and newsstands. The Turnstyle project at Columbus Circle is the newest try of MTA commercial space used as underground commercial street.

The second type of subway commercial space is categorized as underground commercial street, which refers to the underground shop group connected by underground passage near the subway station. They usually feature a strip shape along the passageway of the subway station, and usually connect the subway stations with the nearby shopping centers. In New York, there is only one such underground commercial street——The TurnStyle Underground Market. TurnStyle Underground Market (image 7) is a typical underground commercial street as shown on the map (image 8) opened on the 6th of December 2018.

Image 7: TurnStyle Underground Market at Columbus Circle, New York City, USA

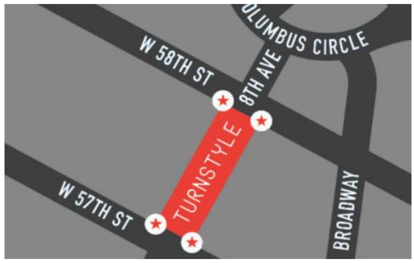


Image 8: The map of TurnStyle on the official website

The most common type of underground commercial space in Shanghai is the ground commercial center, which is also an imperative part of Shanghai's construction goal of building a multidimensional traffic complex. During the common track interval I studied, there are 9 traffic complex shopping malls like Raffels City Shanghai Changning which is connected to line 2, 3, and 4, and Shanghai Global Harbor which is connected to line 3, 4, and 13. However, there is only one such traffic complex in NYC but not included in my focused interval, which is Grand Central. Since its opening on February 2, 1913, Grand Central has been one of the world's most visited tourist attractions and a famous landmark and transportation hub in Midtown Manhattan

d) Expansion Analysis

Table7: Data of Land Usage along Shanghai Subway Line 3 and 4 (factor x5 included)

	Residential units (number*21)	Supporting facilities	Restaurants	Schools	Parks and Museums	Office buildings	Shopping Malls	interval u ridership
	x1	x2	x3	x4	x5	x6	x7	y
Hongqiao Road Station to Yanan Road Station	3780	11	16	1	2	12	0	4031673
Yanan Road Station to Zhongshan Park Station	864	6	5	1	0	10	3	5265825
Zhongshan Park Station to Jinshajiang Road Station	2538	11	8	2	0	11	1	5274609

Jinshajiang Road Station to Caoyang Road Station	756	2	1	0	0	5	1	5397036
Caoyang Road Station to Zhenping Road Station	1512	7	5	1	1	13	0	4793136
Zhenping Road Station to Zhongtan Road Station	648	8	5	1	0	22	1	3401421
Zhongtan Road Station to Shanghai Railway Station Station	1134	9	8	1	0	8	3	11082846
Shanghai Railway Station Station to Baoshan Road Station	2160	4	2	0	1	7	0	10996287
Average	1674.00	7.25	6.25	0.88	0.50	10.63	1.38	6280354.13

I ran the full regression including the parameter x_5 of parks for the data of Shanghai subway (Table 6) and got a statistically significant result as shown in Table 7 and Table 8. After adding the x_5, the coefficient of x_1 (-684.9223) shows a drastic change in value. The big, positive value in the coefficient of parks and museums (x_5:1747861.85) shows that land usage of parks is the reason that the total effect of residence and parks is positive in the earlier regression. Additionally, by referencing the data of annual tourism of Shanghai museums in 2020 which is 9,820,000, I conclude that in Shanghai, parks and museums are a big attraction for tourism, which increases the ridership of the common track subway. However, if I consider the effect of residence alone, it does not boost ridership of the common track subway.

Table 8: Summary Output for Common Track Subway of SHC (Factor x5 included)

Regression Statistics	
Multiple R	1
R Square	1
Adjusted R Square	1
Standard Error	1.7462E-10
Observations	9

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	7	6.36816E+13	9.09738E+12	2.98341E+32	4.4578E-17
Residual	1	3.04932E-20	3.04932E-20		
Total	8	6.36816E+13			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 90.0%*	*Upper 90.0%*
Intercept	2965431.31	2.99655E-10	9.89616E+15	6.433E-17	2965431.31	2965431.31	2965431.31	2965431.31
x1	-684.9223	1.5113E-13	-4.53202E+15	1.40472E-16	-684.9223	-684.9223	-684.9223	-684.9223
x2	2546972.29	9.36609E-11	2.71936E+16	2.34107E-17	2546972.29	2546972.29	2546972.29	2546972.29
x3	-1051472.1	4.59602E-11	-2.28779E+16	2.78268E-17	-1051472.1	-1051472.1	-1051472.1	-1051472.1
x4	-5965021.8	3.1527E-10	-1.89204E+16	3.36473E-17	-5965021.8	-5965021.8	-5965021.8	-5965021.8
x5	1747861.85	2.4091E-10	7.25524E+15	8.77462E-17	1747861.85	1747861.85	1747861.85	1747861.85
x6	-422382.95	1.60962E-11	-2.62412E+16	2.42603E-17	-422382.95	-422382.95	-422382.95	-422382.95
x7	1018848.22	7.05528E-11	1.44409E+16	4.40844E-17	1018848.22	1018848.22	1018848.22	1018848.22

5. **Prospective of Future Urban Subway Construction**

From my study results so far, I feel the strong and continuous influence of culture on many perspectives. In the field of urban subway construction, acknowledging and paying attention to the differences in the culture and society is as important as studying the history of different urban railway construction.

Hence, based on my study I call for several characteristics in the future urban subway construction to improve efficiency and raise ridership:

- New York City:

(1) Employ individual track construction around residential intervals as Ill as school intervals, because individual track design is more punctual and more efficient at relieving the traffic pressure.

(2) Common track is acceptable around the intervals of shops, convenient stores, and shopping malls.

(3) Interval with concentrated restaurants and/or office buildings can still use common track design to save the budget.

- Shanghai City:

(1) Individual tracks can replace common tracks around residential areas, shops and supporting facilities like convenient stores and office buildings.

(2) Common track can be used on the interval with concentrated shopping centers, parks and museums, restaurants, and school areas. The reason for

employing common track on the first two types of land usage is to relieve traffic, while the reason for the construction of the latter two is to lower the budget.

(3) Adjust more encouraging civil laws regarding private companies participating in activities of exploring underground commercial complexes.

6. Conclusion

The interpretation of the comparison result should be avoided without taking cultural or historical context into consideration. By analyzing from the construction background in terms of governmental policies and social-economic conditions to the unignorable impacts of cultural and ethnic discrepancy, the quantitative evaluation of the relationships between the common track subway and various types of surrounding neighborhood is made.

NYC metro, even without the prevalent exercise of subway-orientated underground commercial spaces, still developed strongly positive relationships with some types of the surrounding areas, which according to the result of this study are supporting facilities, schools, and shopping malls. The common track subway of NYC has been releasing the traffic density and functioning well with the fastforward pace of the city, but there also exists the problem of inefficiency in exploration of potential underground commercial space, which is inevitable due to historical city planning. While SHC metro on the other hand, actively utilizes a variety of underground commercial spaces with less common track designs. The subway system also maintained a positive relationship between various land usages. And with a relatively more affordable ticket price and more modern operating system, the common track of SHC incurs higher efficiency in terms of productivity. Yet, the limit of this study falls into the statistical significance of some of the results. Hence there still need future study to confirm the accurate relationships between common track design and the neighborhood areas.

Both two metropolises are catering intense economic activities every day. The subway as one of the most important components of city mobility determines the lifeline of economic growth and therefore is worthy of cities' close attention and continuous investment.

Abbreviation:

NYC: New York City
SHC: Shanghai City
OLS: Ordinary Least Square
IRT: The Interborough Rapid Transit Company
BMT: The Brooklyn-Manhattan Transit
IND: The Independent Subway System
BRT: the Brooklyn Rapid Transit

Conflict of Interest:

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References:

Alla V. Reddy, Alex Lu, and Ted Wang. “Subway Productivity, Profitability, and Performance--A Tale of Five Cities”. Transportation Research Record: Journal of the Transportation Research Board, No. 2143, Transportation Research Board of the National Academies, Washington, D.C., 2010, pp. 48–58. DOI: 10.3141/2143-07. Accessed 10.24.2021.

Feinman, Mark S. Feinman. “History of the Independent Subway”. https://www.nycsubway.org/wiki/History_of_the_Independent_Subway. Paragraph 2. Accessed 2.14.2022.

Fischler, Stan. and Henderson, John. The subway and the city : celebrating a century. Syosset, N.Y. : Frank MerrIll Incorporated c2004. P311. Accessed 2.14.2022.

Image.8:https://www.turn-style.com/. Accessed 6 August 6, 2020.

Intro of Shanghai Metro: https://en.wikipedia.org/wiki/Shanghai_Metro. Accessed 8.7. 2020.

King, David. “Developing Densely: Estimating the Effect of Subway Growth on New York City Land Uses.” Journal of Transport and Land Use, vol. 4, no. 2, 2011,pp.19–32. JSTOR, https://doi.org/10.5198/jtlu.v4i2.185. Accessed 1.Aug.2020.

Liu, Xing. “The design planning of ground commercial space of city subway” 2010. https://Inku.baidu.com/view/b96ef41ac5da50e2524d7f38.html. Accessed 5 Aug. 2020.

Spivack Caroline. “MTA to replace subway newsstands with vending machines across the city18: High-tech vending machines could replace your local subway newsstand”. Mar 4, 2019. https://ny.curbed.com/2019/3/4/18249824/mta-nyct-replace-subway-newsstands- vending-machines. Accessed 4 Aug. 2020.

Sun Xiying and Ju Jun.“The Effect of Shanghai Subway Construction to the Economic Growth of the Surrounding Area”. *Modern Economic Information*. 2016. Issue 33. Accessed 7.21.2022.

Statista.https://www.statista.com/forecasts/1206880/convenience-stores-revenue-in-ne w-york. Accessed 4.1.2023.

Ye, Tong. “Shanghai rail transit on the cover of comprehensive development Practice and exploration”. DOI:10.14052/j.cnki.china.metros.2021.03.012.“

Extracted from Ye’s speech at “Urban Railway Transportation TOD Comprehensive Development Forum ”. https://www.doc88.com/p-34987136003704.html. Accessed 5.1.2022.

Web editor: Hongyu, Bianji. http://en.people.cn/n3/2019/0520/c90000-9579887.html

Tipping Behavior and Delays: Evidence from Chicago’s Transportation Network Providers

Gustavo Pellegrini Dias

University of Chicago

Abstract

In this paper, we consider evidence from 2.2 million rideshare trips hat occurred in the city of Chicago between the years of 2018 and 2022 o examine the theory that service quality influences how much a client ips. In an app ride, the passenger’s experience depends almost entirely on their interaction with the driver. Therefore, if the tipping behavior is letermined by the quality of a server’s work, this relationship is more ikely to appear in rideshare than in most other settings. We run a linear egression between the trips’ deviation from average path time and the ndicator function of tips controlling for fixed effects of the different oute characteristics to estimate the effect of travel time delay on a client’s probability of tipping. Our study finds evidence of a statistically significant negative relationship between trip delays and tips. The findings of our study also suggest that arriving on time does not represent a significant difference for the driver’s annual income, and that the effect of loss aversion on tipping behavior depends on the magnitude of the leviation from the average path time.

1 Introduction

1.1 Research Question and First Thoughts:

The main research question this paper addresses is “How does an increment of a minute delay on a rideshare trip affects the probability hat a passenger will tip the driver?”. We started this paper expecting a statistically and economically significant negative correlation and causal relationship between these variables at the driver level. And though the study confirmed the statistically significance of the negative correlation we expected, the economic implications of our findings were surprisingly lower than expected.

1.2 Of The Paper's Structure

To answer the question posed above, we structured this paper in 6 sections. In order of appearance, they are: 'Introduction', 'Theoretical Framework', 'TNP and Descriptive Data', 'Empirical Framework', 'Results', 'Discussion and Conclusion'. The first section, the 'Introduction', concerns itself with a description of the paper's structure and a brief exposure of the main motivation and ideas relevant to the investigation in this paper. Next, the 'Theoretical Framework' is divided into 3 subsections: a review of the tipping behavior literature, the construction of the causal theory linking tip probability and trip delays, and a description of the control variables selected for our analysis.

The 'TNP and Descriptive Data' section includes a breakdown of the Transportation Network Providers (TNP) database from which we extract all information associated with the ride-hail trips used in our analysis. The second part of this section provides plots and tables that describe the main variables and controls in this study. This includes frequency tables and scatter plots illustrating the co-relationship between variables such as tipping probability and household income.

The second half of the paper starts with the 'Empirical Framework', where we detail the model used to estimate the causal effect of delays on tipping probability. Following this is a section where we report the results of our model. Next, we consider the impact of these findings in the discussion of whether tipping behavior can be modeled as a rational behavior or not. Finally, we present our concluding remarks, emphasizing the importance of observing more diverse and generalizable tipping behavior for the development of the Behavioral Economics field.

1.3 Of The Paper's Main Motivation and Ideas

From an economic standpoint tipping is a very intriguing behavior as it suggests that consumers are opting to increase their costs for no clear increment in their own utility. Tipping therefore poses a challenge to models of consumer behavior that rely on the Rational Choice Theory (RCT). In the RCT's framework, rational individuals are assumed to always make choices that maximize their net benefits, prompting social scientists to ask questions such as "why pay extra?" and "why do people leave money to strangers when they are not legally obligated to do so?" (Azar, 2005; Lynn 2010).

This paper tries to answer these questions by focusing on evidence of tipping from the ride-hail industry. Historically, studies on tipping behavior have centered around evidence from randomized control trials in restaurants. However, the choice of analyzing data from a rideshare service may be more appropriate for testing the relationship between tips and the amount of value generated to a client by a server's labor. The experience in an app ride depends almost fully if not entirely on the interaction between a driver and the passenger. Therefore, if the tip decision is determined in any way by the quality of a server's work, this relationship is much more likely to show itself in ride-hail than in settings such as restaurants.

And, assuming time is the determinant factor in a person's decision to order a ride through an app, there exists a clear measure of a driver's service quality in terms of how many minutes they are able to save their passengers in their commute. We can derive this measure by taking the difference between a trip's travel time for a particular route and the average travel time of trips through that same route. The outcome variable we are interested in this case, is an indicator function of tipping where 0 represents "tip in dollars = 0", and 1 represents "tip in dollars > 0". With these variables at hand, we can run a regression with selection on observables to estimate the effect of trip delays, or service quality on the probability of tipping.

The data which we use to run our analysis comes from the Transportation Network Providers (TNP) database publicly available in the Chicago Data Portal. Particularly, we investigate the subset of ride-hail trips that occurred on weekdays, from 8 AM to 10 AM. This filtering of our data is important not only because we get to observe trips at similar conditions of traffic, but also because, at this date and time, a trip's pickup census tract is more likely to be the tract where the passenger resides. This allows us to crosswalk the TNP data with the results of the 2020 American Community Survey (ACS) with reasonable accuracy. This adds relevant data for our regression controls that should raise the internal validity of our study at a very low cost to its external validity.

In the final section of this paper, we discuss the main finding of our analysis: statistically significant evidence of a negative effect on the likelihood of passenger tipping caused by trip delays, plus some other systematic characteristics of passengers tipping behavior we observe. But for now, we will concern ourselves with detailing the historical theories on tipping behavior, their limitations, and the contributions this paper makes to the field. The next section of this paper contains three subsections. First, we give a review of the literature on tipping behavior, careful construction of the causal theory between tip probability and trip delays, and the reasoning behind each observable selected as a control for our regression.

2 Theoretical Framework

2.1 Background Literature

Traditionally, answers to the question of why people tip have been divided between explanations founded on RCT and explanations founded on the Social Norms Theory (SNT). A counterpart to RCT, the SNT proposes that human behavior is often determined by unwritten rules, that is cultural expectations of how one should behave. These societal expectations overcome individuals' will to maximize their utility, and that would explain the existence of some irrational behaviors such as tipping. In truth, most of the recent literature on tipping behavior concludes that tips are a product of social norms and that economic models will always fail to answer such questions unless they account for the effect of social pressure and feelings in an agent's economic decision (Azar, 2020). Nevertheless, there are still those that claim tipping could have a rational explanation. A particular rational argument proposed by Bodvarsson and Gibson that remains to be tested is that tipping exists because it is an efficient pricing strategy for businesses in which customers have more information than employers about a worker's contribution to a firm's revenues (Bodvarsson and Gibson 1988; Jacob and Page 1980). And from the passenger's perspective, the existence of tips is explained as a rational decision to reward good drivers and punish bad ones, increasing the expected value of their next ride. Tipping, then, would be the result of consumers acting like a firm, monitoring and pricing the marginal product of a worker's labor. Of course, this functional explanation implies that tipping varies systematically according to service quality, server friendliness, and other variables through which a server can add value to a customer's experience. And it is exactly this variation that we intend to test in our analysis.

To understand whether Bodvarsson's and Gibson's theory has any validity, data on the relationship between the probability of tipping and labor-generated value to a client must be analyzed. Previous studies have tried to shed light on why people tip by comparing tip sizes with variables such as service quality (Lynn, 2003), dining-party size (Pearl, 1988), and customer frequency (Lynn and Grassman, 1990). A 2003 paper by Michael Lynn compared randomized service quality surveys in restaurants to tip sizes and found a weak correlation between satisfaction with service and tipping. According to Lynn, evaluations of service quality accounted for less than 2% of the variability in tips expressed as a percentage of the bill. In Thrane et al. 2020, researchers performed a survey experiment with Norwegian students to test peer effects on tipping behavior. They find that students are 12 to 14% more likely to tip when their friends tip. This study also found a significant effect of perceived service quality and tipping, which is inconsistent with the previous findings in Lynn, 2003. A 2021 paper by Brewster et al. tests the effects of wearing a mask on tipping. The authors conclude from their results that wearing a mask is not likely to, on average, have a meaningful effect on how much restaurant customers tip their servers.

But it was bill size, out of all the variables tested by recent research, that showed the strongest correlation with absolute tip. Correlation with bill size was also the most consistent result, showing up as the best predictor of tips in three independent studies (Lynn et al. 1993; Thrane et al. 2020; Bodvarsson et al. 1997). This suggests that a customer's gratuity could be in fact the result of a hidden social agreement in which tips should be a certain percentage of the bill size. That would explain why when the bill size increases, so too does the tip in dollars.

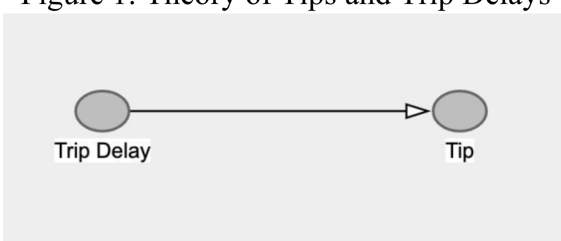
These studies, however, have focused mainly on restaurant settings. There are many issues that arise from this limitation. In a restaurant, it is very common for a waiter to serve more than one customer at once, or for a table to be served by more than one waiter during the same meal. This makes it difficult to measure the amount of value a waiter generated for one single customer. Additionally, a client's experience in a restaurant is much more complex than the waiter or waitress service can account for. There are many other variables at play such as the ambiance of the restaurant and the chef's quality. All of these factors could mean that customers are not actually apt to discern how much value a worker adds to their experience in a restaurant as Bodvarsson 1988 and Jacob 1980 suggested. This raises an important question of whether the persistent tip percentage that these studies observed is really a product of social norms or rather a consequence of clients' lack of information. Perhaps, customers resort to a tipping norm that varies little with their actual experience not because of unwritten social rules, but because they don't have enough information to estimate the value of a worker's labor, just like the firms. This creates a necessity for studies that observe tipping in settings where clients have a better understanding and can quantify the impact of a server's labor on their utility more precisely.

And this is where the choice of ride-hail data enters. Ride-hail is one of the few services that sits at the intersection between industries where firms cannot monitor their worker's activity and clients have full information about the value that a server adds to their experience. Especially when it comes to the value of time, the ride-hail industry concentrates on economic interactions in which the client is fully aware of the time input necessary to perform a service. This is information that consumers don't have access to in most other markets. Consumers do not know how long it takes for a firm to produce a fridge, a television, or a car. But they have a pretty good idea of how long it takes a driver to take them to their home or their workplace. In the next section, we translate the rational choice theories of tipping behavior to the context of ride-hail service, using time as the main measure of service value.

2.2 The Theory of Tips and Delays

The causal theory we want to test in this paper is that delays in rideshare trips cause passengers to not tip their drivers. The main variables that concern our theory are the binary tip and the deviation on average trip duration. The diagram below illustrates the theory of change we propose, where tipping is the dependent variable and deviation on average trip duration is the independent variable.

Figure 1: Theory of Tips and Trip Delays



Ceteris paribus, passengers should derive a greater utility from trips that are faster. Therefore, a good measure of how drivers create value for their passengers' is how much of their passengers' time they can save by covering a trip in less time than expected. Similarly, a good measure of how drivers destroy value to their passengers' experience is how much of their passengers' time they waste by covering a trip in more time than expected.

The causal mechanism that we suggest is the passenger's rational decision to reward good drivers and punish bad ones, increasing the expected value of their next ride. Better drivers are the ones who arrive earlier than expected, thereby generating more value for the passenger, and bad drivers are the ones that arrive later.

Whether someone gave a tip or not to their driver is estimated by an indicator function that assigns 0 to every ride that had a tip equal to $U\$0.00$ and 1 to every ride that had a tip greater than $U\$0.00$. Estimating the deviation on average trip duration associated with each trip requires a bit more work. We start by grouping trips that have a similar path. We create subsets of trips that have a common starting Census tract and a common destination Census tract. Then, we calculate the average trip duration for each of these paths. Finally, we subtract the average trip duration of a path from the total duration of each of the trips that belong to that path subset. This will relate each trip with the difference between their duration and the average duration of a rideshare trip along that path.

This theoretical design needs to account for some important confounding variables. These variables can be divided into two groups: those related to aspects of the trip itself, and those related to passenger's characteristics. In the following section, we will describe all observable confounders we wish to control for in order to improve the internal validity of our analysis.

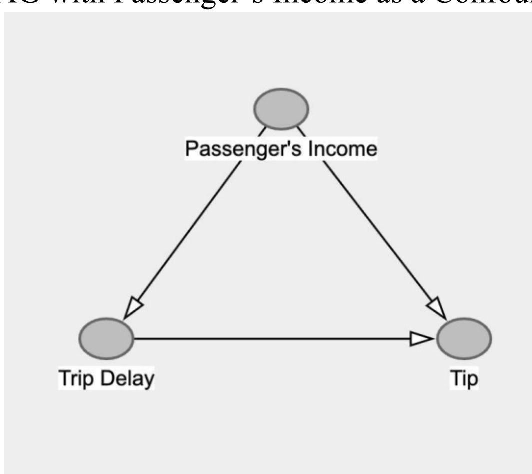
2.3 Controlling for Potential Confounding Variables

Before detailing the confounding variables for which we were able to control, it is important to note the variables that we could not control for. The delay of a trip is not always a result of poor driving, but a product of exogenous circumstances that the passenger is fully informed about. Because of that, these are elements that affect the pricing of the tip, but we simply do not have the data to account for it in this study.

Surrounding traffic, weather conditions, and roadblocks are just a few examples of such variables. Filtering the rides to a short period of time and comparing rides that occurred within very similar roads was our best way to account for them. Having said that, we will move on to the variables that we were able to control.

Passenger's Income: We began this paper by defining tips as voluntary payments that a client makes to a worker after a service is performed. Therefore, there are good reasons to believe that the decision to tip is determined by a passenger's budget to some degree. There are also reasons for us to believe that a passenger's budget can affect their trip's delay. For instance, it could be argued that people who have lower budgets reside in neighborhoods that have less reliable roads, or that drivers want to take extra caution when they drive by.

Figure 2: DAG with Passenger's Income as a Confounder Variable



Income is a variable that does a good job of indicating an individual's budget size. Therefore, our analysis will control for a passenger's income. However, as we will see in the next sections of this paper, the data publicly available on ride-hail trips does not allow us to identify the personal characteristics of passengers, including their income. Instead, we will use Census data to generate an adequate estimate of the income of each passenger.

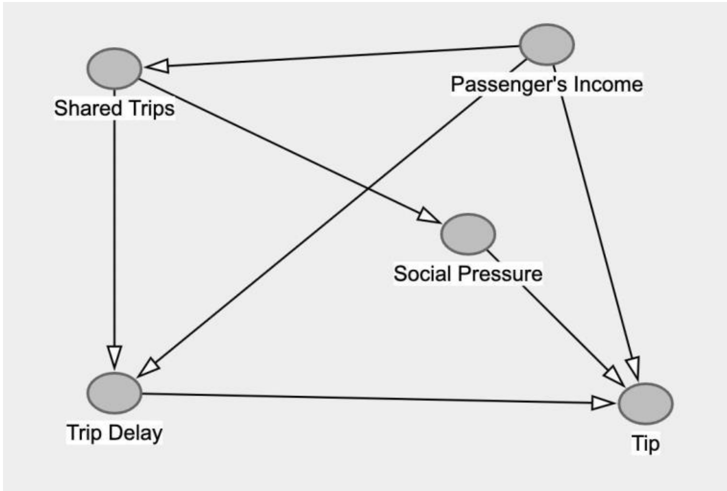
A theoretical argument that permits us to join income data from the Census with the ride-hail database is that we can select rides that occurred only in particular time frames that are characteristic of passengers who reside in the pickup Census tract. For example, trips that happened between 8 AM and 10 AM on weekdays are very likely to be ordered by passengers that actually live in their pickup Census tract. Therefore, the average income data of that Census tract is representative of this passenger's income.

The only problem with this approach is that we will lose a big share of our trip data. Still, because more than 50 million trips occurred during 2020, we will have a significant amount of observations from which we can make relevant inferences even after applying this filter. This is a good trade-off as it allows us to add any Census information to our analysis at the low cost of slightly diminishing our test's external validity.

All of these filters reduce the original database from 50 million trips to around 2.2 million trips. We apply these filters, however, so that we can join mean household income information to our analysis

Shared Trips: A good critique of the theory of delay is that there exists a social pressure mechanism in ride-hail trips that is related to trip delays and affects how much tip a passenger awards their drivers. For example, it could be argued that passengers who share trips are exposed to increased levels of social pressure, which causes them to tip higher. At the same time, shared trips are more likely to have more than one stop, which causes delays to be longer. Beyond that, passengers who have lower incomes are more likely to share trips since it is a more cost-efficient option of transportation than single rides. All of these dynamics are illustrated in the diagram below.

Figure 3: DAG of All Observable Causal Model



There is also a rhetorical value to controlling our results for shared trips. As detailed in the introduction of this paper, one of the reasons why rideshare trips are a better setting than restaurants to test clients' tipping behavior is because it is a less noisy environment and passengers have their experience determined only by the driver's service. By controlling for shared trips we make this assumption stronger since we are essentially controlling for other people in the car. Of course, not every trip that is actually shared is reported as shared on the app, so there is an effect of company in rides that we cannot control for. But controlling for what we observe helps us make this assumption a bit more pertinent.

It is also important to note that there is very little reason to suspect reverse causality. Ride-hail drivers complete a trip with no previous knowledge of the tip size. A passenger could try to influence the duration of their ride while the trip is ongoing by suggesting they will give higher tips if the driver arrives at the final destination earlier. However, this is a rare practice. And even if it does happen, there is no guarantee that the driver will respond to this incentive due to the uncertainty of the passenger honoring this promise.

3 Data and Descriptive Data

3.1 Transportation Network Providers

The data analyzed in this paper comes from the 2018-22 Transportation Network Providers (TNP) database. The TNP is a database made publicly available by the city of Chicago Data Portal. Its unit of analysis is ride-hail trips. In total, there are around 262 million trips recorded in this database. The TNP was constructed in 2018 after updates in Chicago's transportation legislation required ride-hail companies to report trip, driver, vehicle, and compensation data to the city's Commissioner. As stated in the Municipal Code of Chicago, "Each licensee [Uber, Lyft, and Via] shall provide the following data to the Commissioner:

(1)*Trip request data*. A record of each request for a trip made through the licensee's Internet-enabled application or digital platform by a potential passenger;

(2) *Trip data*. A record of each trip that shows where a passenger is picked up and dropped off;

(3) *Driver data*. A record of each of the licensee's drivers who are authorized to pick up passengers using the licensee's Internet-enabled application or digital platform;

(4) *Session data*. A record of each driver session on the licensee's Internet-enabled application or digital platform. A driver's session begins when a licensee's driver activates a mode in the licensee's Internet-enabled application or digital platform, signaling the driver's readiness to receive and respond to trip requests. A driver's session ends when the driver deactivates the mode and is no longer able to receive and respond to trip requests;

(5) *Vehicle data*. A record of each vehicle that is used by each of the licensee's drivers for picking up passengers through the licensee's Internet-enabled application or digital platform;

(6) *Location data*. For every transportation network vehicle and driver combination, location snapshots are captured at specified intervals for all times the driver is in session. Each snapshot shall indicate the vehicle's precise location and corresponding date and time;

(7) *Compensation data*. A record of each of the licensee's drivers who are paid an hourly rate, and any other record needed to capture actual driver pay information that is not reflected in licensee's hourly rate compensation records."

The authenticity of this data relies on the honest reporting of the TNP companies Uber, Lyft, and Via, and on the transparency of the Chicago Data Portal. Reports by these institutions tend to be trustworthy since they all come from either public companies or governments. This makes this a very interesting sample to study since it provides us with very accurate information on trip duration, tips, fares, start location, and end location of trips to perform our analysis. It also contains more than 50 million trips recorded, giving us a very representative sample of all ride-hail trips performed in 2020. The database has 21 variables in total. All the variables are listed below. The list items are in the format "Variable Name (Type): Description".

- Trip ID (Plain Text): A unique identifier for the trip;
- Trip Start Timestamp (Date and Time): When the trip started, rounded to the nearest 15 minutes;
- Trip End Timestamp (Date and Time): When the trip ended, rounded to the nearest 15 minutes;
- Trip Seconds (Number): Time of the trip in seconds;
- Trip Miles (Number): Distance of the trip in miles;
- Pickup Census Tract (Plain Text): The Census tract where the trip began;

- Dropoff Census Tract (Plain Text): The Census tract where the trip ended. This column often will be blank for locations outside Chicago.
- Pickup Community Area (Number): The Community Area where the trip began. This column will be blank for locations outside Chicago.
- Dropoff Community Area (Number): The Community Area where the trip ended. This column will be blank for locations outside Chicago.
- Fare (Number): The fare for the trip, rounded to the nearest $2.50.
- Tip (Number): The tip for the trip, rounded to the nearest dollar. Cash tips will not be recorded.
- Additional Charges (Number): The taxes, fees, and any other charges for the trip.
- Trip Total (Number): Total cost of the trip. This is calculated as the total of the previous columns, including rounding.
- Shared trip Authorized (Checkbox): Whether the customer agreed to a shared trip with another customer, regardless of whether the customer was actually matched for a shared trip.
- Trips Pooled (Number): If customers were matched for a shared trip, how many trips, including this one, were pooled. All customer trips from the time the vehicle was empty until it was empty again contribute to this count, even if some customers were never present in the vehicle at the same time. Each trip making up the overall shared trip will have a separate record in this dataset, with the same value in this column.
- Pickup Centroid Latitude (Number): The latitude of the center of the pickup Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.

- Pickup Centroid Longitude (Number): The longitude of the center of the pickup Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.
- Pickup Centroid Location (Point): The location of the center of the pickup Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.
- Dropoff Centroid Latitude (Number): The latitude of the center of the dropoff Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.
- Dropoff Centroid Longitude (Number): The longitude of the center of the dropoff Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.
- Dropoff Centroid Location (Point): The location of the center of the dropoff Census tract or the community area if the Census tract has been hidden for privacy. This column often will be blank for locations outside Chicago.

3.2 Descriptive Statistics

As stated in the introduction and the theoretical framework, this analysis is to the subset of approximately 11 million ride-hail trips from the TNP database that occurred from 2018 to 2022, on weekdays, from 8 AM to 10 AM. Having established our observations, we must now detail the statistics of the main variables in our regression analysis. They are ’tip’ and ’deviation on average trip duration’, and the control variables are ’shared authorized approved’, and ’mean household income’. Each of these variables is listed below with a brief description and its data type.

- Tip: Indicator function for whether the trip had a tip or not. Data type numeric, binary.

- Deviation on Average Trip Duration (or "Delay"): Difference between total trip duration and average trip duration from the ride's pickup Census tract to its Drop Off Census tract rounded to the nearest minute. Data type numeric, integer.
- Shared Trip Authorized: Indicator function for whether the trip was shared with another rider through the app. Data type numeric, binary.
- Census Tract Household Mean Income: Mean income for each pickup Census tract where a trip started. Income values are rounded to the nearest dollar. Data type numeric, integer.

First, we look at trip delays. The minimum delay found among our sample was -240 minutes. This means that the least delayed trip arrived around 4 hours earlier than the expected path time. The maximum delay was 1274 minutes, meaning that the most a trip was delayed in our sample was 21 hours. Delays of this magnitude are very uncommon and unrealistic, so these are considered to be irregular rides or a rare phenomena (outliers).

Figure 4: Distribution of Travel Time Delays

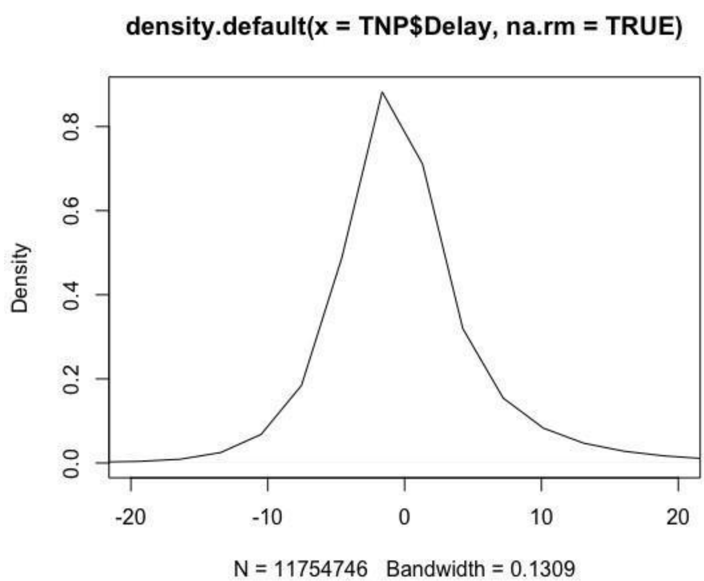


Table 1: Trip Delays (in Minutes)

Min	1st Quartile	Median	Mean	3rd Quartile	Max
-240.116	-2.926	-0.661	0.000	2.002	1266.023

Most trips are concentrated between the intervals of +10 and -10 minutes of delay. This can be seen very clearly in the density graph below. Therefore, a more reliable statistic is the 1st, 2nd, and 3rd quartile, which show that the majority of trips are somewhere 2.5 minutes early or delayed with respect to the average path time, with a very symmetric distribution centered near 0.

Next, we have the variable Tips in percentage. Tips have a very skewed distribution to the left, as we can observe from Figure 5. Most values are concentrated around the minimum of U$0.00. Out of the 11 million trips observed in this study, 81.28% — almost 9 million trips — did not give any tip. So it was expected that the distribution of tips would be skewed to 0. The average tip in percentage was nearly 3%.

Figure 5: Distribution of Tip in Percentage

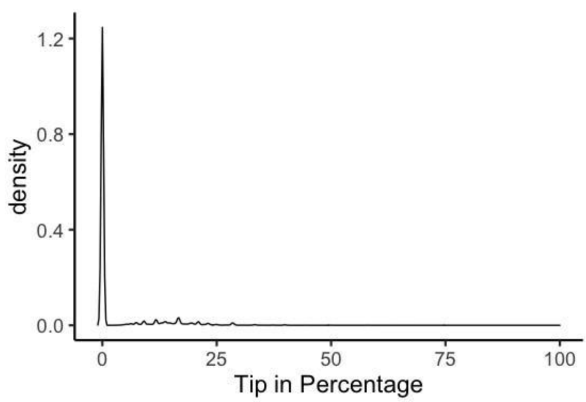
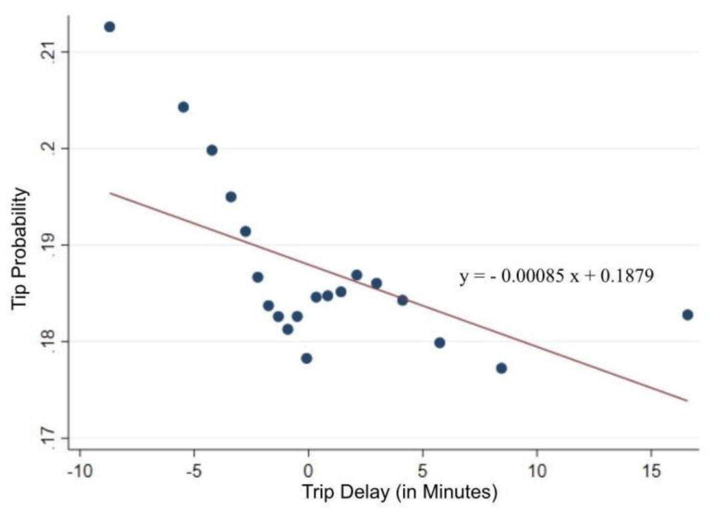


Figure 6 shows us the correlation between our dependent and independent variables: trip probability and trip delay. This scatter plot was created by splitting the 11 million trips into 20 bins with the same number of observations and mapping their average to the plot. The trend line is an OLS regression that uses all the 11 million trips observed. Without any control for confounding variables or fixed effects, the relationship between tipping probability and delay is negative, as expected by the theory of tips and delays. However, it is a very weak relationship if we consider it at the individual trip level. Every 10 minutes a driver is delayed, for instance is associated with a drop in the tip probability of only 0.85%.

Figure 6: Correlation Between Tipping Probability and Trip Delay



Now, we move to the description of our control variables. Our first control variable, Shared trip Authorized, is a binary variable that can only take up the values of 0 and 1. In the sample of trips we observed there was a mean of 0.1353 shared trips. This means that around 13.5% of trips were shared between different passengers. This is likely an underestimation of the number of passengers that actually shared ride-hails due to underreporting from the passenger's end. That is, many passengers share trips but are not willing to report them on the app most likely because they do not want to or cannot share the total cost of the ride with the other passenger.

Our second control variable, the passenger's expected household income, is a continuous variable and has its distribution represented in Figure 7 below.

Figure 7: Frequency of Trips by Passenger's Income

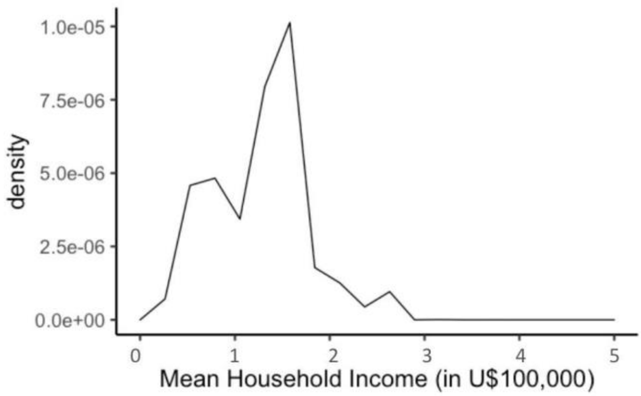
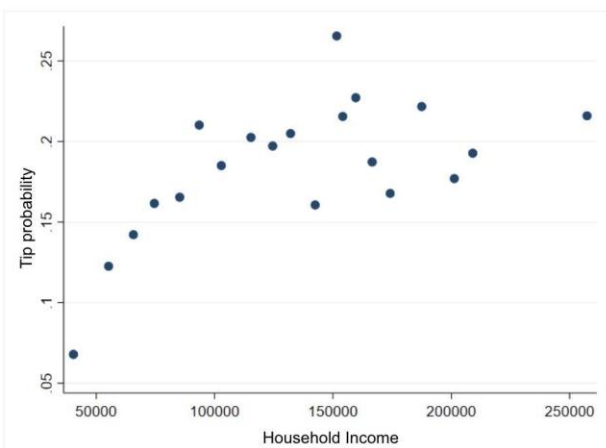


Table 3: Mean Household Income (in Dollars)

Min	1st Quartile	Median	Mean	3rd Quartile	Max
21,704	88,810	133,564	133,321	168,992	448,853

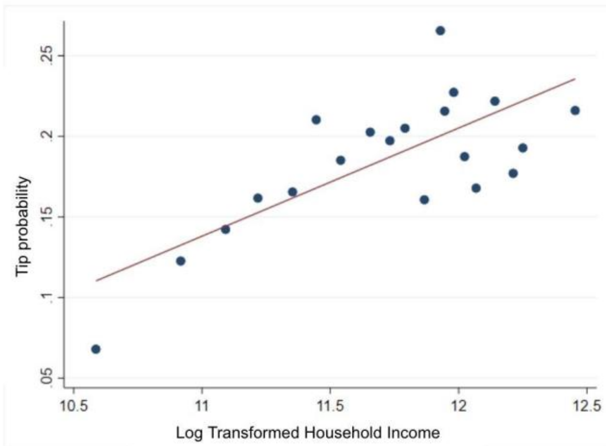
This table shows us that the Census tract that has the minimum mean household income has a mean household income of U$21,704. The maximum has an income of U$448,853. We anticipate there to be significant systematic differences associated with tipping behavior between tracts due to the differences in their mean income. We can find evidence for this suspicion in the next two plots.

Figure 8: Correlation between Tip and Passenger's Expected Income



The scatter plot in Figure 8 suggests that the relationship between tip probability and household income is positive for lower-income households, but then reaches a plateau as it reaches the higher values of household income. Figure 9 shows how by performing a log transformation on household income we arrive at a scatter plot with an approximately linear relationship between tipping and household income.

Figure 9: Correlation between Tip and Log Transformation of Passenger's Expected Income



Because we suspect income influences tipping probability and that it may also influence trip delays, it is important for the external validity of our study to have an estimate of passengers' income as one of our control variables.

The last table in this section is a compilation of all the correlations between the dependent and independent variables and the control variables.

Table 4: Correlation Between Variables of Interest and Control Variables

	Tip Probability	Tip Probability	Delay	Delay
Shared Trip	-0.1279156***	-	3.892113***	-
	(0.0003356)	-	(0.0047529)	-
Mean Household Income	-	4.86e-07***	-	-3.67e-07 ***
	-	(2.16e-09)	-	(3.16e-08)
Observations	11379883	10745414	11381372	10730671

Standard errors in parentheses

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

One interesting statistic is the correlation between Delay and Shared Trip. There is an increment in the trip delay of 3 minutes associated with sharing a trip instead of not sharing a trip. This matches the expectation of the relationship between these variables that we set in the construction of our theoretical model, namely that sharing trips increases the number of stops between pickup and destination, and therefore increases the deviation from average path time. Of course, this is just descriptive data. Nevertheless, it is useful by reaffirming the necessity of this control variable in our model. All of the betas from these correlations are statistically significant. Note, however, that this is not unexpected since we are studying a relatively big dataset.

4 Empirical Framework

The statistical test used to investigate the causal effect associated with trip delay and tips is a linear regression of deviation from average trip duration on whether a trip had tips or not (TipIndicator), controlling for the confounding variables 'shared trips' and 'passenger's income'. The sample regression equation is written below, with TipIndicator as the dependent variable and deviation from average trip duration as the independent variable. The coefficient we are interested in is $\hat{\beta}_1$. $\hat{\beta}_1$

represents the approximate change in *TipIndicator* associated with an additional unit of *Delay* accounting for the controls and route fixed effects.

$$TipIndicator_i = \hat{\beta}_1 Delay_i + \hat{\beta}_2 Shared_i + \hat{\beta}_3 Income_i + RouteFixedEffects + u_i$$

With $i = 1,...,n$.

Our causal identification strategy is achieved through a combination of linear regression and two key assumptions. The first assumption is that the relationship between the dependent and independent variables is in fact linear. The second is the conditional-independence assumption, which is that the common variables that affect delays and tips are observable. The dependence between delay and whether a passenger tips or not can be removed by conditioning on observable variables passenger's income and shared trips. Having both of these assumptions we can reach a new interpretation of $\hat{\beta}_1$ as the approximate change in *TipIndicator* (in percentage) caused by an additional unit of *Delay* (in minutes).

If the causal theory of tips and delays is true, then this test will show us a negative correlation between the probability of tipping and trip delay within the entire 95% confidence interval. This result should be both statistically (p-value < 0.5%) and substantively significant (non-trivial change in probability of tipping promoted by a change in trip delay). But whether this theory is rejected or not, the findings of this paper will contribute to the understanding of why people tip. If we are able to reject this causation theory, then this paper will serve as evidence of the prevalent social norm interpretation of tipping behavior. If we are not able to reject it, then there will still be reasons to suspect that tips are motivated by rational choice.

5 Results

5.1 Testing the Theory of Tips and Delays

The results of our linear regression controlled by both ’shared trips’ and ’passenger’s income’ as well as the route fixed effects can be seen in Table 5 below. $\hat{\beta}_1$, in particular, is equal to -0.000145. This tells us that the likelihood of a passenger tipping decreases by approximately 0.015% with each additional minute of delay. This result is significant at the $p < 0.001$ level. Again, not surprising. And the 95% confidence interval is $(-0.000021, -000124)$, which means that the causal relationship between tipping probability and delay is confidently below zero. That is, it would be extremely rare to observe a null or positive effect of delays on tipping behavior.

Table 5: Regression of Tips and Delay with Fixed Effects

		Tip Probability
Delay		-0.000145 *** (0.0000210)
Shared Trip		*** (0.000369) - 0.0948
Household Income		0 (.)
Constant	0.294*** (0.000432)	
Observations	10652140	Yes

Standard errors in parentheses

$^{*} p < 0.05$, $^{**} p < 0.01$, $^{***} p < 0.001$

Table 6 on the next page contains the same estimated effect plus partial results from OLS regressions with different combinations of controls. The regression controlling only for route fixed effects shows us that for every unitary increment in delay (in minutes) the likelihood of tipping decreases by 0.153%. This was the largest effect of delay on tipping we found. A graph of this linear model can be seen in Figure 10 on page 22.

This stronger negative association, however, seems to dissipate as we control for more variables, and we end up with the weaker negative effect that we originally observed in Table 5. The results of all these regressions are statistically significant (p-value < 0.001). Once more it is important to note that all of these regressions had an unusually big sample size, so other adjustments should be in order in case future work wants to apply this model for other scenarios beyond the specific one that we limit ourselves to in this paper.

An interesting observation is that the effect of delay on tipping behavior drops much more significantly when we introduce control for shared trips than for any other confounding variable or fixed effect. In particular, the effect of delay drops by a factor of 10, approximately. What is more curious is that the effect sharing a trip has on tip probability is negative. This has important consequences for the main arguments that Behavioral Economics have about the effect of being observed by friends or other individuals during our tipping decision. We shall discuss this topic in more depth in the next section of this paper.

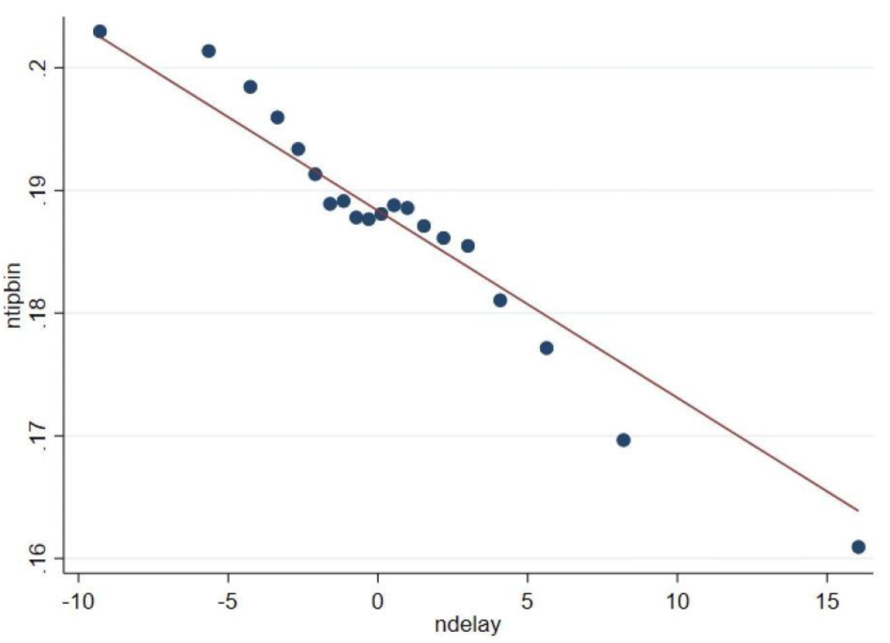
Table 6: Table with All Regressions

	Tip Prob.			
Delay	-0.00153***	-0.00155***	-0.000110***	-0.000145***
	(0.0000200)	(0.0000204)	(0.0000206)	(0.0000210)
Income		0		0
		(.)		(.)
SharedTrip			-0.0962***	-0.0948***
			(0.000362)	(0.000369)
Constant	0.189***	0.187***	0.297***	0.294***
	(0.000113)	(0.000115)	(0.000423)	(0.000432)
Observations	11301507	10652140	11301507	10652140

Standard errors in parentheses
$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

As we anticipated, the next plot illustrates the effect of travel time delays on tipping probability while controlling only for the fixed effects of the routes. To create this plot we divided our sample into 20 bins, each with the same number of observations. Then, we took the average tip probability of the observations in each bin and plotted the data points back to the graph. Controlling for fixed effects of the trip routes we get the trend line that is graphed below. This trend line shows that at 0 minutes of delay, passengers have a nearly 19% chance of giving a tip. If the delay increases by 5 minutes, then the probability of tipping decreases by 0.85%.

Figure 10: Relationship Between Tipping and Delay with Routes FE



Beyond the partial measure of the effect of delays on tipping behavior, this plot also suggests some other characteristics of passengers tipping behavior. First, we can observe how passengers display higher tolerance for delays of around 0 to 3 minutes. This is evidenced by the fact that the tipping probability remains approximately constant between delays of −3 to 3. So the same way that they do not penalize small delays harshly, they also do not reward small early arrivals generously. This plot also gives support to the claim that passengers are less willing to reward drivers for saving their time than they are willing to punish drivers for wasting their time. We can observe this from the fact that trips that arrive earlier than expected have their tip probability increase at a slower rate than trips that arrive later have. For instance, the average tip likelihood of trips that had around 9 minutes of delay is around 17%, or approximately -1.9% than if they had arrived on time. At the same time, trips that arrived around 9 minutes earlier than expected had a tip probability of 20.1%, or approximately 1.2% higher than if they had arrived at the expected time.

6 Discussion and Conclusion

The 3 main findings from the analysis of the 2.2 million samples of trips are:

1. Evidence of a statistically significant negative causal effect of trip delays
on the likelihood of a passenger tipping.
2. Evidence that passengers interpret small delays and early arrivals similarly.
3. Evidence of punishing large delays at a harsher rate than they reward large time-saving outcomes.

Let us begin by considering the impact of the 1st result in our theory of tipping behavior and delays, as well as its impact on the broader service quality and behavior modeling debate. According to the rational theory of tips and delays described in this paper, after controlling for all observables and fixed effects, we should have observed a negative effect of delays in the likelihood of tips. This is exactly what we see. As travel time delay increases, the probability of tipping decreases. If we can trust the statistical significance level and the standard errors of our analysis, then we have enough evidence to reject the theory that travel time delays cause passengers to tip more or have no effect on passengers' tipping behavior. And as we suggested at the beginning of the paper, this is a result that supports the idea of a rational model that can explain the tipping behavior, despite what is mainly believed in the Behavioral field.

It is also true, however, that this is a relatively small effect at the level of individual trips. Arriving at the destination 10 minutes earlier or 10 minutes later represents a change of approximately 0.3% in the likelihood of a driver receiving a tip. The substantive significance of this finding, therefore, is not as clear. At the industry level, however, this estimate gains more significance. The yearly average number of trips in the period between 2018 and 2022 was 52.4 million TNP trips in Chicago. Out of these 52 million, only around 9 million give tips assuming there are no delays per ride. An average delay of 2 minutes, however, would cost the industry 1.62 million tipping rides. And this has a substantive significance for the income of individual drivers.

The 2nd finding gives us hints at what a rational model of tipping would look like in the context of our empirical analysis. By providing evidence that passengers interpret small delays and early arrivals similarly, our 2nd finding suggests that the linear model used in this paper cannot properly address what seems to be a systematic tipping behavior. Future modelings of tip behavior would have to take this into account.

The 3rd finding also gives us characteristics that a more predictive tipping model would have to take into account. To be more precise, the model would have to consider the fact that great losses are punished more harshly than very early arrivals. But more importantly, our 3rd finding sheds light on a topic that is very dear to the field of Behavior Economics: loss aversion. It is often claimed by behavioral economists that negative events have a larger weight than positive events. According to these scientists, individuals have a bias to respond to losses more poignantly than they do to gains. The original evidence of this phenomenon was reported by Kahneman and Tversky (Econometrica 47:263–291, 1979).

However, recent critics suggest that loss aversion proponents have over-interpreted these findings. While commenting on Kahneman and Tversky’s work, Echiam 2019 states that "Specifically, the early studies of utility functions have shown that while very large losses are overweighted, smaller losses are often not." (Psychological Research 83, 1327-1339, 2019). Our finding is consistent with this characterization of loss aversion as we also do not find evidence for this bias in small delays, but only in larger delays. And evidence of loss aversion limited only to large delays is not supportive of behavioral explanations for tipping behavior. In fact, it is quite the opposite. Avoiding large losses more than valuing large gains is compatible with rational behavior. Large losses, be they money or time losses, have the potential to destroy your life. It is not a cognitive bias to take extra caution when a decision can put your life at risk, make you lose your house, or harm your family. No matter the upside, it makes sense to avoid extreme and definitive losses.

In conclusion, the analysis of tipping contained in this paper adds to methodological and substantial debates within the Tipping Behavior literature and the broader Behavioral Economics field. Our study contributes to the field by finding a new environment to conduct our analysis, moving away from the restaurant market. It innovates in analytical methods by conducting a study on data with large datasets, which are not common in the field. And all these innovations come with improvements to the quality and confidence of our results.

There are, of course, limitations to our analysis. The lack of personally identifiable data on both passengers and drivers associated with the trips, for example, prevents us from controlling for other relevant variables such as driver's app rating, passenger's age, or driver ethnicity. And our main findings, though consistent with previous literature, are also limited and must be submitted to further inquiries in order to improve our understanding of tipping and other behaviors claimed to be irrational.

References

Conlisk, Sarah. "Tipping in crises: Evidence from Chicago taxi passengers during COVID-19." Journal Of Economic Psychology (2022): 102475.

Azar, Ofer H. "Why pay extra? Tipping and the importance of social norms and feelings in economic theory." The Journal of Socio-Economics 36, no. 2 (2007): 250-265.

Azar, Ofer H. "The social norm of tipping: A review 1." Journal of Applied Social Psychology 37, no. 2 (2007): 380-402.

Azar, Ofer H. "What sustains social norms and how they evolve?: The case of tipping." Journal of Economic Behavior Organization 54, no. 1 (2004): 49-64.

Lynn, Michael. "Restaurant tips and service quality: A weak relationship or just weak measurement." International Journal of Hospitality Management 22, no. 3 (2003): 321-325.

Lynn, Michael, George M. Zinkhan, and Judy Harris. "Consumer tipping: A cross-country study." Journal of Consumer Research 20, no. 3 (1993): 478-488.

Saunders, Stephen G., and Michael Lynn. "Why tip? An empirical test of motivations for tipping car guards." Journal of Economic Psychology 31, no. 1 (2010): 106-113..

Chapman, Gretchen B., and Jennifer R. Winquist." The magnitude effect: Temporal discount rates and restaurant tips." Psychonomic Bulletin Review 5, no. 1 (1998): 119-123.

Azar, Ofer H. "The economics of tipping." Journal of Economic Perspectives 34, no. 2 (2020): 215-36.

Thrane, Christer, and Erik Haugom. "Peer effects on restaurant tipping in Norway: an experimental approach." Journal of Economic Behavior Organization 176 (2020): 244-252.

Brewster, Zachary W., and Kenneth Gourlay. "The “Mask Effect” on the tips that customers leave restaurant servers." International Journal of Hospitality Management 99 (2021): 103068.

Bodvarsson, Örn B., and William A. Gibson. "Economics and restaurant gratuities: Determining tip rates." American Journal of Economics and Sociology 56, no. 2 (1997): 187-203.

CHICAGO, IL., MUNICIPAL CODE OF CHICAGO ch. 9, § 115 (2022).

City of Chicago, 2022. Transportation Network Providers – Trips [data file]. Chicago Data Portal. Retrieved from:

https://data.cityofchicago.org/Transportation/ Transportation-Network-Providers-Trips/m6dm-c72p.

Theory References

Kahneman, Daniel, and Amos Tversky. “Prospect Theory: An Analysis of Decision under Risk.” Econometrica, vol. 47, no. 2, 1979, pp. 263–91. JSTOR, https://doi.org/10.2307/1914185. Accessed 29 Sep. 2022.

Echiam, E. Acceptable losses: the debatable origins of loss aversion. Psychological Research 83, 1327–1339 (2019). https://doi.org/10.1007/s00426-0181013-8

Data Set References

Hegde S, Abkarian H, Mahmassani HS. Characterizing Ride-Hailing Driver Attrition and Supply in the City of Chicago Through the COVID-19 Pandemic.
Transp Res Rec. 2022 Aug 24:03611981221117542. doi: 10.1177/03611981221117542. PMCID: PMC9412150.

Credit, Kevin, Gustavo Dias, and Brenda Li. "Exploring neighbourhoodlevel mobility inequity in Chicago using dynamic transportation mode choice profiles." Transportation research interdisciplinary perspectives 12 (2021): 100489.

Correia, Sergio. 2017. "Linear Models with High-Dimensional Fixed Effects:
An Efficient and Feasible Estimator" Working Paper. http://scorreia.com/research/hdfe.pdf

Noah Constantine, Sergio Correia, 2021. reghdfe: Stata module for linear and instrumental-variable/GMM regression absorbing multiple levels of fixed effects. https://ideas.repec.org/c/boc/bocode/s457874.html

Impact of Implementing Weighted School Funding on High School Educational Attainment

Alice Martynova
Massachusetts Institute of Technology

Abstract

Since 1993, many large school districts across the United States have shifted away from deploying federal funds to schools based on uniform staffing formulas and have instead adopted weighted school funding (WSF). WSF provides a fixed-dollar amount to schools for each student type, with larger increments going to students from low-income backgrounds, those with special needs, and/or those who are English-language learners. In this study, I used publicly available NCES data to study the impact of WSF on high school graduation rates, dropout rates, and pupil-per-teacher ratio. The difference-in-difference empirical strategy finds that WSF has limited statistically significant impact on any of these educational attainments. These results suggest that WSF's effects still need to be further studied to fully understand the power and drawbacks of this new and emerging funding schematic.

1 Introduction

Over the last two decades, many large school districts across the United States have shifted away from deploying federal funds to districts based on uniform staffing formulas to allocating funds to individual schools within the district based on the particular mix of students at each school. This new funding strategy, known as weighted student funding (WSF), deploys a fixed-dollar amount to schools for each student type with larger increments going to students from low-income backgrounds, those with special needs, and/or those who are English-language learners. It is worth noting that WSF does not change the total amount of money a district receives; rather, it alters how the funds are distributed amongst the schools within the district.

New funding methods, like WSF, have the potential to fight the poverty cycle, reduce inequality, and have significant effects on student educational outcomes (Johnson and Jackson 2019) [9]. This paper examines how WSF affects high school educational attainments, particularly graduation rates, dropout rates, and pupil-per-teacher ratio. Although testing this hypothesis is challenging due to the limited number of schools that have adopted WSF and the limited quantity of public data, the available data provides for robust empirical tests to better understand WSF's potential.

To understand the effect on high school district graduation rates, dropout rates, and pupil-per-teacher ratios, this paper implements a staggered difference-in-difference for each educational attainment to compare outcomes at control schools that never implemented WSF to treated districts that implemented WSF between 1995-2018. The treated group consists of the 27 WSF school districts documented in the U.S. Department of Education report. The control districts are chosen from the NCES annual table of "Selected statistics on enrollment, teachers, dropouts, and graduates in public school districts enrolling more than 15,000 students." By only selecting districts from this category, I guarantee districts have similar sizes and are nationally representative which leads to more robust results.

The study concludes positive but mostly insignificant effects of WSF on graduation rates and pupil-per-teacher ratio and inconclusive negative effects on dropout rates. Following treatment, the pupil-per-teacher ratio and graduation rates remain unchanged relative to the pre-treatment mean and both effects are statistically insignificant. In addition to NCES documented graduation rates, this paper introduces a new statistic, "pseudo-graduation" rate, which is calculated by dividing the number of graduates by total district enrollment. WSF increases pseudo-graduation rates by about 0.3 percentage points which is an overall 6% increase, but this is only at the 10% significance level and must be interpreted with caution. Finally, WSF appears to decrease dropout rates by 1 percentage point which is an overall 11% decrease. Nonetheless, this result must be interpreted with caution as the dropout rate regression does not satisfy the parallel trend assumption (further discussed in Sections 4 and 5). Although there is no apparent sizable and significant effect, WSF does not negatively affect any educational attainment at the district level which questions theories that WSF has an overall harmful impact.

The current WSF literature only focuses on describing the WSF model and studying its impact on standardized test scores; furthermore, many WSF studies use datasets that limit the robustness and interpretation of results. For instance, a National Study by the U.S. Department of Education focuses on describing WSF policy, its intended changes and benefits, and details school districts that adopted WSF before 2018 (Levin, Manship, Hurlburt, and Atchison 2019) [6]. However, it does not quantify the effect on educational attainments. Another paper focuses on understanding the financial details of WSF at the district level, specifically the unique weight formulation of each district and whether the formulations are aligned with WSF's goal of increasing equity (Roza et.al 2019) [4]. While this study begins to scrape at the surface of understanding the academic outcomes of WSF, the Edunomics report notes that the state-level results should be interpreted with caution since WSF districts tend to be different than others in their state in both enrollment size and student composition. Moreover, the effects of WSF cannot be isolated from the effects of other policies implemented around the same time. As mentioned previously, my paper produces more robust results by only selecting treatment and control districts with over 15,000 students from a nationally representative sample.

The rest of the paper is organized as follows. Section 2 provides an in-depth WSF policy debrief. Section 3 presents the data. Section 4 describes the empirical strategy. Section 5 presents the results. Section 6 presents a summary discussion. Section 7 details future work, and Section 8 concludes the paper.

2 Policy Background

Historically U.S. school districts distribute federal funds to schools through tangible resources rather than allocating specific dollar amounts to individual schools. These traditional uniform staffing allocation systems typically determine the number of teachers, school administrators, and other types of staff for each school based on its total student enrollment. However, many educators and researchers have noted that these systems can contribute to and increase inequity amongst schools, especially those with higher concentrations of at-risk students may not receive additional resources to meet their complex needs (Rubenstein, Schwartz, and Stiefel 2006) [8].

To mitigate these inequities, the WSF Federal Government program allows districts to deploy a fixed-dollar amount to schools for each student type with larger increments going to students from low-income backgrounds, with special needs, and/or who are English-language learners. Under the WSF approach, districts may allocate resources more effectively to meet the specific needs of each of their school's students.

Policymakers from the federal government to the district level are always researching and creating new programs and funding methods to improve public education. Districts choose to adopt WSF to increase equity, transparency, flexibility, and school-level autonomy to focus on improving student outcomes (Roza et.al 2019) [4]. WSF has been around since 1995 and over the past 2 decades, 27 school districts have implemented WSF with these goals in mind. This paper sets out to understand whether WSF indeed improved student educational outcomes. This is relevant today as Biden plans to double funding for K-12 education through the "Build Back Better" plan as schools struggle to successfully emerge out of the pandemic and help students meet standards following the past year of virtual learning (Camera 2021) [3]. Understanding the effects of WSF can help schools and the federal government use their budget effectively.

3 Data

To study the impact of implementing WSF on high school educational attainment through a difference-in-difference model, I need funding data at the district level to understand which districts implemented WSF as well as school district performance data. Both datasets are further detailed below.

To identify control and treated districts, I will rely on the findings of existing WSF literature. The U.S. Department of Education made a detailed 2019 WSF report (Levin, Manship, Hurlburt, and Atchison 2019) which includes a table of 27 well-documented districts that have implemented WSF and have continued to use it, along with the year in which they implemented it. The Roza et al. (2019-Present) Georgetown report provides a similar table of 18 districts that have implemented WSF, and these 18 districts align with the 27 districts provided by the U.S. Department of Education. I use both lists to develop my treatment group. However, both WSF papers, anonymize schools that did not implement WSF increasing the difficulty of creating a control group. Through direct discussion with Hannah Jarmalowski, a Research Fellow at Georgetown Edunomics Lab, she explained that there are very few districts that have implemented WSF and districts that have are documented in the literature. I use the limited existing literature to create a thorough table of WSF implementing school districts (Table 1) and have confidence that unlisted districts have never implemented WSF.

To explore multiple levels of educational attainment, the main data resource will be the NCES, the National Center for Education Statistics. The NCES provides annual tables of "Selected statistics on enrollment, teachers, dropouts, and graduates in public school districts enrolling more than 15,000 students" from 1995-2018. The pupil per teacher ratio is one of the only variables available every year from 1995-2018. It is important to note that the ratio itself is not a measurement of educational attainment, but in the literature, lower pupil per teacher ratio is correlated with higher educational achievement (Jackson, Rucker and Persico 2015) [2]. The NCES tables also contain high school dropout rates by district from 1996-2009. Although this does not cover up to 2018, there are 9 schools that adopted WSF around 2002 and dropout rates can be observed for those sub-selected districts. The NCES also documents high school graduation rates by district from 2007 to 2018, and this data can be used for the 10 schools that adopted WSF between 2007 and 2018. Note intuitively it should be possible to get graduation rates from 1996-2009 by using 1-dropout rate, but for 2007 and 2008 in which both graduation and dropout rates are available, graduation rates are not equivalent to 1-dropout rates.

Due to changes in data collection methods, it is difficult to find consistent data measurements over the past 25 years. The NCES does provide the number of high school graduates at the district level from 1995 to 2009, but this raw number is unusable because it does not separate number of graduates from national migration changes and general population growth. In addition to the number of graduates, the NCES provides the total enrollment count at every district. As a rough estimation, I divide the number of graduates by total enrollment to get a "pseudo-graduation" rate from 1995 to 2009.

To perform a robust staggered difference-in-difference, the data must be divided into treated and control groups using information from the U.S. Department of Education report and NCES. The treated group for each educational attainment will be selected from the 27 WSF school districts documented in the U.S. Department of Education report. I do not use all 27 school districts currently implementing WSF as the Minneapolis School District implemented WSF in 1993 and the Prince William County Public Schools implemented WSF in 1994, but there is insufficient NCES data prior to 1995. Atlanta Public Schools and Shelby County Schools districts implemented WSF in 2018, but NCES has yet to upload the needed data beyond 2018. Following these adjustments, the treated group is selected from a pool of 23 districts. The control districts will be chosen from the NCES yearly table of "Selected statistics on enrollment, teachers, dropouts, and graduates in public school districts enrolling more than 15,000 students." Only districts with consistent data for the respective time-period for each attainment will be chosen.[1] All the districts will be from this table since most WSF implementing districts are large urban school districts, and the NCES

[1]For pupil-per-teacher ratio, actual graduation rate, and pseudo-graduation rate, only districts with data for every year will be chosen. The exception is the dropout rate controls because dropout data is not available for every year for any treated district. This will limit the interpretation of the dropout rate results which is further discussed in Section 6.

only provides district-level statistics on districts with more than 15,000 students.

Summarizing the NCES data reveals the number of treated and control observations for every educational attainment. Table 2 summarizes the pupil-per-teacher ratio, dropout rate, graduation rate, and pseudo-graduation rate by control and treatment group. Notice that dropout rate has the greatest number of observations (total and by control/treated) even though it does not cover the full period from 1995-2018 because it does not omit treated or control districts that are missing data for any year in between 1996-2009. On the other hand, the other 3 measurements only include data for districts with measurements for every year. This is necessary because there are no treated districts that had dropout data for every year between 1996 and 2009. This effects the interpretation of dropout regression results which will be further discussed in Section 6, the discussion section. Figure 1 visually summarizes the data by graphing arbitrarily chosen districts treated in the same year versus control schools for each educational attainment. Even before running the empirical tests, this figure hints at two findings: parallel trends are likely to be unsatisfied for dropout rates and results for all educational attainments are likely to be small and minimal in effect.

4 Methods

To understand the effect of WSF on various high school educational attainments at the district-level, this study will rely on the difference-in-difference method to compare outcomes at control districts that never implemented WSF and treated districts that implemented WSF between 1995-2018.

Difference-in-difference is the best method, given the available data and nature of WSF implementation across districts, to estimate the educational effects of WSF. However, among the 23 treated school districts, many districts were treated at different times. This makes it difficult to perform a simple regression and a traditional difference-in-difference. Thus, I propose the following regression which should find coefficients on the pre-treated periods are statistically insignificant and hence demonstrate parallel trends leading into the treatment. The coefficients on post-treated periods will show the effect of WSF on the specific measurement of educational attainment:

$$E_{d,t} = \alpha_d + \delta_t + \sum_{y=0}^{T_1} \gamma_y D_{d,y} + \sum_{y=T_0}^{-2} \gamma_y D_{d,y} + \epsilon_{d,t} \quad \text{(Equation 1)}$$

Where $E_{d,t}$ is the educational outcome for a district d at time t. α_d and δ_t are the district and year fixed effects respectively. ϵ_{dt} is the error term. T_0 and T_1 in the summation are, respectively, the lowest lag year and highest lead year to consider surrounding the treatment period. $D_{d,y}$ is a dummy variable that is equal to 1 if the observation's period relative to district d's first treated period is the same value as y; otherwise the dummy is equal to 0 and is 0 for all never-treated observations. The regression coefficients are the γs which are for each year leading and lagging the treatment. Note the -1 is omitted from the summation to avoid multicollinearity and serves as the point of reference.

Equation 1 describes a dynamic regression which will give detailed insight into the effect of WSF on the educational attainment every year after treatment. However, for simplicity of understanding the overall effects of WSF, I will also run a static regression (Equation 2).

$$E_{d,t} = \alpha_d + \delta_t + \beta * (POST_t * TREAT_d) \quad \text{(Equation 2)}$$

In Equation 2, I regress the outcome for district d in year t on a dummy variable that is the interaction between $POST_T$ (year t is after WSF has been implemented in that district) and $TREAT_d$ (district d is a district in which WSF has been or will be implemented). Like in Equation 1, α_d and δ_t are the district and year fixed effects respectively.

Graphing the γ coefficients from Equation 1 will show the sign and size of the treatment, but to be able to effectively interpret these results, several assumptions need to be satisfied. First, the allocation of intervention must not be determined by the outcome; meaning if an increase in educational attainments is found following the implementation of WSF, it is due to the new funding scheme rather than prior characteristics or other novel changes of the school district. This assumption is satisfied because the 23 WSF implementing school districts and the control group are nationally representative. Potential educational attainment changes can be attributed to WSF because it is unlikely multiple schools passed similar policies other than WSF at the same time and achieved similar educational results.

Additionally, there must be no spillover effects from treated to untreated school districts. Historically, school districts are very isolated, and students within one district are within the same city and their education is unaffected by the policies of nearby districts. Furthermore, there have been numerous peer-reviewed, economic studies that have compared various school districts in the same area using a difference-in-difference model (Harris and Larsen 2018) [5].

The most important assumption to satisfy is the parallel trend assumption. As in most economic studies, it is impossible to observe the treatment group in the absence of treatment. Thus, I will show the γ coefficients leading into treatment in Equation 1 are zero indicating parallel trends into treatment. Graphing these coefficients in Figures 2-5 for all districts across all years reveals the coefficients on the pre-period dummies are statistically indistinguishable from 0. These findings are further discussed in Section 5.

5 Results

This study considers the impact of WSF on high school district pupil-per-teacher ratio, actual and pseudo-graduation rates, and dropout rates. I run the dynamic regression described in Equation 1 and the static regression described in Equation 2 for the selected control and treated districts while accounting for district and year fixed effects. The rest of this section will present the results from the static regression followed by the dynamic regression.

The static regression reveals positive alas only marginally significant effects on pupil-per-teacher ratios, actual and pseudo-graduation rates, and harder-to-interpret negative effects on dropout rates. Table 3 shows these raw results of the static regression and illustrates pupil per teacher ratio and graduation rate coefficients are slightly positive but statistically insignificant at the 5% and even 10% level. The effect of WSF on the dropout rate is negative and statistically significant at the 10% level. However, Figure 5, a graph of coefficients on dropout rates from the dynamic regression, clearly shows that parallel trends are unsatisfied for dropout rates, thus these results are not robust. Most notably, the pseudo-graduation rate appears to be slightly positive and to be statistically significant at the 10% level. However, pseudo-graduation is a measurement created for this study and is difficult to interpret. It will be further discussed in Section 6.

Dynamically regressing on pupil-per-teacher ratio leads to coefficients of negligible size leading into and lagging out of treatment. Referencing Figure 2, the confidence intervals on the regression coefficients for every lead year cover 0. However, the lagging coefficients also cover 0 and do not seem to have a constant trend which signals that WSF does not have a significant effect on the pupil-per-teacher ratio. Figure 2 was created using Appendix Table A1 which includes raw coefficients and standard errors.

While the pupil-per-teacher regression satisfies parallel trends, the dynamic graduation rate regression shows not all leading coefficients cover 0 in their 95% confidence interval (Figure 3 created using Appendix Table A2). This is likely due to the limited number of treatment schools and smaller time-period compared to the pupil-per-teacher data. Due to the noise of these results, the actual graduation rate results are unusable in identifying the effect of WSF.

WSF appears to have a noticeable effect on pseudo-graduation rates at the 10% significance level. Pseudo-Graduation was calculated from 1995-2009 for 4 treated districts and 137 control districts. Starting with satisfying the parallel trends assumption, all leading treatment coefficients in Figure 4 have confidence intervals that cover 0. This helps support the parallel trend assumption leading into treatment. In this case, the lagging coefficients appear to have an upward trend that becomes slightly significant around 6 years after treatment. Figure 4 was created using Table A3 attached in the appendix which includes raw coefficients and standard errors.

As discussed previously, the dropout rate coefficients are difficult to interpret as they do not satisfy the parallel trend assumption (Figure 5). It is important to note that after treatment, the regression coefficient confidence intervals do follow a negative trend, however, the coefficients continue to cover 0 indicating an absence of a statistically significant effect of WSF on dropout rates.

6 Discussion

Overall, WSF has limited impact in size and significance on high school pupil-per-teacher ratio, actual and pseudo-graduation rates, and dropout rates. As described in the results section, the coefficient on pupil-per-teacher ratio is close to 0 and statistically insignificant. This is not immensely surprising because as noted in the Section 1 and 2, WSF does not increase the total sum of money a district receives. Even though some higher-risk schools within a district may receive additional funding through WSF to invest in more teachers, at the district level and nationally WSF has limited impact on the pupil-per-teacher ratio.

The effect on actual graduation rates is close to null which is unsurprising given the literature on the challenges of improving high school graduation rates. Following treatment, the mean graduation rates rise for treated districts from 62.89 to 63.056 (Table 3) which is a close to 0 effect and statistically insignificant. Again, this is not immensely surprising, as high school graduation rates are historically difficult to improve even through programs targeted at improving graduation rates (Abele and Iver 2011) [1]. Furthermore, graduation rates do not fully satisfy parallel trends making the interpretation less robust (Figure 3). This is likely because there are only 5 treated districts which increases the noise.

Dropout rates slightly decrease following WSF, but it is imminent to remember that dropout rates fail to satisfy parallel trends. Districts that implement WSF appear to decrease dropout rate by about 1% compared to districts that do not implement WSF which is a relative 11% decrease of the pre-treatment mean, at the 10% statistically significant level. However, this result is not robust as the dropout rate regression does not satisfy the parallel trend assumption (Figure 5). Without this vital assumption, there is no definitive conclusion. The data for dropout rates was not panel data which likely increased the noise of the data leading into treatment. If more data is acquired, parallel trends can be satisfied, and a definitive effect of WSF on dropout rates can be identified.

Finally, I find after implementing WSF, district pseudo-graduation rates increase from 4.32 to 4.616 which is a 6% increase, but there are many limitations to this result. First, it is at the 10% significance level and should be approached with caution. Furthermore, this educational attainment measurement was made for this paper due to limited publicly available district level data. The original goal of the pseudo-graduation rate measurement was to support the results of the effect of WSF on standard graduation rates. However, the pseudo-graduation result should not be fully discarded and rather further studied. Remember pseudo-graduation is equal to the number of graduates divided by total enrollment within a district. Since I found no effect of WSF on graduation rates, WSF increasing pseudo-graduation rate could indicate WSF leads to a decrease in total enrollment within a district. This could signal a decrease in high school enrollment which is not necessarily an adverse effect. For example, decreasing high school enrollment within the studied schools could imply migration of families to less urban and crowded schools.

Ultimately WSF has no sizable and significant effects on pupil-per-teacher ratio, dropout rates, graduation rates, and even pseudo-graduation rates. However, even this finding should not be discarded. One major critique of WSF is that it reallocates money from higher-income students to those who are qualified for WSF funding which could negatively impact more privileged students. However, the 95% confidence interval of every coefficient covers 0 which indicates that WSF does not harm the general student population.

7 Future Work

The inconclusive results of this study indicate a need to continue understanding WSF's effect at the level of students directly targeted by WSF. Due to time constraints, I was unable to also explore the effects of WSF at the level of students who are English-Language learners, have disabilities, or come from low-income backgrounds. After exploring literature and data sets from the Equality of Opportunity project, I identified two promising data sets: the EDFacts data set and Neighborhood Characteristics by County.

The EDFacts data set details the percentage of students in every district who score above proficient on their state's ELA and Math standardized test from 2009-2018, broken down by low income, disability, and English language learner status. The large and complex EDFacts data set needs to be thoroughly processed and separated by control and treatment districts, about 10 treated districts in the given period. The empirical method for standardized testing will follow the same dynamic regression described in Equation 1 in Section 4. Although it is disappointing that the effect of WSF on standardized testing must be left as a future study, there are many drawbacks in current literature that hindered this study from focusing on standardized testing. First, the Georgetown WSF study already explores the effect of WSF on standardized testing. As the purpose of this study was to expand upon WSF's overall effects, I chose to put full focus into exploring other educational attainments. Another reason this study did not focus on testing is over 40 states changed their standardized tests in 2010 with the adoption of national common core increasing the difficulty of isolating the effect of WSF from drastic changes in standardized testing (Polleck and Jeffery 2017) [7]. However, I can try to mitigate this effect by adding a fixed state effect.

It is also important to study the effect of WSF on pupil-per-teacher ratio, actual and pseudo graduation rates, and dropout rates at the low-income level using the Equal Opportunity data source Neighborhood Characteristics by County. This data set details the percentage of low-income county residents. However, this data set is by county, so I would only use control and treated districts that cover full counties and assume that the county-level and district level percentage of low-income backgrounds are similar. The regression will follow a similar format to Equation 1 from Section 4, but with an additional variable $I_{d,y}$, representing the low-income population percentage in district d and year y (Equation 3). The γ coefficient will find the isolated effect on districts implementing WSF, ρ coefficient will represent the isolated effect on continuous income levels, and σ, our main coefficient of interest, finds the interaction for every year for varying income levels.

$$E_{d,t} = \alpha_d + \delta_t + \sum_{y=0}^{T_1} \gamma_y D_{d,y} + \sum_{y=0}^{T_1} \rho_y I_{d,y} + \sum_{y=T_0}^{-2} \gamma_y D_{d,y} + \sum_{y=T_0}^{-2} \rho_y I_{d,y}$$

$$+ \sum_{y=T_0}^{-2} \sigma_y D_{d,y} * I_{d,y} + \sum_{y=0}^{T_1} \sigma_y D_{d,y} * I_{d,y} + \epsilon_{d,t} \quad \text{(Equation 3)}$$

8 Conclusion

Weighted School Funding has been around for over 2 decades and over 20 districts have implemented the funding policy to solve inequities between students by allocating additional funds to students from low-income backgrounds, who are english-language learners, or who have a disability. However, WSF is largely unstudied, and little is known about its effects on educational attainments. Using available public data, I studied the effect of WSF on pupil-per-teacher ratio, graduation rates, pseudo-graduation rates, and dropout rates.

Although overall WSF has limited impact on these educational attainments or produces inconclusive results, I discover WSF has no apparent negative effect and must be further studied. First, the mostly null effects of WSF indicate that WSF appears to not harm students from privileged backgrounds which was one of the only policy concerns. This study also emphasizes that researchers have barely scraped the surface of thoroughly understanding WSF. As described in Section 7, there are already 2 potential analyses; however, there are even more undiscovered empirical tests that can further the understanding of WSF such as the effect of WSF on college enrollment, primary school attainments, etc. WSF is implemented by some of the largest and most innovative school districts like New York City and Boston. As WSF continues to spread nationally, it is crucial that policymakers and educators take a more practical approach and study WSF to weigh the benefits and drawbacks of this funding policy.

References

[1] Martha Abele and Mac Iver. The challenge of improving urban high school graduation outcomes: Findings from a randomized study of dropout prevention efforts. *Journal of Education for Students Placed at Risk*, pages 167–184, 2011.

[2] Rucker C. Johnson C. Kirabo Jackson and Claudia Persico. The effects of school spending on educational and economic outcomes: Evidence from school finance reforms. *The Quarterly Journal of Economics*, pages 157–218, 2015.

[3] Lauren Camera. Biden's budget significantly boosts k-12 education spending. *U.S. News*, 2021.

[4] Marguerite Roza et. al. Lessons learned: Weighted student funding. 2020.

[5] Douglas N. Harris and Matthew F. Larsen. The effects of the new orleans post-katrina market-based school reforms on medium-term student outcomes. 2019.

[6] Steve Hurlburt Drew Atchison Ryoko Yamaguchi Adam Hall Stephanie Stullich Jesse Levin, Karen Manship. Districts' use of weighted student funding systems to increase school autonomy and equity: Findings from a national study. 2019.

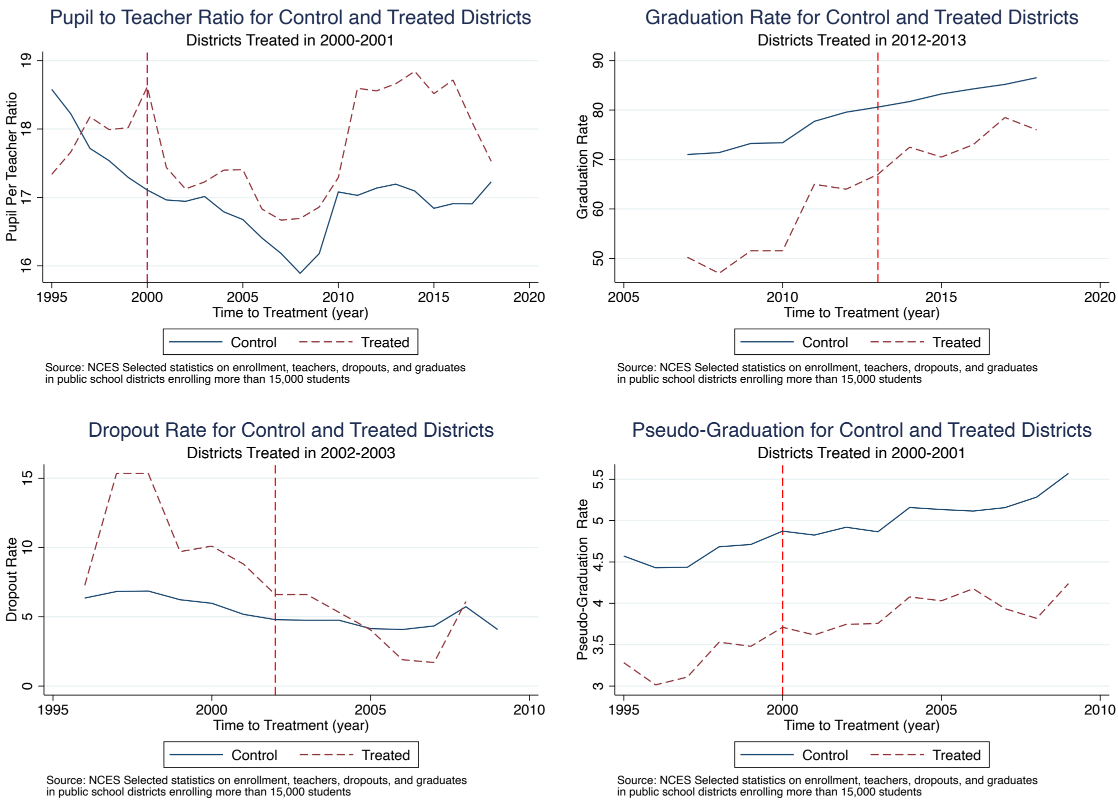
[7] Jill V. Jeffery Jody N. Polleck. Common core standards and their impact on standardized test design: A new york case study. *The High School Journal*, pages 1–26, Fall 2017.

[8] Amy Ellen Schwartz Ross Rubenstein and Leanna Stiefel. Rethinking the intradistrict distribution of school inputs to disadvantaged students. 2006.

[9] C. Kirabo Jackson Rucker C. Johnson. Reducing inequality through dynamic complementarity: Evidence from head start and public school spending. *American Economic Journal: Economic Policy*, pages 310–349, 2019.

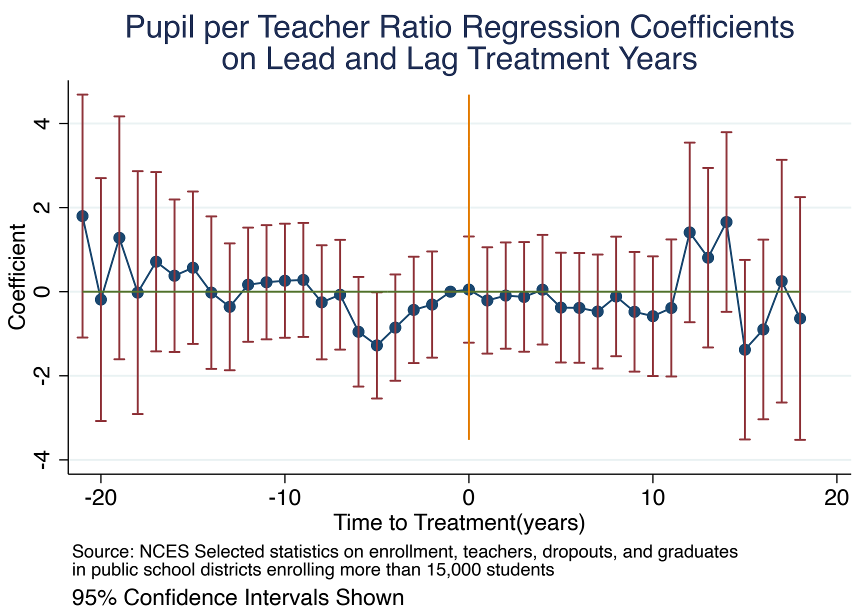
Figures and Tables

Figure 1: Educational Attainment Trends Over Time



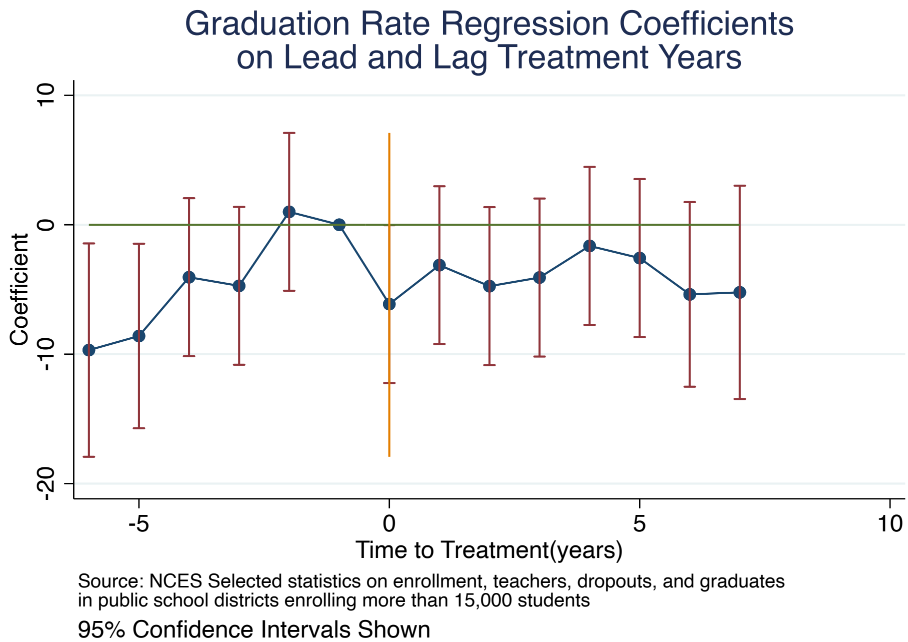
Note: Pupil Per Teacher Ratio from 1995-2018 for 144 control districts versus 2 districts treated in 2000-2001 school year. Graduation Rate from 2007-2018 for 138 control districts versus 1 district treated in 2012-2013 school year. Dropout rates were calculated from 1996-2009 for 343 control districts versus 3 districts treated in 2002-2003 school year. Pseudo-Graduation was calculated from 1995-2009 for 272 control districts versus 2 district treated in 2000-2001 school year using $\frac{\text{\# of high school graduates within district}}{\text{total enrollment within the district}} * 100$. Treated districts were identified using Georgetown Edunomics WSF Report (Roza et. al 2019-Present).

Figure 2



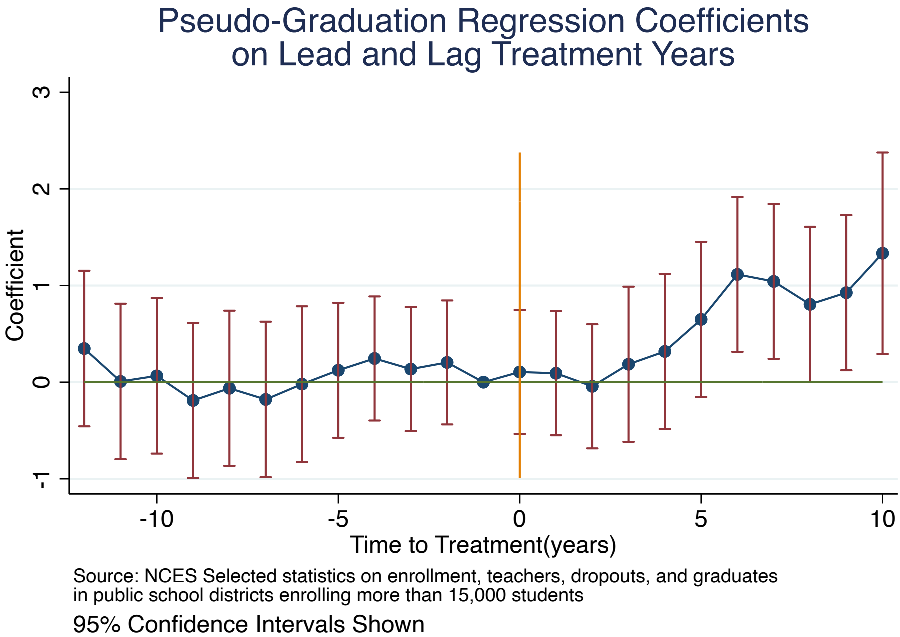
Note: Regression coefficients with confidence intervals on lead and lag years for Pupil per teacher ratio from 1995-2018 for 8 treated districts and 144 control districts. Note the year before treatment has been omitted to avoid multicollinearity and have a relative time reference. The graph was created using Appendix Table A1.

Figure 3



Note: Regression coefficients with confidence intervals on lead and lag years for Graduation Rates from 2007-2018 for 5 treated districts and 272 control districts. Note the year before treatment has been omitted to avoid multicollinearity and have a relative time reference.The graph was created using Appendix Table A2.

Figure 4

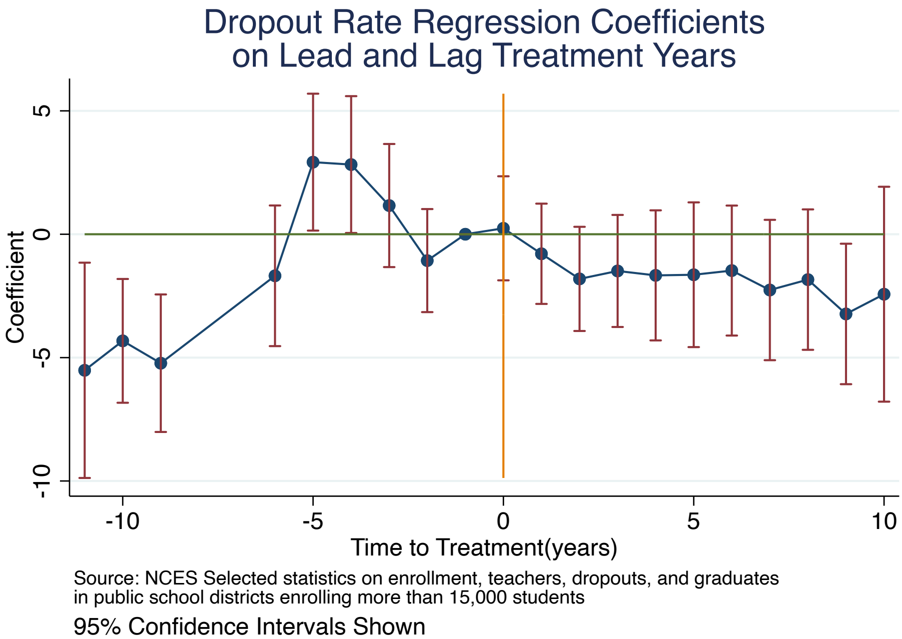


Note: Regression coefficients with confidence intervals on lead and lag years for Pseudo-Graduation rate:

$$\frac{\#\text{ of high school graduates within district}}{\text{total enrollment within the district}} * 100$$

Pseudo-Graduation was calculated from 1995-2009 for 4 treated districts and 137 control districts. Note the year before treatment has been omitted to avoid multicollinearity and have a relative time reference. The graph was created using Appendix Table A3.

Figure 5



Note: Regression coefficients with confidence intervals on lead and lag years for Dropout Rates from 1996-2009 for 9 treated districts and 343 control districts. Note the year before treatment has been omitted to avoid multicollinearity and have a relative time reference. The graph was created using Appendix Table A4.

Table 1: Districts Implementing WSF as of 2019

District Names	State	Year Adopted	Enrollment	Number of Schools	Poverty Rate	Urbanicity
Minneapolis Public Schools	MN	1993–94	36,793	86	24%	City
Prince William County Public Schools	VA	1994–95	87,793	92	9%	Suburb
Cincinnati Public Schools	OH	1999–2000	34,227	54	33%	City
Houston Independent School District	TX	2000–01	215,627	283	31%	City
Milwaukee School District	WI	2000–01	75,749	158	34%	City
San Francisco Unified School District	CA	2002–03	58,865	116	12%	City
St. Paul Public School District	MN	2002–03	37,698	103	27%	City
Hawaii Department of Education	HI	2006–07	181,995	289	10%	Suburb
Denver Public Schools	CO	2007–08	90,235	189	20%	City
New York City Public Schools	NY	2007–08	981,667	1,579	26%	City
Poudre School District	CO	2007–08	29,527	53	9%	City
Baltimore City Public Schools	MD	2008–09	83,666	182	31%	City
Douglas County School District	CO	2008–09	66,896	89	2%	Suburb
Falcon School District 49	CO	2010–11	20,561	22	8%	City
Boston Public Schools	MA	2011–12	53,885	120	28%	City
Charlotte-Mecklenburg Schools	NC	2011–12	146,211	164	17%	City
Newark Public School District	NJ	2011–12	40,889	65	33%	City
Prince George's County Public Schools	MD	2012–13	128,936	207	12%	Suburb
Adams 12 Five Star Schools	CO	2013–14	39,287	53	10%	Suburb
City of Chicago School District 299	IL	2013–14	387,311	591	27%	City
Cleveland Municipal School District	OH	2013–14	39,410	101	43%	City
Metro Nashville Public Schools	TN	2015–16	85,598	154	23%	City
Jeffco Public Schools	CO	2015–16	86,731	165	7%	Suburb
Santa Fe Public Schools	NM	2015–16	13,265	33	20%	City
Indianapolis Public Schools	IN	2016–17	31,371	67	41%	City
Atlanta Public Schools	GA	2018–19	51,500	89	33%	City
Shelby County Schools	TN	2018–19	114,487	208	34%	City

Note: Table reads: Minneapolis Public Schools adopted a WSF system in the 1993–94 school year, enrolls 36,793 students, has 86 schools, a poverty rate of 24 percent, and is located in a city.

Sources: Information gathered from U.S. Department of Education report, *Districts' Use of Weighted Student Funding Systems to Increase School Autonomy and Equity:Findings From a National Study*.

Note: Poverty rates are based on the 2016 Census Small Area Income Poverty Estimate (SAIPE) data for school districts.

Table 2: Summary Statistics

	Sum	Mean	SD	Min	Max	N
Pupil per Teacher 1995-2018						
Control	58,880	17.05	3.12	9	57	3,456
Treated	3,242	16.88	1.87	12	22	192
Total	62,121	17.04	3.07	9	57	3,648
Pseudo-Graduation Rate 1995-2009						
Control	10,172	4.93	0.95	0	9	2,070
Treated	258	4.29	1.23	2	7	60
Total	10,429	4.91	0.97	0	9	2,130
Graduation Rate 2007-2018						
Control	258,454	79.09	11.76	35	100	3,264
Treated	4,110	68.51	9.72	37	83	60
Total	262,564	78.90	11.81	35	100	3,324
Dropout Rate 1996-2009						
Control	23,768	4.95	3.49	0	33	4,802
Treated	689	8.20	3.96	1	21	84
Total	24,457	5.01	3.53	0	33	4,886

Note: Description: Pupil per teacher ratio from 1995-2018 for 8 treated districts and 144 control districts. Pseudo-Graduation Rate which is $\frac{\#\text{ of high school graduates within district}}{\text{total enrollment within the district}}$. Pseudo-Graduation was calculated from 1995-2009 for 4 treated districts and 138 control districts. Graduation Rates in percentage form from 2007-2018 for 5 treated districts and 272 control districts. Dropout rates from 1996-2009 for 9 treated districts and 343 control districts. Treated districts were identified using Georgetown Edunomics WSF Report (Roza et. al 2019-Present) and U.S. Department of Education WSF Report (Levin, Manship, Hurlburt, and Atchison 2019).

Table 3: Static Regression Coefficients

	Pupil Per Teacher	Graduation	Pseudo-Graduation	Dropout
Coefficient	0.0414	0.166	0.296	-1.004
	(0.18)	(1.52)	(0.17)	(0.57)
	[0.26]	[0.11]	[1.76]	[-1.78]
Mean Pre-Treatment	16.65	62.89	4.32	9.14
Observations	3,648	3,324	2,130	4,886

Note: Standard errors in parentheses and t-statistics in brackets. Static regression coefficients for every educational attainment. Note that pseudo-graduation is $\frac{\text{\# of high school graduates within district}}{\text{total enrollment within the district}} * 100$.Table of β coefficients Based on Equation 2. Each regression covers unique set of years and has its own set of control and treated schools based on which districts have data: Pupil per teacher ratio from 1995-2018 for 8 treated districts, Graduation Rates from 2007-2018 for 5 treated districts, Pseudo-Graduation from 1995-2009 for 4 treated districts, and Dropout Rates from 1996-2009 for 9 treated districts.

Appendix

Appendix Tables

Table A1: Leading and Lagging Coefficients for Pupil Per Teacher Ratio Dynamic Regression

	(1) Coefficient
lead/lag year=-21	1.768
	(1.47)
lead/lag year=-20	-0.212
	(1.47)
lead/lag year=-19	1.253
	(1.47)
lead/lag year=-18	-0.0495
	(1.47)
lead/lag year=-17	0.685
	(1.09)
lead/lag year=-16	0.354
	(0.93)
lead/lag year=-15	0.545
	(0.93)
lead/lag year=-14	-0.0487
	(0.93)
lead/lag year=-13	-0.648
	(0.77)
lead/lag year=-12	-0.0144
	(0.69)
lead/lag year=-11	0.220
	(0.69)
lead/lag year=-10	0.0362
	(0.69)
lead/lag year=-9	0.275
	(0.69)
lead/lag year=-8	-0.136
	(0.69)
lead/lag year=-7	-0.0745
	(0.67)
lead/lag year=-6	-0.885
	(0.67)
lead/lag year=-5	-1.264*
	(0.64)
lead/lag year=-4	-0.642
	(0.65)
lead/lag year=-3	-0.432
	(0.65)
lead/lag year=-2	-0.307
	(0.64)
lead/lag year=-1	0
	(.)
lead/lag year=0	0.0511
	(0.64)
lead/lag year=1	-0.281
	(0.65)
lead/lag year=2	-0.0913
	(0.65)
lead/lag year=3	-0.120
	(0.67)
lead/lag year=4	-0.0774
	(0.67)
lead/lag year=5	-0.374
	(0.67)
lead/lag year=6	-0.381
	(0.67)
lead/lag year=7	-0.464
	(0.69)
lead/lag year=8	-0.0997
	(0.73)
lead/lag year=9	-0.323
	(0.73)
lead/lag year=10	-0.569
	(0.73)
lead/lag year=11	-0.378
	(0.83)
lead/lag year=12	1.426
	(1.09)
lead/lag year=13	0.824
	(1.09)
lead/lag year=14	1.675
	(1.09)
lead/lag year=15	-1.362
	(1.09)
lead/lag year=16	-0.881
	(1.09)
lead/lag year=17	0.267
	(1.47)
lead/lag year=18	-0.621
	(1.47)
Constant	17.01***
	(0.63)
Observations	3648

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Raw regression coefficients with standard errors on lead and lag years for Pupil per teacher ratio from 1995-2018 for 8 treated districts and 144 control districts. Note the year before treatment has been omitted to avoid multicollinearity and have a relative time reference. This table was used to create Figure 2.

Table A2: Leading and Lagging Coefficients for Graduation Rate Dynamic Regression

	(1)
	Coefficient
lead/lag year=-6	-9.674*
	(4.19)
lead/lag year=-5	-8.585*
	(3.63)
lead/lag year=-4	-4.043
	(3.10)
lead/lag year=-3	-4.713
	(3.10)
lead/lag year=-2	1.002
	(3.10)
lead/lag year=-1	0
	(.)
lead/lag year=0	-6.132*
	(3.10)
lead/lag year=1	-3.117
	(3.10)
lead/lag year=2	-4.739
	(3.10)
lead/lag year=3	-4.068
	(3.10)
lead/lag year=4	-1.625
	(3.10)
lead/lag year=5	-2.564
	(3.10)
lead/lag year=6	-5.371
	(3.63)
lead/lag year=7	-5.211
	(4.19)
Constant	84.98***
	(3.07)
Observations	3324

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Raw regression coefficients and standard errors in parenthesis on lead and lag years for Graduation Rates from 2007-2018 for 5 treated districts and 272 control districts. Note the year before treatment has been omitted to avoid multicollinearity and have a relative time reference. These values were used to create Figure 3.

Table A3: Leading and Lagging Coefficients for Pseudo-Graduation Rate Dynamic Regression

	(1) Coefficient
lead/lag year=-12	0.348
	(0.41)
lead/lag year=-11	0.00444
	(0.41)
lead/lag year=-10	0.0655
	(0.41)
lead/lag year=-9	-0.189
	(0.41)
lead/lag year=-8	-0.0631
	(0.41)
lead/lag year=-7	-0.179
	(0.41)
lead/lag year=-6	-0.0200
	(0.41)
lead/lag year=-5	0.123
	(0.36)
lead/lag year=-4	0.245
	(0.33)
lead/lag year=-3	0.135
	(0.33)
lead/lag year=-2	0.204
	(0.33)
lead/lag year=-1	0
	(.)
lead/lag year=0	0.106
	(0.33)
lead/lag year=1	0.0923
	(0.33)
lead/lag year=2	-0.0428
	(0.33)
lead/lag year=3	0.186
	(0.41)
lead/lag year=4	0.318
	(0.41)
lead/lag year=5	0.650
	(0.41)
lead/lag year=6	1.115^{**}
	(0.41)
lead/lag year=7	1.043^{*}
	(0.41)
lead/lag year=8	0.806^{*}
	(0.41)
lead/lag year=9	0.926^{*}
	(0.41)
lead/lag year=10	1.334^{*}
	(0.53)
Constant	4.798^{***}
	(0.32)
Observations	2130

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Raw regression coefficients with standard errors on lead and lag years for Pseudo-Graduation rate:

$$\frac{\#\text{ of high school graduates within district}}{\text{total enrollment within the district}} * 100$$

Pseudo-Graduation was calculated from 1995-2009 for 4 treated districts and 137 control districts. Note the year before treatment has been omitted to avoid multicollinearity and have a relative time reference. This table was used to create Figure 4.

Table A4: Leading and Lagging Coefficients for Dropout Rate Dynamic Regression

	(1) Coefficient
lead/lag year=-9	-5.515*
	(2.22)
lead/lag year=-8	-4.321***
	(1.28)
lead/lag year=-7	-5.227***
	(1.42)
lead/lag year=-6	-1.684
	(1.45)
lead/lag year=-5	2.922*
	(1.41)
lead/lag year=-4	2.824*
	(1.41)
lead/lag year=-3	1.163
	(1.27)
lead/lag year=-2	-1.068
	(1.06)
lead/lag year=-1	0
	(.)
lead/lag year=0	0.240
	(1.07)
lead/lag year=1	-0.792
	(1.03)
lead/lag year=2	-1.810
	(1.08)
lead/lag year=3	-1.488
	(1.16)
lead/lag year=4	-1.667
	(1.34)
lead/lag year=5	-1.642
	(1.49)
lead/lag year=6	-1.472
	(1.34)
lead/lag year=7	-2.260
	(1.45)
lead/lag year=8	-1.839
	(1.45)
lead/lag year=9	-3.231*
	(1.45)
lead/lag year=10	-2.431
	(2.22)
Constant	4.806***
	(1.07)
Observations	4886

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Note: Raw regression coefficients with standard errors on lead and lag years for Dropout Rates from 1996-2009 for 9 treated districts and 343 control districts. Note the year before treatment has been omitted to avoid multicollinearity and have a relative time reference. This table was used to create Figure 5.

CUER

www.cornelleconomicssociety.com