Effects of Ridership with Implementation of Fare-Free Public Transportation

Ethan Cherne
Glennon Langan
University of St. Thomas
May 2020

Introduction

As municipalities grapple with increased urbanization and traffic congestion that wastes time for workers commuting, public transit seems to be the solution. Public transit solves many issues including, but not limited to, lack of road infrastructure, pollution, rush hour traffic, and lack of bike and walk accommodating roadways. The United States is facing urban sprawl where cities and megacities alike are making the choice to invest in public transit. What some cities are trying to do is increase ridership by implementing a fare-free (FF) transit system. This does not require passengers to pay anything and would be funded by the local government authority. This research studies the various effects on ridership for public transit districts in the United States. The hypothesis for this study is:

H_0: *Fares* ≥ 0

H_A: *Fares* < 0

The variable fares is the main component of the study. A negative relationship is expected because of the success of fare free programs around the globe. Higher fares would disincentivize passengers to choose buses and trains over other modes of transit. The implications of this study should be used for policy making in the public sector as for individual districts to decide whether or not to implement a fare free program. Although this study is a very large sample size and close to population parameters, as all the public transit districts in the United States are required to report data to this database, it is in the best interest of policy makers to use this study with caution. This study is, by no means, a conclusive.

Literature Review

The four pieces we chose were all academic and from varied geographies. In order to study public transportation well, research must be done outside of the United States. Countries with a longer history of urban planning have better public transit. These areas are namely, Europe and East Asia. The large metropolitans in these regions have advanced transit systems that carry millions of people to and from their destinations every year.

The first article is a study done to help policy makers pass a bill that would allow students in Los Angeles county to get a free bus pass that would be available 24/7. The most notable thing about this study was the possible outcomes that the policymakers had predicted. The benefits include increased attendance, the school district saves money on bus services, financial freedom for families, better learning environments. Since the students would no longer have to pay for their fare, the bus company would lose out on one fifth of its revenue. On the upside of this deal, the long terms benefits would pay off by students being able to invest more time into their education. The paper concluded that the program would have benefits across many sectors, although the serious expenditures was left unresolved.

The Los Angeles case highlights the many of the affects that this study does not account for. This is mainly due to the structure of the study. At the time of writing, the researchers' goal was to run a cost-benefit analysis of such a program in Los Angeles but did not have any data to conclude whether this program increased ridership. Nonetheless, this case can be a strong case for FF programs.

The second study was done in a developing city named Gaoping in the Shanxi Province of China. The fare-free demonstration was a system wide program for that city. The main modes of transportation were car, illegal motorcycle rides, biking and walking. Their prediction was that more people will take the bus, which will reduce traffic and use of illegal motorcycle taxis. The outcome was that traffic was changed at a negligible rate in the sample that it was not considered to be an effective way to reduce traffic. On the other hand, motorcycle taxis decreased in use and many of the drivers reported having to look for new jobs.

The Gaoping study was the most quantitative of the four articles chosen for review. There were real numbers on the actual effects of FF transit. There was a sharp increase in the use of public transit showing a more complicated relationship in prices. The Gaoping study was able to calculate the elasticity of demand for transit, finding transit to be an inelastic service. The researchers kept this in mind but did not have enough data to include a demand model for this study.

A more philosophical approach to this subject is written by Victor Wallis. The author uses logical rhetoric to compare the use of elevators to buses in saying that they are both forms of transportation, but we pay for one and not the other. Further points are made that there are countless benefits that are not considered in the policy making process, which will translate to laws not getting passed that are in favor of public transit. Free transit comes down to an issue of equality. It takes an overhaul in political and social views for a community, let alone a society, to adopt free and accessible transportation.

There were not many methods that could be extracted from this journal. If anything, this would be more helpful to read over on behalf of a policy maker, as it includes the qualitative benefits that are not included in this study. Mainly, this article points out there isn’t much literature on the subject, which leads to gaps in data.

The final article looks at the social justice issues surrounding public transit in Calgary, Alberta. Out of the four articles, this is the most comprehensive in terms of scope and data. The point made here is to maximize the utility of individuals by giving them the money that they would have spent on transportation but instead are choosing to take free transit. This system incentivizes citizens to not use cars, which in turn reduces traffic and pollution. Political arguments from both sides were expressed on the literature, where the author states that no one group seemed to be louder than the other. This paper is another example of the many benefits of the program that are held back by public opinion and the political process.

These four articles helped create the foundation for this study. Variables became recognizable and some were determined more important than others. While the literature was informative, it wasn't able to fit the data to what was expected. The data picked for research had preselected variables that did not necessarily align with important factors of FF programs. Discrepancies in the data will be discussed in later sections.

The Model

This study ultimately aims to answer the question of whether fare free programs increase or decrease ridership for transit authorities in the United States. To test this, ridership is identified as the dependent variable. The model will show how ridership changes with variation in any of the x variables. The researchers employed a classical linear regression model to come up with the following equation as the model:

$$\textit{Fig 1.}\ UPT_i = \beta_0 + \beta_1 Fares_i + \beta_2 UZA_i + \beta_3 SaPop_i + \beta_4 OEFY_i + \beta_5 ATLFY_i + \beta_6 ACTFY_i + \beta_7 SASQM_i + \varepsilon_i$$

Figure one is the original model with alteration to the variables.

$$\textit{Fig. 2}\ lnUPT_i = \beta_0 + \beta_1 lnFares_i + \beta_2 UZA_i + \beta_3 SaPop_i + \beta_4 lnOEPC_i + \beta_5 ATLFY_i + \beta_6 ACTFY_i + \beta_7 SASQM_i + \varepsilon_i$$

Figure two is the final model which is a semi-log variant of the original. The explanatory variable is Unlinked Passenger Trips (UPT) which will be interchangeably used with ridership. UPT is the best variable to describe ridership because it counts every time a passenger steps on a train or bus without relation to a previous trip. Therefore, the actual number of vehicles used to get from point

A to point B will inflate the number of stand-alone trips made without regards the number of vehicles used. The first independent variable is Fares, which is the dollar amount collected as payment. The researchers predict that this will have a negative correlation with UPT. It is assumed that the higher the price of tickets, the less passengers will take public transit. The next variables Urbanized Zone Area Population Density (UZA), Service Area Population (SaPop), and Service Area Square Miles (SASQM) have to do with the population and land of the districts. The researchers predict that these variables will have a positive correlation with UPT because more people in the population would suggest a larger demand for transit and ridership. In Figure 1, the variable Operating Expenses Fiscal Year (OEFY) reports the operating expenses for the respective transit authority. The researchers predict that OEFY should have a positive effect on UPT because the authorities with more expenses have larger districts and more vehicles, therefore completing more trips. In Figure 2, the variable Operating Expenses per Capita (OEPC) is a variant of OEFY but is divided by the SaPop. This function helps reduce collinearity and will be explained in later sections. The researchers predict that this variable will have a positive effect much like OEFY, but now controlled for population. The final two variables are in regard to the quantitative aspects of the actual trips. Average Trip Length (ATLFY) is the average distance, in miles, that UPT's are in that district. Average Cost Per trip (ACTFY) is the dollar amount it costs for the transit authority to provide the service per person. The researchers predict that ACTFY will have a negative effect on UPT because it the costs might be transferred onto the passenger. This aligns with the previous prediction of fares negatively effecting UPT. Finally, the researchers predict that ATLFY will have a negative effect on UPT because longer trips will cost the consumer more, decreasing the demand for UPT.

Many of the variables in this model are not predicted to be highly correlated or significantly change the UPT but they could help in policy decision making. Variables involving the ~~population~~population, or the aspects of passenger trips will help policy makers determine if a future program will be able to apply this model to their population. The variables that the researchers are the most focused on are Fares and Operating Expenses in their relation to UPT. These variables most closely adhere to the hypothesis.

Data Description

The data collected for this model is from the United States Department of Transportation. This data is collected for the sole purpose if policy making under the Secretary of Transportation. The first page of the data is known as the "Master", and it is where this study takes it sample from. At the time of the study, the most recent monthly data that was collected for January 2020. The reporting individuals in the dataset are transit districts, which come in many forms. These districts are typically metropolitan areas but can encompass entire counties and transit authorities for universities. All districts are to provide data for predetermined categories. The first variable in the dataset is the mode of transportation for that district. The modes include passenger trains, monorails, buses, rapid transit (buses with faster routes), ferry boats, and more. The way the data was reported was that a district would be represented for as many iterations as there were different types of transit. For example, Metro Transit, the district that operates in Minneapolis, St. Paul, and surrounding cities, is recorded as three individuals; motorbus, commuter bus, and light rail. For the simplification of the model, the individuals were only districts that had passenger trains or buses. This limited the individuals from roughly 2,100 to around 700 from the sample. Fare free programs were identified by choosing districts that collected $0 in fares for the most recent fiscal year. There were 40 individuals that were identified as fare free. Further variables in the data set were those used in the model.

An ANCOVA model was considered for this data. The model would have compared the 40 FF programs to the rest of the programs. The value of 0 signified FF and 1 signified paid fare programs. After running a preliminary regression, researchers found this model statistically insignificant in terms of correlation. The R^2 computed was less than 0.001. Further justification and explanation will be included in the Interpretations section.

The other pages of the study are named Unlinked Passenger Trips, Vehicle Revenue Miles, Vehicle Revenue Hours, and Vehicles Operated in Maximum Service. These datasets go into extreme detail of their prospective topics. The study only considered the “Master” and “Unlinked Passenger Trips” as sources of data as they relate to the hypothesis. The UPT dataset was considered unfit for the study because there were significant gaps in the data. The sample was a timeseries data where many of the districts we had chosen for modes of transit had large gaps. Data was collected monthly since 2002 until January of 2020. Of the 700 identified individuals, only 30 were considered fare free districts, as opposed to the 40 fare free programs from the Master sample. Further complications were found when figuring out how to interpret the data. One proposed method was to compare difference in the number of Unlinked Passenger Trips from the most recent date to the earliest date available. The problem was that for the 30 FF districts, many of them appeared to be only in operation for a few years at a time, which would result in a decrease. This method was unreliable and not considered further.

Once the data was organized, the researchers decided on using the Master sample to regress the variables Unlinked Passenger Trips, Fares FY, Urbanized Area Population/Square miles, Service Area Population, Operation Expenses FY, Average Trip Length FY, Average Cost per

Trip FY, Service Area Square Miles, Urbanized Area Square Miles, and Urbanized Area Population. Many of these variables were variations of others and were omitted in the final model for simplification.

The regression of this model was preformed using Stata/IC 16.1. The distribution of values for the variables were estimated and the summary statistics of the data set used for this study can be found below in Table 1.

Variable	Obs	Mean	Std. Dev.	Min	Max
UPT	728	1.32e+07	1.04e+08	0	2.63e+09
FFY	735	2.05e+07	1.45e+08	0	3.50e+09
UZA	735	2980.554	1457.511	809.9046	6999.422
SaPop	735	1130982	1914787	0	1.12e+07
OEFY	735	5.69e+07	2.47e+08	0	5.07e+09
ATLFY	728	9.134733	11.81933	0	92.7719
ACTFY	727	9.288074	9.591512	.8746	139.9655
SASQM	735	642.1238	1216.606	0	15355
OEPC	734	50.46788	68.10385	0	720.3764

Table 1: Summary statistics for all variables involved in the estimated model. Number of observations (Obs), mean, standard deviation (Std. Dev.), minimum (Min), and maximum (Max)

Many of the variables have a minimum of zero, which translates to right skewed distributions. Under the central limit theorem, the distributions should be normal. Standard deviations of the variables also demonstrate the skewness of the data. For example, average trip length has a mean of 9.1, a standard deviation of 11.8 and a maximum of 92.8. Those statistics show that most of the trips for these districts are less than ten miles.

The raw regression of results for all variables in the sample can be found in Table 2. These results show strong correlation of UPT to FFY, SaPop, OEFY, and ATLFY. The model explains 92% of the variation in UPT, denoted by the R^2 value. This value is not as high as the final model, where the explanatory variable and some of the independent variables have their natural logs taken.

Natural logs can help control for skewness in these variables. The final model has more significant variables than the raw regression.

Source	SS	df	MS			
Model	7.1780e+18	10	7.1780e+17	Number of obs	=	727
Residual	6.0989e+17	716	8.5181e+14	F(10, 716)	=	842.68
				Prob > F	=	0.0000
				R-squared	=	0.9217
				Adj R-squared	=	0.9206
Total	7.7879e+18	726	1.0727e+16	Root MSE	=	2.9e+07

UPT	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
FFY	.6309452	.0258107	24.45	0.000	.5802715	.6816188
UZA	1971.111	1247.001	1.58	0.114	-477.1048	4419.327
SaPop	-3.982817	.9036679	-4.41	0.000	-5.756973	-2.208662
OEFY	.0449467	.0167206	2.69	0.007	.0121193	.077774
ATLFY	-337140.5	104470.7	-3.23	0.001	-542246.1	-132035
ACTFY	-87889.83	124977.3	-0.70	0.482	-333255.6	157476
SASQM	66.18975	1041.057	0.06	0.949	-1977.7	2110.079
UZASQM	2256.008	3663.838	0.62	0.538	-4937.143	9449.158
UZAPop	-.9138707	.8404936	-1.09	0.277	-2.563997	.7362559
OEPC	-37306.13	19558.32	-1.91	0.057	-75704.63	1092.377
_cons	2993343	3856235	0.78	0.438	-4577536	1.06e+07

Table 2: Raw regression of all variables in the sample.

Model Estimation and Interpretation of Results

This study was able to avoid significant issues with multicollinearity, tested by correlation between independent variables and variance inflation factor. However, given the timeline and used dataset this estimation has issues of heteroskedasticity, tested by White’s test (p-value < 0.001) and Breush-Pagan/Cook-Weisburg test (p-value < 0.001). Additionally, model misspecification issues were found with the Ramsey test (p-value < 0.001) and the link test (p-value < 0.001).

The unrestricted model was found to be statistically significant ($F_{7, 681}$ = 1614.10, F-value < 0.0001). The goodness of fit, R^2, was found to be 0.94. The regression report from the model can be found below in Table 2.

Source	SS	df	MS			
				Number of obs	=	689
				F(7, 681)	=	1614.10
Model	2974.13729	7	424.876756	Prob > F	=	0.0000
Residual	179.257916	681	.263227483	R-squared	=	0.9432
				Adj R-squared	=	0.9426
Total	3153.39521	688	4.58342327	Root MSE	=	.51306

Table 2: Regression report including error terms, degrees of freedom, F-statistics and R^2

While the entire model was significant, there was one individual variable from the model that through a t-test were found to be insignificant alone. This variable was the urbanized area population density (UZA). The variables that were found to be significant were the natural log of fares (p-value < 0.001), service area population (p-value < 0.001), natural log of operating costs per capita (p-value < 0.001), average trip length (p-value < 0.001), average cost per trip (p-value < 0.001), and service area square miles (p-value = 0.005). These results are also found in Table 3 below.

lUPT	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
lFFY	.6083921	.0168401	36.13	0.000	.5753273	.6414569
UZA	.0000162	.0000155	1.05	0.295	-.0000142	.0000467
SaPop	1.61e-07	1.61e-08	10.00	0.000	1.29e-07	1.92e-07
lOEPC	.3248274	.0220735	14.72	0.000	.2814871	.3681677
ATLFY	-.0325572	.0019086	-17.06	0.000	-.0363047	-.0288097
ACTFY	-.0277546	.0024295	-11.42	0.000	-.0325248	-.0229844
SASQM	.0000526	.0000185	2.84	0.005	.0000162	.0000889
_cons	4.73574	.1861716	25.44	0.000	4.370201	5.10128

Table 3: Regression results for the variables of the model. Slope coefficients, standard error, t-statistic, p-values, and 95% confidence intervals.

This regression estimation found that an increase in fares was correlated with increased unlinked passenger trips (p-value < 0.001). The coefficient was found to be 0.608, which means on average this model predicts that when fares increase by 1%, unlinked passenger trips decrease by 0.6%. This appears to be a contradiction of hypothesis and the literature, but it is important to consider how the data is currently organized and how the model is currently constructed. The fares are the total fares in USD collected by the transit authority in the fiscal year. Because of this, assuming two transit authorities had the same pricing, the one with more trips would be expected to have higher fares. This model does not consider the distinction between fare-free and fare collecting programs because the introduction of the dummy variable resulted in an estimate that explained very little error in the data. Additionally, this model is a modified semi-log model that captures how changes in the independent variables are related to percent changes in the dependent, unlinked passenger trips. These conditions will be true for all future variables discussed in this report. The literature from observations in Gaoping, China (Shen & Zheng, 2015) and estimations in L.A., USA (Gase et al., 2014) both supported the idea that by eliminating fare charges on public transit the number of trips would increase. This study's model is not be able to reliably capture the effect of fares on unlinked passenger trips in the same way since this study utilizes cross sectional data from the USA and those studies looked at data over time for specific municipalities. Demand for fare-free programs have been found to be inelastic, which the coefficient for the natural log of fares agrees with (Shen & Zheng, 2015).

The basis of this study's hypothesis for service area population was that the larger the population the transit authority services, the more unlinked passenger trips since it would be expected more people would use the transit system. In this regression this was found to be true, with an increased population being correlated with number of trips (p-value < 0.001). The estimated coefficient was 1.61e-7. Based on this model estimation for everyone person increase in population, on average, unlinked passenger trips would be expected to increase by 0.000000161%. A more relatable number would be that on average the model predicts if a population were increased by 10 million, trips would increase by 1.61%. This is consistent with the reasoning that more people will result in more trips.

The operating expenses per capita was found to have a statistically significant positive correlation with unlinked passenger trips (p-value < 0.001). The estimated coefficient was 0.325. This result predicts that when operating expenses per capita increases by one percent, on average, UPT will increase by 0.325%. This becomes increasingly difficult to interpret in terms of causality because how much of an increase in unlinked passenger trips drives an increase in operating expenses vs. how much of an increase in expenses drives an increase in trip, which was a difficulty in constructing this model. An increased amount of expenses correlates with passenger trips but ultimately the amount of passenger demand dictates the operating expenses of a transit authority. In the analysis of data, issues were found with multicollinearity and the variance inflation factor which led to the replacement of total operating expenses with operating expenses per capita. Operating expenses are important for policy decisions surrounding fare-free programs as transit expenses are one of the biggest drawbacks, so it is important to consider how the two relate.

The regression of average trip length found a negative correlation with unlinked passenger trips (p-value < 0.001). The slope coefficient was -0.0325. In words, for each 1-mile increase in average trip length, on average, the percent of trips decreases by 0.0325%. This suggests that a larger number of passengers is correlated with transit that offers much more transfers/destinations.

Average cost per trip was found to have a significant positive correlation with unlinked passenger trips (p-value< 0.001). The estimated coefficient was -0.0277. This result estimates that for each dollar increase in average cost per trip, the percent of unlinked passenger trips decreases by 0.0277%. This result agrees more with our initial hypothesis that increased cost of a transit ride would decrease trips because of demand but this is the trend for the transit authority, not the consumer.

The regression of service area square miles was found to have a significant positive correlation with unlinked passenger trips (p-value < 0.005).

The constant for this model was found to be statistically significant (p-value < 0.001). The coefficient was 4.74 and using the anti-log this model predicts for all independent variables at zero the number of unlinked passenger trips would be almost 114. This is not a reportable result in the context of this model because it is not possible to have all variables at zero for a transit authority.

The significant results were found for the model,

$$lnUPT_i = \beta_0 + \beta_1 lnFares_i + \beta_2 UZA_i + \beta_3 SaPop_i + \beta_4 lnOEPC_i + \beta_5 ATLFY_i + \beta_6 ACTFY_i + \beta_7 SASQM_i + \varepsilon_i$$

In addition, significant results were found for the variables of natural log of fares, service area population, natural log of operating expenses per capita, average trip length, average cost per trip, and service area square miles. Results such as fares seem to contradict this studies hypothesis and the cited literature that fare free increases demand, but the authors of this study believe this is not yet discernable given the way the model is constructed.

Conclusion

The model showed that at this time, while there is a positive correlation between fares and passenger trips, a causal relationship cannot be established with confidence. The findings of this study contradict the cited literature and intuitive relationship of demand for transit trips. The largest difference between the cited literature and this study is the scope of the data. The literature looked at data over time in a specific city while this study looked to estimate the relationship between unlinked passenger trips and data related to transit, with an emphasis on fares.

In further research a better constructed model would allow the utilization of dummy variables to discern between the fare-free and fare collecting programs to see if the fare-free programs attract consumers that might otherwise purchase their own vehicle or use another mode of transport when facing fares. This model has multiple flaws as discussed at the beginning of the interpretation section but through further refinement of the model data could be used to more accurately predict the relationship between unlinked passenger trips and fares.

This study is a testament to the importance of time and relevant variables. One of the great obstacles of this work was designing the model that made the most sense to the authors as well as attempting to limit issues related to the model. In the end we only were able to prevent multicollinearity by creating an operating expenses per capita that ended up being log transformed, but we were able to report all results of the diagnostics to represent those issues that could affect the model.

This study did not find a general trend across operating areas in the United States that would support a fare-free program in all areas. Based on these results from the country-wide data used in this study and the literature cited, the application of a fare free system would need to be determined on a case-by-case basis based on data over time.

Bibliography

Federal Transit Administration, Monthly Module Adjusted Data Release. (2020, April 1). https://www.transit.dot.gov/ntd/data-product/monthly-module-adjusted-data-release

Gase, L. N., Kuo, T., Teutsch, S., & Fielding, J. E. (2014). Estimating the costs and benefits of providing free public transit passes to students in los angeles county: Lessons learned in applying a health lens to decision-making. International Journal of Environmental Research and Public Health, 11(11), 11384-97. Retrieved from https://login.ezproxy.stthomas.edu/login?url=https://search-proquest-com.ezproxy.stthomas.edu/docview/1626513183?accountid=14756

Oldridge, P. (2012). The implications of fare-free transit: A canadian case study (Order No. MR83835). Available from ProQuest Dissertations & Theses Global. (1022986840). Retrieved from https://login.ezproxy.stthomas.edu/login?url=https://search-proquest-com.ezproxy.stthomas.edu/docview/1022986840?accountid=14756

Shen, J. X., & Zheng, S. K. (2015). Fare-free public transit service: experience from Gaoping city of China. Advances in Transportation Studies, 3–12. https://doi-org.ezproxy.stthomas.edu/10.4399/978885488881401

Wallis, V. (2019). Free Public Transit. Monthly Review: An Independent Socialist Magazine, 71(7), 59–62. https://doi-org.ezproxy.stthomas.edu/10.14452/MR-071-07-2019-11_5