Impact of Fares on Passenger Transit Trips

Benjamin R. Davison

University of Kentucky

Abstract

In this paper, we explore the relationship between average transit fares and unlinked passenger trips (UPT) per person based on data from the National Transit Database (NTD), which carries 1,102 agency-mode combinations from across the United States. Various forms of this relationship are explored with respect to statistical properties of the data, and a model is defined. From this relationship, due to the nature of fares and trips as a price-good combination, we derive an estimate for elasticity under the assumption of a linear demand curve, given the wide sample size of (price, quantity) estimates. A potential population effect is also explored in the relationship between total fares and total passenger trips. In addition to understanding this general relationship, we examine the differences between both various transit modes, such as rail (LR) vs. bus (MB), and types of service, referring to whether an agency-mode combination is directly operated (DO) or privately contracted (PT).

Impact of Fares on Passenger Transit Trips

The relationship between price and quantity has remained a classical economic question for all of time; as new markets arise, these ideas are re-applied to new situations to test validity of theory or direction of effect. These tests are driven by new data that is essential to testing economic questions. Here, we examine this relationship in public transit to see if fares (AF) have any effect on quantity demanded of unlinked passenger trips (UPT). To measure the strength of such a relationship, we can use price elasticity formulas to determine the strength of direction given that fares represent prices and passenger trips represent quantities. In our context, we can reapply classical elasticity formulas for a given model, assumed to be linear (Van Zandt):

$$Q = \beta_0 + \beta_1 P + u$$

$$UPT = \beta_0 + \beta_1 AF + u$$

$$Elasticity_i = \left(\frac{dQ}{dP}\right) * \left(\frac{P}{Q}\right) = -\beta_1 * \left(\frac{UPT_i}{AF_i}\right)$$

Notice that, if we use logged variables, we can simplify this formula based on the proxy of the natural log for a percentage change:

$$\ln(Q) = \beta_0 + \beta_1 \ln(P) + u$$

$$Elasticity = -\beta_1$$

The negative sign here accounts for the negative slope of the demand curve assumed here. Thus, if we estimate a regression $\ln(UPT) = \beta_0 + \beta_1 \ln(AF)$, we can interpret the coefficient β_1 as an elasticity measurement. This principle will be applied in this paper as a method of measuring the strength of the relationship between prices and quantities in a transit context, as we seek to determine the sensitivity of potential passengers to changes in fares. Though we do not have explicit demand curves for individual transit agencies, we have one (fare, trips) relationship for each agency + mode (such as rail or bus) combination in our data, sourced from the National Transit Database. This creates a wide enough dataset to estimate a market curve.

Data

The data used here includes 1,102 agency-mode combinations (hereby referred to as "services"), reported by transit agencies across the United States to the National Transit Database (maintained by the Federal Transit Administration). Each service used in this dataset represents data from the 2019 calendar year, ensuring that each price value is equivalent. Of course, there are differences in costs of living across the United States, which generally explains the wide distribution in average fares (Figure 1). The wide tails here are explained by this and the various transit modes; some modes included are "demand response," which indicate that the service is explicitly generated from a request (such as calling a taxi, for example).

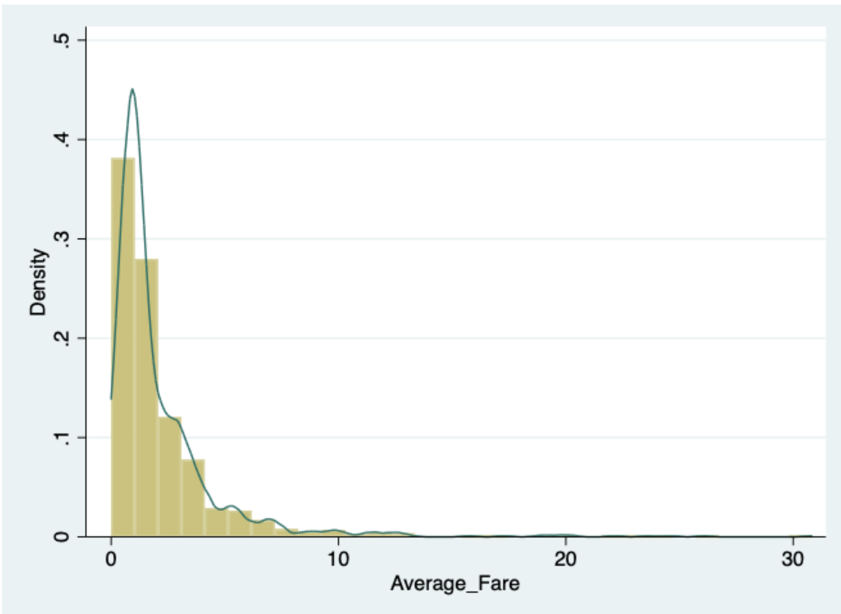

Figure 1

We also examine the distribution of passenger trips, normalized by person (Figure 2). This normalization is used for both total fares and total passenger trips (the original columns in the dataset) to explicitly account for wide variances in population and to create consistency.

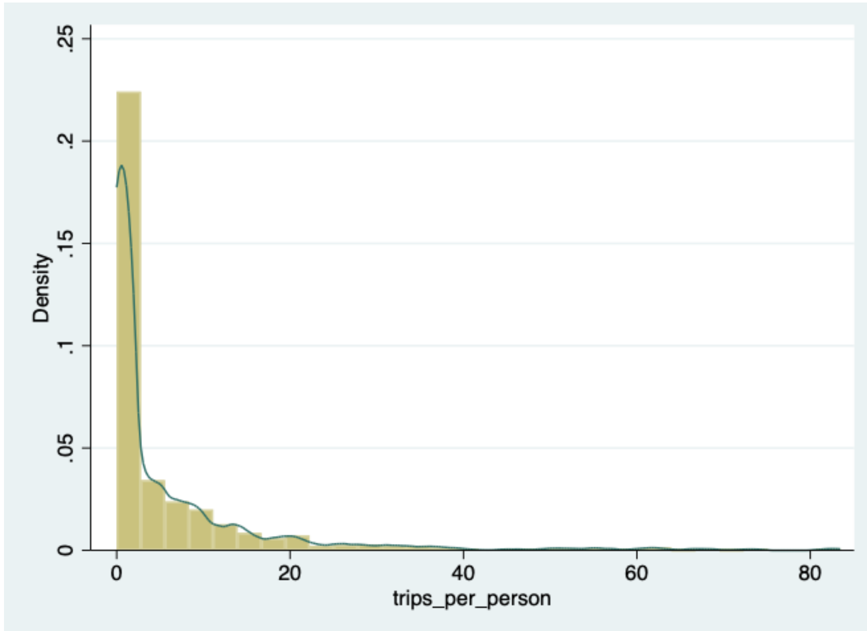

Figure 2

It should be glaringly obvious from examining both Figures 1 & 2 that our variables have heavy skews; to account for this, we log-transform both variables to obtain a more normal distribution (justified later in model estimation). In addition to modelling this relationship between fares and trips, we will also examine the potential effect of the population of the service area (sa_pop), Summary statistics for the main variables in this study are shown in Table 1.

Table 1	(1) N	(2) mean	(3) sd	(4) min	(5) max
sa_pop	1,102	936,893	1.685e+06	12,954	1.164e+07
unlinked_trips	1,102	6.063e+06	2.916e+07	1,158	6.916e+08
fares	1,102	1.042e+07	5.644e+07	0	8.461e+08
average_fare	1,102	2.162	2.848	0	30.84
trips_per_person	1,102	5.949	11.51	0.00220	83.45

Two other modifications/investigations explored here relate to the possible differences in the effect/elasticity between modes and types of service. “Modes” included in this dataset are rail (such as heavy rail or subway systems, denoted “LR”), bus (denoted “MB”), and demand response, discussed earlier (denoted “DR”). Due to the wide variety of services included under the DR category, only bus and rail are focused on in this study. The reasoning behind this comparison is that the customer bases for rail vs bus transit services are likely to vary; often, rail services connect locations using longer distances with fewer alternatives, so the relationship for rail services may be relatively inelastic compared to bus trips, which typically connect using shorter distances (thus, more alternatives). Correspondingly, bus services have lower fares as shown in the means for key variables by mode (Table 2).

Table 2	(1) DR-N	(2) mean	(3) LR-N	(4) mean	(5) MB-N	(6) mean
average_fare	441	2.866	98	2.552	563	1.542
trips_per_person	441	0.401	98	11.67	563	9.299
log_trips_per_person	441	-1.316	98	1.239	563	1.162
log_average_fare	441	0.746	98	0.438	563	0.0581

For types of service, the NTD differentiates services as either directly operated by the transit property (DO) or operated by a private third party (PT). It is hypothesized that this differentiation represents a wide variety of factors that may effect elasticity. For example, “PT” services may cater to a very different customer base than DO services, or have specific characteristics that make its services less replaceable. To represent this in models, we use a dummy variable (tos_private) which indicates (DO = 0, PT = 1):

$$\ln(UPT) = \beta_0 + \beta_1 \ln{(AF)} + tos_{private} + u$$

This will complicate our elasticity estimation, but, ultimately, we only make this distinction to see if there is any significant difference between the two types of service (and, if so, the direction of the difference). Because DR modes represent a massive chunk of our dataset, we opt to not use a dummy variable between MB and LR (bus and rail), because this clouds the estimation of the difference between the two modes. Instead, we run separate regressions for total (all modes), LR only, and MB only, with the expectation of examining the difference in coefficients between the LR and MB regressions. We examine the means of key variables by type of service in Table 3.

Table 3	(1) N-DO	(2) mean	(3) N-PT	(4) mean
average_fare	600	2.002	502	2.353
trips_per_person	600	8.097	502	3.381
log_trips_per_person	600	0.658	502	-0.397
log_average_fare	600	0.252	502	0.506

We note that many services were excluded from the dataset to reflect consistencies (such as ensuring all reporting periods were 2019) and to eliminate clearly irrelevant data (such as niche modes with low sample sizes, or services with fares = 0). Due to the simplicity of the question, OLS is used for each model; we also use robust standard errors in all model estimation, with the hope of creating consistency and automatically battling any heteroskedasticity.

Empirical Results

As mentioned earlier, it is clear that logging of primary variables is necessary based on the skewed distributions. We establish the non-logged and logged models now, and compare results to justify the use of the logged model:

$$UPT = \beta_0 + \beta_1 AF + u$$

$$\ln(UPT) = \beta_0 + \beta_1 \ln(AF) + u$$

In Figures 3 and 4, we compare the scatterplots between these two models, and include a line to represent the selected model in Figure 4. It is clear that, without any log transformation, we have a very non-linear and high variance model. In applying a transformation, our model is simplified, and elasticity estimation is simpler (as explained in the introduction).

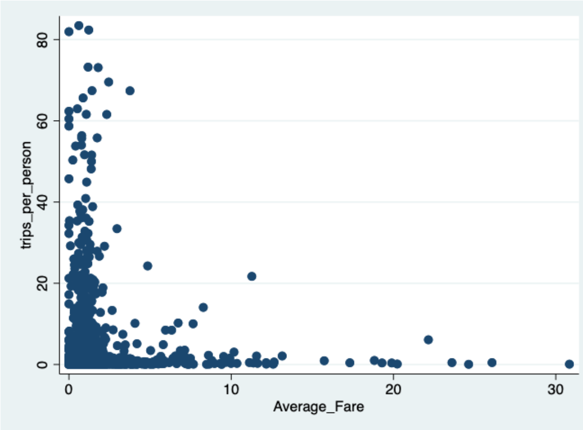

Figure 3

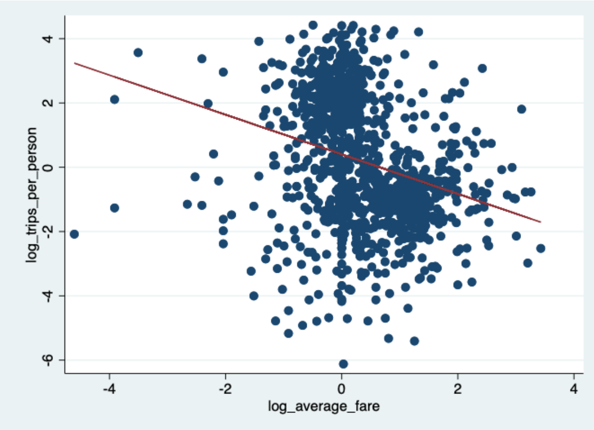

Figure 4

Results from the two regressions are now shown in Table 4. The robust standard error indication and significance level legend apply to all future tables in this study.

Table 4	(1) trips_per_person	(2) log_trips_per_person
average_fare	-0.694***	
	(0.0992)	
log_average_fare		-0.616***
		(0.0713)
Constant	7.448***	0.404***
	(0.463)	(0.0700)
Observations	1,102	1,102
R-squared	0.029	0.078

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Due to the reasons stated earlier, this logged model (2, from Table 4) is used going forward in adding new techniques and effects. Immediately, we make note of the 0.616 elasticity value (using the derivation in the introduction), representing a generally inelastic good. This effect is statistically significant, and verifies a clear relationship between price and quantity in this transit context. Although the R^2 value is generally low, we acknowledge that it is inevitable that there is a multitude of effects that will determine the number of trips taken per person for a given service, often specific to a region or service area. For example, it may be hypothesized that customers in a dense urban location such as New York City will ride transit at a tremendously higher rate than a rural area with just a small bus service. To test this hypothesis (as an aside from our elasticity estimation), we try and create a stronger model for unlinked_trips by looking at a more macro level and not normalizing by person, and instead including service area population in Table 5.

Table 5	(1) unlinked_trips
sa_pop	0.713
	(0.774)
fares	0.397***
	(0.115)
Constant	1.262e+06**
	(514,233)
Observations	1,102
R-squared	0.619

Interestingly, the effect of service area population is not statistically significant in modelling unlinked trips, while fares remains a statistically significant regressor. The R^2 value is tremendously higher than the models in Table 4; however, this is because the normalization per person modifies the variances enough to change the model. For reference, the R^2 of the model

$$unlinked\ passenger\ trips = \beta_0 + \beta_1 total\ fares + u$$

is 0.617, although this is just an unnormalized version of Table 4 (1). Ultimately, from Table 5, we conclude that service area population has little effect on passenger trips when fares are included in the model.

Now, we differentiate between modes and types of service to examine if our model coefficients change. Tables 6 and 7 show the results of our regressions discussed earlier (Table 6 compares LR vs. MB, and Table 7 uses the dummy variable approach to differentiate DO vs. PT). Additionally, Figures 5 and 6 show the scatter plots and model fits of each approach.

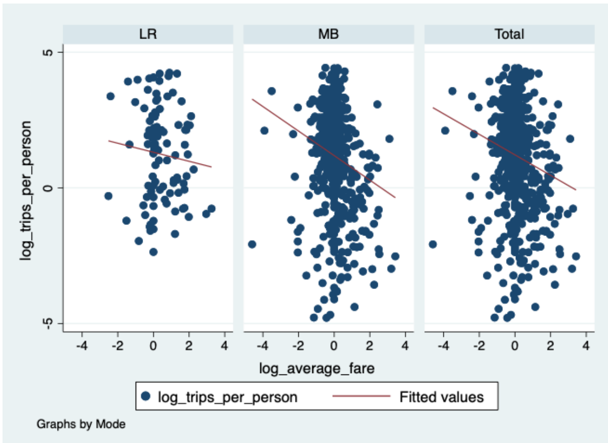

Figure 5

Figure 6

Table 6	(1) LR	(2) MB
log_average_fare	-0.166	-0.452***
	(0.177)	(0.123)
Constant	1.312***	1.189***
	(0.204)	(0.0784)
Observations	98	563
R-squared	0.009	0.039

Table 7	(1) log_trips_per_person
tos_private	-0.916***
	(0.115)
log_average_fare	-0.545***
	(0.0702)
Constant	0.795***
	(0.0836)
Observations	1,102
R-squared	0.129

Interestingly, in differentiating mode, we see that LR has a statistically insignificant coefficient on fares, indicating that fares may have little effect on trips. This may be due to small sample size, but it is also worth noting this implies an elasticity statistically indistinct from zero, or perfect inelasticity. We hypothesized earlier that rail is probably more inelastic due to the fewer alternatives available to customers, so this finding is in line with our reasoning, where buses clearly have a more elastic relationship between price and quantity demanded, albeit with a weak R^2. From Figure 5, we visualize this difference, where the model line for LR clearly is flatter, and explains less variance than MB.

Finally, we observe the difference in type of service. Immediately, we see that our dummy variable indicating 1 = PT, has a statistically significant coefficient, implying that there is a statistically significant difference between DO and PT. This gap of almost a -1 shift on the y-intercept is clearly visualized in Figure 6; it is clear that there are less trips per person for PT than DO. This dummy variable approach naturally increases our R^2 due to the inclusion of new variables, but this slightly changes the elasticity compared to our primary model.

Conclusion

Ultimately, in our primary study, we conclude that, in examining transit trips vs. average fares in the United States, transit services have a generally low price elasticity according to our derivation and model of about 0.616. When adding in an effect for population, we see virtually no change in effect when accounting for the lack of normalization. Additionally, we find that rail modes are much more inelastic than bus modes, likely due to the difference in alternatives available for each mode. Finally, we compare directly operated services against third-party services and find that while elasticity generally remains the same, privately operated services have less trips in general, shifting our original model down when differentiating.

References:

Federal Transit Administration. (2021, February). *National Transit Database (Monthly)* [Dataset]. https://www.transit.dot.gov/ntd/data-product/monthly-module-raw-data-release

Federal Transit Administration. (2017). *National Transit Database Policy Manual*. https://www.transit.dot.gov/sites/fta.dot.gov/files/docs/2017%20NTD%20Policy%20Manual.pdf

Van Zandt, T. (2012, August). *Firms, Price, and Markets* [Lecture Notes]. INSEAD. https://faculty.insead.edu/vanzant