*The Dartmouth Undergraduate Journal of Politics, Economics, and World Affairs*
*Volume. I, Issue. III, 5-23, June 2021*

# The Cost of Subway Delays: A Counterfactual Welfare Analysis of Boston's T

ADAM DEAN

Brandeis University

ABSTRACT: Boston's subway system, the T, is an important artery for transportation in and around the city. However, it is the oldest subway system in the United States, and, as a result, is in dire need of upgrades. This paper employs a welfare analysis to calculate the economic cost of the T's lack of reliability, while also comparing the T's reliability rate to other transit systems around the world. With a counterfactual estimate of a 95 percent reliability rate versus the pre-pandemic 88.47 percent reliability rate, the difference in welfare is found to be between $54 million and $163 million annually. Thus, long overdue improvements to the T would have a significant impact on the overall welfare of the Boston metropolitan area, and serve as a great economic benefit to all stakeholders.

## Introduction

In 2019, President Donald Trump and Democratic congressional leaders Chuck Schumer and Nancy Pelosi agreed to pursue a $2 trillion infrastructure plan[1]. The display of bipartisanship illustrated the widespread agreement from across the political spectrum that America's deteriorating infrastructure was in need of an upgrade. But like most infrastructure talk in recent years, the proposal failed to gain traction as lawmakers felt uneasy about the large costs and finding potential funding sources. State and federal legislators currently remain in stalemate about who will foot the bill for upcoming infrastructure projects across the country.

Boston's subway, the T, is the oldest subway system in North America, thus making it a prime candidate for an infrastructure upgrade[2]. In 2019, the Massachusetts Bay Transportation Authority (MBTA) released a 25-year investment plan that would modernize the regional transportation system, including the T, to best serve the needs of the region in the future. Like most transit agencies in the US, the MBTA cannot cover its costs, so shortfalls are financed by the state government. During the winter of 2015, brutal storms slammed the northeast, causing massive delays and even the complete suspension of certain transit lines for days. The poor performance brought public attention to MBTA's issues. For example, at the time, almost one-fourth of the MBTA budget went towards servicing debt, with billions of dollars in infrastructure backlog. Following the 2015 winter disaster, the Governor of Massachusetts created a special panel to get the "MBTA Back on Track." The panel successfully recommended a state-funded capital program to begin working through the billion-dollar backlog. However, with the onset of the COVID-19 pandemic, ridership has been decimated, threatening to impact the solvency of the agency[3]. As a result, funds have been shifted away from capital projects towards the immediate budget shortfall.

The pandemic will push back the day when the MBTA works through its infrastructure backlog and starts modernizing the T for the future. This will result in a continuance of the status quo costs even when ridership returns to normal. Policymakers and the MBTA have many cost-benefit analyses that explain the impacts of potential projects available to them. But being presented with just the costs and benefits of a particular project can allow policymakers to overlook the present costs delays. For every year policymakers avoid solutions to fix reliability problems, riders continue to deal with delays, which have real economic costs.

In this paper, I calculate the welfare difference between the pre-pandemic T and a hypothetical T with an improved reliability rate. This paper is a lower bound for the economic cost of subway delays because I only analyze the short-term welfare implications to consumers. In the conclusion, I will briefly explain why the long-term welfare effects are likely to be larger. I also do not explore the impact on road traffic from displaced riders nor what improved reliability would mean for employers. Additionally, I do not analyze how a more reliable public transit system would impact the environment and social equity.

In my estimate, I employ the consumer choice logit model, as created by Daniel McFadden, to estimate welfare differences. To construct my utility function, I use ridership, reliability, and line-specific data from the MBTA. With a counterfactual estimate of 95 percent reliability versus the pre-pandemic figure of 88.47 percent, I estimate the difference in the market's welfare to be between $54 million and $163 million annually. This figure is meant to be used as the beginning of a welfare discussion because I estimate welfare implications only for consumers and not for the several other parties that are impacted by delays. Furthermore, these findings are not meant to advocate for immediate infrastructure funding, but rather to remind policymakers of the cost of maintaining the status quo.

This paper will proceed as follows. Section 2 provides background on public transit, the MBTA, and the use of consumer surplus as a welfare metric. Section 3 provides a literature review which summarizes existing studies on this topic. Section 4 shows where the data comes from and the limitations of the data. Section 5 describes the empirical strategy which employs the logit model. Section 6 presents the findings of the model and the discussion. Section 7 concludes the paper.

## Background

### Benefits of Public Transit

Public transit is an important asset for urban areas. This is because it is seen as beneficial to the economy, society, and environment. According to a 2020 report from the American Public Transportation Association, robust public transit allows consumers to save on vehicle ownership and maintenance, reduces traffic congestion and needed space for parking, and allows businesses access to a larger labor market. Further, effective public transit allows a variety of people access to the job market, schooling, healthcare, and leisure activities. Because they do not need to spend on a car, public transit provides individuals across the socioeconomic spectrum access to moving across a city. This benefit can also be extended to disabled people who are unable to drive, in cases where public transit is accessible. Lastly, public transportation's efficient use of resources benefits the environment. Hundreds of countries across the world are trying to reduce their carbon dioxide emissions to limit the impacts of climate change. Public transit's ability to move hundreds of people within a single subway car is a highly efficient use of resources. Fewer cars lead to reduced pollution.

## The T

The MBTA is the fifth-largest public transportation system in the United States. It is also the oldest. Its divisions include rapid transit (the T), a bus system, the commuter rail, and a ferry. In 2019, weekday ridership of all MBTA services averaged 1.3 million people. The transit authority is a state entity that falls under the control of the Massachusetts Department of Transportation (MassDOT) and ultimately is overseen by the Governor.

The subway stretches across the Boston metropolitan area, including stations at Fenway Park, Logan International Airport, Harvard University, and Braintree, the suburb south of the city. The commuter rail feeds into North and South Stations in Boston and provides transfer points to the subway. The most ridden lines in order from greatest to least are the: Red Line, Orange Line, Green Line, and Blue Line. The system has had an 88.47 percent reliability rate since 2016. The clear laggard in reliability is the Green Line, as seen in Table 1.

The Green Line traces its origins back to the late 19th century when streetcars crisscrossed Boston in the same area the Green Line does today. The streetcars were first designated as a subway in 1897, when the Tremont Street tunnel was created and part of the streetcars' route was put underground. The Green Line vehicles are still essentially streetcars, traveling above and below the streets. In contrast, the Red, Orange, and Blue Line vehicles are exclusively heavy rail, meaning that they are heavier trains that operate on devoted right-of-way tracks. It seems that the Green Line's reliability issues are in part due to two main factors: above-ground traffic and tunnel entry and exit. Projected arrival and departure times include the knowledge that the Green Line may have to stop for traffic, yet delays remain substantial. Further discussion about reliability and its calculation is later in this paper.

At least every five years, the MBTA is required by the federal government to conduct a systemwide survey that gathers demographic, travel, and fare data. Among the pertinent findings to this paper from the 2015-17 survey were purpose and substitution data[4]. Home-based work and home-based school trips made up 71.6 percent and 5.7 percent of trips, respectively. There are also data about alternative transportation options for people whose first choice was the T. If they were unable to take the T, most would opt for another MBTA service (~45%), while the second-most popular choice was driving alone (~24%).

**Table 1: Summary Statistics of The T's Lines**

| Line | Reliability | Avg. Weekday Ridership (2019) | % of Ridership | Length of Track (mi) | End to End Time (min) | Year Founded |
|---|---|---|---|---|---|---|
| Red | 90.09% | 237,000 | 34% | 45 | ~54 | 1912 |
| Orange | 91.95% | 207,000 | 30% | 22 | ~40 | 1901 |
| Green | 77.05% | 177,000 | 26% | 46 | ~59 | 1897 |
| Blue | 94.70% | 71,000 | 10% | 12 | ~23 | 1904 |

Source: MBTA Blue Book Open Data Portal

## The State of the MBTA

In April 2015, the special assessment panel commissioned by Governor Charlie Baker released its report calling the "catastrophic winter breakdowns symptomatic of structural problems that require fundamental change in virtually all aspects of the MBTA." The executive report exposed the shortcomings of the organization's unsustainable operating budget, instability, and lack of customer focus. The report also revealed interesting insights about the MBTA's infrastructure issues.

A factor in the MBTA's rising operational costs comes from the aging infrastructure. Across seven peer transit organizations across the country studied by the panel's report, the MBTA had the oldest fleet of train cars. For example, the Red Line still uses vehicles made in the 1960s.

Repeated capital underinvestment was one of the nine major findings in the report. While additional funding from the state, and potentially the federal government, will be needed to finance large capital projects, the MBTA has a history of diverting funds allocated to capital to operating expenses. In 2015, the public transit authority spent $66.5 million earmarked for capital projects on employee salaries instead. The commission found that from 2009 to 2015, the MBTA spent only $2.3 billion of the $4.5 billion it had planned to spend on capital construction. Billions of dollars in missing capital spending over just this six-year span helps to explain why the MBTA faces such a substantial infrastructure backlog.

The ramifications of reallocating capital funds can be seen in the MBTA's capital backlog. The backlog is likely incomplete due to an inventory system that is not up to date or comprehensive. Nonetheless, in 2015, the MBTA said it had a service backlog of $6.7 billion to bring the equipment back to a State of Good Repair (SGR), a rating metric threshold for equipment quality. The report notes that due to the inadequate inventory system and underestimation of expenses, the backlog total is "unquestionably" higher. Items that directly impact reliability — such as vehicles, bridge signals, stations, track, and power—make up 85 percent of the backlog.

Another finding of the report was that when the MBTA does spend money on capital projects, it faces an elongated timeline. A recent morale boost for riders has been the occasional sighting of new train cars on the Red and Orange Lines. But the process behind the procurement of these vehicles illustrates the inefficiencies experienced by the MBTA. In 1994, 74 Red Line cars were due for retirement, and in 2004, 120 Orange Line cars were due for retirement. Despite having this information, the MBTA did not submit its first draft for buying new cars until 2008. It took another six years for the board to approve a builder for the cars. In 2020, the first non-test vehicle was supposed to be deployed, and in 2022, the MBTA should have the last car in its order delivered. If there are no further delays, it would mark 28 years from when the Red Line cars were supposed to be retired to when the old vehicles will be completely replaced. This lack of initiative to keep up with infrastructure maintenance is part of the reason why the T is in disarray.

The report recommended that all board members of the MBTA resign. Six of the seven ultimately did. But the resignation of past leadership and the implementation of new oversight have not been an instant fix for the organization. In 2018, three years after the report's publication, the MBTA re-calculated a $10.1 billion figure for the maintenance and modernization of capital, a $3.4 billion increase from the initial estimate.

## Consumer Surplus

An increase in reliability of the T would not directly put money in the pockets of the residents of the Boston metro region, but it would benefit them nonetheless. Consumer surplus is an economic principle that attempts to quantify this benefit, and this paper relies heavily on evaluating consumer surplus. The definition of consumer surplus is the difference between what consumers would be willing to pay for a product and what they actually pay. For example, if the price of a product decreased, with no changes in other factors, the consumer surplus would rise. Alternatively, if the quality of a product improved and the price stayed the same, the consumer would likely be willing to pay more for the product than before; this would also increase consumer surplus. The latter example is what this paper sets out to estimate.

## Literature Review

The literature of consumer discrete choice begins with the work of Daniel McFadden, the 2000 co-laureate of the Nobel Prize in Economics. *The Measurement of Urban Travel Demand* (1974) is where McFadden created the logit model to empirically analyze consumer behavior in BART (Bay Area Rapid Transit). I use McFadden's model, as described by Kenneth Train's textbook, to make my consumer surplus estimates for the T.

Outside of McFadden's work, much of the economic literature falls short of actually estimating costs to consumers from delays or subpar access. Instead, most economic studies focus on different components of a large-scale welfare calculation.

Van Oort (2016) is one of the few papers that quantifies the costs to passengers. By studying a new tram line in Utrecht, Netherlands, the paper found benefits resulting from reliability increases accounted for two-thirds of total benefits. The estimate was done through simulating a counterfactual estimation but did not use a logit model. The analysis also claimed its findings were important to convince policymakers to support the project.

There are various papers that study the relationship between general welfare, public transportation, and transit efficiency. Benezech and Coulombel (2013) calculated the marginal effect of reliability on expected travel costs due to changes in wait times, schedule delays, and congestion. Baum-Snow and Khan (2005) evaluated the extent to which urban rail network expansions in US cities have spurred new ridership and estimated partial welfare gains that came from traffic reduction and savings in car ownership. Lobo and Couto (2015) analyze the operational performance of European metro systems and determine what factors contribute to good transit performance. Another study[5] calculated welfare losses not from reliability changes, but from fare hikes on the Madrid Metro. Fare hikes resulted in a welfare loss of 3.66% in income for low- and medium-income households, while the richest suffered a 1.5% reduction in welfare.

The New York City Comptroller (2017) released a rudimentary calculation that estimated large-scale economic costs from delays with extensive data. By looking at ridership, magnitude of delays, and assuming a $34 hourly average wage, the report put the city's losses in 2016 in the range of $170 million to $389 million. The costs were justified because of lost productivity for businesses and lost wages for workers. Economic loss estimates were also given for each line of the Metropolitan Transit Authority (MTA).

There is substantial literature for estimating welfare gains from social media through the use of the logit model and willingness-to-accept experiments: Alcott et al. (2018), Brynjolfsson et al. (2019), Corrigan et al. (2018), Mosquera et al. (2018), and Sunstein (2019). One study found the welfare difference in the US between not having Facebook in 2003 to having it in 2017 was $231 billion[6].

There have been several transportation-delay welfare studies in the airline industry as well. One paper found that flight delays on airlines negatively affect consumer demand and increase average fares[7]. The study found that a 10 percent reduction in delays would yield a consumer surplus gain of $1.48 per passenger; a 20 percent reduction in delays would yield $3.06 per passenger. The paper was not completely comprehensive, however, as it was a partial equilibrium analysis. Welfare could change as passengers shift from other transportations sectors, like cars. Yimga (2017) found that on average, the welfare costs of airline delays to consumers at their final destination to be $1.38, $1.07 and $0.91 per minute in short-, mid- and long-haul markets, respectively. Several studies also look at the relationship between airline delays and consumer demand, but do not estimate welfare changes[8].

## Data

### MBTA Data

Nearly all of the data I use in this paper come from the MBTA Open Data Portal, created in 2015 in response to the winter disaster. The first dataset I analyze concerns reliability, which includes specific figures on reliability, date, line, and a peak indicator. In a partnership with MIT, the MBTA developed an Origin-Destination-Transfer (ODX) model that estimates reliability by analyzing Charlie Card (the T's subway pass) data. Predicating the origin of a rider is simple as it is where a Charlie Card is first swiped. But since the T does not require swipes to exit a station, the destination is predicted by identifying the next location where the same Charlie Card is swiped. Transfers are predicted using a similar methodology with spatial and logical checks. From the sample of Charlie Cards that it analyzes, the model extrapolates the reliability metric for riders across the system.

The reliability metric is calculated as the number of people unaffected by delays divided by the number of total riders. For example, if 900 people out of 1,000 were unaffected by delays, then the reliability of that line would be 90 percent for the given time period.

There is criticism about who qualifies as being unaffected by delays[9]. Subway trains arrive on scheduled intervals, or headways. If a subway train is supposed to arrive every five minutes, it is said to have a five-minute headway. A common understanding of subway delays would classify any train that arrives after its scheduled time as being delayed and the passengers boarding that train as having been affected. But the MBTA measures the unaffected portion of reliability as the number of riders that board a train within the scheduled headway time starting when they enter the station. This interpretation means that a person could board a delayed train and be counted as being unaffected by delays.

Figure 1 illustrates this counting phenomenon. Suppose a train had a five-minute headway, but it came to a station two minutes late, making the headway seven minutes. In a typical interpretation of delays, no matter when a rider arrived at the station, anyone boarding the late train would be deemed affected. However, the MBTA would not include every rider as having faced a delay and would opt instead to analyze when riders arrived at the station to determine if they were impacted. Suppose for the same five-minute headway train that arrived two minutes late, there were two riders who entered the station planning to ride the train. Rider 1 arrived at the one-minute mark of the five-minute headway interval, and Rider 2 arrived at the three-minute point. Since the train arrived two minutes late, Riders 1 and 2 will have waited six and four minutes for the train, respectively. Rider 1 will be counted as having been affected by delays, but not Rider 2, even though they both ultimately board on a train that was late to the station. Yes, Rider 2 does wait less than the scheduled headway time of five minutes. But had the train been on time, Rider 2 would have saved two minutes in wait time. The criticism of the metric is that it underestimates delays because it does not count everyone who would have saved time if a given train arrived as scheduled.

A second issue from the MBTA definition of reliability is that the model assumes that everyone who waits for a train during rush hour is able to get on one. Due to overcrowding, this is not always true.

However, despite the shortcomings of the metric, the reliability metric still gives valuable insight about the performance of the T.

**Figure 1: MBTA Reliability Definition**


**Source: MBTA Data Blog**

Two other data sets from the MBTA that I employ are the gated entries and monthly ridership data. The ridership data is at the monthly level because the MBTA gathers data through multiple sources (passenger counters, fare collection, counts, etc.) and adjusts the tally at the end of the month to account for fare evaders and system errors. Because my reliability data is at the daily level, this created a data mismatch. To solve this problem, I used gated entry data, which is also at the daily level. The proportionality of the gated entry data did not align with the share of ridership for each line at the monthly level because the green line does not collect gated entries at overground stations. To address this issue, I generated a new daily ridership variable that adhered to the proportionality of the ridership data at the monthly level and included the day-to-day fluctuations in magnitude from the gated entry data. This daily ridership variable is not precisely what ridership was on that day, but it is close enough that no integrity issues exist.

## Other Subways' Reliabilities

As previously mentioned, the reliability of the T is 88.47 percent, with the Red, Orange, Green, and Blue Lines having reliabilities of 90.09, 91.95, 77.05, and 94.70 percent, respectively. To estimate a counterfactual in reliability, I wanted to place its reliability rates in context with other systems around the world. Two difficulties emerged: many subway systems do not easily share performance data and reliability definitions can vary by system. Nonetheless, I was able to procure figures for a handful of subway systems that had similar definitions of reliability. In the results section, I will discuss why I choose to focus on a counterfactual reliability of 95 percent.

**Table 2: Reliability Rates of International Subway Systems**

| *Urban Area* | *Average Weekday Ridership* | *Reliability Rate (%)* |
|---|---|---|
| Boston | 692,000 | 88.6 |
| Berlin | 1,500,000 | 96.3 |
| Glasgow | 52,000 | 95.0 |
| Hong Kong | 4,962,000 | 99.9 |
| New York City | 5,700,000 | 84.0 |
| Paris | 4,160,000 | 98.2* |
| São Paulo | 5,300,000 | 99.3* |
| Seoul | 13,000,000 | 99.9 |

| | | |
|---|---|---|
| Singapore | 7,035,000 | 98.5 |

**Additional Data and Notes**

Later in this paper, I discuss instrumenting for reliability because of endogeneity issues in my regression. One of the instruments I use for this is weather data (precipitation levels and average temperatures) from the National Weather Service.

The period for all data starts in 2016 and continues to the present. However, due to the pandemic, the last month in the data is February 2020. The analysis excludes weekend data to allow for certain regression controls and to simplify the analysis.

To calculate utility, I needed to determine the T's market share. To do this, I had to produce an estimate of the market size for potential T riders. The US Census has done analyses that estimate the daytime populations of Suffolk County (the county that Boston resides) and Somerville (a suburb of Boston). The 2010 analyses found that Suffolk Country grows around 33 percent during the day and the neighboring suburbs decrease by a similar factor. I collected population data for Suffolk County and Boston's surrounding cities that are not in Suffolk County but still have easy access to the T (Cambridge, Somerville, Brookline, Newton, Medford, Malden, Everett). Lastly, I transformed these populations by the proportions found by the US Census analyses and added the figures together to get a market size of 1,440,000 people.

## Variables Summary Statistics

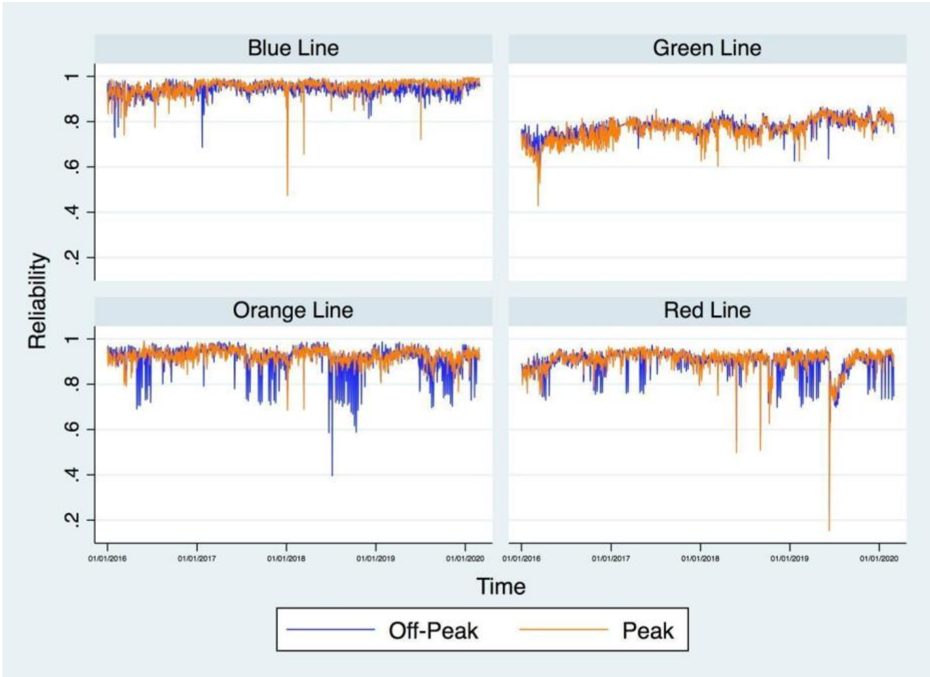
### Table 3: Summary Statistics of Variables

| *Variable* | *Obs* | *Mean* | *Std. Dev.* | *Min* | *Max* |
|---|---|---|---|---|---|
| Price | 10,254 | 2.256627 | 0.0794851 | 2.1 | 2.4 |
| Precipitation | 10,254 | 0.1238268 | 0.2967151 | 0 | 2.68 |
| TempAvg | 10,254 | 52.3741 | 17.26458 | 0 | 90 |
| *Reliability* | | | | | |
| Blue Line | 2,602 | 0.9470299 | 0.0324076 | 0.4745002 | 0.9962027 |
| Green Line | 2,563 | 0.7705131 | 0.0444417 | 0.428605 | 0.8695235 |
| Orange Line | 2,561 | 0.9195957 | 0.0470082 | 0.3962716 | 0.9908824 |
| Red Line | 2,528 | 0.9009723 | 0.0543583 | 0.1541655 | 0.9720004 |
| Aggregate | 10,254 | 0.8847025 | 0.0815763 | 0.1541655 | 0.9962027 |
| *Ridership* | | | | | |
| Blue Line | 2,602 | 67,909.36 | 12,892.55 | 6,515.416 | 147,666.6 |
| Green Line | 2,563 | 188,382.5 | 43,356.02 | 16,767.78 | 449,132 |
| Orange Line | 2,561 | 199,585.4 | 39,657.03 | 10,112.37 | 293,992.2 |
| Red Line | 2,528 | 239,053.1 | 48,087.32 | 13,141.93 | 323,482.5 |
| Aggregate | 10,254 | 173,102.1 | 74,757.27 | 6,515.416 | 449,132 |

### Table 4: Reliability Instruments Correlation Matrix

| *Variable* | *Reliability* | *Precipitation* | *Temp Avg* | *Previous Day* | *Inverse Reliability* |
|---|---|---|---|---|---|
| Reliability | 1 | | | | |
| Precipitation | -0.0188 | 1 | | | |
| Temp Avg | -0.006 | -0.001 | 1 | | |
| Previous Day | 0.0873 | 0.0439 | 0.0004 | 1 | |
| Inverse Reliability | 0.951 | -0.0188 | -0.006 | 0.0873 | 1 |

**Figure 2: Reliability Rates by Line**




## Empirical Strategy

I employ Daniel McFadden's logit model, as advocated by Kenneth Train in *Discrete Choice Models in Simulation,* to estimate the consumer surplus gain from improved reliability on the T. The logit model is useful in situations where consumers make a discrete choice (i.e. only choose one item out of several alternatives). For this context, I am looking at the probability of consumers choosing to ride the T versus all other alternatives, such as driving a car or even another MBTA service.

In the model, a person, *n*, chooses among alternatives, *j*. The conditional indirect mean utility, excluding the i.i.d. (identically independently distributed) error term, obtained from alternative choice *j* is represented by *V* in equation 1. *V* includes factors that I am able to observe such as the impact of prices, reliability, and weather. The error term, $\varepsilon$, contains unobservable factors that affect utility and it is treated as random.

$$U_{nj_t} = V_{j_t} + \varepsilon_{nj_t} \quad (1)$$

Train describes the logit choice possibilities, as derived by Daniel McFadden, in equation 2. The equation represents the possibility that decision maker *n* chooses alternative *i* over all other choices *j*. The equation states that the probability the decision maker chooses choice *i* is the probability that the utility yielded by *i* is greater than that of *j*. Through further derivation and substitution, Train rewrites the probability of decision maker *n* choosing alternative *i* over all other choices *j* in equation 3. Equation 3 includes the outside good, whose mean utility is normalized to 0.

$$P_{ni} = Prob(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj}) \qquad (2)$$

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}} = \frac{e^{V_{it}}}{1+e^{V_{it}}} \qquad (3)$$

Consumer surplus is utility expressed in dollars. In this logit model, this is the dollar value of the decision that maximizes utility for consumers. Equation 4 shows that the expected consumer surplus for a decision maker, *n*, is equal to the expected value of the maximizing utility choice divided by$\alpha$, the marginal utility of income (the change in a person's utility resulting from a one-unit change of their income).

$$E(CS_n = \frac{1}{\alpha} E[max_j(V_{nj} + \varepsilon_{nj})] = \frac{log\left(1+exp(V_{jt})\right)}{\alpha} \qquad (4)$$

From this logit framework, it is quite simple to calculate the consumer surplus using the log-sum term, according to Train. The log-sum term is the log of the denominator of equation 3. Thus, the logit model allows me to estimate the change in consumer surplus between the status quo of the T and a more reliable counterfactual by calculating the two utilities and their log-sums. The difference between the two log-sums is the consumer surplus currently lost by consumers as the T remains behind on maintenance and technological upgrades. This consumer surplus is just for one person, an aggregate number is calculated by multiplying individual consumer surplus by the market size.

From the data provided by the MBTA open data portal, I am able to estimate a regression that predicts the utility that is needed for the logit model. The model is as follows:

$$mean\ utility_t * = \beta_{0_t} + \beta_1 Price_t + \beta_2 Reliability_t + \beta_3 Peak_t + \gamma(Controls)_t + \xi_t$$

where *utility* is the natural log of the market share minus the natural log of the outside share (market share plus outside share is equal to 1)[10], *price* is the price of a subway ticket without a pass, *reliability* is the value produced by the MBTA's ODX model, *peak* is an indicator equal to 1 if it is peak time, and $\gamma$ represents the various controls used. In every regression, I use time-fixed effects at the day, week, and year level and fixed effects by line. There is an endogeneity issue with this model because ridership is thought to have an effect on reliability. For example, larger crowds can create door-closing delays. To account for this problem, I instrument for reliability using several different variables. These include precipitation levels, average temperature, the reliability of the previous day, and the reliability of the inverse peak period.

## Results and Discussion

In this section, I review the results of my logit model regression and the consumer surplus calculation. Table 1 presents the results of the three regressions, which are differentiated based on what instruments are included for reliability.

Initially, I included an interaction variable of reliability and peak to account for the difference in reliability during peak and nonpeak hours. But after further investigation, the statistical variation between reliability during on- and off-peak hours was near zero. This appeared odd, as one would think that increased ridership during peak hours would put extra stress on a subway system. However, peak hours only represent a small portion of operating hours for the T, while the off-peak period makes up the majority of the day. It is likely that the high incidence of delays during the peak period is matched by the length of the off-peak period in terms of total delays. This understanding seems to be further supported by the lack of statistical significance for the peak variable in all three regressions.

Because none of the four instruments employed are strong, regressions 1 and 2 differ on the magnitude of the variable effects. Table 5 shows reliability regressed on the instruments. Despite all of the instruments having a theoretical case for being strong, inverse reliability is the only statistically significant, nonzero instrument. But since reliability does not have significant statistical variation between peak and non-peak periods, inverse reliability is highly correlated with reliability. This makes inverse reliability a poor instrument as well.

Although regressions 1 and 2 differ on the magnitude of their coefficients, they both have similar effects for the variables. As expected, the results show an increase in reliability would have a positive effect on utility. The reliability coefficients are statistically significant at the 95 percent confidence level.

From 2016 to 2019, the T had two fare increases. The three different price levels during the period of the study allow the model to take into account the effect of price on the market share interaction. Typical endogeneity issues are unlikely to exist because fare hikes were planned and announced far in advance. For both regressions, price has a negative effect on utility and those findings are statistically significant. The peak variable has a slight negative effect in the regressions, but it is not statistically significant. This could be because of the previously discussed disproportionate lengths of the peak and off-peak periods.

It is unclear which regression is closer to the true values of the utility function. regression 2 has similar effects to regression 3, which does not include any instrumentals. But a strong instrument might yield effects closer to regression 1. Thus, I use both regressions to create my consumer surplus findings.

**Table 5: Logit Model Utility Regression**

| | Dependent Variable: Utility | | |
|---|---|---|---|
| **Variable** | **(1)** | **(2)** | **(3)** |
| | 12.16616 | 0.4857369 | 0. 4999682 |
| Reliability | (6.6) | (5.17) | (6.56) |
| | -1.52854 | -0.1807926 | -0. 1778105 |
| Price | (-5.84) | (-2.27) | (-2.24) |
| | -0.1035653 | -0.0624026 | -0.062207 |
| Peak | (-.19) | (-.75) | (-.57) |
| Day of Week Fixed Effects | Yes | Yes | Yes |
| Month Fixed Effects | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes |
| | -9.415116 | -2.299249 | 0 |
| Constant | (-7.65) | (12.2) | |
| Observations | 10,524 | 10,524 | 10,524 |
| $R^2$ | 0.0531 | 0.0041 | .0039 |
| ***Instruments*** | | | |
| Temperature | Yes | No | No |
| Precipitation | Yes | No | No |
| previous day | Yes | No | No |
| inverse reliability | No | Yes | No |

Note: Line fixed effects are controlled for in the panel selection of the regression.

**Table 6: Instrument Test**
**Dependent Variable: Reliability**

| Variable | Coefficient | t | F-Statistic |
|---|---|---|---|
| Precipitation | -0.0052141 | -1.92 | 3.69 |
| Temp Average | -0.0000369 | -0.79 | 0.63 |
| Previous Day | 0.000013 | 7.09 | 50.25 |
| Inverse Reliability | 0.9510001 | 283.74 | 80507.85 |

**Table 7: Consumer Surplus Calculations for Different Potential Reliabilities**

| Regression | Counterfactual Reliability | Consumer Surplus (annually) |
|---|---|---|

| | | |
|---|---|---|
| 1 | 95% | $163,439,990 |
| 2 | 95% | $54,177,113 |
| 1 | 98% | $331,271,700 |
| 2 | 98% | $80,097,196 |

Table 2 shows the consumer surplus calculations for four variations: two variations for regressions 1 and 2, and two variations for different counterfactual estimates. The consumer surplus calculation is described earlier in *Empirical Strategy*. The figures shown in the table are the yearly difference in consumer surplus between the status quo and a counterfactual reality. This consumer surplus gain is not just for current riders, but also for non-riders who theoretically switch over to the T from enhanced utility.

I focus on a counterfactual reliability of 95 percent for a few reasons. The T's current reliability is at 88.47 percent, so an increase to 95 percent would represent a significant 6.6 percentage point increase. Additionally, this does not seem completely unreasonable given the MBTA's documented level of underinvestment and the growing infrastructure backlog. The Blue Line has a reliability of 94.7 percent, for example. It is the newest line, so it has the smallest infrastructure backlog and makes the 95 percent target more realistic. Comparatively, the Red Line still uses train cars that were supposed to be retired over two decades ago, and has other outdated equipment in other parts of the system. If the T were able to work through its backlog and start modernizing its equipment, that would yield increased reliability as well.

Further, in Table 2, all of the international systems listed have a reliability of over 95 percent, with most of them close to 99 percent. It is unfair to make a direct comparison to international systems given the US's historical reliance on cars and a large lack of funding given to public transit compared to other countries. Nonetheless, the international systems demonstrate that near-perfect reliability is not impossible, especially given that most of those systems carry more passengers than the T.

The subway systems of Asia are much newer than the T and are considered the best in the world. Thus, a comparison to the T cannot be made. However, the European systems trace their origins back to the late 19th or early 20th century, the same period as the establishment of the T. Given that these systems have reliability rates above 95 percent and that the T's Blue Line itself is very close to that figure, 95 percent does represent a reasonable figure. Even still, I include a 98 percent counterfactual reliability to serve as a thought experiment, but it is a threshold that is unrealistic for the T given its current circumstances.

Using regression 1, I find that lost consumer surplus due to substandard reliability costs Boston metro residents over $160 million annually. If counterfactual reliability were upped to 98 percent, that figure rises to over $330 million. Using regression 2, I find a $54 million loss in welfare at a 95 percent counterfactual, and an $80 million loss for a 98 percent reliability rate. Using the 95 percent reliability level, from 2016 to 2019, the welfare loss to residents is between $215 million to $640 million.

## Conclusion

This paper uses a logit model to estimate the welfare loss of residents of the Boston metropolitan area due to the unreliability of Boston's subway, the T. By finding the difference in consumer surplus between the status quo and a counterfactual reality of improved reliability, I estimate the annual welfare losses to be in the range of $54 million to $163 million.

This estimate is a lower bound for a few reasons. The consumer choice logit model predicts short-term welfare changes as reliability shifts day to day. In the long run, consumer welfare losses to poor reliability are likely higher. If a rider is dissatisfied with reliability during a given week, they are unlikely to go out and buy a car the next. But if reliability issues persist over months, they have more time to switch to other transportation alternatives. If the T is someone's first choice transportation method, then transitioning to another method, such as buying a car or bike, reduces their welfare.

This estimate also does not consider the effects on third parties. It does not evaluate what the benefits of reduced road congestion would be. The commuter rail is the primary public transportation option for people that live in suburbs further from the city; they drive cars at higher rates than people that live within the immediate urban area of Boston. But, the commuter rail feeds into two stations in central Boston, where many commuters transfer to the T to get to their final destination. Thus, the reliability of the T still has effects on commuter rail riders. Lastly, there are accessibility, social equity, and environmental benefits from improved reliability that are not within the scope of this paper.

These consumer surplus estimates should not be used alone to advocate for upgrading the T because cost-benefit analyses are beneficial for understanding a broad sense of the effects of a specific project. However, my analysis should be considered in the cost side of the cost-benefit analyses read by policymakers. By having a substandard subway system in 2019, the Boston metro area lost tens of millions of dollars, if not hundreds of millions. Unless policymakers choose to let the system run without upgrades and essentially dissolve, large infrastructure upgrades will eventually happen. Every year until those infrastructure upgrades begin to improve service, the public will absorb the costs of unreliability. The pandemic has reduced ridership and has even made cutting subway service an option for meeting budget shortfalls. The losses of 2019 will not be the same as 2020 because of the pandemic. Nonetheless, the underlying principle of this paper will have relevance when the pandemic ends and ridership begins to return to more normal levels.

This paper may even be more relevant at the end of the pandemic when the economy begins to recover. Infrastructure spending may again be thrust into political discussions as a way of delivering stimulus to the economy. Ultimately, however, it will be up to policymakers to work through political obstacles that have held up large-scale infrastructure spending in recent years. If they do not, then the costs of the status quo will persist.

## Notes

1. New York Times 2019
2. Encyclopedia Britannica 2017
3. WBUR 2020
4. MBTA 2018

5. Burguillo et al. 2017
6. Brynjolfsson et al. 2019
7. Britto et al. 2011
8. Abrahams, 1983; Anderson & Kraus, 1981; Ippolito, 1981
9. Casale and Odayappan 2017
10. Berry 1994

## References

Abrahams, M. (2002, August 29). A service quality model of air travel demand: An empirical study. Retrieved from http://www.sciencedirect.com/science/article/abs/pii/0191260783900079

Anderson, J., & Kraus, M. (1981). Quality of Service and the Demand for Air Travel. *The Review of Economics and Statistics, 63*(4), 533-540. doi:10.2307/1935849

Baum-Snow, N., & Kahn, M. (2000, May 25). The effects of new public projects to expand urban rail transit. Retrieved from http://www.sciencedirect.com/science/article/pii/S0047272799000857

Benezech, V., & Coulombel, N. (2013, October 17). The value of service reliability. Retrieved from http://www.sciencedirect.com/science/article/pii/S0191261513001690

Berry, S. (1994). Estimating Discrete-Choice Models of Product Differentiation. *The RAND Journal of Economics, 25*(2), 242-262. Retrieved March 29, 2021, from http://www.jstor.org/stable/2555829

Britto, R., Dresner, M., & Voltes, A. (2011, November 20). The impact of flight delays on passenger demand and societal welfare. Retrieved from https://www.sciencedirect.com/science/article/abs/pii/S136655451100134

Burguillo, M., Romero-Jordán, D., & Sanz-Sanz, J. (2017, February 24). The new public transport pricing in MADRID metropolitan area: A welfare analysis. Retrieved from http://www.sciencedirect.com/science/article/pii/S0739885917300264

Casale, M., & Odayappan, D. (2017). How Reliable Is The T: Needed Improvements to the MBTA’s Method for Measuring Subway Reliability on the Back on Track Performance Dashboard. Retrieved from masspirg.org/sites/pirg/files/reports/MASSPIRG%20Ed%20Fund%20-%20How%20Reliable%20is%20the%20T%20(Report,%20Oct%202017).pdf

Cochrane, E., Karni, A., & Rappeport, A. (2019, April 30). Trump and Democrats Agree to Pursue $2 Trillion Infrastructure Plan. Retrieved from http://www.nytimes.com/2019/04/30/us/politics/trump-infrastructure-plan.html

Economic Impact of Public Transportation Investment: 2020 Update. (2020). Retrieved from https://www.apta.com/wp-content/uploads/APTA-econ-impact-transit-investment-2020-ES.pdf

The History of the T. (n.d.). Retrieved from http://www.mbta.com/history

Integrated Report: 2019. (2019). Retrieved from http://www.metro.sp.gov.br/en/pdf/integrated-report-2019.pdf

Lobo, A., & Couto, A. (2015, June 5). Technical efficiency of european metro systems: The effects of operational management and socioeconomic environment. Retrieved from https://link.springer.com/article/10.1007/s11067-015-9295-5
Ippolito, R. (1981). Estimating Airline Demand with Quality of Service Variables. *Journal of Transport Economics and Policy, 15*(1), 7-15. Retrieved from http://www.jstor.org/stable/20052594

MBTA Back on Track: An Action Plan to Transform the MBTA. (2015). Retrieved from http://www.mbta.com/mbta-back-on-track

MBTA Faces 'Existential' Budget Crisis Amid Pandemic, According To New Report. (2020, June 15). Retrieved from www.wbur.org/bostonomix/2020/06/15/mbta-budget-deficit-coronavirus

MBTA Systemwide Passenger Survey, 2015-17. (2018). Retrieved from www.ctps.org/apps/mbtasurvey2018/accessible.html

McFadden, D. (2002, March 26). The measurement of urban travel demand. Retrieved from http://www.sciencedirect.com/science/article/pii/0047272774900036

NYC Subway Dashboard. (2020). Retrieved from http://dashboard.mta.info/

Oort, N. (2016, March 01). Incorporating enhanced service reliability of public transport in cost-benefit analyses. Retrieved from https://link.springer.com/article/10.1007/s12469-016-0121-3

Quality Campaign S-Bahn PLUS. (2020). Retrieved from https://sbahn.berlin/en/about-us/quality-drive

RATP Group 2018 Results: a Year of Profitable Growth and Investments to the Benefit of Passengers. (2018). Retrieved from www.ratpdev.com/sites/default/files/annexes/communiques/RATP%20Group%202018%20results.pdf

The results are in. (2018, July 3). Retrieved from www.spt.co.uk/2018/04/seal-of-approval-from-subway-passengers

Seoul Metropolitan Rapid Transit Annual Report. (2012). Retrieved from www.metropolis.org/sites/default/files/seoul_metropolitan_rapid_transit_english.pdf

Stringer, S. The Economic Cost of Subway Delays. (2017). Retrieved from www.comptroller.nyc.gov/reports/the-economic-cost-of-subway-delays

Tikkanen, A. (Ed.). (2017, August 31). Subway. Retrieved from http://www.britannica.com/technology/subway

Train, K. (2009). Discrete Choice Methods with Simulation (2nd ed.). Cambridge: Cambridge University Press. doi:10.1017/CBO9780511805271

Yimga, J. Airline on-Time Performance and Its Effects on Consumer Choice Behavior. (2017, June 29). Retrieved from www.sciencedirect.com/science/article/pii/S0739885917300148