Analyzing route choice preferences in response to planned public transport disruptions

A study of the Washington DC metro smart card data

by

Karin van den Berg

Chair:	Prof. dr. O. Cats
Daily Supervisor:	Dr. J. Gao
Second Supervisor:	Dr. S. van Cranenburgh
External committee member:	Dr. Y. Xin
Faculty:	Faculty of Transport Infrastructure & Logistics, TU Delft

Student number: 4938933
Date: 17-04-2025
To be publicly defended on: 01-05-2025

TU Delft

Preface

This thesis marks the end of my two-and-a-half-year journey in the Transport, Infrastructure, and Logistics program at TU Delft. It's been a challenging but rewarding adventure, and I'm incredibly grateful to everyone who supported me along the way.

First, I would like to sincerely thank my daily supervisor, Jie. Your guidance, encouragement, and thoughtful advice gave me confidence, especially during moments of doubt. I've grown so much throughout this process thanks to your support. I also want to thank Sander for his invaluable help in tackling the modeling challenges I faced. After each meeting with you I always had many new suggestions to move forward with, which helped tremendously. Oded, your feedback during the milestone meetings and your suggestions had a big impact on improving my work. I also really appreciated how all of my supervisors were always quick to respond and approachable for questions, which was something that made this process much smoother and more enjoyable.

One of my personal goals for this thesis was to improve my coding skills, and I'm proud to say I've learned a lot in that area, which is something I'll carry with me into future projects.

To my friends, especially those I studied with at the Fellowship during exam periods—thank you for the motivation, support, and good company. I'm also very grateful to my parents for always providing a place of calm and comfort during weekends. This meant I could recharge for each new week ahead. And last but certainly not least, to my boyfriend Matthijs—thank you for always being there, for your constant support, patience, and for believing in me throughout this entire journey even though sometimes I was not so sure about things myself.

I hope you enjoy reading this thesis and find it both interesting and insightful.

Karin van den Berg
Delft, April 2025

Executive Summary

This thesis investigates how public transport passengers' route choice preferences change in response to a planned disruption, using Automated Fare Collection (AFC) data from the Washington DC metro system. The study aims to understand not only whether passengers change their routes, but also which factors drive these changes and whether the adjustments persist over time. In doing so, it also assesses the strengths and limitations of using revealed preference data alone to model route choice behavior in the context of service disruptions.

A detailed case study was conducted around a selected planned disruption that affected multiple route alternatives between a key origin-destination (OD) pair. By dividing the analysis into a pre-disruption period and three post-disruption periods, the study captures both short-term and long-term behavioral adaptations. Descriptive analyses show that while travel times and wait times remained relatively stable across periods, passengers shifted from direct routes to those with more transfers but shorter in-vehicle times. This indicates a change in preferences, where minimizing travel time appeared to take precedence over avoiding transfers.

However, discrete choice models such as the Multinomial Logit (MNL) and Mixed Logit (ML) yielded counterintuitive results, with some positive coefficients for time-related attributes. These inconsistencies were largely attributed to the strong correlation between in-vehicle time and the number of transfers, as well as the presence of dominated alternatives in the choice set. The ML model allowed for random taste variation and showed a statistically better fit than the MNL model, but practical improvements were limited and the results remained difficult to interpret.

Although the discrete choice models encountered challenges in isolating the exact factors driving these changes—mainly due to the strong correlation between in-vehicle time and the number of transfers—the broader behavioral patterns were clear. Even though some coefficients had unexpected signs, the shift from positive to negative coefficients for in-vehicle time suggests a change in preferences. Whether this change is due to increased sensitivity to travel time, a decreased reluctance to transfer, or a combination of both could not be fully determined.

The study also highlights that while AFC data is highly valuable for detecting real-world changes in travel behavior, it has important limitations when used alone. The absence of perceptual factors such as comfort, reliability, and route familiarity makes it difficult to fully explain why some passengers make seemingly suboptimal choices. Additionally, the high degree of overlap between available routes in the metro network limited the ability to observe distinct trade-offs between route alternatives.

From a policy perspective, these findings suggest that disruptions should not only be seen as operational challenges, but also as potential moments where lasting changes in travel patterns can occur. If passengers become more sensitive to travel time after experiencing a disruption, it becomes important for agencies to ensure that fast and convenient alternatives remain available afterward, to support passenger satisfaction and avoid the risk of losing travelers to other modes.

Finally, while this study focused on a single disruption in a single metro network, it provides a foundation for broader research. Future studies should aim to apply similar analyses to different networks, with longer and more diverse route options, and explore whether disruption characteristics such as duration influence the persistence of behavioral changes. Combining AFC data with additional perception-based information could also offer deeper insights into the factors driving passengers' adaptation processes.

Contents

Nomenclature

In this section you can find abbreviations that are often used in this thesis

Abbreviations

Abbreviation	Definition
WMATA	Washington Metropolitan Area Transit Authority
SP	Stated Preference
RP	Revealed Preference
AFC	Automatic Fare Collection
AVL	Automatic Vehicle Location
MNL	Multinomial Logit
ML	Mixed Logit
PSL	Path Size Logit
AIC	Akaike Information Criterion
BIC	Bayesian Information Criterion

1 Introduction

As cities around the world struggle with increasing traffic congestion and rising emissions, public transport plays a critical role in ensuring sustainable urban mobility (Lako and Gjevori, 2023). Efficient and reliable transit systems reduce dependence on private vehicles, which in turn helps to reduce overcrowding and lower carbon emissions. However, maintaining public transport as an attractive alternative to driving requires ensuring high levels of passenger satisfaction with service quality (Tuan et al., 2022).

A key factor influencing passenger satisfaction is reliability, as unpredictable service can discourage people from using public transport regularly (Soza-Parra et al., 2019). One of the main threats to reliability are disruptions, which can lead to increased travel times, missed connections, and overcrowding. These disruptions not only affect overall service quality but also influence passengers' transit route choices, forcing them to adapt their travel behavior in response to service changes due to the disruption.

Research has already been conducted on passengers' public transport route choices during disruptions, often using discrete choice models (Dixit et al., 2023). These models help reveal transit riders' preferences by assessing the relative importance of various travel attributes, offering policymakers valuable insights into how passengers make route choices in response to disruptions.

Many earlier studies on public transport route choice analysis were based on stated preference (SP) data (Shires et al., 2019, Li et al., 2020, Zhu et al., 2017), where passengers respond to hypothetical situations. However, SP surveys may not fully capture actual passenger behavior, potentially leading to biased estimates (Mo et al., 2022a). Later studies therefore turned to revealed preference (RP) data from travel surveys to estimate actual route preferences (Marra and Corman, 2023). Although RP surveys offer more accurate data, they are often costly and limited in sample size (Ingvardson et al., 2024). With the rise of large-scale disaggregate datasets, such as automatic fare collection (AFC) data, RP data can now be collected more efficiently, enabling a more accurate and large scale analysis of route choice behavior (Berggren et al., 2022).

AFC data has already been used in analyzing public transport users' choices during disruptions. For instance, Marra and Corman (Marra and Corman, 2023) used AFC data to examine how network disturbances affect passengers' route choices, while Eltved et al. (Eltved et al., 2021) analyzed the impacts of long-term service disruptions on travel behavior using smart card data. Whilst AFC data has been used to analyze passenger behavior during and after disruptions, it has rarely been applied to study route choice preferences specifically in the context of disruptions. Most existing studies rely on a combination of RP and SP surveys (Rahimi et al., 2020, Li et al., 2020), making it unclear whether AFC data alone can sufficiently capture route choice preferences. One exception is the study by Mo et al. (Mo et al., 2022b), which uses AFC data to examine mode choice preferences during disruptions rather than route choices. This means there is a gap in studies using only AFC data to look at route choice preferences in response to a disruption.

Another gap in the literature is that most research focuses on route choice during the disruption itself, rather than what happens after the disruption ends. While some studies, such as Eltved et al. (Eltved et al., 2021), have examined route choice before and after a disruption, their dataset only covered a period of three weeks before and after, making it difficult to determine whether changes in preferences persist over time.

This study aims to fill these gaps by extending the analysis period from weeks to several months before and after a disruption, allowing for a better understanding of whether changes in route choice preferences persist over time. Additionally, this research relies solely on AFC data to analyze changes in route choice preferences, a methodology that has not yet been applied to the Washington DC metro system. These contributions provide valuable scientific insights into how public transport users adapt to planned disruptions over extended periods and demonstrate the feasibility of using AFC data alone to infer route choice preferences.

Beyond its academic relevance, this research also has important practical implications. The Washington Metropolitan Area Transit Authority (WMATA), a key stakeholder in this study, plays a crucial role in promoting public transport as a sustainable alternative to private vehicles. Unlike unexpected disruptions, planned disruptions provide transit agencies with the opportunity to proactively manage service changes, as they can adjust mitigation strategies, improve passenger communication, and even modify disruption plans in response to anticipated behavioral shifts. By understanding how passenger preferences evolve during and after planned disruptions, this research can help refine policies, optimize service planning, and develop strategies that better retain and attract riders. These insights ultimately support the government's broader sustainability goals, such as reducing congestion and emissions by increasing public transport usage.

1.1. Research objective and Questions

Following the gaps in literature the objective of this research is to analyze transit users' route choice preferences in response to planned disruptions, using smart card data from the Washington DC metro network. Based on this objective the following research question was constructed.

How do public transport passengers' route choice preferences change in response to a planned public transport disruption?

To answer this main questions the following sub-questions need to be answered.

- *Which disruptions can be used to analyze changes in route choice preferences?*
- *What are the main factors influencing passengers' route choices in response to planned public transport disruptions?*
- *How do route choice preferences evolve over time during the post-disruption period?*
- *How suitable is revealed preference data for analyzing changes in route choice preferences in response to a planned public transport disruption?*

1.2. Thesis Structure

This thesis is organized as follows: Chapter 2 reviews the existing literature, providing the state of the art in the field. Chapter 3 outlines the methodology used to address the research questions. In Chapter 4, this methodology is applied to data from the Washington D.C. metro network. Chapter 5 presents the results, followed by Chapter 6, which discusses the findings, draws conclusions, and offers insights for future research.

2 Literature review

To find the knowledge gap in the current body of literature it is important to first establish the state of the art on planned public transport disruption literature. To determine this, first a search strategy was adopted to find articles related to public transport disruptions and how people tend to behave during these disruptions. Based on these articles there are certain themes that are relevant to zoom in on more. First the state of the art on disruptions in literature will be discussed. This entails the type of disruption, the duration of the disruption, the research focus and whether or not there is an analysis after the disruption. Once we know the state of the art we can look at the data sources that have been used in the different studies and after that the study methods that were used will be outlined. Based on these themes the gap in the literature will be established in the literature discussion.

2.1. Search Strategy

To find articles that have to do with the behavior of passengers during a public transport disruption there are three topics upon which search words were based. The first topic is *planned disruptions in public transport*, to evaluate what articles are already written in the context of transit user behavior during planned public transport disruptions.

We have a smart card dataset available for this thesis, which is a form of Revealed preference data. Therefore it was decided to also look at literature on the use of revealed preference data in public transport behavior research.

Based on these topics search words were constructed which can be found in table 2.1. The articles that were found that had to do with passenger behavior during public transport disruptions were selected and can be found in this table. The search engines that was used for this literature review was Sciencedirect/Elsevier and the TU Delft Library website.

Table 2.1: Search words for disruptions in public transport

Topic	Search words	Articles found
Planned disruptions in public transport	Planned AND disruption AND public AND transport Planned AND disruptions OR disturbances AND travel AND behavior OR behavior	Marra and Corman, 2023, Yap and Cats, 2022, Eltved et al., 2021, Deng et al., 2022, Arslan Asim et al., 2021, Li et al., 2020, Shires et al., 2019, Yap et al., 2018, Yap and Cats, 2021a
Revealed preference in public transport	”Revealed preference” AND public AND transport	Adelé et al., 2019
Smart card data in public transport	”smart card data” OR AFC AND public AND transport OR transit AND disruption OR disturbance	Bernal et al., 2016, Sun et al., 2016, van der Hurk, 2015, Wang et al., 2024, Mo et al., 2022a, Mo et al., 2022b, Liu et al., 2021, Nazem et al., 2018

There are also articles that were found via snowballing. The article of Zhu et al. (Zhu et al., 2017) was found via backward snowballing from the article of Eltved et al. (Eltved et al., 2021).

The articles of Rahimi et el. (Rahimi et al., 2019, Rahimi et al., 2020) were also found via backward snowballing, this time via the article of Mo et al. (Mo et al., 2022b).

The articles that were found and shown in table 2.1 were analyzed and will be discussed based on the following topics: public transport disruptions, the data source that is used in each article and the study methods that are used in these papers.

2.2. Disruptions in public transport systems

A disruption in a public transport system is an event that prevents a bus, tram, metro, or train from following its standard route. Over the past decade, the focus on managing public transport disruptions has shifted from an operations-oriented perspective to a passenger-oriented approach (van der Hurk, 2015, Zhu et al., 2017, Krishnakumari et al., 2020). This shift means that disruptions are no longer assessed solely in terms of operational delays; instead, their effects on passengers have become a key consideration. In particular, recent studies have analyzed the impact of disruptions on passengers, with a focus on changes in their route or mode choices (Marra and Corman, 2023, Eltved et al., 2021, Deng et al., 2022).

Besides route and mode choice, some studies also look at how disruptions affect overall demand in public transport systems (Yap and Cats, 2022, Yap et al., 2018, van der Hurk, 2015). While research covers different aspects of disruptions, most studies still focus on how passengers adjust their route and mode choice behavior when disruptions occur.

Studies on disruptions differ not only in their research focus but also in the type of disruption they examine. Disruptions can be either unplanned, such as those caused by technical failures or accidents, or planned, such as those due to scheduled maintenance. The nature of a disruption heavily influences how passengers react, particularly in adjusting their route or mode choice. The impact of a planned disruption is smaller than if this same disruption would occur unplanned, due to the ability for passengers to anticipate their route or mode change. However, planned disruptions usually last longer than unplanned disruptions, which means the accumulated disruption impact can be far greater for planned disruptions (Yap and Cats, 2021a). This can also lead to a loss of public transport ridership on the affected routes, leading to revenue losses.

A key advantage of planned disruptions, however, is that they provide public transport authorities with the opportunity to implement mitigation strategies such as alternative routes, shuttle services, or improved communication. This makes them particularly useful for policy-oriented studies, as they allow researchers and transit agencies to assess how well different mitigation measures work. Understanding how passengers adapt to planned disruptions can help improve contingency planning and long-term infrastructure strategies.

There is a fairly balanced amount of research on the effects of planned and unplanned disruptions in the literature examined for this study. Examples of studies on unplanned disruptions include Marra and Corman (Marra and Corman, 2023) and Li et al. (Li et al., 2020), while research on planned disruptions can be found in Eltved et al. (Eltved et al., 2021) and Zhu et al. (Zhu et al., 2017).

A key takeaway from the literature is that most studies focus on passenger behavior or demand prediction only during the disruption itself (Wang et al., 2024). While this makes sense for short-term disruptions, as they are unlikely to have lasting effects, it is also valuable to examine the impact of longer-term disruptions after they have ended. However, only a small number of studies extend their analysis beyond the disruption period, such as those by Eltved et al. (Eltved et al., 2021, Shires et al. Shires et al., 2019, and Nazem et al. Nazem et al., 2018).

Another notable trend in the literature is that most disruptions studied are short-term, typically lasting no more than a day (Arslan Asim et al., 2021, Bernal et al., 2016). When assessing post-disruption effects, longer-term disruptions—which last at least several weeks—are more relevant, as they are more likely to influence passenger behavior even after normal operations resume. Despite this, research on the long-term effects of disruptions remains limited.

2.3. Data sources for public transport disruption studies

The literature not only highlights differences in the types and durations of disruptions but also reveals significant variation in the data sources used to study them. Many earlier studies on public transport route choice analysis were based on stated preferences (SP) data for which surveys are used. Like the study of Arslan et al. (Arslan Asim et al., 2021) that used SP surveys to analyze transit users' mode choice behavior during light rail transit short term planned service disruptions. However, SP surveys require passengers to respond to hypothetical situations, which may not reflect the actual travel choices of passengers and might therefore result in biased estimates (Mo et al., 2022a).

Later studies therefore started using revealed preference (RP) data in the form of travel surveys to reflect the actual choice passenger made. Like the study of Marra and Corman (Marra and Corman, 2023) that used travel surveys to analyze how different network disturbances affect public transport passengers, regarding the chosen route and the travel cost. Or the study of Zhu et al. (Zhu et al., 2017) which, based on travel surveys distributed before and after disruptions, identified three types of behavioral changes. Using surveys allows for a very detailed analysis, they are however costly and often limited in sample size (Ingvardson et al., 2024).

However, with the use of smart cards there is also a new way of collecting revealed preference data that can be used. The usage of smart cards allows for automatic fare collection (AFC). The AFC system is a system which automatically collects the entrance and/or exit of each passenger in the network, allowing to infer their trips. AFC data can be used to analyze and predict the way passengers make choices during disruptions. This is a more reliable way of collecting data than via surveys because the data reflects real behavior without relying on memory or subjective interpretation. It also eliminates the disadvantage of the limited sample size that is the case for RP surveys. Smart card data often covers a large number of users, providing a broad and diverse dataset for analysis. So with the coming of smart cards and its AFC data there is now a vast amount of RP data available for a wide range of analysis within public transport planning and modelling.

AFC data is already used to analyze and predict the way passengers make choices during disruptions. One of the papers investigating public transport route choice under disturbances from AFC data is (Yap et al., 2018). Focusing on four planned disruptions, they identified that the in-vehicle time of rail-replacing services and the associated waiting time are perceived worse than the ones of normal services.

While AFC data provides valuable insights into passenger movements, it has limitations. One major drawback is the lack of descriptive statistics about the passengers using smart cards, as privacy regulations often prevent access to personal data. Additionally, AFC data does not capture subjective factors such as perceived comfort, safety, or satisfaction—elements that can play a crucial role in route choice decisions. In discrete choice modeling, additional variables are often included to enrich the analysis, but when relying solely on AFC data, researchers are limited to the attributes recorded by the system. As a result, some important factors influencing route choice may be overlooked.

This issue does not arise in Revealed Preference (RP) and Stated Preference (SP) surveys, as these methods allow researchers to collect both passenger demographics and perception-based data. For example, Rahimi et al. (Rahimi et al., 2019) conducted a combined RP-SP survey to analyze rail users' responses to a subway disruption in Toronto. The RP section gathered information about passengers' real-life experiences with unplanned rail disruptions, while the SP section presented hypothetical disruption scenarios and asked respondents how they would react—such as canceling their trip or switching modes. Additionally, the survey collected demographic details such as age and gender, which AFC data alone does not provide.

Given these differences, combining smart card data with survey data—where possible—could help address the limitations of AFC data and provide a more complete picture of passenger behavior during disruption.

Beyond the type of data used, the location where AFC data is collected also plays a crucial role in shaping study conclusions. Differences in public transport networks, passenger demographics, and travel patterns mean that findings from one region may not be directly transferable to another. The data used in this study originates from Washington, D.C., a transit system that has not been widely analyzed in the context of disruptions. To the author's knowledge, the only study that has previously used this data to examine disruptions is the study of Yap and Cats (Yap and Cats, 2021a), which focused on predicting unplanned disruptions and their impact on passenger delays.

2.4. Study methods for analyzing and predicting transit users' choice behavior during disruptions

In addition to differences in disruption types and data sources, studies also vary in the methods used to analyze and predict transit users' choice behavior during disruptions. Broadly, these methods can be categorized into aggregate and disaggregate approaches. Aggregate-level analysis examines overall trends and patterns across the transit system, providing high-level insights into system performance. This approach is particularly useful for policymakers when assessing the broader impact of disruptions. In contrast, disaggregate-level analysis focuses on individual travel behavior, considering factors such as personal characteristics, route choices, and decision-making processes. This allows for a more detailed understanding of how different passenger groups respond to disruptions. Different methods are employed within both aggregate and disaggregate analyses. The following section provides an overview of the key study methods identified in the literature.

2.4.1. Aggregate study methods

Aggregate-level methods focus on identifying general patterns, trends, and system-wide responses to public transport disruptions. The following subsections describe the aggregate methods most commonly found in the literature.

Data analysis

Data analysis is an aggregate-level method widely used to examine transit user behavior during disruptions, identifying patterns, trends, and key influencing factors. One of the most common approaches involves using AFC data, which provides large-scale, objective insights into passenger movements. Studies such as those from Liu et al. (Liu et al., 2021), Nazem et al. (Nazem et al., 2018), and Zhu et al. (Zhu et al., 2017) have leveraged AFC data to analyze how travel patterns shift during disruptions.

Beyond overall trends, data analysis can also help categorize different forms of passenger behavior. For example, the study of Sun et al. (Sun et al., 2016) used AFC data to classify passengers into three behavioral groups: continue, detour, or leave the system. These categorizations offer a deeper understanding of how disruptions influence individual decision-making.

To analyze relationships, trends, and behavioral responses in greater detail, various statistical methods have been applied in disruption studies. One commonly used method is regression analysis, which examines the relationship between disruptions and ridership. For example, Bernal et al. (Bernal et al., 2016) applied regression models to evaluate how service disruptions impact passenger demand. Another statistical approach is the Accelerated Failure Time (AFT) model, which analyzes the time until specific events occur. Rahime et al. (Rahimi et al., 2019) used an AFT model to investigate passengers' waiting time tolerance during a disruption, providing insights into how long passengers are willing to wait before adjusting their travel plans.

In addition to regression-based models, other studies have explored categorical relationships in passenger behavior. Multiple Correspondence Analysis (MCA), for example, has been used to analyze relationships between categorical variables. The study of Adelé et al. (Adelé et al., 2019) applied MCA to study how suburban train users responded to disruptions, uncovering behavioral patterns among different groups of commuters. Another method used to estimate population-level behaviors is statistical inference modeling. In the study by Mo et al. Mo et al., 2022a, statistical inference techniques were applied to predict the mean and variance of the number of passengers who share the same behavioral response to disruptions.

While regression analysis and AFT modeling focus on quantitative relationships and timing, MCA and statistical inference explore categorical trends and broader behavioral patterns. Together, these methods provide valuable insights into how passengers react to disruptions, helping transit agencies develop more effective mitigation strategies and service improvements.

Simulation and/or modelling

Simulation techniques are aggregate-level methods used to analyze passenger flows, predict changes in demand, and test service adjustments during disruptions. They help researchers and planners understand the impact of different scenarios and improve public transport operations through data-driven decisions.

In the context of passenger behavior during disruptions, simulation has been employed as an analytical method in several studies. Deng et al. (Deng et al., 2022) utilized an agent-based model to predict passenger flow distribution in a planned metro station service disruption scenario. Similarly, Yap et al. (Yap et al., 2018) developed a public transport ridership prediction model to estimate ridership levels under four different disruption types, demonstrating how simulation can help anticipate the impact of disruptions on public transport usage. Additionally, Wang et al. (Wang et al., 2024) applied simulation techniques to model service disruptions, restoration processes, and various disruption scenarios caused by natural disasters.

These studies highlight how simulation and modeling are already being used to analyze passenger behavior during public transport disruptions and underscore their potential as valuable tools for transit planning and management.

Machine learning

Machine learning is another aggregate-level method used to analyze passenger behavior during public transport disruptions. By processing large datasets, machine learning models can uncover patterns in travel behavior, predict demand changes, and assess the impacts of disruptions. These techniques offer valuable insights into how passengers adapt their travel choices in response to different disruption scenarios.

An example of machine learning applied in transit research is the study by Yap and Cats (Yap and Cats, 2022), which used machine learning techniques to predict passenger demand during a planned closure. Similarly, Yap and Cats (Yap and Cats, 2021a) employed machine learning to predict disruptions and their passenger delay impacts for different disruption types. Their study developed a supervised learning approach to estimate how frequently different types of disruptions occur at various stations within a public transport network and to predict the resulting passenger delays. This method allows for station-specific and time-period-specific predictions without requiring extensive empirical disruption observations for each location and time period.

While the previously mentioned studies used machine learning for prediction, it can also be applied for clustering analysis. K-means clustering, for example, is a machine learning technique that partitions a dataset into a predetermined number of clusters by grouping similar data points together while keeping dissimilar points separate. The algorithm iteratively assigns data points to the nearest cluster centroid and recalculates the centroids based on the mean of each cluster. Eltved et al. (Eltved et al., 2021) used k-means clustering to group passengers based on their travel behavior before and after a disruption. This approach allowed them to observe how different passenger groups adjusted their travel patterns following the disruption, while also comparing these changes to reference lines unaffected by disruptions to account for general travel trends.

2.4.2. Disaggregate study methods

Disaggregate-level methods focus on understanding individual passenger behavior, preferences, and decision-making processes during disruptions. Discrete choice modeling is the primary disaggregate method identified in the reviewed literature.

Discrete choice modelling

Discrete choice modeling is a disaggregate-level method that examines individual preferences for travel attributes and to predict passenger flows in response to network changes (Dixit et al., 2023). By estimating the relative valuation of different travel attributes, discrete choice models provide insights into how passengers make route or mode choices under varying conditions. Most discrete choice models are based on the Random Utility Maximization (RUM) framework, which assumes that individuals choose the alternative that provides them with the highest utility among the available options. The utility of each alternative is determined by observable attributes, such as travel time and cost, as well as unobserved factors that influence decision-making. While discrete choice models share a common foundation in RUM, they differ in their specifications and utility functions. The following section discusses the most frequently applied discrete choice models in the literature.

- **Multinomial Logit (MNL)**

  One of the most widely-used discrete choice methods is the Multinomial Logit (MNL) model. It estimates the probabilities of selecting each alternative from a set of options, based on the attributes or characteristics of the alternatives and and individual specific factors (McFadden, 1974). It has the property that the relative probabilities of each pair of alternatives are independent of the presence or characteristics of all other alternatives. This property, known as the independence of irrelevant alternatives (IIA), implies that the introduction or improvement of any alternative will have the same proportional impact on the probability of each other alternative Koppelman and Wen, 1998. Through the estimation of parameters using maximum likelihood estimation techniques, the MNL model provides insights into decision-making processes in various domains. In the context of transit disruptions, the MNL model can be applied to analyze passenger preferences for travel attributes and how these preferences influence their choices when faced with alternative travel options.

  Several studies have applied the Multinomial Logit (MNL) model to analyze passenger behavior during transit disruptions. Marra and Corman (Marra and Corman, 2023) quantified the impact of disturbances on individual trips by developing a metric for service degradation and analyzing how disruptions influence passengers' route choices. Similarly, Shires et al. (Shires et al., 2019) used MNL to examine how passengers respond to planned engineering-based disruptions, assessing whether they change their route, switch modes, or forgo travel altogether. In the context of unplanned service disruptions, Rahimi et al. (Rahimi et al., 2019) employed MNL to investigate passengers' waiting tolerance and identify the factors influencing their behavior when facing unexpected delays.

- **Mixed Logit (ML)**

  The Mixed Logit (ML) model is a flexible and widely used statistical model for analyzing discrete choices. As an extension of the Multinomial Logit (MNL) model, it offers several advantages. One key benefit is that it does not exhibit the Independence of Irrelevant Alternatives (IIA) assumption, allowing for more realistic modeling of choice behavior. Additionally, ML accounts for correlations in repeated choices made by the same individual, making it well-suited for panel data (Algers et al., 1998, Chorus, 2019).

ML overcomes three major limitations of the standard logit model. First, it allows for random taste variation, meaning individuals may have different sensitivities to travel attributes. Second, it enables unrestricted substitution patterns across choices, meaning alternatives are not assumed to be independent in how they compete. Third, it accounts for correlations in unobserved factors over time, making it particularly useful for datasets with multiple observations per individual (Chorus, 2019).

An example of ML being applied in transit research is the study by Yap et al. (Yap et al., 2020), which employs a mixed logit model with panel effects to evaluate crowding in urban tram and bus travel. Since their dataset contains multiple route choice observations from the same smart card number (representing the same individual), using an ML model corrects for possible correlations between choices made by the same passenger. By extending the standard MNL model to a panel-data ML model, the study accounts for within-individual variation in route choices and improves the accuracy of crowding valuation.

Another study by Rahimi et al. (Rahimi et al., 2019) applies a random parameter multinomial logit model to account for heterogeneity across observations and panel effects. Their analysis reveals that a wide range of factors, including socio-demographic attributes, personal attitudes, trip characteristics, and built environment factors, significantly influence passenger behavior during unplanned transit disruptions. Moreover, the study finds that the effect of service recovery time varies depending on the type of transit service affected, with rail users being more sensitive to recovery times than bus users. These findings provide valuable insights for transit agencies aiming to improve service quality, enhance user satisfaction, and strengthen transportation resilience.

Given the variety in estimated parameters and the observed heterogeneity in passenger responses, the ML model proves to be a valuable approach for analyzing transit behavior during disruptions. By capturing unobserved heterogeneity and accounting for repeated choices, ML provides a more comprehensive understanding of how passengers respond to service changes. This makes ML a valuable tool for transit agencies to inform policy decisions and improve disruption management strategies.

- **Nested Logit (NL)**

  The Nested Logit (NL) model is one of the most widely recognized extensions of the Multinomial Logit (MNL) model, allowing for interdependence between alternatives within the same group. It is based on the idea that some alternatives share common characteristics and can be grouped into nests. Within a nest, the error terms of alternatives may be correlated, whereas the error terms across different nests remain uncorrelated (McFadden, 1978). This relaxation of the MNL model structure helps address violations of the Independence of Irrelevant Alternatives (IIA) assumption, making it more suitable when alternatives are not completely independent.

  An example of NL being applied in transit research is the study by Mepparambath et al. (Mepparambath et al., 2023), which uses a nested logit model to calibrate an integrated taxi and transit mode and route choice model. The study explores behavioral interdependencies between taxi and transit options, testing both a two-level nested logit model and a cross-nested logit model. While this application is not specific to disruptions, the methodology is relevant to mode and route choice modeling and could be extended to study passenger behavior during disruptions. However, NL is particularly useful when analyzing mode choice distinctions rather than exclusively focusing on route choices, as it captures the structural relationships between different transport modes more effectively.

- **Latent Class Model (LCM)**

  The Latent Class Discrete Choice Model (LCM) recognizes that individuals within a population may have different preferences and behavior patterns. Unlike standard models that assume a homogeneous decision-making process, LCM allows for unobserved (latent) classes or segments within the population, each characterized by distinct preferences and behaviors. Individuals are probabilistically assigned to these classes based on their observed choices (van Cranenburgh, 2021). Separate sets of parameters are estimated for each class, enabling the model to capture heterogeneity more effectively than the MNL model by identifying distinct groups of decision-makers with unique characteristics.

An example of this approach is the study by Li et al. (Li et al., 2020), which applies a latent class model to analyze behavioral heterogeneities in metro passengers' travel plan choices during unplanned service disruptions with uncertainty. By segmenting passengers into different latent classes, the study reveals how distinct groups respond differently to disruptions, providing valuable insights into passenger decision-making under uncertainty.

- **Path Size Logit (PSL)**

  The Path Size Logit (PSL) model is a state-of-the-art approach for modeling route choice in public transport (Nielsen et al., 2021, Yap and Cats, 2021b). It builds upon the Multinomial Logit (MNL) model by introducing an additional path size correction term, which penalizes correlated alternatives within a choice set. This adjustment accounts for the fact that passengers may perceive routes with overlapping segments as less distinct alternatives, reducing the Independence of Irrelevant Alternatives (IIA) assumption typically found in MNL models.

  Given its ability to account for route overlap, PSL is particularly relevant for public transport networks, where many alternative paths share common segments. This makes it a useful tool for analyzing passenger behavior in transit disruptions, as it provides a more realistic representation of route choices compared to standard MNL models.

2.4.3. Summary of approaches to analyze passenger behavior during disruptions

As demonstrated, both aggregate-level and disaggregate-level methods are used to analyze and predict passenger behavior during public transport disruptions. Aggregate approaches, such as data analysis, simulation, and machine learning, are suited for identifying system-wide patterns and forecasting the broader impacts of disruptions. Disaggregate methods, particularly discrete choice modeling, focus on individual decision-making, capturing preferences and behavioral differences among passengers. Together, these methods show that a wide range of approaches can and have been used to analyze and predict passenger behavior during public transport disruptions.

2.5. Discussion and Conclusion

From the literature review, it becomes evident that there is a substantial body of research on transit users' behavior during service disruptions. These studies utilize various data sources, including revealed preference (RP) and stated preference (SP) surveys, as well as smart card (AFC) data, and apply different methodological approaches. A summary of these studies is provided in Tables 2.2 and 2.3, which highlight key research themes, data sources, and study methods.

Table 2.2: Summary of Disruption Types, Duration, and Data Sources in Public Transport Studies

Study	Geographical Context	Disruption Type		Disruption Duration			Disruption Analysis		Data Source		
		Planned	*Unplanned*	*< 1 day*	*1-7 days*	*> 7 days*	*During*	*After*	*SP survey*	*RP survey*	*Smart card*
Marra and Corman, 2023	Zürich, Switzerland		x	x			x				x
Yap and Cats, 2022	Amsterdam, Netherlands	x			x		x				x
Eltved et al., 2021	Copenhagen, Denmark	x				x	x	x			x
Deng et al., 2022	Shanghai, China	x		x			x				x
Arslan Asim et al., 2021	Calgary, Canada	x		x			x		x		
Li et al., 2020	Guangzhou, China		x	x			x		x	x	
Shires et al., 2019	London, England	x					x	x	x		
Adelé et al., 2019	Paris, France		x	x			x			x	
Yap et al., 2018	The Hague, Netherlands	x					x				x
Wang et al., 2024	Hangzhou, China		x			x	x				x
Mo et al., 2022a	Chicago, USA		x	x			x				x
Mo et al., 2022b	Chicago, USA		x	x			x				x
Liu et al., 2021	-		x	x			x				x
Nazem et al., 2018	Montreal, Canada	x				x	x	x			x
Bernal et al., 2016	Chicago, USA	x		x			x			x	x
Sun et al., 2016	Beijing, China		x	x			x				x
van der Hurk, 2015	Netherlands		x	x			x				x
Zhu et al., 2017	Washington D.C., USA	x			x	x	x	x	x	x	
Rahimi et al., 2019	Chicago, USA		x	x			x		x	x	
Rahimi et al., 2020	Chicago, USA		x	x			x		x	x	
Yap and Cats, 2021a	Washington D.C., USA		x	x			x				x
Current Study	**Washington D.C. USA**	**x**				**x**		**x**			**x**

Table 2.3: Summary of Methodologies and Research Focus in Public Transport Disruption Studies

Study	**Study Method**					**Research Focus**				
	Discrete choice modelling	*Data analysis*	*Simulation or Modelling*	*Machine learning*	*Cluster modelling*	*Route choice*	*Mode choice*	*Demand prediction*	*Waiting Tolerance*	*Delay impacts*
Marra and Corman, 2023	x					x				
Yap and Cats, 2022				x				x		
Eltved et al., 2021					x	x	x			
Deng et al., 2022			x			x				
Arslan Asim et al., 2021	x						x			
Li et al., 2020	x					x	x			
Shires et al., 2019	x					x				
Adelé et al., 2019		x			x	x	x			
Yap et al., 2018								x		
Wang et al., 2024			x							
Mo et al., 2022a		x				x	x			
Mo et al., 2022b	x						x			
Liu et al., 2021		x				x	x			
Nazem et al., 2018		x					x			
Bernal et al., 2016		x					x			
Sun et al., 2016		x				x				
van der Hurk, 2015			x					x		
Zhu et al., 2017		x				x	x			
Rahimi et al., 2019		x							x	
Rahimi et al., 2020	x					x	x			
Yap and Cats, 2021a				x						x
Current Study	**x**					**x**				

The review of existing literature on public transport disruptions reveals several key insights regarding disruption types, data sources, and study methodologies. Many studies focus on passenger behavior during disruptions, particularly in terms of route and mode choices (Marra and Corman, 2023, Eltved et al., 2021, Deng et al., 2022), while others examine demand prediction (Yap and Cats, 2022, Yap et al., 2018) or waiting tolerance and delay impacts (Rahimi et al., 2019, Yap and Cats, 2021a). However, a common trend in the literature is that most studies focus on passenger behavior during the disruption itself, with limited attention given to how passengers adjust their travel choices after the disruption ends. While some studies, such as Eltved et al. (Eltved et al., 2021), analyze pre- and post-disruption behavior, their short observation period (three weeks before and after) makes it unclear whether passenger route choices revert to pre-disruption patterns or remain permanently altered.

Another key observation is the variation in data sources used in disruption studies. Early studies relied on stated preference (SP) surveys (Shires et al., 2019, Li et al., 2020, Zhu et al., 2017), which, while useful for understanding hypothetical decision-making, may not fully capture actual passenger behavior, leading to potential biases in route choice estimations (Mo et al., 2022a). Later studies incorporated revealed preference (RP) data through travel surveys (Marra and Corman, 2023), providing more realistic insights into passenger decision-making. However, RP surveys are costly and often limited in sample size (Ingvardson et al., 2024). With the increasing availability of AFC data, researchers now have access to large-scale, real-world datasets that enable more detailed analyses of passenger behavior. AFC data has been utilized in studies examining passenger behavior during disruptions (Marra and Corman, 2023, Eltved et al., 2021), but very few studies have applied AFC data specifically to analyze route choice preferences during and after disruptions. Most studies investigating route choice preferences rely on RP or SP surveys, and even when AFC data is used, it is typically focused on mode choice rather than route choice (Mo et al., 2022b). This presents an important research gap in using AFC data to analyze passenger route choice preferences during and after disruptions, particularly when relying solely on AFC data, which does not include socio-demographic characteristics or subjective factors like perceived comfort and safety.

Furthermore, another key gap in the literature is the lack of research on route choice preferences in the Washington, D.C. metro system. The geographical context plays a crucial role in shaping study conclusions, as differences in public transport networks, passenger demographics, and travel patterns mean that findings from one region may not be directly transferable to another. Despite the availability of AFC data in Washington, D.C., this system has not been widely analyzed in the context of disruptions. To the author's knowledge, the only study that has examined disruptions in this region using AFC data is Yap and Cats (Yap and Cats, 2021a), which focused on predicting unplanned disruptions and their impact on passenger delays, rather than route choice preferences.

To address the identified research gaps, this study focuses on analyzing passenger route choice preferences during and after planned disruptions using AFC data from the Washington, D.C. metro system. Unlike previous studies that primarily rely on RP and SP surveys, this research aims to assess whether AFC data alone can be used to infer route choice preferences, even in the absence of descriptive and subjective factors. By leveraging smart card data, this study offers a large-scale, real-world dataset, providing insights into actual passenger behavior rather than stated intentions.

Furthermore, this study extends the post-disruption analysis period from weeks to several months, allowing for a more comprehensive understanding of whether passenger behavior returns to pre-disruption patterns or remains altered over time.

Additionally, this study contributes to bridging the geographical gap by focusing on the Washington, D.C. metro system, a network that has not been widely studied in the context of disruptions. Since public transport networks, demographics, and travel patterns vary across cities, findings from other regions may not be directly applicable to Washington, D.C. This study will provide empirical insights tailored to this transit system, offering valuable information for local policymakers and transit agencies.

Based on these considerations, the aim of this research is to analyze transit users' route choice preferences before and after planned disruptions, using AFC data from the Washington, D.C. metro network.

3 Methodology

This section introduces the designed methodology for this research. A detailed visualisation of the methodology is shown in figure 3.1.

This study aims to analyze the impact of planned disruptions on passenger route choice preferences through a discrete choice modeling approach. The methodology is structured into a series of systematic steps to ensure a comprehensive analysis of pre- and post-disruption behavior. The process begins with Disruption Identification (Step 1), where significant disruptions are identified. Once a disruption is identified, the analysis moves to choosing the Pre- and Post-Disruption Period (Step 2), where appropriate time frames before and after the disruption are selected to capture any changes in passenger preferences regarding route choice.

Following this, an Affected Origin-Destination (OD) Pair is selected (Step 3) to narrow down the analysis to routes most impacted by the disruption. This OD pair forms the basis for generating a choice set for both the Pre- and Post-Disruption Period (Step 5). This choice set reflects the different route options available to passengers during the selected periods, providing a detailed dataset for the subsequent modeling steps. However, to be able to generate this choice set, we first need to identify the relevant attributes available to us (Step 4).

After the choice set generation several discrete choice models are estimated (Step 6). These are the Multinomial Logit (MNL) model, a Mixed Logit model to account for panel data and a Path Size Logit (PSL) model that accounts for the overlap between routes. After estimating these models, the 'best' model is selected based on a trade-off between its fit and complexity (Step 7).

Finally, the study conducts a beta coefficient analysis (Step 9) to assess how the estimated model parameters change between the pre- and post-disruption periods. This comparison provides insights into shifts in passenger preferences, highlighting the disruption's effect on route choice preferences.

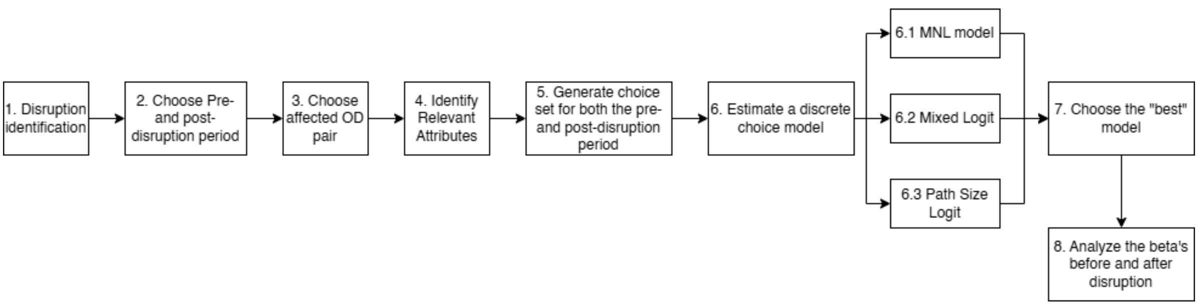


Figure 3.1: Flowchart of methodology

3.1. Disruption Identification Process

The first step in this research is to identify disruptions that significantly impact passenger behavior and meet specific criteria. This step is essential for setting a strong foundation for analyzing how travel patterns shift before, during, and after disruptions. To do this, a set of criteria must first be established to ensure the selection of disruptions that provide meaningful insights into passenger behavior. These criteria draw inspiration from the work of Fariba Tavakoli (Tavakoli, 2024), who used a similar dataset for disruption analysis.

Once the criteria are set, the next task is to identify suitable disruptions from the planned disruptions file. This file provides information on the date, time, and a description of the disruption.

Disruption Identification Criteria

The first step is to identify disruption criteria. The criteria and their explanation can be found below:

- *The disruption should occur both on week and weekend days*

  This criterion ensures that the analysis captures passenger behavior throughout different travel patterns seen on weekdays and weekends. Weekday travel often reflects commuting patterns (work and school), while weekend travel can be more leisure-oriented. Including both provides a comprehensive understanding of how passengers adapt to disruptions under varying demand conditions.

- *The disruption should last several days in a station or a set of adjacent stations*.

  A disruption that spans several days allows for more robust data collection and analysis of how passenger behavior adapts over time. It provides sufficient time to observe adjustments in travel patterns and allows passengers to settle into new routines or route choices during the disruption. Focusing on a single station or a set of adjacent stations ensures that the disruption's impact is concentrated in a specific area, making it easier to analyze changes in travel behavior

- *There must be at least one month prior to the disruption, as a pre-disruption period, and a minimum of 3 months following the disruption for the post-disruption analysis*

  Having a defined pre-disruption period of at least one month ensures a baseline of passenger behavior before the disruption occurs, providing a point of comparison. A post-disruption period of at least three months allows for the analysis of long-term changes in travel behavior after the disruption has ended. This longer duration is necessary to capture any residual effects of the disruption and ensure that the observed changes are stable and not just temporary adjustments.

- *There should be no other disruption lasting more than several hours in the affected area during both the pre- and post-disruption period.*

  This criterion ensures that the observed changes in passenger behavior can be attributed solely to the disruption being studied. If other disruptions occur in the same area, it becomes challenging to isolate their effects and understand how passengers are responding specifically to the disruption of interest. Eliminating other significant disruptions ensures the clarity and accuracy of the analysis.

- *There should be other route options to take during the disruption.*

  The presence of alternative routes is crucial for a study focused on route choice behavior. It allows for an analysis of how passengers change their travel routes when their preferred route is disrupted. A disruption where detours or alternative routes are not possible would not provide the necessary variation in choice behavior, making it less suitable for understanding how passengers adjust their travel decisions.

Disruption identification from planned disruptions file

The disruption should be identified from the planned disruptions file, which includes the start date, time of day, and message for each disruption. Disruptions that span multiple days have the word "Thru" in the message column. By initially selecting these disruptions, we can then determine if they meet the additional criteria, before selecting a final disruption for analysis.

3.2. Selection of Pre- and Post-Disruption Periods

Selecting appropriate time frames for the pre- and post-disruption periods is crucial for accurately analyzing the impact of a disruption on passenger behavior. These periods serve as a reference to understand changes in travel patterns and route choices caused by the disruption, helping to isolate the effects of the disruption from other external factors that could influence travel behavior.

Pre-Disruption Period

The pre-disruption period is selected to represent typical passenger behavior before the disruption. A minimum of one month is chosen for this period to ensure that a stable baseline is established. This time frame provides enough data to capture regular travel patterns, accounting for any day-to-day variations or weekly trends. By understanding the baseline behavior, the analysis can more accurately detect shifts in travel patterns caused by the disruption. Additionally, a longer pre-disruption period helps to control for potential seasonal effects, ensuring that the data reflects typical conditions before any disruption occurred. So it would therefore be preferable to have a pre-disruption period of 2 months. The selected pre-disruption period will therefore be between 1 and 2 months.

Post-Disruption Period

The post-disruption period needs to be long enough to observe how passenger behavior evolves after the disruption has ended. For this study, a post-disruption period between 3-5 months is selected to capture both short-term adjustments and longer-term changes in travel patterns. This extended period allows for the analysis of immediate reactions to the end of the disruption, as well as any gradual shifts as passengers settle into new routines or potentially revert to their pre-disruption behavior.

To gain a deeper understanding of how passenger preferences may change over time, the post-disruption period will be analyzed in two ways: first, as a single comprehensive post-disruption period, and second, by dividing the period into separate months. This approach helps to assess whether any initial changes in route choice behavior gradually revert to pre-disruption patterns or if new preferences persist throughout the entire post-disruption period.

It is also important to allow sufficient time for passengers to adjust their travel patterns after the disruption. Immediately following the disruption, passengers may continue using alternative routes before gradually returning to their preferred routines. By including a longer post-disruption period, the analysis can capture both the initial adjustments and any longer-term changes in route choices, ensuring a comprehensive assessment of the disruption's impact on travel behavior.

3.3. Choose affected OD pair

The selection of the affected Origin-Destination (OD) pair is a critical step in understanding how disruptions influence passenger route choices. This selection is particularly important because the chosen OD pair forms the basis for generating the choice sets used in the discrete choice models. The objective is to identify OD pairs where at least one of the available routes passes through the disrupted segment of the network. This ensures that the disruption directly affects travel options between the selected origin and destination, allowing for a meaningful analysis of route choice adjustments. The process of selecting the affected OD pair involves the following steps:

1. *Identification of all OD pairs*

   The initial step is to identify all possible OD pairs within the transit network. An OD pair represents a journey from a specific origin station to a destination station. By mapping out all OD pairs, the analysis covers a comprehensive range of travel movements within the affected area, ensuring that no potential journeys are overlooked.

2. *Identification of Routes for Each OD Pair*

   Once the OD pairs are identified, the next task is to determine all routes that passengers use between each OD pair. This includes both direct routes and those that involve transfers between lines. Understanding the full range of routes available between each OD pair helps identify the different travel paths that might be impacted when a disruption occurs.

3. *Selection of OD Pairs with Multiple Route Options*

The analysis focuses on OD pairs that have multiple route options between the origin and destination. These pairs are ideal for studying route choice behavior, as they offer alternative routes that passengers might switch to if one is disrupted. OD pairs with only a single route are not suitable for this analysis, as there would be no variability in route choice when that route is disrupted.

4. *Selection of OD pairs with transfer variability*

   The analysis also places emphasis on OD pairs where the available routes differ in the number of transfers required. Transfer variability is important because, even when multiple routes exist between an origin and destination, these routes can sometimes follow the same tracks or overlap significantly. In such cases, the analysis might not capture the full extent of route diversity if differences in the number of transfers are not considered. By focusing on OD pairs with a variety of transfer options this step ensures that the study includes routes with genuinely different travel paths. This distinction allows for a more nuanced understanding of how disruptions influence route choice behavior, as passengers may weigh transfer convenience against travel time and other factors when selecting among routes with different transfer counts.

5. *Selection of the Final OD Pair for Analysis*

   The final step involves selecting an OD pair, from the remaining candidates, where at least one of the available routes passes through the disrupted segment of the network and has a high volume of passenger observations. This ensures that the selected OD pair is directly affected by the disruption, providing a meaningful context for analyzing how passengers adjust their route choices. By prioritizing an OD pair with a large number of observations, the analysis gains a richer dataset, allowing for a more comprehensive exploration of route choice behaviors and yielding more statistically robust results. Initially the OD pair going over the disrupted section with the highest amount of observations will be selected. If this OD pair proves unsuitable during further analysis, the next candidate with the highest number of passenger observations will be considered, continuing in this manner until a suitable OD pair is identified.

By following this systematic process, the chosen OD pair is one with multiple available routes with transfer variability that goes over the disrupted section. This allows for generating a meaningful choice set in the next step of the process.

3.4. Identify Relevant Attributes

To generate a choice set, we first need to know which attributes to include in that choice set. That is why this section aims to identify the relevant attributes.

There are many different factors that affect passengers route choice in public transport. It is well-known from literature that time and cost are very important for public transport travellers. Previous studies have included various time components, namely in-vehicle time, access/egress time, and transferring time, as the main descriptors of passengers' route choices (Nielsen et al., 2021, Jánošíková et al., 2014). Also, there is often a general penalty for transfers. Next to that, fare and/or frequency of the lines are also commonly used parameters in route choice models. Also personal characteristics, like age and gender, are often affecting passenger public transport route choice (Grison et al., 2017).

These all would be good attributes to include in this study. We however are dependent on the available data in the dataset. Since the data uses anonymous smart card data, there are no personal characteristics included. This means we cannot use personal characteristics in this study.

The inclusion of attributes such as time and fare components will be determined based on data availability during the case study phase during which the data will be analyzed.

3.5. Choice Set Generation

The process of generating a choice set is crucial for modeling route choice behavior, as it defines the range of route options available to passengers between a selected Origin-Destination (OD) pair. This choice set is generated separately for the pre-disruption and post-disruption period(s), allowing for a comparison of travel behavior across these periods. The goal is to create a realistic set of route alternatives by incorporating all routes that passengers have actually chosen while also finding a balance in the size of the choice set.

On one hand, a small choice set may not be able to reproduce the chosen route, and consequently may not capture the individual's behavior and preferences. On the other hand, a very large choice set might cover the chosen route but could lead to misinterpretation of the estimated route choice coefficients, as relatively few alternatives are actually perceived by individuals (Yao and Bekhor, 2020). Therefore, the choice set generation process aims to include all meaningful alternatives that passengers could realistically consider, without overwhelming the analysis with routes that are unlikely to be perceived as viable options.

3.5.1. Data Selection, Filtering, and Cleaning

The data preparation process involves three main steps: selecting the relevant data, filtering out invalid entries, and ensuring the dataset is clean and consistent for analysis.

Data Selection

The first step involves selecting data for the chosen OD pair across two time frames—before and after the disruption. This ensures that the dataset captures passenger travel behavior during both periods. The analysis focuses on passengers who are present in both time frames, allowing for a direct comparison of their route choices. This is done by selecting passengers whose Card ID number is in both datasets. The result is a dataset that includes all relevant trips for these travelers across the two periods.

Data Filtering

Once the data is selected, it needs to be filtered to remove entries that may distort the analysis. The following entries are considered invalid and are removed:

- Transactions where tap-in and tap-out data are identical.
- Metrobus entries, as the focus is on Metrorail trips.
- Transactions with missing tap-in and/or tap-out information.
- Transactions where the tap-out time is earlier than the tap-in time.
- Transactions with journey times longer than three hours, as the longest paths in the network are a bit over two hours.
- Transactions with journey times shorter than two minutes, since the shortest paths in the network are around two minutes.
- Duplicate entries

Data Cleaning and Attribute Preparation

After filtering the data, further cleaning ensures the dataset is free from incomplete or inconsistent entries. This involves:

- Retaining only records with complete information for key attributes.
- Calculating additional route-specific attributes, like the number of transfers required for each route, to better understand the convenience of different travel options.
- Correcting any inconsistencies in route descriptions to ensure uniformity across all records.

By following these steps, the resulting dataset is robust, accurate, and well-suited for analyzing the impact of network disruptions on passenger route choices.

3.5.2. Choice set generation logic

The next step in the analysis involves identifying potential route alternatives for each journey between the observed OD pairs. Since this study is based on smart card data, the observed routes provide a foundation for determining non-chosen alternatives, which together form the full choice set. To ensure the choice set represents realistic route options, the following steps are applied:

- The dataset with chosen routes includes only those passengers who are present in both the pre-disruption and post-disruption datasets, ensuring a consistent sample for comparison.

- Routes that are chosen less than 1% of the time are excluded from the dataset. This step helps to focus on more realistic and frequently chosen routes, ensuring that the analysis is not skewed by rarely selected alternatives.

To ensure that attribute values (e.g., travel time, wait time, number of transfers) reflect realistic travel conditions, values are derived from observed AFC data rather than hypothetical assumptions. Since passengers make decisions based on expected travel conditions, attribute values for both chosen and non-chosen alternatives are calculated based on historical travel conditions, using the following methodology:

- Time interval logic

  Since travel conditions fluctuate throughout the day, the ideal approach would be to calculate average attribute values per hour. However, in some cases, there may be insufficient data within a single hour to produce a reliable average. To maintain realism while ensuring more robust estimates, attribute values are instead calculated based on broader time frames. This allows for a better balance between capturing variations in travel conditions and having a sufficient number of observations to derive meaningful averages. The time frames used are:

  – Morning peak: 6am - 9am
  – Midday: 10am - 3pm
  – Evening rush: 4pm - 7pm
  – Evening: 8pm - 12pm

  Each observation is assigned to a time frame based on when the trip occurred. The average attribute values for each route are calculated separately for each time frame, ensuring that non-chosen alternatives reflect conditions that would have been expected at the time of travel.

- Fallback logic for missing data If attribute values for a non-chosen alternative are missing in a specific time frame, the system selects values from the closest comparable time frame rather than applying broad averages. For instance, if data is unavailable for the midday time frame, values from the evening time frame are used as a substitute.

Choice Set Format for Discrete Choice Modeling

Once the choice set is prepared, it is formatted for use in discrete choice modeling software like Biogeme. Each observation is structured to include:

- One chosen alternative, representing the route that the passenger selected.
- Several non-chosen alternatives, representing other viable routes that the passenger could have taken.
- A binary variable indicating whether a route was chosen (`CHOSEN` = 1 for the selected route and `CHOSEN` = 0 for alternatives).

This structured choice set allows for detailed modeling of route choice behavior, providing insights into how passengers respond to changes in network conditions, such as disruptions. By using observed data and applying systematic fallback methods, the choice set represents a realistic range of alternatives for passengers, ensuring a robust foundation for the subsequent analysis.

3.6. Set up of the discrete choice models

Discrete choice models are used to explain or predict an individual's selection from a set of two or more discrete alternatives, making them well-suited for analyzing route choices.

At the core of these models is a utility function, which represents the attractiveness of each alternative in the choice set. This function is typically expressed as a linear combination of observed attributes, such as travel time, cost, or number of transfers, along with corresponding coefficients, known as betas (β), which indicate the relative importance of each attribute.

The utility of an alternative consists of two parts: a systematic component, which is based on the observed attributes and their coefficients, and a random component, which accounts for unobserved factors that influence individual choices but are not included in the model. The probability of choosing a particular alternative depends on how its utility compares to the utilities of all other available options.

While it is possible to include Alternative Specific Constants (ASCs) in the utility function to account for inherent preferences for certain alternatives, this approach is less suitable for this study. Since passengers may choose different routes within the same trip, it is unlikely that they have a fixed inherent preference for a specific combination of routes. Instead, their choices are more likely influenced by varying travel conditions and route attributes rather than a persistent bias toward a particular route option.

While all discrete choice models share the same fundamental structure based on utility maximization, they differ in how they handle specific aspects such as correlation between alternatives, repeated choices, and route overlap. Since it is not immediately clear which discrete choice model will best capture the patterns in the data and provide the best model fit, this study will estimate and compare three different models: the Multinomial Logit (MNL) model, the Panel Logit model, and the Path Size Logit (PSL) model. All models will be estimated using Biogeme in Python to determine which one provides the most accurate representation of passenger route choice behavior.

3.6.1. Multinomial Logit (MNL) Model

The Multinomial Logit (MNL) model is one of the most commonly used models for analyzing choices between multiple alternatives. It provides a straightforward way to examine how route attributes like affect choice probabilities. One key feature of the MNL model is that it assumes the choice between two options is not influenced by other available alternatives, known as the Independence of Irrelevant Alternatives (IIA). This makes the MNL model relatively simple to use, as it calculates the probability of choosing an option based on how its attributes compare to those of the other options.

In the MNL model, the utility of each alternative i for individual n is given by:

$$U_{ni} = V_{ni} + \epsilon_{ni}$$

Where ϵ_{ni} is the random error term, capturing unobserved influences on the choice and V_{ni} is the systematic utility that depends on the observed attributes of alternative i (such as travel time, cost, or number of transfers). The systematic utility V_{ni} is a linear combination of the observed attributes and their respective coefficients:

$$V_{ni} = \beta_1 X_{1ni} + \beta_2 X_{2ni} + \cdots + \beta_k X_{kni}$$

In this equation, x_{kni} represents the observed attributes of alternative i, such as travel time, cost, or the number of transfers. The coefficients β_k indicate the relative importance of each attribute, reflecting how much they contribute to the overall utility of an alternative.

In the MNL model, the probability that individual n chooses alternative i from a choice set is given by the following formula:

$$P_{ni} = \frac{e^{V_{ni}}}{\sum_j e^{V_{nj}}}$$

While the MNL model's assumption of IIA simplifies the analysis, it can be a limitation when alternatives share common characteristics. Despite this, the MNL model remains a popular choice for understanding how factors like travel time or cost influence route choice preferences, particularly when alternative-specific effects such as the ASC are included. And it is a good basic model to start with.

Transfer penalty

In this study, a transfer penalty approach is also considered to address the potential correlation between in-vehicle time and the number of transfers. Since preliminary analysis suggested that routes with more transfers often have shorter in-vehicle times, explicitly including both attributes in the utility specification could lead to multicollinearity. To avoid this, the disutility of transfers can be incorporated indirectly by applying a fixed penalty to the in-vehicle time of routes that require a transfer.

The process for selecting an appropriate transfer penalty is as follows. First, several different penalty values will be tested by adding them to the in-vehicle time for alternatives involving a transfer. For each tested value, a Multinomial Logit (MNL) model will be estimated, and its final log-likelihood will be recorded. The penalty value that results in the best model fit — indicated by the highest final log-likelihood — will be selected for use in the final model. This approach allows the effect of transfers to be captured without introducing multicollinearity into the utility function.

3.6.2. Mixed Logit (ML) Model with panel data

The Mixed Logit (ML) model is considered because it can account for differences in individual preferences and does not rely on the strict Independence of Irrelevant Alternatives (IIA) assumption, which is a limitation of the Multinomial Logit (MNL) model. The notation follows the framework from Algers et al. (Algers et al., 1998).

In a choice situation, an individual i chooses an alternative j at time t based on the utility of that alternative:

$$U_{ijt} = \beta_i x_{ijt} + \epsilon_{ijt} \tag{3.1}$$

where:

- x_{ijt} represents the characteristics of the alternative, such as travel time or cost.
- β_i is a set of coefficients that measure how much an individual values each attribute. Unlike in the MNL model, where these values are the same for everyone, the ML model allows them to vary between individuals.
- ϵ_{ijt} is a random error term that captures other factors influencing the choice that we do not observe.

In the MNL model, the coefficients β are fixed for everyone. However, in the Mixed Logit model, we assume that people have different preferences, and these coefficients follow a probability distribution:

$$\beta_i \sim f(\beta|\theta) \tag{3.2}$$

where θ describes the characteristics of the distribution, such as the mean and standard deviation. For example, some individuals might value shorter travel times more than others, and the ML model accounts for this variation by estimating both a mean preference (β) and a standard deviation (σ). The standard deviation captures the extent to which individual preferences deviate from the average, allowing the model to account for heterogeneous tastes.

Given an individual's specific preferences β_i, the probability of choosing alternative j is calculated using the standard logit formula:

$$L_{ij}(\beta_i) = \frac{e^{\beta_i x_{ij}}}{\sum\limits_j e^{\beta_i x_{ij}}} \tag{3.3}$$

However, since we do not directly observe each individual's preferences β_i, we need to account for all possible values that β_i could take, based on the assumed distribution:

$$L_{ij}(\theta) = \int \frac{e^{\beta_i x_{ij}}}{\sum\limits_j e^{\beta_i x_{ij}}} f(\beta_i|\theta) d\beta_i \tag{3.4}$$

This means we compute the probability of choosing an alternative by considering all possible preference variations in the population.

Since we have panel data, meaning that each individual makes multiple choices over time, we must account for the fact that choices from the same person are related. Instead of looking at each choice independently, we consider the probability of observing an entire sequence of choices:

$$S_i(\beta_i) = \prod_t L_{ij(t)}(\beta_i) \tag{3.5}$$

Since we do not know the true value of β_i, we integrate over all possible values:

$$S_i(\theta) = \int \prod_t L_{ij(t)}(\beta_i) f(\beta_i|\theta) d\beta_i \tag{3.6}$$

This integral does not have a simple solution, so we approximate it using simulation techniques.

Because the Mixed Logit model involves random draws from a probability distribution to estimate choice probabilities, each time the model is run, the results could slightly change. To ensure that results are consistent and reproducible, a random seed is used in the Python implementation. This ensures that the same random numbers are used each time the model is estimated, preventing variations caused by different random draws.

The Mixed Logit model provides a more flexible way to analyze route choices compared to the MNL model. By allowing preferences to vary between individuals and incorporating repeated choices, it can better capture real-world decision-making. However, because it requires simulation-based estimation, it is computationally more complex. In this study, the ML model is evaluated alongside other models to determine the best approach for understanding route choice behavior.

In relation to the research questions, the Mixed Logit model is used to explore whether differences in how travelers value route attributes could have influenced the changes in route choice preferences after the disruption. Although this study does not focus directly on individual-specific behavior, identifying whether travelers responded differently is important. It helps determine whether the observed changes reflect a broad shift in preferences or if they are driven by variation between individuals. Understanding this is important for correctly interpreting the disruption's impact on route choice preferences.

3.6.3. Path Size Logit (PSL) Model

The Path Size Logit (PSL) model is an extension of the Multinomial Logit (MNL) model designed to account for the correlation between overlapping routes in a choice set. A key assumption of the standard MNL model is the Independence of Irrelevant Alternatives (IIA) property, which implies that the relative probabilities of choosing two alternatives remain unaffected by the presence of additional options. However, when multiple routes share significant portions of their paths, this assumption is violated, leading to an overestimation of the probability of selecting similar routes.

To address this issue, the PSL model introduces a path size factor, which adjusts the utility of each alternative based on the extent of its overlap with other routes in the choice set. This ensures that alternatives sharing a large number of links receive a penalty, reducing their relative attractiveness. The notation follows the framework from Duncan et al. (Duncan et al., 2020).

The systematic utility V_i in the PSL model extends the traditional utility function by incorporating the path size term γ_i, which captures the distinctiveness of each route:

$$V_i = \beta' X_i + \beta_{PS} \ln(\gamma_i) \tag{3.7}$$

where:

- β_{PS} is the coefficient associated with the path size factor, indicating the weight given to the path size adjustment. A higher value of β_{PS} means that the model places more importance on penalizing overlapping routes.

- $\ln(\gamma_i)$ is the natural logarithm of the path size term γ_i, which ensures that the penalty for route overlap is applied proportionally. Since γ_i is always between 0 and 1, taking the logarithm results in a negative term that reduces the utility of highly overlapping routes. The more a route overlaps with others, the smaller the value of γ_i, leading to a stronger penalty.

The path size term γ_i for route i is computed as follows:

$$\gamma_i = \sum_{a \in A_i} \frac{t_a}{c_i} \frac{1}{\sum_{k \in R} \delta_{a,k}} \tag{3.8}$$

where:

- γ_i measures how unique a route is compared to others in the choice set.
- The denominator $\sum_{k \in R} \delta_{a,k}$ represents the total number of routes that share link a. If many routes use the same link, this term increases, reducing γ_i and applying a stronger penalty.

The probability of choosing an alternative in the PSL model follows the same general structure as the MNL model, with the only modification being the inclusion of the path size factor. Since the probability formula remains unchanged except for this additional adjustment, the PSL model can be considered a direct extension of the MNL framework, incorporating a correction for route similarity.

The PSL model provides an effective way to account for overlapping routes in route choice modeling, addressing a major limitation of the MNL model. By introducing a path size factor, it ensures that alternatives with significant overlap are penalized, improving the realism of the estimated choice probabilities. This approach is particularly useful in public transport networks where multiple routes often share infrastructure.

The PSL model is considered in this study because it offers a way to address potential overlap between different route alternatives. Since it is likely that multiple routes between the selected OD pair share common track segments, overlap could bias the results if not properly accounted for. By including a path size factor, the PSL model adjusts for similarities between alternatives, ensuring that estimated route choice preferences more accurately reflect true traveler behavior. This is important for answering the research questions because it improves the ability to detect genuine changes in route choice preferences following the disruption, without confounding effects from overlapping infrastructure.

3.7. Model Selection

The process of selecting the best model involves comparing the fit of the previously discussed models using several criteria, including the log-likelihood, Akaike Information Criterion (AIC), and Bayesian Information Criterion (BIC). While the log-likelihood measures the model's goodness of fit, AIC and BIC penalize overfitting, helping to identify the model that best balances predictive accuracy and complexity

Log-Likelihood Comparison

The log-likelihood values of each model are compared to assess how well each model captures the observed choices. A higher (less negative) log-likelihood indicates that the model is better at explaining the decision-making behavior of passengers.

Rho-Square and Likelihood Ratio test

The Rho-square (ρ^2) statistic is often used to assess model fit in discrete choice models. A Rho-square value between 0.2 and 0.4 is considered to indicate a model with good explanatory power (Lam and Xie, 2002).

Another important measure for model selection is the Likelihood Ratio (LR) test, which tests whether a more complex model provides a significantly better fit than a simpler model. It is calculated as (King, 1989):

$$LR = -2(LL_{\text{restricted}} - LL_{\text{unrestricted}})$$

where:

- $LL_{\text{restricted}}$ is the log-likelihood of the restricted (simpler) model.
- $LL_{\text{unrestricted}}$ is the log-likelihood of the more complex model.

The likelihood ratio statistic follows a chi-square (χ^2) distribution, and its significance is evaluated based on the degrees of freedom equal to the difference in the number of parameters between the two models.

Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC)

In addition to log-likelihood, the AIC (Akaike, 1974) and BIC (Stone, 1979) are calculated to evaluate the trade-off between model fit and complexity. These criteria adjust for the number of estimated parameters in the model, penalizing more complex models to avoid overfitting:

$$\text{AIC} = -2 \times LL_{\text{model}} + 2K$$

$$\text{BIC} = -2 \times LL_{\text{model}} + K \times \ln(N)$$

where:

- LL_{model}: Log-likelihood of the model on the estimation sample.
- K: Number of estimated parameters.
- N: Number of observations.

Lower values of AIC and BIC indicate a better balance between model fit and complexity, with BIC applying a stronger penalty for model complexity than AIC. These measures help prevent overfitting and guide the selection of a model that generalizes well to unseen data.

Final Model Selection

The final model is selected based on a combination of log-likelihood, ρ^2, the likelihood ratio test, AIC, and BIC. The model with the highest log-likelihood, a reasonable ρ^2 value, and the lowest AIC and BIC is preferred. However, statistical significance from the likelihood ratio test is also considered to ensure that added complexity is justified by an improvement in model fit.

The selected model is then used for further analysis to interpret how the disruption has influenced route choice behavior and to draw conclusions about changes in passenger preferences before and after the disruption.

4 Case Study

This case study focuses on the Washington D.C. Metrorail system and demonstrates how the methodology outlined in Chapter 3 is applied to analyze passenger route choice preferences before and after planned disruptions. Using smart card data from the Washington Metropolitan Area Transit Authority (WMATA), the case study highlights the key steps of disruption identification, pre- and post-disruption period selection, and route choice modeling. This analysis supports the main goal of this research: to understand how disruptions affect transit users' route choice preferences.

The data that is provided for this study is provided by WMATA through the Smart Public Transport Lab at TU Delft. Data from the Washington DC Metrorail is very interesting for research because it provides transit service for more than 600,000 customers a day throughout the Washington DC area, making it a critical component of the region's transportation infrastructure. Its system is the second busiest in the United States and consists of six color-coded rail lines: Red, Orange, Silver, Blue, Yellow, and Green, which together form a network spanning over 98 stations and more than 129 miles of track Washington Metropolitan Area Transit Authority, 2024. The layout of the system can be seen in Figure 4.1.

The data utilized in this study is a comprehensive set of records from August 2019 to December 2022. This period captures a range of operational conditions, including regular service patterns, planned maintenance activities, and unforeseen disruptions. The data sources include the following that will be used in this study:

1. **Automated Fare Collection (AFC) Data**: This dataset provides detailed records of individual passenger journeys, including tap-in and tap-out data, time stamps, card IDs, and travel duration. It is therefore suitable for tracking passenger movement patterns across the network, allowing for an analysis of route choice behavior before, during and after service disruptions. It can thus help in identifying how passengers adjust their travel paths in response to disruptions.
2. **Disruption Documentation**: The dataset also includes a detailed log of all planned and unplanned service disruptions during the study period. This documentation provides specific information about each disruption, including its start and end times, affected stations, impacted train lines, and a description of the event's nature. This allows for understanding how disruptions impact passenger behavior and the network performance.

For this case study the methodology from the previous chapter will be used. Therefore, the first section will focus on identifying a suitable disruption for further analysis. Then the pre- and post-disruption period will be determined. The identification of the affected OD pair is next. The a section will be dedicated to data preparation and lastly the attributes that will be used in the later stages will be identified.

4.1. Disruption Identification

The first step in the case study involves identifying disruptions that meet the criteria outlined in the methodology section on disruption identification. Adopting this approach, ensures that we focus on disruptions most likely to influence passenger behavior. To begin, disruptions lasting several days

Figure 4.1: Washington DC Metro Network

were identified by searching for entries containing the word 'Thru' in the detail column of the dataset. This initial search yielded a selection of disruptions, which were further screened to ensure that they occur on both weekdays and weekends. This process resulted in a list of 15 disruptions. Each of these disruptions fell within a suitable time range, allowing for the definition of both a pre- and post-disruption period. Next, the feasibility of detours or alternative routes during each disruption was assessed. This narrowed the selection to four disruptions that met all the criteria, specifically:

Table 4.1: Final disruptions

Date	Line	Affected Stations
21/06/2020 - 27/06/2020	Green	Between L'Enfant Plaza and Shaw Howard University
28/06/2020 - 02/07/2020	Green	Between Mt. Vernon Square and U street
06/07/2020 - 18/07/2020	Green	Between U street and Fort Totten
19/07.2020 - 25/07/2020	Red	Between Judiciary Square and Rhode Island Avenue

The final criterion is that there should be no other disruptions lasting more than several hours in the affected area during both the pre- and post-disruption periods. However, in this case, all observed disruptions are relevant when traveling between Gallery Place and Fort Totten, as they occur at adjacent stations and impact the same sections of the line. For this reason, we will treat these disruptions as a single, continuous disruption affecting the area. As a result, the disruption period was defined as lasting from 21/06/2020 to 25/07/2020, with the disruption occurring consistently along the Green or Red Line between Gallery Place and Fort Totten. The identified disruption takes place due to trackwork and platform repairs.

4.2. Pre-and post disruption period

With the disruption now identified, the pre- and post-disruption periods can be established. The pre- and post-disruption periods are defined based on the criteria outlined in section 3.2 of the methodology. The selected pre-disruption period provides a stable baseline of passenger behavior, while the post-disruption period captures both short-term adjustments and potential longer-term changes in route choices. Post-disruption period 1 reflects passengers' immediate responses once the disruption has ended, providing insight into short-term adjustments in route choice. Post-disruption period 2 represents an intermediate phase, where it becomes possible to observe whether passengers maintain their new habits or begin reverting to their original routes. Finally, post-disruption period 3 captures longer-term behavior and indicates whether passengers have settled into a new, stable pattern of travel.

In the pre-disruption period there are no other big disruptions on the selected line, meaning we can use the preferred pre-disruption period of two months. In the post-disruption period there are some disruptions on the selected part of the Green Line. In the period from 31/10/2020 until 22/11/2020 there is a planned disruption message each evening. It would therefore not seem suitable to take this period into account for the post-disruption analysis. From the preferred 3-5 month post disruption period we will therefore go for a 3 month period, excluding the time in which there are other disruptions on the line. This leads to the selection that can be found in table 4.2.

Table 4.2: pre- and post disruption periods

Period	Start date	End date	Duration
Pre-disruption	21/04/2020	21/06/2020	2 months
Post-disruption total	25/07/2020	25/10/2020	3 months
Post-disruption 1	25/07/2020	25/08/2020	1 month
Post-disruption 2	26/08/2020	25/09/2020	1 month
Post-disruption 3	26/09/2020	25/10/2020	1 month

4.3. Identification of affected OD pair

In this section, we apply the methodology's approach to OD pair selection, as described in Section 3.3. By narrowing down OD pairs to those with multiple route options and transfer variability, the analysis targets cases where passengers are likely to have adjusted their routes during the disruption.

After identifying all OD pairs in the network (over 4,000 in total), the next step is to determine the routes available for each pair. Each route can consist of a single line or multiple lines, identified by their color codes [GR (Green), RD (Red), YL (Yellow), SV (Silver), OR (Orange), BL (Blue)]. For instance, a passengers' route can be labeled as BL for a direct route or as RD > YL or BL > SV > GR, where the ">" symbol indicates a transfer between different lines.

With the routes identified, each OD pair is categorized based on whether it has one or multiple route options. The analysis focuses on those OD pairs that offer multiple route options across all three datasets used for the disruption period. This selection process reduces the number of OD pairs to around 1,600.

By also adding the transfer variability criteria we are left with 169 OD pairs that have both multiple routes and a variability in their amount of transfers.

To narrow down the selection further, an OD pair is needed where at least one route passes through the disrupted section of the network, specifically the Green Line between Gallery Place and Fort Totten. When ordering the remaining OD pairs based on the amount of passengers between them, we get the following first 10 OD pairs, that can be found in table 4.3.

In the table we get the number of trips between each OD pair over the disrupted line in the period between 21/04/2020 and 25/10/2020.

Table 4.3: Possible OD pairs

From	To	# Trips
Columbia Heights	Southern Avenue	6658
Columbia Heights	Minnesota Avenue	2930
Fort Totten	Pentagon	2774
Foggy Bottom	Greenbelt	2501
Prince George's Plaza	Foggy Bottom	2497
Fort Totten	Congress Heights	2228
Foggy Bottom	West Hyattsville	2167
Prince George's Plaza	Waterfront	2008
Fort Totten	Suitland	1956
Prince George's Plaza	Ballston	1309

When analyzing the routes between the first OD pair of Columbia Heights and Southern Avenue it becomes evident that there is no significant route variation. When removing the routes that are chosen less than 1% of the time, there is actually no route variation left. This means the first OD pair is not suitable to use in further analysis.

The second OD pair between Columbia Heights and Southern Avenue has the same shortcomings as the first OD pair and is therefore also not suitable for further analysis.

When looking at the third OD pair from Fort Totten to Pentagon, there is route variation. However, the route that is chosen most of the time is not one that is actually existing. It is only the Green line, without any transfers. The problem is that it is not possible to reach Pentagon station without any transfers. This OD pair therefore also does not seem suitable.

The fourth OD pair between Foggy Bottom and Greenbelt is the next route that is analyzed. This route has different options that are actually chosen and transfer variability. Therefore, this OD pair seems suitable for further analysis. This means the OD pair that is used for estimating the discrete choice models will be the OD pair of Foggy Bottom and Greenbelt.

4.4. Data preparation

The data of the OD pair Foggy Bottom - Greenbelt is filtered and cleaned based on the criteria described in the methodology. Furthermore there is also another data preparation step that we decided to include. Regarding this specific OD pair, we have chosen to reassign all instances of the yellow line (YL) to the green line (GR) for the specific OD pair between "Foggy Bottom" and "Greenbelt." This decision is based on the assumption that the Yellow Line does not provide a realistic travel route for this OD pair. While the dataset may show YL as part of a journey, in reality, the Yellow Line does not serve Greenbelt directly, and taking the Yellow Line would involve a walk of approximately 8 miles between the nearest Yellow Line stop and Greenbelt station.

Such a walking distance is impractical for typical metro passengers, making this route an unrealistic alternative. Therefore, we are consolidating YL into GR to reflect the more accurate and feasible routing options between these two stations. This assumption helps ensure that the route choice set consists of realistic travel alternatives, leading to more reliable results in the subsequent modeling and analysis.

This leaves us with the following routes between this OD pair, which will be continuously used during the rest of this thesis:

- Route 1: BL > GR
- Route 2: OR > GR
- Route 3: SV > GR

- Route 4: OR > RD > GR
- Route 5: BL > RD > GR
- Route 6: SV > RD > GR

4.5. Identified attributes

Based on the WMATA data we were able to identify the attributes that we will take into account for estimating the discrete choice models. These are the following:

- **Veh_sec**: This is the total in-vehicle time for the passengers of this stage, expressed in seconds.
- **Wait_sec**: This is the cumulative platform wait time of all rail rides in this stage, expressed in seconds.
- **Transfers**: This attribute is inferred from the route taken, with the number of transfers identified as the number of routes taken, minus 1.

There were more attributes in the dataset that were considered; however, they were not chosen due to various reasons, which are explained below:

- Denied

  The attribute `Denied` represents the number of times a passenger was inferred to have been denied boarding. However, since this value was zero in the vast majority of cases, there was little variation to observe, making it impossible to identify a meaningful trade-off.

- Max_ppc

  This represents the maximum number of passengers per car (ppc) experienced during a fare stage. Given that each metro car has approximately 60 seats, and the highest recorded `max_ppc` value in the dataset was only 37.2, it suggests that seating availability was not a limiting factor, as all passengers likely had a seat. The limited impact of this variable makes it an unsuitable attribute for modeling.

- Reliability

  Reliability could have been measured based on the delay column and potentially represented as On-Time Performance (OTP). However, the delay data was found to contain many inconsistencies and incorrect values, making it unreliable for analysis. Due to the lack of accurate delay measurements, this attribute was excluded from the model.

- Fare

  The fare values in the dataset were often zero, and even when nonzero, the differences in fare between alternative routes were minimal. Since fare variation between options was very small, it would not have had a significant impact on the model's utility function. Additionally, passengers with transit passes or employer-funded transit benefits may not consider fare as a decisive factor, further reducing its relevance in modeling route choice.

- Frequency

  The frequency of a route is based on the route with the lowest service frequency in the dataset. However, this value was the same across all routes in the choice set, meaning there was no variation in this attribute between alternatives. Since frequency does not differentiate route options, it was excluded from the utility function.

4.6. Conclusion

This case study demonstrates the practical application of the methodology developed in Chapter 3. By applying the methodology to the Washington D.C. Metrorail system, we were able to identify disruptions, define appropriate pre- and post-disruption periods, select affected OD pairs, prepare the data for model estimation and and determine the attributes to be used in the discrete choice models.

5

Results

This chapter looks at how the disruption affected route choice preferences. It starts by examining changes in key route attributes and travel patterns, including trends in travel and wait times, differences across routes, and overall ridership levels. Then, it explores route choice behavior using discrete choice models, showing how factors like travel time, transfers, and wait times influenced decisions and how these preferences shifted due to the disruption.

5.1. Analysis of Route Attributes

This section focuses on analyzing transit data to explore how travel attributes (e.g., travel time, wait time) evolved across periods and their potential influence on route choices. The goal is to provide descriptive insights into route-specific patterns and establish a foundation for the next part of the analysis.

5.1.1. Travel time

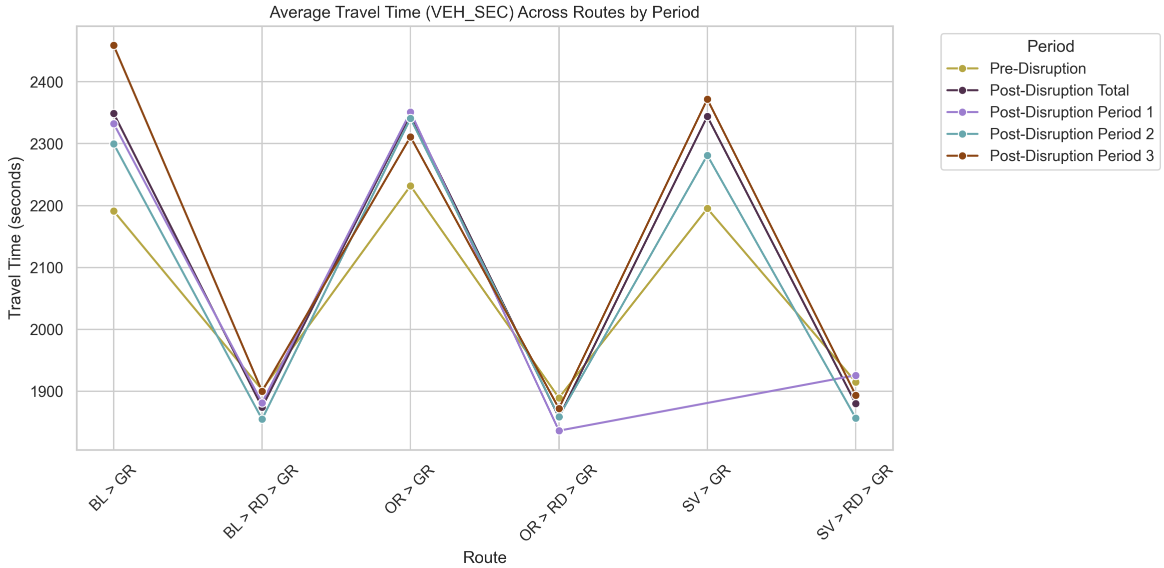


Figure 5.1: Average travel time per route across periods

The graph of average travel times (Figure 5.1) shows clear distinctions between the different routes across all periods. Transfer-heavy routes, such as *BL > RD > GR* and *SV > RD > GR*, consistently have shorter average travel times compared to direct routes like *SV > GR* and *BL > GR*. This patterns stays like that across all different periods. The average travel time for the transfer-heavy routes stays consistent across periods. In contrast, the average travel time for the direct routes increases after the disruption. This shows that transfer-heavy routes not only had shorter travel times, but also remained more consistent throughout the disruption period.

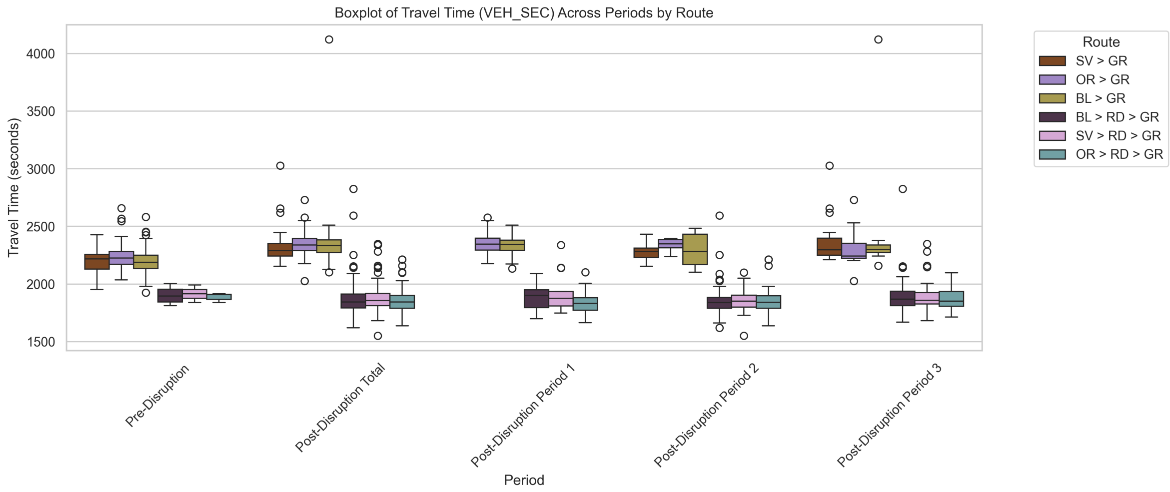


Figure 5.2: Boxplot travel time per route across periods

The boxplots in Figure 5.2 provide further insights into the variability of travel times across different routes and periods. Before the disruption, travel times for more direct routes, such as BL > GR, exhibit greater variability, while transfer-heavy routes appear more stable. However, this could be partly due to the fact that transfer-heavy routes were chosen less frequently in the pre-disruption period, resulting in less observed variation. After the disruption, travel time variability decreases for most routes, particularly in the total post-disruption period and Post-Disruption Period 1, as shown by the narrower interquartile ranges. In Post-Disruption Period 3, however, variability increases slightly for some routes, suggesting a potential shift back toward pre-disruption patterns. Overall however all routes have pretty similar travel time variability.

5.1.2. Wait time

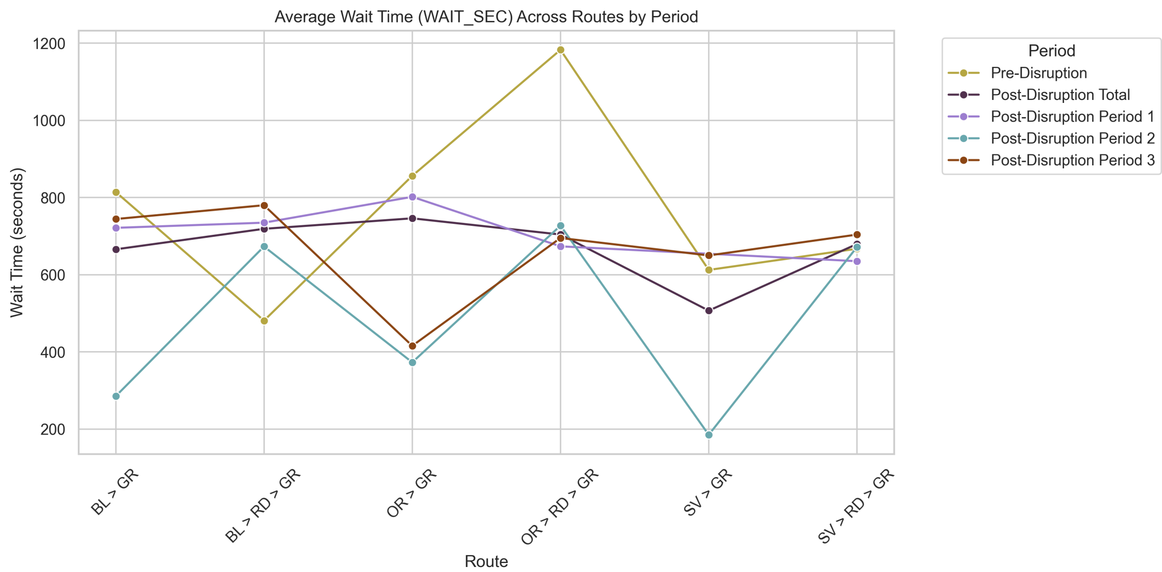


Figure 5.3: Average wait time per route across periods

Figure 5.3 shows the average wait times per route across all periods. Overall, there is no clear trend in how wait times change over time. Some routes experience increases while others show decreases, and this varies across periods. The variation in wait time appears to be route-specific rather than following a consistent pattern linked to the disruption.

One noticeable outlier is the high average wait time for *OR > RD > GR* in the pre-disruption period. This is likely due to the low number of observations for that route in that period, which makes the average more sensitive to extreme values. Aside from this, the differences in wait times across routes remain relatively modest, and no single route consistently stands out across all periods. This suggests that, unlike travel time, wait time was less clearly impacted by the disruption in a consistent or systematic way.

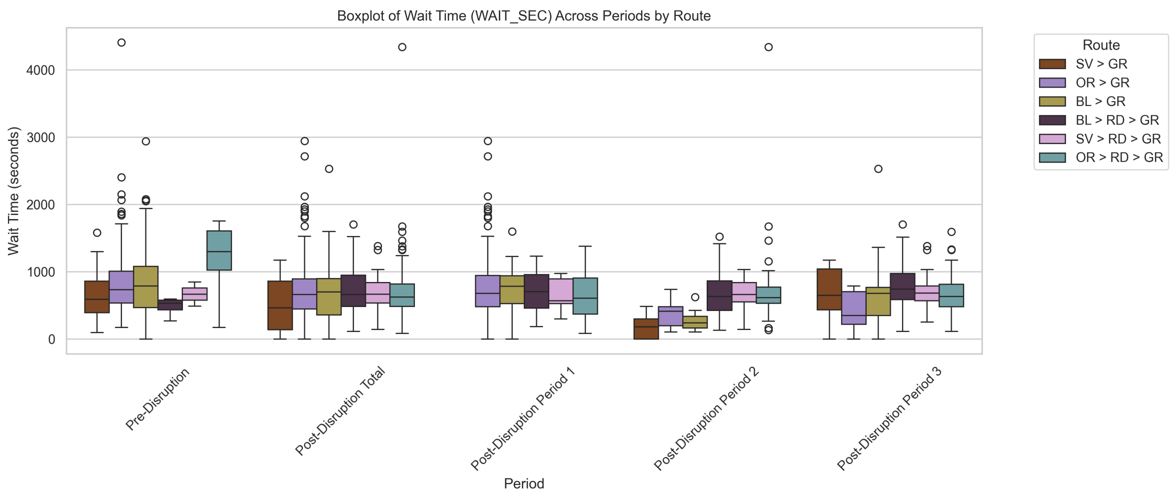


Figure 5.4: Boxplot wait time per route across periods

The distribution of wait times, as depicted in Figure 5.4, provides additional insights into variability across routes and periods. Pre-disruption, routes such as *OR > RD > GR* show high variability, indicated by the wide range of values and routes as *OR > GR* show high variability by their numerous outliers. Other routes, such as *SV > RD > GR* and *BL > RD > GR* exhibit narrower distributions. As mentioned before, the notably high wait time for *OR > RD > GR* in the pre-disruption period stands out as an outlier, likely due to the low number of observations for this route during that time frame. The few recorded observations tend to have higher wait times, which explains this anomaly.

Post-disruption, wait time distributions do not seem to change that much. There is some variability across routes. However, when we look at post disruption period 3, the wait time distributions have stabilized quite a bit. The role of wait times in influencing route choice preferences is further explored in the discrete choice models.

5.2. Analysis of Route Choice Changes

This section shifts focus from route attributes to route choices, analyzing changes in ridership patterns and passenger behavior across periods. By examining aggregate trip counts, route transitions, and individual preferences, this section provides a detailed understanding of how passengers adapted their route choices after the disruption.

5.2.1. Total Trip Counts Across Different Periods

Figure 5.5 shows the total trip counts per route across five distinct periods: the pre-disruption period, the post-disruption total period, and three individual post-disruption sub-periods. The x-axis represents each route, while the colored bars reflect the number of trips taken during each period.

The visualization highlights changes in route usage over time. During the pre-disruption period, the routes *BL > GR* and *OR > GR* had the highest usage, while multi-transfer routes like *BL > RD > GR*, *OR > RD > GR*, and *SV > RD > GR* were used much less frequently. After the disruption, ridership on these multi-transfer routes increased significantly, suggesting a shift in behavior as passengers adapted

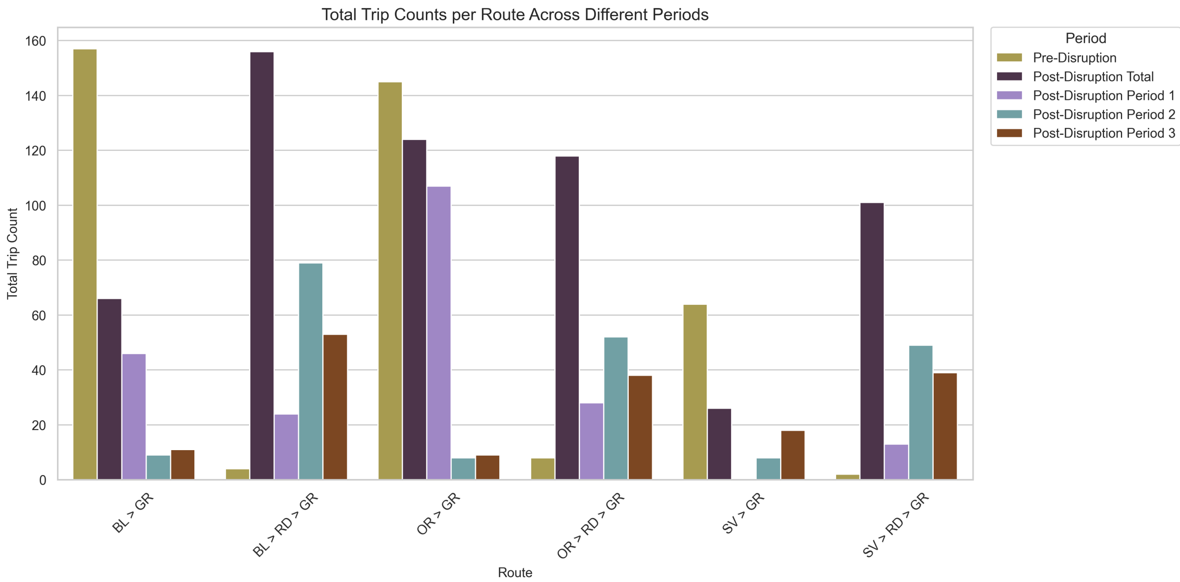


Figure 5.5: Trip counts per route

their travel patterns.

Looking at the post-disruption sub-periods, ridership on the multi-transfer routes remains higher than before the disruption, indicating that many passengers continued using these alternatives even after normal service resumed. Meanwhile, usage of single-transfer routes like *SV > GR* declined slightly over time. This pattern suggests that some travelers may have permanently adjusted their route choice preferences following the disruption.

5.2.2. Sankey Diagram of Route Choices

While total trip counts provide a high-level overview of how passengers adapted their route choice to the disruption, they do not capture how the changes in ridership translate into shifts between specific routes. A Sankey diagram is used to visualize the flow of passengers across routes from the pre-disruption to the post-disruption period. This visualization can be found in figure 5.6. Each node on the left represents a route taken during the pre-disruption period, while each node on the right corresponds to a route in the post-disruption period. The width of the flows between nodes reflects the relative proportion of passengers who transitioned between routes, scaled to account for the higher total number of routes taken post-disruption (586 total trips) compared to the pre-disruption period (381 total trips).

From the diagram, we observe several trends. Direct routes, such as *BL > GR* and *OR > GR*, saw a significant decline in total passenger volumes during the post-disruption period, with many passengers shifting to alternative routes. Transfer-heavy routes, such as *BL > RD > GR*, *SV > RD > GR*, and *OR > RD > GR*, gained a larger share of passengers, absorbing much of the demand from disrupted direct routes. Passengers redistributed themselves across the network, often moving from direct to multi-transfer routes. For example, flows from *BL > GR* and *SV > GR* shifted toward *BL > RD > GR* and *SV > RD > GR*, respectively.

While transfer-heavy routes gained stability or growth in passenger volumes, some routes experienced an overall reduction in ridership, as seen by the thinner nodes for *BL > GR* and *OR > GR* post-disruption. This redistribution highlights a complex adjustment to changing service conditions.

The diagram provides valuable insights into how passenger flows changed during the disruption, setting the stage for the discrete choice model analysis. The model will examine the extent to which travel time, wait time, and transfers influenced these transitions.

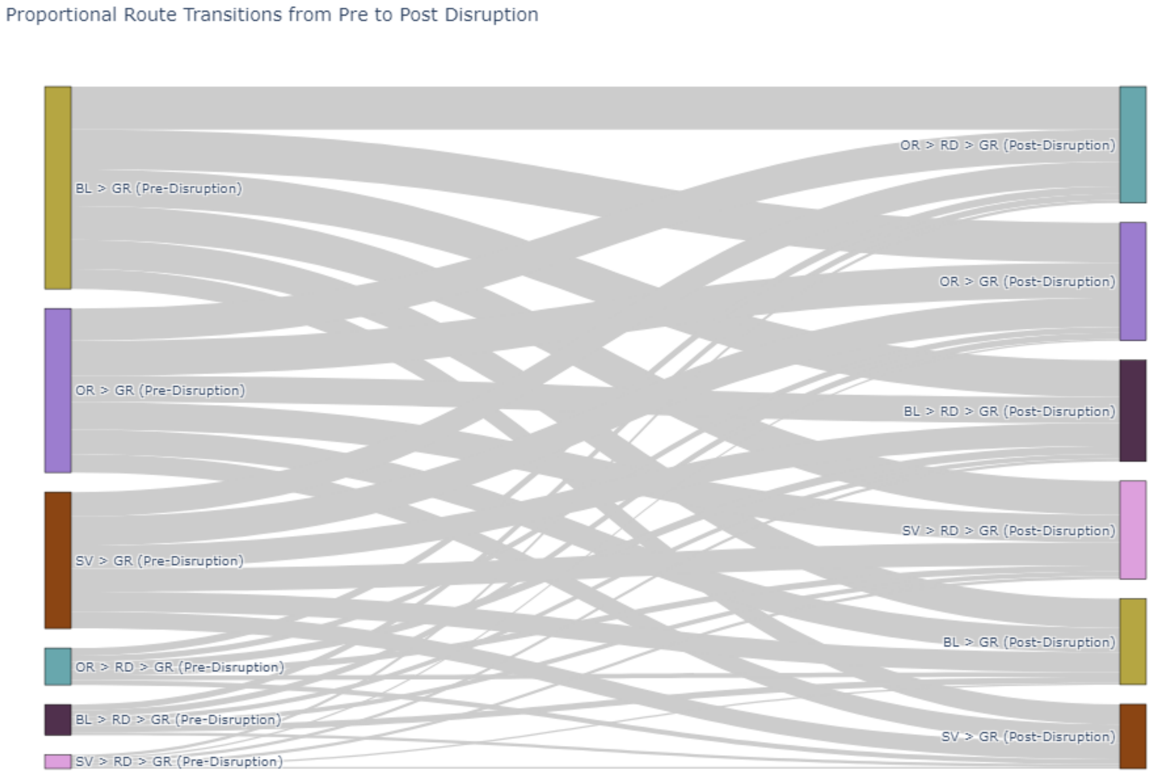


Figure 5.6: Sankey diagram of route changes

5.2.3. Route choices per passenger

The Sankey diagram provided an overview of how passengers redistributed themselves across routes during the disruption, highlighting aggregate flows between routes. However, these aggregate trends do not capture how individual passengers adjusted their route choices. To explore this, the next section uses heatmaps to analyze route choice preferences and changes at the individual level, offering deeper insights into passenger behavior before and after the disruption.

Figure 5.7 illustrates the route choices before the disruption. The intensity of the color corresponds to the frequency with which each passenger used specific routes. Passengers such as Passenger 11 and Passenger 35 relied heavily on the *BL > GR* route, as indicated by the darker blue shading. Other routes, such as *SV > GR* and *BL > RD > GR*, were used less frequently by most passengers. Overall, this heatmap reveals a strong reliance on direct routes, with lower utilization of multi-transfer options during the pre-disruption period.

Figure 5.8 shows the route choices after the disruption. Unlike the pre-disruption period, passengers display a greater reliance on transfer-heavy routes such as *BL > RD > GR* and *SV > RD > GR*, as seen in the darker red shading for passengers like Passenger 13 and Passenger 35. Direct routes such as *BL > GR* see significantly lighter shading for most passengers, indicating a reduced usage compared to the pre-disruption period. This suggests a substantial redistribution of route choice preferences post-disruption, with passengers adapting their choices to the altered network conditions. Another notable observation is that passengers do not consistently prefer a single route. Those with multiple observations tend to choose different routes, suggesting that route choice is not driven by fixed habits.

Figure 5.9 visualizes the changes in route choices by comparing post-disruption route frequencies with pre-disruption frequencies. Positive changes are shown in shades of red, while negative changes are

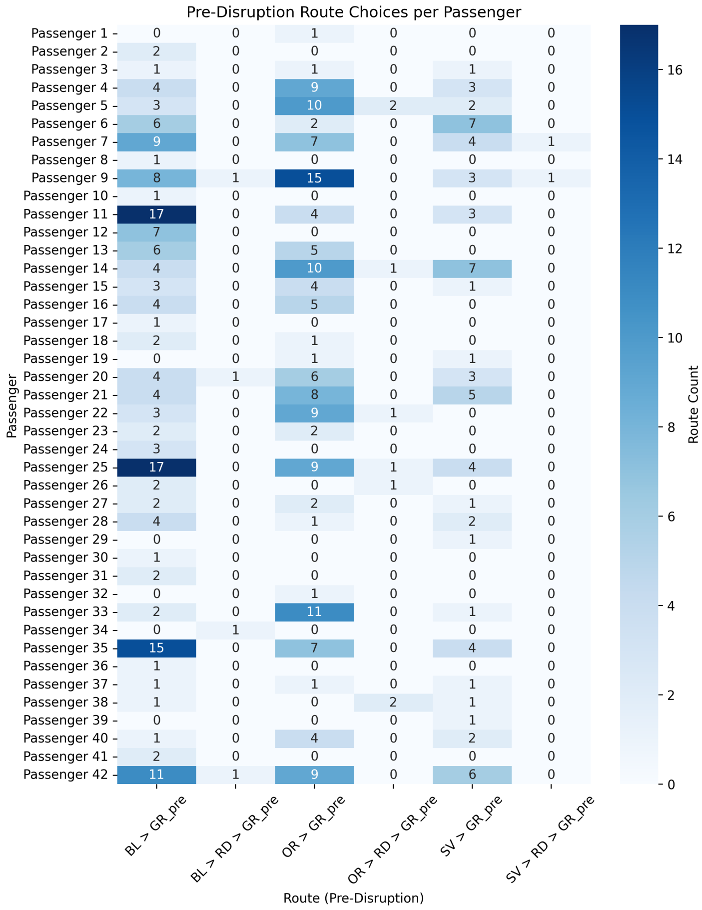


Figure 5.7: Heatmap of Pre-Disruption Route Choices per Passenger

displayed in shades of blue. Passengers such as Passenger 13 and Passenger 35 show significant increases in their use of *BL > RD > GR*, as indicated by the darkest red shading. Conversely, routes like *BL > GR* show widespread reductions in usage, with many passengers, such as Passenger 11 and Passenger 25, exhibiting negative changes marked by blue shading. This heatmap highlights the varying degrees of passenger adaptability, with some passengers making significant shifts to new routes and others maintaining more consistent behavior.

5.2.4. Summary

The combined analysis of route choices through aggregate trip counts, the Sankey diagram, and individual-level heatmaps provides a comprehensive understanding of passenger behavior in response to the disruption. The trip counts reveal significant shifts in ridership levels, highlighting that passengers redistributed their travel across routes, with transfer-heavy routes gaining popularity while direct routes saw a decline. The Sankey diagram complements this by visualizing proportional transitions between routes, emphasizing how groups of passengers adapted their choices by switching to alternatives with additional transfers or greater reliability.

The heatmaps offer a granular perspective, showing that individual passengers responded differently to the disruption. While some passengers made substantial shifts to new routes, others retained a mix of pre- and post-disruption choices. These variations underscore the complexity of decision-making processes and highlight the influence of factors such as travel time, wait time, and transfers.

Together, these analyses demonstrate the need for a discrete choice modeling approach, such as an MNL model, to quantify the impact of specific route attributes on passenger decisions. By analyzing travel behavior at both the aggregate and individual levels, the groundwork has been established to explore the factors driving route choices in a structured and explanatory framework. The MNL model

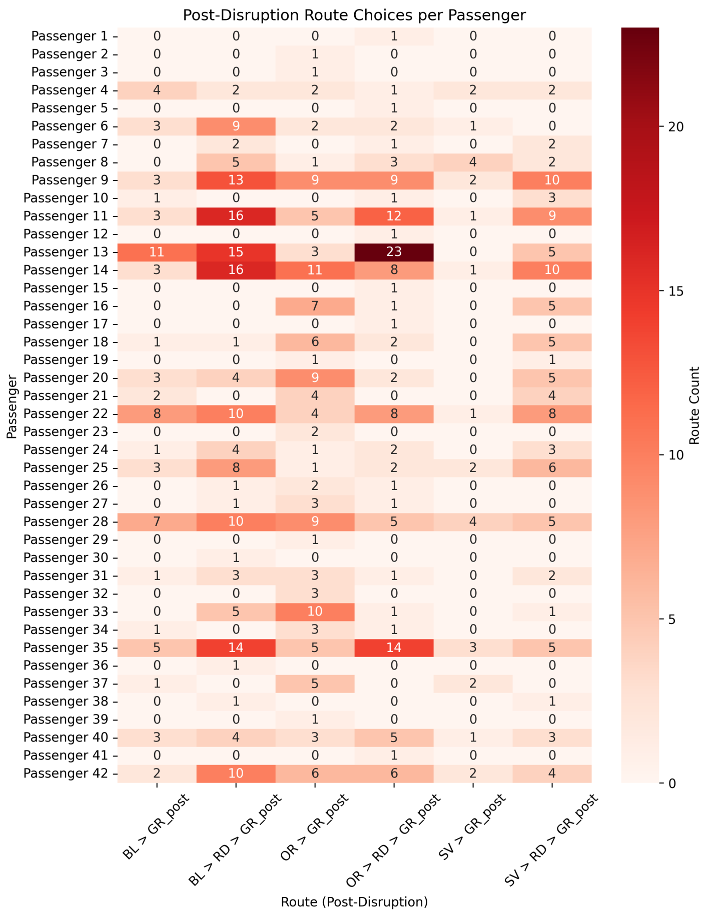


Figure 5.8: Heatmap of Post-Disruption Route Choices per Passenger

will enable us to measure the relative importance of travel time, wait time, and transfers in shaping route choice preferences, both before and after the disruption, providing deeper insights into passenger decision-making.

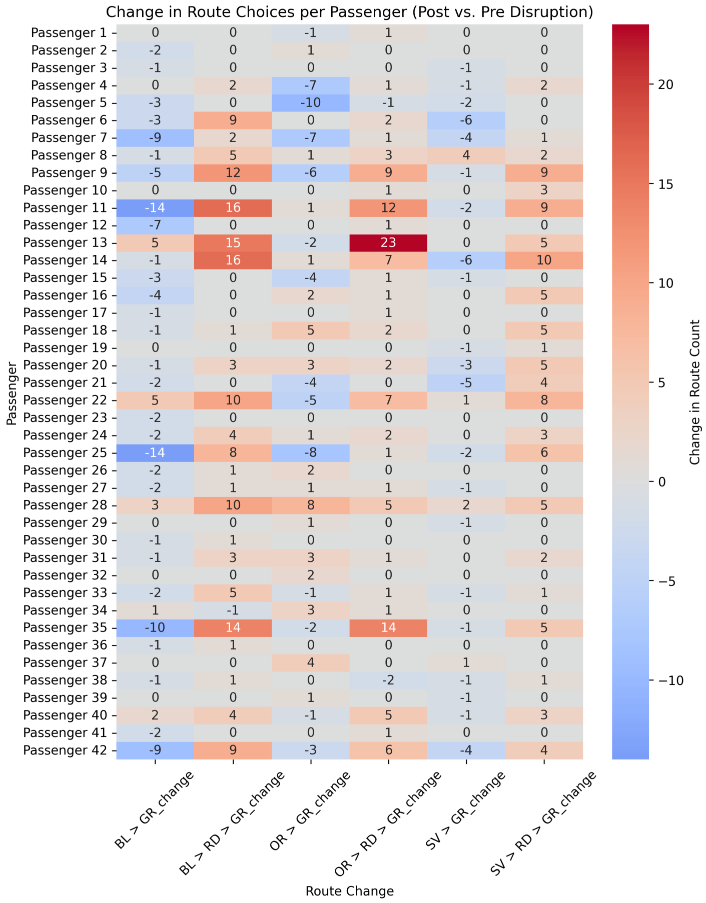


Figure 5.9: Heatmap of Changes in Route Choices per Passenger (Post vs. Pre-Disruption)

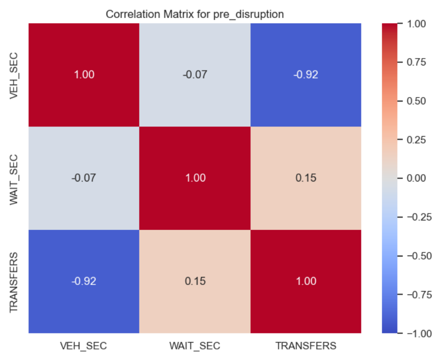
5.3. MNL results

Before estimating a MNL model, it is important to examine the correlation between attributes, as highly correlated variables can lead to multicollinearity. This poses a problem in MNL models because it affects parameter estimation. When two or more independent variables are strongly correlated, the model struggles to estimate their individual effects, resulting in unreliable and unstable estimates. For the data used, the correlation between the variables is displayed in figure 5.10.

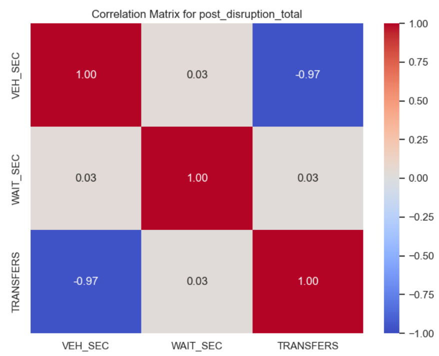
As we can see there is a strong negative correlation between in-vehicle time and the number of transfers, meaning that travel time is generally shorter when there are more transfers, and vice versa. Including both attributes in the MNL model would introduce multicollinearity, making it difficult to estimate their individual effects. To avoid this issue, only in-vehicle time was included in the model. This choice was further justified by the fact that in-vehicle time exhibits greater variation in the data. Unlike the number of transfers, which has only two distinct values, in-vehicle time fluctuates more across different trips, capturing a broader range of travel experiences. This greater variability makes in-vehicle time a more informative attribute, allowing for a more meaningful estimation of its effect on route choice.

The estimation results from the MNL model with only in-vehicle time and wait time in table 5.1 and 5.2 show that the final log-likelihood (LL) of the model is greater than the null LL, which means the model provides a better fit than a random model. Also, the rho-square of 0.114 indicates a semi reasonable model fit, as McFadden suggests that ρ^2 values between 0.2 and 0.4 represent an excellent fit McFadden, 1977.

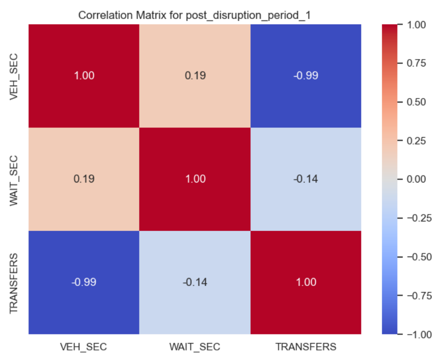
For the estimated parameters a negative coefficient suggests that an increase in the attribute decreases the probability of choosing an alternative, whereas a positive coefficient suggests an increase in probability. Time related attributes are expected to be negative, since a longer wait or travel time has a



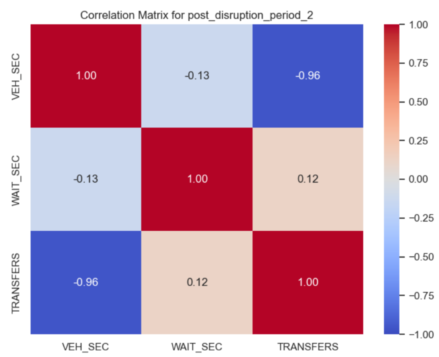
(a) Pre-Disruption



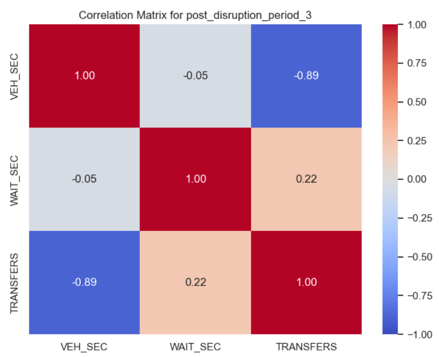
(b) Post-Disruption Total



(c) Post-Disruption Period 1



(d) Post-Disruption Period 2



(e) Post-Disruption Period 3

Figure 5.10: Correlation Heatmaps for data used in the MNL model

negative impact on utility. However, when looking at these results we see that there are quite a lot of positive coefficients that are significant (p-value < 0.05). Like for example the beta for wait time in the pre- and post disruption total period.

Typically, the in-vehicle time coefficient is expected to be negative, as travelers generally prefer shorter travel times. However, the positive coefficient observed in some periods is counterintuitive. This can be attributed to the strong negative correlation between transfers and in-vehicle time, where routes with more transfers often have shorter travel durations due to optimized connections or more direct paths. Since transfers were not included as a separate variable in the model, their effect may have been absorbed into the in-vehicle time coefficient, leading to an unexpected sign reversal.

To address this issue, two alternative model specifications were tested:

Table 5.1: Model Estimation Results MNL

Statistic	Value
Number of estimated parameters	10
Sample size	1577
Null log-likelihood	-2728.461
Final log-likelihood	-2416.265
Rho-square	0.115
Rho-square-bar	0.111
Akaike Information Criterion (AIC)	4852.530
Bayesian Information Criterion (BIC)	4906.163

Table 5.2: Estimated Parameters MNL

Parameter	Value	Rob. Std Err	Rob. t-test	Rob. p-value
BETA_VEH_POST_1	0.170982	0.019461	8.785678	0.000000
BETA_VEH_POST_2	-0.246755	0.026767	-9.218534	0.000000
BETA_VEH_POST_3	-0.171042	0.026250	-6.515931	0.000000
BETA_VEH_POST_TOTAL	-0.054247	0.011350	-4.779524	0.000002
BETA_VEH_PRE	0.601684	0.045771	13.145656	0.000000
BETA_WAIT_POST_1	0.051765	0.038428	1.347084	0.177953
BETA_WAIT_POST_2	-0.009663	0.053035	-0.182209	0.855419
BETA_WAIT_POST_3	-0.041786	0.035506	-1.176862	0.239251
BETA_WAIT_POST_TOTAL	0.093753	0.017437	5.376732	0.000000
BETA_WAIT_PRE	0.043243	0.021738	1.989264	0.046672

- An interaction term between in-vehicle time and the number of transfers was included to account for the potential correlation between these variables. As presented in Appendix A.1, the inclusion of the interaction term did not substantially improve model performance. Most interaction effects were statistically insignificant, and the estimated coefficients for in-vehicle time remained largely positive, suggesting that this approach did not fully resolve the underlying issue.
- Another alternative specification involved applying a transfer penalty to in-vehicle time, capturing the disutility of making transfers without introducing transfers as a separate variable. These results are presented in Appendix A.2. However, determining an appropriate penalty value proved challenging, as no single value significantly improved model fit. While the in-vehicle time coefficient became negative under some penalty values, the coefficient for wait time remained counterintuitive in certain cases. This suggests that additional underlying factors or data limitations may be influencing the results.

These experiments confirmed that strong attribute correlation posed a fundamental modeling challenge that could not be easily resolved through re-specification.

5.3.1. Domination

Further investigation revealed that the constructed choice set contained dominated alternatives: routes that were consistently worse across all attributes but were still frequently chosen (Table 5.3). This is a problem in discrete choice models since they operate under the premise that individuals are utility maximizers, selecting the option that provides the highest perceived benefit. However, as seen in Table 5.3, the second route, OR > GR, is the most frequently chosen among the routes with only one transfer, despite performing worst across all considered attributes. This inconsistency suggests that factors beyond the observed attributes may be influencing travelers' choices, leading to counterintuitive positive estimates for some coefficients. The presence of dominated alternatives distorts the estimated utility functions, making it difficult to derive meaningful behavioral insights from the model.

To partially address domination, a simplified choice set was constructed by merging routes with the same number of transfers (Appendix A.3). While this approach reduced some inconsistencies, the strong negative correlation between in-vehicle time and transfers persisted. Despite this, the simplified

Table 5.3: Choice set evening rush post-disruption

Route	Transfers	In vehicle time (in seconds)	Wait seconds (in seconds)	Max person per car	Observed count
BL > GR	1	2317	584	9.2	36
OR > GR	1	2334	713	12.3	70
SV > GR	1	2313	477	6.8	16
BL > RD > GR	2	1858	563	9.9	81
OR > RD > GR	2	1860	565	10.8	74
SV > RD > GR	2	1868	585	9.5	47

model showed some improvement: the wait time coefficients were no longer strongly positive and were either negative or statistically insignificant. This suggests that merging routes reduced distortions in model estimation. However, collapsing routes into just two categories limited the model's ability to capture more nuanced variations in route choice preferences, making it less sensitive to potential differences between individual routes. Although this approach provided useful insights into the role of route structure in model estimation, it did not fully resolve all issues, highlighting the need to consider additional factors that may influence route choice.

5.3.2. Other attributes

Beyond in-vehicle time and wait time, the possibility that habitual behavior influenced route choice was also explored. If travelers consistently selected the same route out of habit, this might help explain some of the observed deviations from utility-maximizing behavior. However, as shown in Figure 5.11, no strong patterns of habitual behavior were observed. Instead, many travelers frequently switched between different routes, suggesting that inertia was not a dominant factor affecting decisions. This finding supports the idea that other unobserved factors, such as perceived comfort or reliability, are more likely to explain the unexpected model results.

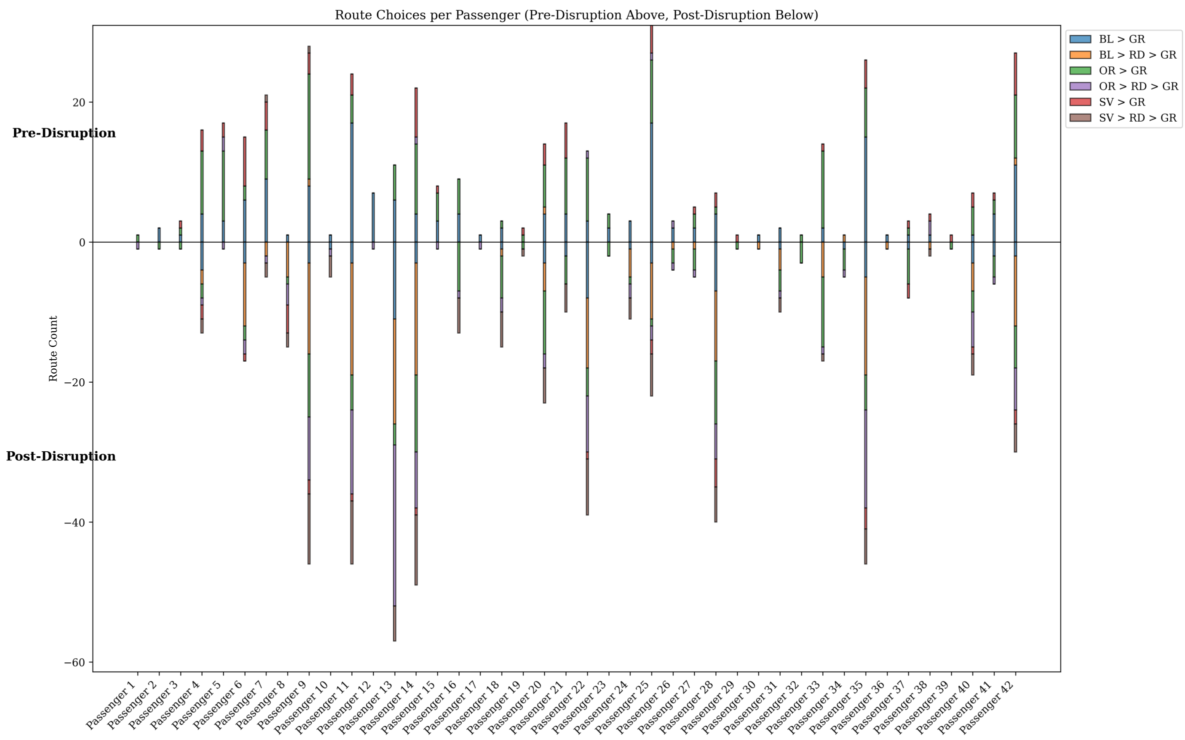


Figure 5.11: Route choices per person

5.4. Mixed Logit Model Results

To better understand the unexpected and inconsistent findings observed in the MNL model, a Mixed Logit (ML) model was estimated. Although this research does not primarily focus on individual-specific behavior, it is important to assess preference heterogeneity because variation across travelers could influence the overall patterns seen in the data. In particular, if passengers perceive attributes like in-vehicle time and wait time differently, then averaging their behavior in a simple MNL model could produce misleading or counterintuitive results. Estimating a Mixed Logit model allows us to test whether the inconsistencies found earlier could be explained by underlying differences in how travelers respond to key attributes.

The overall estimation results of the ML model are shown in Table 5.4, and the estimated parameters are shown in Table 5.5.

Table 5.4: Estimation Report Mixed Logit

Statistic	Value
Number of estimated parameters	20
Number of respondents	42
Sample size	1577
Initial log-likelihood	-2784.947
Final log-likelihood	-2393.276
Rho-square (initial model)	0.141
Rho-square-bar (initial model)	0.133
Akaike Information Criterion (AIC)	4826.552
Bayesian Information Criterion (BIC)	4933.832

Table 5.5: Estimated Parameters Mixed Logit

Parameter	Value	Std Err	z-test	p-value
VEH_SEC_PRE	0.670900	0.125000	5.386000	0.000000
VEH_SEC_POST_TOTAL	-0.025700	0.076000	-0.340000	0.734000
VEH_SEC_POST_1	0.201400	0.088000	2.284000	0.022000
VEH_SEC_POST_2	-0.272600	0.146000	-1.862000	0.063000
VEH_SEC_POST_3	-0.161400	0.087000	-1.859000	0.063000
WAIT_SEC_PRE	0.107100	0.080000	1.345000	0.179000
WAIT_SEC_POST_TOTAL	0.163200	0.221000	0.739000	0.460000
WAIT_SEC_POST_1	0.140800	0.257000	0.547000	0.584000
WAIT_SEC_POST_2	0.009600	0.184000	0.052000	0.958000
WAIT_SEC_POST_3	-0.053000	0.233000	-0.228000	0.820000
Sigma VEH_SEC_PRE	-0.067700	0.326000	-0.207000	0.836000
Sigma VEH_SEC_POST_TOTAL	0.073600	0.043000	1.699000	0.089000
Sigma VEH_SEC_POST_1	0.110700	0.093000	1.195000	0.232000
Sigma VEH_SEC_POST_2	-0.151200	0.171000	-0.883000	0.377000
Sigma VEH_SEC_POST_3	0.125600	0.136000	0.923000	0.356000
Sigma WAIT_SEC_PRE	0.146200	0.179000	0.819000	0.413000
Sigma WAIT_SEC_POST_TOTAL	-0.135200	0.120000	-1.124000	0.261000
Sigma WAIT_SEC_POST_1	0.254600	0.266000	0.959000	0.338000
Sigma WAIT_SEC_POST_2	-0.019500	0.537000	-0.036000	0.971000
Sigma WAIT_SEC_POST_3	0.025300	0.350000	0.072000	0.942000

Based on the ML results, none of the sigma coefficients are statistically significant. This indicates that there is no strong evidence of preference heterogeneity in the dataset: travelers appear to perceive in-vehicle time and wait time relatively similarly.

The estimated mean coefficients for in-vehicle time still vary across periods, with some unexpected positive values—particularly in the pre-disruption and early post-disruption periods. Since no significant heterogeneity is detected, these unexpected results are unlikely to be explained by differences in individual sensitivities.

In addition to exploring preference heterogeneity, a Path Size Logit (PSL) model was also considered to address the issue of overlapping routes. The idea was to account for the similarity between alternatives by introducing a path size factor.[1] However, because the path size factors offered no meaningful differentiation between the routes, PSL modeling was ultimately not pursued.

5.5. Comparison MNL and ML model

To assess whether the ML model provides a significantly better fit than the MNL model, a Likelihood Ratio (LR) test was conducted. This test compares the log-likelihood values of the two models, where the MNL model serves as the restricted model (L_0), and the ML model serves as the unrestricted model (L_1). The LR test statistic is computed as follows:

$$\text{LR} = -2 \times (L_0 - L_1) = -2 \times (-2393.276 + 2416.265) = 45.978 \tag{5.1}$$

The LR statistic follows a chi-square (χ^2) distribution, which is used to test whether a more complex model provides a significantly better fit than a simpler one. The number of degrees of freedom in this test are equal to the number of additional parameters estimated in the ML model compared to the MNL model. In this case, the ML model includes 10 extra parameters, which account for the estimated standard deviations (σ values) of the random coefficients. To determine whether the improvement in model fit is statistically significant, we compare the LR statistic to a critical chi-square value. For 10 degrees of freedom and a significance level of $\alpha = 0.05$, the critical chi-square value is 18.31. If the LR statistic is greater than this value, we can reject the null hypothesis (H_0) that the MNL model is sufficient and we can conclude that the ML model provides a significantly better fit than the MNL model.

However, despite the statistical improvement in model fit, the AIC (Akaike Information Criterion) and BIC (Bayesian Information Criterion) values suggest that the ML model does not substantially enhance model performance. The AIC for the MNL model is 4852, compared to 4826 for the ML model, showing only a marginal improvement. Meanwhile, the BIC for the MNL model is 4906, while the BIC for the ML model is 4933, which is actually slightly worse. Since BIC penalizes additional parameters more heavily than AIC, the higher BIC value for the ML model suggests that the added complexity does not justify the improvement in model fit.

Furthermore, the ML model still produces unexpected coefficient signs. While the ML model accounts for heterogeneity, the lack of substantial preference variation across individuals, combined with these inconsistencies in coefficient signs, suggests that it may not be the most appropriate model for this dataset.

Given the minimal improvements in log-likelihood, AIC, and BIC, the Mixed Logit model does not appear to offer significant advantages over the simpler MNL model. While estimating the ML model was important to check whether hidden differences in passenger preferences could explain the unexpected findings in the MNL model, no significant preference heterogeneity was detected. This suggests that the unexpected results are more likely due to structural limitations in the dataset rather than variation across travelers. Therefore, although the statistical test confirms a better fit, the practical implications of the added complexity remain limited.

[1] The path size factors, calculated based on both overlapping links and overlapping stations, were found to be nearly identical across all routes. This indicated that the available alternatives were not meaningfully distinguishable based on overlap. As a result, the Path Size Logit model was not estimated. A full explanation of the path size factor computation is provided in Appendix A.4.

6 Discussion and Conclusion

This study examined how public transport users' route choice preferences changed in response to a planned disruption, using AFC data from the Washington DC metro system. The research aimed to address gaps in the literature by analyzing whether passengers' preferences for route attributes shifted after the disruption and whether these changes persisted over time. In addition, the study assessed whether AFC data from WMATA alone is sufficient for analyzing route choice behavior during disruptions, or whether additional data sources are needed to fully capture passenger decision-making.

The descriptive analysis showed that travel times and wait times remained relatively stable across different periods. However, despite this stability in measured attributes, passengers' route choices changed significantly. Before the disruption, travelers tended to prefer direct routes, even when they had longer travel times. After the disruption, passengers increasingly chose routes with more transfers but shorter in-vehicle times. This shift suggests a change in route choice preferences, with travelers placing greater importance on minimizing in-vehicle time and showing a greater willingness to tolerate transfers.

To further investigate these changes, discrete choice models were estimated. However, the modeling results revealed unexpected findings: some models produced positive coefficients for travel time and wait time, contrary to theoretical expectations. These counterintuitive results are likely explained by the strong negative correlation between in-vehicle time and the number of transfers, which made it difficult to separately identify their effects. Additionally, the presence of dominated alternatives — routes that were worse across all measured attributes but were still chosen — complicated the estimation process, indicating that unobserved factors also played a role in passenger decisions.

Beyond the basic Multinomial Logit (MNL) model, a Mixed Logit (ML) model was estimated to account for individual differences in preferences. While the ML model provided a statistically better fit than the MNL model, practical improvements were limited, and the problem of counterintuitive coefficient signs remained. Similarly, a Path Size Logit (PSL) model was tested to account for route overlap, but the path size factors were nearly identical across routes, meaning that overlap did not meaningfully differentiate alternatives.

While these modeling and data challenges complicate the interpretation of individual coefficients, the broader patterns observed still provide valuable insights into how passenger behavior and preferences evolved in response to the disruption. To structure the interpretation of the findings, the next sections reflect on the research sub-questions.

6.1. Reflection on research questions

To answer the main question — How do public transport passengers' route choice preferences change in response to a planned public transport disruption — four sub-questions were constructed. These covered which disruptions were suitable for analysis, what factors influenced route choices, how preferences changed over time, and whether AFC data was suitable for studying this kind of behavior.

1. Which disruptions can be used to analyze changes in route choice preferences?

Finding a suitable disruption to analyze wasn't straightforward. Many of the disruptions in the dataset were too short, didn't offer clear alternative routes, or overlapped with other events, making it hard to isolate their effects. After going through all options, one planned disruption was selected that clearly affected multiple routes between a key OD pair and didn't overlap with other major changes. This was the disruption occurring consistently along the Green or Red Line between Gallery Place and Fort Totten lasting from 21/06/2020 to 25/07/2020. Even though this meant focusing on just one case, it provided a clear and useful opportunity to look at how passengers adjusted their behavior.

This also shows one of the limitations of using AFC data for this kind of analysis, which is that there just aren't that many disruptions that meet all the necessary conditions. Still, the selected disruption offered a solid basis for exploring how route choice preferences can change in response to a disruption.

2. What are the main factors influencing passengers' route choices in response to planned public transport disruptions?

This study focused on in-vehicle time, the number of transfers, and wait time as key factors that influence route choices in the Washington DC metro network. However, a strong negative correlation between in-vehicle time and the number of transfers made it difficult to estimate their individual effects in the model. In this network, routes with more transfers often had shorter in-vehicle times, which created multicollinearity. Because of that, transfers were not included directly in the model, and only in-vehicle time was used. This likely distorted the results, since the model may have picked up the effect of transfers through the travel time coefficient.

Before the disruption, passengers mostly chose direct routes, even when those had longer travel times. After the disruption, many shifted to transfer-heavy routes with shorter in-vehicle times, suggesting that travel time became more important. However, due to the correlation between transfers and travel time, it is difficult to determine whether this shift was driven primarily by a desire to reduce travel time, by a greater willingness to transfer, or both.

Wait time was also included in the model, but its influence proved difficult to interpret. In some cases, wait time coefficients were unexpectedly positive or not statistically significant. This may be due to the presence of dominated alternatives in the choice set or to missing or unreliable data for certain time periods.

Other potentially important factors like delay and crowding could not be included because the available variables were either inaccurate or incomplete. And as with most AFC-based studies, attributes like comfort, perceived reliability, or familiarity were not available at all.

In conclusion, while the descriptive analysis shows that route choice preferences clearly changed after the disruption, the model could not identify a single main factor driving this change. Due to the correlation between in-vehicle time and transfers, it is not possible to say whether passengers were responding more to time savings or to changes in transfer behavior. The influence of wait time is also unclear. These limitations highlight the challenges of modeling route choice behavior with AFC data alone.

3. How do route choice preferences evolve over time during the post-disruption period?

The analysis indicates that route choices shifted in response to the disruption and that these changes persisted afterward. In the later post-disruption periods, many passengers continued to choose faster, multi-transfer routes instead of returning to the more direct options they had used before. This suggests that passengers did not simply revert to their previous habits once the disruption ended.

Even though the estimated coefficients for in-vehicle time had unexpected signs, there was still a notable change: the coefficients shifted from positive values in the pre-disruption period to negative values in the post-disruption periods. This indicates that passengers' preferences changed following the disruption, even though it remains unclear whether this was due to an increased sensitivity to in-vehicle time, a reduced reluctance to transfer, or a combination of both. While the strong correlation between in-vehicle time and the number of transfers complicates interpretation, it can be concluded that the disruption triggered a lasting change in route choice preferences.

This is an interesting finding because a systematic review by Noureldin and Diab Noureldin and Diab, 2024 showed that most existing studies on disruptions focus on mode changes and that much less is known about changes in route choice. This study therefore adds new evidence that disruptions can also lead to lasting changes in passengers' route choice preferences.

4. How suitable is revealed preference data for analyzing changes in route choice preferences in response to a planned public transport disruption?

The use of AFC data in this study proved very suitable for detecting changes in route choice behavior in response to the planned disruption. It allowed for tracking real passenger behavior over a long period and provided insights into how travel patterns shifted after the disruption. These are things that would be difficult to achieve with stated preference (SP) data alone. However, while AFC data was effective in capturing that a change occurred, it proved more difficult to identify exactly which factors were driving these changes.

One major limitation was that AFC data does not include perceptual factors like comfort, reliability, or familiarity with a route, which likely influence decision-making but are not directly observable. This issue became apparent when some passengers consistently selected dominated alternatives, suggesting that factors outside the available travel attributes affected their choices. Another difficulty was the strong negative correlation between in-vehicle time and the number of transfers, which made it hard to separate the effects of individual route attributes in the discrete choice models.

These limitations show that while AFC data is useful for capturing real-world changes in behavior, it can be difficult when the aim is to understand the underlying preferences that drive these changes. In future studies, combining AFC data with additional perception-based data might offer a way to better explain why passengers make certain route choices, especially in complex networks where unobserved factors are likely to play a role.

Nevertheless, AFC data can be useful for analyzing route choice preferences. For example, Yap et al. Yap et al., 2020 successfully estimated a discrete choice model using only AFC data to analyze the impact of crowding on route choices. This shows that AFC data can be sufficient to estimate passenger preferences, provided that the key explanatory variables are well captured and there is enough variation between alternatives. Therefore, while this study encountered limitations, the findings do not imply that AFC data is unsuitable by itself. Instead, researchers should carefully assess the structure of the network, the availability of relevant variables, and the potential for multicollinearity when determining whether AFC data alone is sufficient for their analysis.

6.2. Scientific and Societal Contributions

This study contributes to both the academic understanding of transit behavior and the practical challenges faced by transport agencies in managing disruptions.

Scientific contributions

From a scientific perspective, this research adds to the body of literature that uses revealed preference (RP) data to study passenger behavior in the context of disruptions. While earlier studies often relied on stated preference (SP) surveys or limited RP datasets, this study shows that smart card data alone can capture real behavioral changes over time, offering a view of how route choice preferences can shift in response to network disruptions.

At the same time, the study highlights important limitations of using AFC data in isolation. Although large-scale behavioral shifts were clearly observable, it proved difficult to pinpoint the exact factors driving these changes. This was mainly due to strong correlations between key route attributes, limited variation between alternatives, and missing perceptual factors like comfort or reliability.

As such, the study contributes to the scientific literature by emphasizing that researchers who want to use AFC data alone must carefully check for correlation between variables and ensure that enough meaningful and independent variables are available to capture travelers' decision-making processes accurately. Without this, it becomes very difficult to reliably interpret changes in route choice preferences.

Overall, this study provides methodological insights for future research on transit behavior during disruptions, showing both the potential and the limitations of relying exclusively on smart card data.

Societal contributions

From a societal and policy perspective, the findings of this study are valuable for transit agencies planning and managing service disruptions. The analysis shows that route choice preferences are not fixed: disruptions can trigger changes that persist well after normal service is restored. This highlights that planned disruptions can be used not only to manage temporary changes, but also to reshape travel patterns in lasting ways.

The results suggest that passengers may have become more sensitive to travel time after the disruption, although the exact drivers behind the change in preferences remain uncertain. If greater sensitivity to in-vehicle time is indeed part of the behavioral shift, it is important for agencies to ensure that fast travel options remain available after the disruption ends. If post-disruption services offer slower routes or longer travel times, there is a risk that more time-sensitive passengers will become dissatisfied and may ultimately abandon public transport altogether. Maintaining attractive, time-efficient options in the recovery phase would then be critical to retain riders and supporting long-term satisfaction.

However, the study also shows that without a better understanding of what drives these changes, it is difficult to translate insights into concrete policy actions. Therefore, future efforts should aim to integrate passenger feedback or perception data to support more targeted and effective interventions.

6.3. Limitations and Future Research

While this study provides valuable insights into route choice behavior in response to a planned disruption, there are several limitations that should be acknowledged. These limitations also point toward areas where future research can build on and improve the current approach.

- **Data limitations:** The study relied solely on AFC data, which, although rich in actual behavior, lacks information on perceptual factors such as comfort, reliability, or familiarity. These unobserved factors likely influenced passengers' choices, particularly since dominated alternatives were selected.
- **Multicollinearity between route attributes:** A strong negative correlation between in-vehicle time and the number of transfers made it difficult to disentangle the individual effects of each attribute. As a result, the interpretation of the discrete choice models was complicated, limiting the ability to identify the precise drivers behind changes in route choice preferences.
- **Limited generalizability:** This study focused on a single disruption event within one transit system. While the selected disruption provided a clear and valuable case study, the findings may not generalize to other types of disruptions (e.g., unplanned disruptions) or to transit networks with different structures, frequencies, or passenger demographics.
- **Limited variation between alternatives:** Although multiple routes were technically available between the selected OD pair, the high degree of overlap meant that in practice only two truly distinct alternatives existed. Most routes shared large sections of track and had similar travel times and transfer patterns. While this did not prevent estimation, having a third distinctly different route would have allowed for a richer analysis of how passengers trade off different route attributes. The limited variation made it more difficult to observe clear differences in route choice preferences, especially when trying to disentangle the effects of travel time and transfers.

Future research

Based on the limitations identified in this study, several directions for future research can be proposed.

First, future studies should consider combining AFC data with additional data sources, such as stated preference (SP) surveys or qualitative passenger feedback. This would allow researchers to capture important unobserved factors like perceived comfort, reliability, and familiarity, which are not available in AFC data but are likely to influence route choice decisions.

Second, it would be valuable to apply a similar analysis to other public transport networks. Conducting studies in networks with longer routes, greater variation between alternatives, and different structural characteristics—such as suburban rail or regional train systems—could provide further insights into how disruption-induced changes in preferences occur and whether the findings from this study are generalizable.

Once the analysis approach is well-established, an interesting extension would be to investigate whether the duration of a disruption affects the persistence of preference changes. Analyzing networks with multiple disruptions of varying lengths could help assess whether longer disruptions reinforce behavioral changes more strongly than shorter ones.

6.4. Conclusion

This study set out to answer the question: How do public transport passengers' route choice preferences change in response to a planned public transport disruption?

The findings show that planned disruptions can lead to lasting changes in route choice preferences. In the case of the Washington DC metro network, a clear shift was observed. Before the disruption, passengers often chose direct routes, even if these involved longer in-vehicle times. After the disruption had ended and normal service was restored, many travelers chose faster, transfer-heavy routes instead of returning to the more direct options they had previously used. This suggests that passengers reassessed their travel options following the disruption and adjusted their preferences accordingly.

Although the discrete choice models faced challenges in isolating the exact factors driving these changes, largely due to the strong correlation between in-vehicle time and transfers, the broader behavioral shift is clear. Even though some coefficients had unexpected signs, the fact that the coefficients for in-vehicle time shifted from positive before the disruption to negative afterward indicates that passengers' preferences changed. Whether this change was due to a greater sensitivity to travel time, a reduced reluctance to transfer, or a combination of both cannot be fully determined. However, it can be concluded that the disruption lead to a change in route choice preferences.

The study also highlights that while AFC data is very useful for observing real-world behavioral shifts, it has important limitations when used alone. The absence of perceptual factors such as comfort and reliability made it difficult to fully explain why some passengers made seemingly suboptimal choices. Future studies could strengthen this type of analysis by combining AFC data with additional information about travelers' perceptions and motivations.

Overall, this study shows that route choice preferences are adaptable and can change in response to disruptions. A planned disruption prompted many passengers to reconsider their travel options and adopt different routes, and these new patterns persisted after normal service was restored. This highlights that disruptions should not only be seen as temporary challenges but also as moments where long-term changes in travel behavior can occur. For transit agencies, this means that it is important to anticipate how disruptions may shift passenger preferences, and to ensure that fast, convenient travel options remain available to support satisfaction and retention after the disruption ends.

Although this research focused on a single disruption in one metro network, it provides a foundation for broader investigations. Future research should examine how different types of disruptions, variations in network structure, and differences in disruption duration influence the extent and persistence of changes in route choice preferences. Expanding this analysis across multiple cases would help to better understand how planned disruptions shape passenger preferences and how transit agencies can respond to these changes.

References

Adelé, S., Tréfond-Alexandre, S., Dionisio, C., & Hoyau, P.-A. (2019). Exploring the behavior of suburban train users in the event of disruptions. *Transportation Research Part F: Traffic Psychology and Behaviour*, *65*, 344–362. https://doi.org/10.1016/j.trf.2019.08.009

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, *19*(6), 716–723. https://doi.org/10.1109/TAC.1974.1100705

Algers, S., Bergström, P., Dahlberg, M., & Lindqvist Dillén, J. (1998). *Mixed logit estimation of the value of travel time* (tech. rep. No. 15). Uppsala University. https://www.econstor.eu/handle/10419/82839

Arslan Asim, M., Weiss, A., Kattan, L., & Wirasinghe, S. C. (2021). Transit Users' Mode Choice Behavior During Light Rail Transit Short-Term Planned Service Disruption. *Transportation Research Record*, *2675*(10), 711–722. https://doi.org/10.1177/03611981211012421

Berggren, U., D'Agostino, C., Svensson, H., & Brundell-Freij, K. (2022). Intrapersonal variability in public transport path choice due to changes in service reliability. *Transportation*, *49*(6), 1517–1547. https://doi.org/10.1007/s11116-021-10218-z

Bernal, M., Welch, E. W., & Sriraj, P. S. (2016). The effect of slow zones on ridership: An analysis of the Chicago Transit Authority "El" Blue Line. *Transportation Research Part A: Policy and Practice*, *87*, 11–21. https://doi.org/10.1016/j.tra.2016.02.007

Chorus, C. (2019). Statistical analysis of choice behaviour: Mixed Logit.

Deng, Y., Liu, W., Zhang, Q., & Zhang, Z. (2022). Implementation and Application of a Stochastic Agent-based Model for Passenger Flow Distribution Predicting in Planned Metro Station Service Disruption Scenario. *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2252–2257. https://doi.org/10.1109/ITSC55140.2022.9922254

Dixit, M., Cats, O., van Oort, N., Brands, T., & Hoogendoorn, S. (2023). Validation of a multi-modal transit route choice model using smartcard data. *Transportation*, *51*, 1809–1829. https://doi.org/10.1007/s11116-023-10387-z

Duncan, L. C., Watling, D. P., Connors, R. D., Rasmussen, T. K., & Nielsen, O. A. (2020). Path size logit route choice models: Issues with current models, a new internally consistent approach, and parameter estimation on a large-scale network with gps data. *Transportation Research Part B: Methodological*, *135*, 1–40. https://doi.org/10.1016/j.trb.2020.02.006

Eltved, M., Breyer, N., Ingvardson, J., & Nielsen, O. (2021). Impacts of long-term service disruptions on passenger travel behaviour: A smart card analysis from the Greater Copenhagen area. *Transportation Research Part C: Emerging Technologies*, *131*. https://doi.org/10.1016/j.trc.2021.103198

Grison, E., Burkhardt, J.-M., & Gyselinck, V. (2017). How do users choose their routes in public transport? the effect of individual profile and contextual factors. *Transportation Research Part F: Traffic Psychology and Behaviour*, *51*, 24–37. https://doi.org/https://doi.org/10.1016/j.trf.2017.08.011

Ingvardson, J. B., Thorhauge, M., Nielsen, O. A., & Eltved, M. (2024). Public transport route choice modelling: Reducing estimation bias when using smart card data. *Transportation Research Part A: Policy and Practice*, *179*, 103929. https://doi.org/10.1016/j.tra.2023.103929

Jánošíková, L., Slavík, J., & Koháni, M. (2014). Estimation of a route choice model for urban public transport using smart card data. *Transportation Planning and Technology*, *37*(7), 638–648. https://doi.org/10.1080/03081060.2014.935570

King, G. (1989). *Unifying political methodology: The likelihood theory of statistical inference* [p. 84]. Cambridge University Press.

Koppelman, F. S., & Wen, C. (1998). Alternative nested logit models: Structure, properties and estimation. *Transportation Research Part B: Methodological*, *32*(5), 289–298. https://doi.org/10.1016/S0191-2615(98)00003-4

Krishnakumari, P., Cats, O., & van Lint, H. (2020). Estimation of metro network passenger delay from individual trajectories. *Transportation Research Part C: Emerging Technologies*, *117*. https://doi.org/10.1016/j.trc.2020.102704

Lako, A., & Gjevori, S. (2023). Urban public transport as a basis for sustainable mobility development in the transition from private vehicles to urban public transport: A case study of elbasan. *International Journal of Sustainable Development and Planning*, *18*(11), 3609–3615. https://doi.org/10.18280/ijsdp.181126

Lam, S. H., & Xie, F. (2002). Transit path-choice models that use revealed preference and stated preference data. *Transportation Research Record*, *1799*, 58–65. https://doi.org/10.3141/1799-08

Li, B., Yao, E., Yamamoto, T., Tang, Y., & Liu, S. (2020). Exploring behavioral heterogeneities of metro passenger's travel plan choice under unplanned service disruption with uncertainty. *Transportation Research Part A: Policy and Practice*, *141*, 294–306. https://doi.org/10.1016/j.tra.2020.09.009

Liu, T., Ma, Z., & Koutsopoulos, H. N. (2021). Unplanned Disruption Analysis in Urban Railway Systems Using Smart Card Data. *Urban Rail Transit*, *7*(3), 177–190. https://doi.org/10.1007/s40864-021-00150-x

Marra, A. D., & Corman, F. (2023). How different network disturbances affect route choice of public transport passengers. A descriptive study based on tracking. *Expert Systems with Applications*, *213*. https://doi.org/10.1016/j.eswa.2022.119083

McFadden, D. (1974). Conditional logit analysis of qualitative choice behavior. *Frontiers in econometrics*, 105–142.

McFadden, D. (1977). *Quantitative methods for analyzing travel behaviour of individuals: Some recent developments* (tech. rep. No. 707). Cowles Foundation Discussion Papers. https://elischolar.library.yale.edu/cowles-discussion-paper-series/707

McFadden, D. (1978). Modeling the choice of residential location. *Transportation Research Record*, (673). https://trid.trb.org/View/87722

Mepparambath, R. M., Soh, Y. S., Jayaraman, V., Tan, H. E., & Ramli, M. A. (2023). A novel modelling approach of integrated taxi and transit mode and route choice using city-scale emerging mobility data. *Transportation Research Part A: Policy and Practice*, *170*. https://doi.org/10.1016/j.tra.2023.103615

Mo, B., von Franque, M., Koutsopoulos, H. N., Attanucci, J., & Zhao, J. (2022b). Impact of unplanned long-term service disruptions on public transit systems. *IEEE Open Journal of Intelligent Transportation Systems*, *3*, 551–569. https://doi.org/10.1109/OJITS.2022.3199108

Mo, B., Koutsopoulos, H. N., & Zhao, J. (2022a). Inferring passenger responses to urban rail disruptions using smart card data: A probabilistic framework. *Transportation Research Part E: Logistics and Transportation Review*, *159*. https://doi.org/10.1016/j.tre.2022.102628

Nazem, M., Lomone, A., Chu, A., & Spurr, T. (2018). Analysis of travel pattern changes due to a medium-term disruption on public transit networks using smart card data. *Transportation Research Procedia*, *32*, 585–596. https://doi.org/10.1016/j.trpro.2018.10.019

Nielsen, O. A., Eltved, M., Anderson, M. K., & Prato, C. G. (2021). Relevance of detailed transfer attributes in large-scale multimodal route choice models for metropolitan public transport passengers. *Transportation Research Part A: Policy and Practice*, *147*, 76–92. https://doi.org/https://doi.org/10.1016/j.tra.2021.02.010

Noureldin, M., & Diab, E. (2024). Impacts of long-term transit system disruptions and transitional periods on travelers: A systematic review. *Discover Cities*, *1*. https://doi.org/10.1007/s44327-024-00015-5

Rahimi, E., Shamshiripour, A., Shabanpour, R., Mohammadian, A., & Auld, J. (2019). Analysis of transit users' waiting tolerance in response to unplanned service disruptions. *Transportation Research Part D: Transport and Environment*, *77*, 639–653. https://doi.org/10.1016/j.trd.2019.10.011

Rahimi, E., Shamshiripour, A., Shabanpour, R., Mohammadian, A., & Auld, J. (2020). Analysis of Transit Users' Response Behavior in Case of Unplanned Service Disruptions. *Transportation Research Record*, *2674*(3), 258–271. https://doi.org/10.1177/0361198120911921

Shires, J. D., Ojeda-Cabral, M., & Wardman, M. (2019). The impact of planned disruptions on rail passenger demand. *Transportation*, *46*(5), 1807–1837. https://doi.org/10.1007/s11116-018-9889-0

Soza-Parra, J., Raveau, S., Muñoz, J. C., & Cats, O. (2019). The underlying effect of public transport reliability on users' satisfaction. *Transportation Research Part A: Policy and Practice*, *126*, 83–93. https://doi.org/https://doi.org/10.1016/j.tra.2019.06.004

Stone, M. (1979). Comments on model selection criteria of akaike and schwarz. *Journal of the Royal Statistical Society. Series B (Methodological)*, 276–278.

Sun, H., Wu, J., Wu, L., Yan, X., & Gao, Z. (2016). Estimating the influence of common disruptions on urban rail transit networks. *Transportation Research Part A: Policy and Practice*, *94*, 62–75. https://doi.org/10.1016/j.tra.2016.09.006

Tavakoli, F. (2024, August). *Investigating the long-term impact of disruptions on passenger travel behavior using afc data* [Master's thesis]. Delft University of Technology. https://repository.tudelft.nl/record/uuid:57813e26-10a9-446e-a799-45fd7167b0a9

Tuan, V. A., Van Truong, N., Tetsuo, S., & An, N. N. (2022). Public transport service quality: Policy prioritization strategy in the importance-performance analysis and the three-factor theory frameworks. *Transportation Research Part A: Policy and Practice*, *166*, 118–134. https://doi.org/https://doi.org/10.1016/j.tra.2022.10.006

van Cranenburgh, S. (2021). Latent class discrete choice models for travel behavior research.

van der Hurk, E. (2015). *Passengers, Information, and Disruptions*. https://repub.eur.nl/pub/78275/

Wang, Y., Zhao, O., & Zhang, L. (2024). Modeling urban rail transit system resilience under natural disasters: A two-layer network framework based on link flow. *Reliability Engineering & System Safety*, *241*. https://doi.org/10.1016/j.ress.2023.109619

Washington Metropolitan Area Transit Authority. (2024). Metrorail service information. https://www.wmata.com/service/rail/

Yao, R., & Bekhor, S. (2020). Data-driven choice set generation and estimation of route choice models. *Transportation Research Part C: Emerging Technologies*, *121*. https://doi.org/https://doi.org/10.1016/j.trc.2020.102832

Yap, M., & Cats, O. (2022). Analysis and prediction of ridership impacts during planned public transport disruptions. *Journal of Public Transportation*, *24*. https://doi.org/10.1016/j.jpubtr.2022.100036

Yap, M., & Cats, O. (2021a). Predicting disruptions and their passenger delay impacts for public transport stops. *Transportation*, *48*(4), 1703–1731. https://doi.org/10.1007/s11116-020-10109-9

Yap, M., & Cats, O. (2021b). Taking the path less travelled: Valuation of denied boarding in crowded public transport systems. *Transportation Research Part A: Policy and Practice*, *147*, 1–13. https://doi.org/10.1016/j.tra.2021.02.007

Yap, M., Nijënstein, S., & van Oort, N. (2018). Improving predictions of public transport usage during disturbances based on smart card data. *Transport Policy*, *61*, 84–95. https://doi.org/10.1016/j.tranpol.2017.10.010

Yap, M., Cats, O., & van Arem, B. (2020). Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A: Transport Science*, *16*(1), 23–42. https://doi.org/10.1080/23249935.2018.1537319

Zhu, S., Masud, H., Xiong, C., Yang, Z., Pan, Y., & Zhang, L. (2017). Travel Behavior Reactions to Transit Service Disruptions: Study of Metro SafeTrack Projects in Washington, D.C. *Transportation Research Record*, *2649*(1), 79–88. https://doi.org/10.3141/2649-09

A Results

This appendix presents the methods that were explored during the research but were not directly included in the final results. These approaches were tested to assess their potential contribution to the analysis, but for various reasons they were ultimately not incorporated into the main findings. Documenting these methods provides insight into the decision-making process behind the chosen approach and highlights alternative strategies that were considered along the way.

A.1. Interaction effect

To address the multicollinearity between the number of transfers and in-vehicle time, an interaction term between transfers and in-vehicle time was introduced instead of removing transfers altogether. Table A.1 presents the overall model fit statistics, and Table A.2 shows the estimated coefficients. The model achieves a Rho-square of 0.123, which is comparable to earlier models without the interaction term.

Table A.1: Model Estimation Results MNL with interaction effect

Statistic	Value
Number of estimated parameters	20
Sample size	1577
Null log-likelihood	-2728.461
Final log-likelihood	-2392.240
Rho-square	0.123
Rho-square-bar	0.116
Akaike Information Criterion (AIC)	4824.480
Bayesian Information Criterion (BIC)	4931.746

From the results it becomes evident that most of the estimated interaction effects between transfers and in-vehicle time are not statistically significant, as indicated by the high p-values. This suggests that including the interaction term does not substantially improve the model's explanatory power regarding the relationship between transfers and in-vehicle time. The estimated coefficients for in-vehicle time and transfers separately also vary across periods, with some remaining positive. Also the betas for transfers that are significant are all highly positive. Although the interaction was intended to capture potential correlation effects, the results indicate that the unexpected signs for in-vehicle time and transfers cannot be explained by this interaction.

Thus, while introducing the interaction effect was a theoretically valid approach to addressing multicollinearity, it did not significantly resolve the observed counterintuitive estimation results.

A.2. Transfer penalty

The unexpected positive sign for in-vehicle time in some periods suggests that the effect of transfers may have been absorbed into the in-vehicle time coefficient. To address this, an adjusted in-vehicle

Table A.2: Estimated Parameters MNL with interaction effect

Parameter	Value	Rob. Std Err	Rob. t-test	Rob. p-value
BETA_INT_VEH_TRANS_POST_1	0.000208	0.001142	0.181738	0.855788
BETA_INT_VEH_TRANS_POST_2	0.008979	0.369694	0.024288	0.980623
BETA_INT_VEH_TRANS_POST_3	-0.000498	0.000794	-0.627144	0.530565
BETA_INT_VEH_TRANS_POST_TOTAL	-0.000527	0.000576	-0.913993	0.360721
BETA_INT_VEH_TRANS_PRE	-0.278837	0.279812	-0.996515	0.319000
BETA_TRANSFERS_POST_1	4.330822	1.702636	2.543598	0.010972
BETA_TRANSFERS_POST_2	1.245630	13.959036	0.089235	0.928895
BETA_TRANSFERS_POST_3	1.595732	1.604861	0.994312	0.320071
BETA_TRANSFERS_POST_TOTAL	2.283053	0.952007	2.398148	0.016478
BETA_TRANSFERS_PRE	6.302853	9.253691	0.681118	0.495797
BETA_VEH_POST_1	0.724433	0.219702	3.297343	0.000976
BETA_VEH_POST_2	-0.062851	0.707560	-0.088827	0.929219
BETA_VEH_POST_3	0.030441	0.199578	0.152526	0.878772
BETA_VEH_POST_TOTAL	0.234299	0.120195	1.949316	0.051258
BETA_VEH_PRE	0.432921	0.380031	1.139174	0.254631
BETA_WAIT_POST_1	0.030753	0.041850	0.734834	0.462440
BETA_WAIT_POST_2	-0.007315	0.055746	-0.131227	0.895595
BETA_WAIT_POST_3	-0.039767	0.035877	-1.108427	0.267678
BETA_WAIT_POST_TOTAL	0.096979	0.018115	5.353441	0.000000
BETA_WAIT_PRE	0.074348	0.024783	2.999916	0.002701

time variable was introduced, incorporating a transfer penalty to account for the disutility of making transfers. This approach allows the model to reflect the inconvenience of transfers while avoiding the multicollinearity issues that arise when including transfers as a separate variable. The goal was to better capture travelers' true preferences and improve model interpretability.

Various transfer penalty values were tested separately for the pre- and post-disruption periods, with the corresponding final log-likelihood values presented in Table A.3.

Table A.3: Final log-likelihood values for different transfer penalties

Transfer penalty time frame 1	Transfer penalty time frame 2	Final LL
X	X	-2416.266
300	300	-2640.258
600	600	-2480.627
300	600	-2672.960
600	300	-2447.925
700	300	-2442.287
600	200	-2430.973
700	200	-2425.335
700	100	-2418.795
800	200	-2422.512
800	100	-2415.972
800	0	-2413.162
900	0	-2411.533
1500	0	-2408.456

While the introduction of a transfer penalty did not significantly improve the model fit in terms of log-likelihood, it did help correct the expected negative sign for in-vehicle time, making the model more behaviorally plausible. Based on these results, a penalty range of 800–900 seconds (13–15 minutes) for pre-disruption and 100–200 seconds (1.5–3 minutes) for post-disruption was selected, as this ensured that the in-vehicle time coefficient remained negative while reflecting the perceived inconvenience of transfers. The estimated parameters for a transfer penalty of 800 seconds (pre-disruption) and 100 seconds (post-disruption) are shown in Table A.4.

Table A.4: Estimated Parameters with Robust Standard Errors

Parameter	Value	Rob. Std Err	Rob. t-test	Rob. p-value
BETA_VEH_POST_1	-0.011929	0.001724	-6.920923	0.000000
BETA_VEH_POST_2	0.021448	0.002317	9.258074	0.000000
BETA_VEH_POST_3	0.013837	0.002145	6.450786	0.000000
BETA_VEH_POST_TOTAL	0.004997	0.000948	5.273258	0.000000
BETA_VEH_PRE	-0.004186	0.000344	-12.181111	0.000000
BETA_WAIT_POST_1	0.215022	0.015689	13.705272	0.000000
BETA_WAIT_POST_2	-0.006862	0.054128	-0.126767	0.899125
BETA_WAIT_POST_3	-0.034881	0.034631	-1.007223	0.313827
BETA_WAIT_POST_TOTAL	0.093456	0.017592	5.312464	0.000000
BETA_WAIT_PRE	0.079289	0.022605	3.507586	0.000452

As shown in the results, the in-vehicle time coefficient is now negative in most periods, except for Post-Disruption Period 1. Further refinements, such as applying different penalties for each post-disruption period, could potentially resolve this issue. However, determining an appropriate penalty remains challenging, as there is no clear theoretical basis for selecting an optimal value. Additionally, while the sign of in-vehicle time improved, the wait time coefficients remain counterintuitive. This suggests that the transfer penalty alone does not fully resolve the issue, and other unobserved factors or data inconsistencies may still be influencing the model results.

A.3. MNL 2 routes

Since the routes with the same number of transfers contained dominated alternatives, a modified approach was tested in which routes with an equal number of transfers were merged into two broad categories: routes with one transfer and routes with two transfers. This adjustment aimed to simplify the choice set and reduce inconsistencies caused by dominated alternatives.

A.3.1. Correlation

Before re-estimating the model, we first examined whether the strong negative correlation between in-vehicle time and number of transfers persisted after merging the routes. Figure A.1 presents correlation heatmaps across different time periods.

The results confirm that even after merging routes, a strong negative correlation remains between in-vehicle time and transfers.

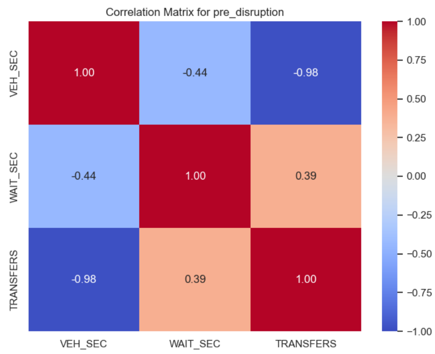
A.3.2. MNL Results

Given the persistent correlation, we estimated a simplified Multinomial Logit (MNL) model using only in-vehicle time and wait time as explanatory variables. The results are summarized in Tables A.5 and A.6.

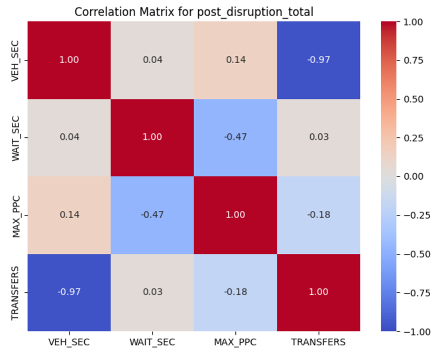
Table A.5: Estimation Report combined routes

Statistic	Value
Number of estimated parameters	10
Sample size	1562
Initial log-likelihood	-1082.326
Final log-likelihood	-754.296
Rho-square	0.303
Rho-square-bar	0.294
Akaike Information Criterion (AIC)	1528.593
Bayesian Information Criterion (BIC)	1582.130

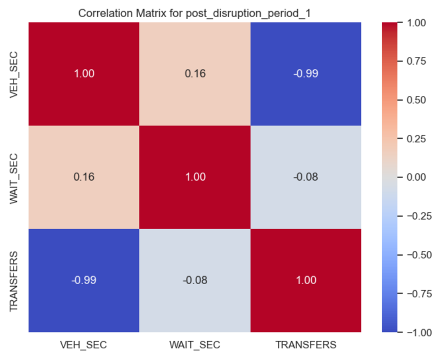
The model fit of this MNL model is relatively strong, with a rho-square value of 0.303, indicating reasonable explanatory power despite the simplified choice set. Compared to the previous model with six distinct routes, the wait time coefficients are no longer strongly positive. In most cases, they are either



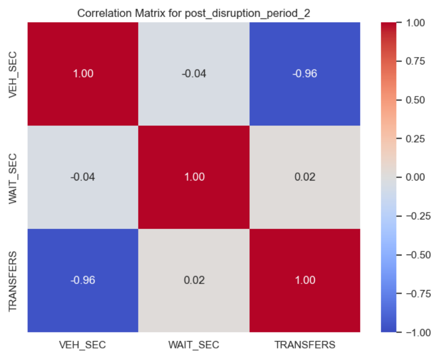
(a) Pre-Disruption



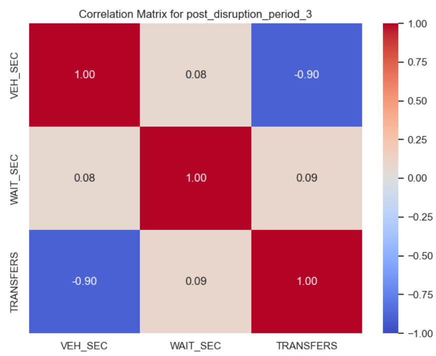
(b) Post-Disruption Total



(c) Post-Disruption Period 1



(d) Post-Disruption Period 2



(e) Post-Disruption Period 3

Figure A.1: Correlation Heatmaps for combined routes

negative or statistically insignificant, addressing one of the key issues observed earlier.

While merging routes helped resolve some inconsistencies, the approach also introduced limitations. By collapsing routes into just two categories, the model lost explanatory power, making it less capable of capturing route-specific variations that may still be relevant to travelers' decision-making. The strong negative correlation between in-vehicle time and transfers remained present, suggesting that structural factors in the dataset are driving these effects. Although the wait time coefficients now behave more consistently with theoretical expectations, the trade-off in losing route-level detail raises concerns about how well this simplified model represents actual route choice behavior.

Overall, while merging routes provided some improvements in interpretability and helped reduce distortions in coefficient estimation, it did not fully address all issues. The results suggest that additional

Table A.6: Estimated Parameters for Combined Routes

Parameter	Value	Rob. Std Err	Rob. t-test	Rob. p-value
BETA_VEH_POST_1	0.111033	0.018844	5.892172	0.000000
BETA_VEH_POST_2	-0.247088	0.027427	-9.008924	0.000000
BETA_VEH_POST_3	-0.159024	0.024281	-6.549355	0.000000
BETA_VEH_POST_TOTAL	-0.067005	0.010917	-6.137395	0.000000
BETA_VEH_PRE	0.684800	0.066580	10.285366	0.000000
BETA_WAIT_POST_1	-0.001726	0.057154	-0.030194	0.975912
BETA_WAIT_POST_2	-0.109021	0.084565	-1.289196	0.197330
BETA_WAIT_POST_3	0.056456	0.074234	0.760504	0.446953
BETA_WAIT_POST_TOTAL	0.010878	0.035602	0.305532	0.759961
BETA_WAIT_PRE	0.115290	0.076673	1.503653	0.132671

factors, may still be influencing route choice and warrant further investigation.

A.4. Path Size Logit

A Path Size Logit (PSL) model was also explored. The path size factor was determined using the equations outlined in the methodology. These path size terms quantify the distinctiveness of routes, penalizing routes that share a significant number of links with other alternatives. Since the goal was to assess the degree of overlap between links across different station pairs, the first step was to calculate the distances between stations. These distances were estimated using the station coordinates provided in the dataset. The resulting distances between stations are presented in Tables A.7 and A.8.

Table A.7: Distance between stations routes with 1 transfer

Station 1	Station 2	Distance (km)
Foggy Bottom-GWU	Farragut West	0.811
Farragut West	McPherson Sq	0.721
McPherson Sq	Metro Center	0.520
Metro Center	Federal Triangle	0.505
Federal Triangle	Smithsonian	0.636
Smithsonian	L'Enfant Plaza	0.642
L'Enfant Plaza	Archives	0.958
Archives	Gallery Place	0.488
Gallery Place	Mt Vernon Sq	0.910
Mt Vernon Sq	Shaw-Howard U	0.770
Shaw-Howard U	U St	0.613
U St	Columbia Heights	1.390
Columbia Heights	Georgia Ave-Petworth	1.168
Georgia Ave-Petworth	Fort Totten	2.551
Fort Totten	West Hyattsville	2.914
West Hyattsville	Hyattsville Crossing	1.590
Hyattsville Crossing	College Park-U of Md	2.830
College Park-U of Md	Greenbelt	3.873

Based on these distances, the path size factors for overlapping links were determined. These results are presented in Table A.9. As we can see, these values are very similar across all routes, indicating that the degree of overlap among the different paths is fairly uniform. This suggests that no particular route stands out as significantly more distinct or more overlapped than others within this set. The similarity in path size factors implies that the network structure leads to comparable levels of route sharing, meaning travelers choosing between these alternatives are exposed to similar levels of overlap in terms of shared links. Consequently, the impact of the path size factor on route choice behavior may be limited, as all routes are penalized to a similar extent for their shared links.

We also examined a path size factor based on overlapping stations to assess whether this approach

Table A.8: Distance between stations routes with 2 transfers

Station 1	Station 2	Distance (km)
Foggy Bottom-GWU	Farragut West	0.811
Farragut West	McPherson Sq	0.721
McPherson Sq	Metro Center	0.520
Metro Center	Gallery Place	0.810
Gallery Place	Judiciary Sq	0.480
Judiciary Sq	Union Station	0.892
Union Station	NoMa-Gallaudet U	0.977
NoMa-Gallaudet U	Rhode Island Ave	1.677
Rhode Island Ave	Brookland-CUA	1.472
Brookland-CUA	Fort Totten	2.149
Fort Totten	West Hyattsville	2.914
West Hyattsville	Hyattsville Crossing	1.590
Hyattsville Crossing	College Park-U of Md	2.830
College Park-U of Md	Greenbelt	3.873

Table A.9: Path Size Factors for Overlapping Links

Route	Path Size Factor
BL > GR	0.240833
OR > GR	0.240833
SV > GR	0.240833
BL > RD > GR	0.231573
OR > RD > GR	0.231573
SV > RD > GR	0.231573

might better capture route distinctiveness. The results, presented in Table A.10, show very similar values across all routes, much like the path size factors for overlapping links. This indicates that the extent to which routes share stations is relatively uniform, with no particular route standing out as significantly more or less overlapped than others.

Table A.10: Path Size Factors for Overlapping Station

Route	Path Size Factor
BL > GR	0.210526
OR > GR	0.210526
SV > GR	0.210526
BL > RD > GR	0.238596
OR > RD > GR	0.238596
SV > RD > GR	0.238596

Given the similarity in path size factors across routes, both for overlapping links and overlapping stations , the PSL model was ultimately not used. The lack of variation in path size factors suggests that the distinctiveness of routes is not well captured by these measures, as all routes receive nearly identical penalties. Since the primary purpose of incorporating a path size factor is to account for route overlap and correct for correlation among alternatives, its effectiveness is limited when all routes are penalized to a similar extent. As a result, adding a path size factor would not meaningfully improve model estimation or provide additional behavioral insights.

Analyzing route preferences before and after planned public transport disruptions

Karin van den Berg
Msc Student Transport, Infrastructure and Logistics
Delft University of Technology
4938933 - K.vandenberg-3@student.tudelft.nl

***Abstract*—Public transport disruptions can significantly affect passengers' travel behavior, yet little is known about how route choice preferences evolve after a disruption ends. This study investigates how passengers' route choices change in response to a planned disruption, using Automated Fare Collection (AFC) data from the Washington DC metro system. Unlike previous research that often relies on stated preference surveys or focuses only on the disruption period, this study analyzes several months of pre- and post-disruption behavior to assess whether changes in preferences persist over time.**

Descriptive analysis reveals that although travel times and wait times remained relatively stable, passengers shifted from preferring direct routes to favoring routes with more transfers but shorter in-vehicle times after the disruption. Discrete choice models, including Multinomial Logit and Mixed Logit models, were estimated to explore these shifts, but the results showed unexpected coefficient signs, likely due to strong multicollinearity between travel time and transfers, and the presence of dominated alternatives. While the Mixed Logit model improved the model fit slightly, practical interpretability remained limited.

The findings suggest that disruptions can lead to lasting behavioral changes, with passengers reassessing their travel options rather than returning to previous habits. Although AFC data is valuable for detecting such shifts, it alone is insufficient to fully capture the factors driving route choice behavior, highlighting the need for complementary perception-based data in future research. From a policy perspective, understanding these behavioral adaptations can help transit agencies design better service recovery strategies that sustain ridership after disruptions, supporting broader goals of promoting sustainable urban mobility.

I. Introduction

As cities around the world struggle with increasing traffic congestion and rising emissions, public transport plays a critical role in ensuring sustainable urban mobility (Lako and Gjevori, 2023). Efficient and reliable transit systems reduce dependence on private vehicles, which in turn helps to reduce overcrowding and lower carbon emissions. However, maintaining public transport as an attractive alternative to driving requires ensuring high levels of passenger satisfaction with service quality (Tuan et al., 2022).

A key factor influencing passenger satisfaction is reliability, as unpredictable service can discourage people from using public transport regularly (Soza-Parra et al., 2019). One of the main threats to reliability are disruptions, which can lead to increased travel times, missed connections, and overcrowding. These disruptions not only affect overall service quality but also influence passengers' transit route choices, forcing them to adapt their travel behavior in response to service changes due to the disruption.

Research has already been conducted on passengers' public transport route choices during disruptions, often using discrete choice models (Dixit et al., 2023). These models help reveal transit riders' preferences by assessing the relative importance of various travel attributes, offering policymakers valuable insights into how passengers make route choices in response to disruptions.

Many earlier studies on public transport route choice analysis were based on stated preference (SP) data (Shires et al., 2019, Li et al., 2020, Zhu et al., 2017), where passengers respond to hypothetical situations. However, SP surveys may not fully capture actual passenger behavior, potentially leading to biased estimates (Mo et al., 2022a). Later studies therefore turned to revealed preference (RP) data from travel surveys to estimate actual route preferences (Marra and Corman, 2023). Although RP surveys offer more accurate data, they are often costly and limited in sample size (Ingvardson et al., 2024). With the rise of large-scale disaggregate datasets, such as automatic fare collection (AFC) data, RP data can now be collected more efficiently, enabling a more accurate and large scale analysis of route choice behavior (Berggren et al., 2022).

AFC data has already been used in analyzing public transport users' choices during disruptions. For instance, Marra and Corman (Marra and Corman, 2023) used AFC data to examine how network disturbances affect passengers' route choices, while Eltved et al. (Eltved et al., 2021) analyzed the impacts of long-term service disruptions on travel behavior using smart card data. Whilst AFC data has been used to analyze passenger behavior during and after disruptions, it has rarely been applied to study route choice preferences specifically in the context of disruptions. Most existing studies rely on a combination of RP and SP surveys (Rahimi et al., 2020, Li et al., 2020), making it unclear whether AFC data alone can sufficiently capture route choice preferences. One exception is the study by Mo et al. (Mo et al., 2022b), which uses AFC data to examine mode choice preferences during disruptions rather than route choices. This means there is a gap in studies using only AFC data to look at route choice preferences in response to a disruption.

Another gap in the literature is that most research focuses on route choice during the disruption itself, rather than what happens after the disruption ends. While some studies, such as Eltved et al. (Eltved et al., 2021), have examined route choice before and after a disruption, their dataset only covered a period of three weeks before and after, making it difficult to determine whether changes in preferences persist over time.

This study aims to fill these gaps by extending the analysis period from weeks to several months before and after a disruption, allowing for a better understanding of whether changes in route choice preferences persist over time. Additionally, this research relies solely on AFC data to analyze changes in route choice preferences, a methodology that has not yet been applied to the Washington DC metro system. These contributions provide valuable scientific insights into how public transport users adapt to planned disruptions over extended periods and demonstrate the feasibility of using AFC data alone to infer route choice preferences.

Beyond its academic relevance, this research also has important practical implications. The Washington Metropolitan Area Transit Authority (WMATA), a key stakeholder in this study, plays a crucial role in promoting public transport as a sustainable alternative to private vehicles. Unlike unexpected disruptions, planned disruptions provide transit agencies with the opportunity to proactively manage service changes, as they can adjust mitigation strategies, improve passenger communication, and even modify disruption plans in response to anticipated behavioral shifts. By understanding how passenger preferences evolve during and after planned disruptions, this research can help refine policies, optimize service planning, and develop strategies that better retain and attract riders. These insights ultimately support the government's broader sustainability goals, such as reducing congestion and emissions by increasing public transport usage.

II. Research objective and Questions

Following the gaps in literature the objective of this research is to analyze transit users' route choice preferences in response to planned disruptions, using smart card data from the Washington DC metro network. Based on this objective the following research question was constructed.

How do public transport passengers' route choice preferences change in response to a planned public transport disruption?

To answer this main questions the following sub-questions need to be answered.

- *Which disruptions can be used to analyze changes in route choice preferences?*
- *What are the main factors influencing passengers' route choices in response to planned public transport disruptions?*
- *How do route choice preferences evolve over time during the post-disruption period?*
- *How suitable is revealed preference data for analyzing changes in route choice preferences in response to a planned public transport disruption?*

As cities around the world struggle with increasing traffic congestion and rising emissions, public transport plays a critical role in ensuring sustainable urban mobility (Lako and Gjevori, 2023). Efficient and reliable transit systems reduce dependence on private vehicles, which in turn helps to reduce overcrowding and lower carbon emissions. However, maintaining public transport as an attractive alternative to driving requires ensuring high levels of passenger satisfaction with service quality (Tuan et al., 2022).

A key factor influencing passenger satisfaction is reliability, as unpredictable service can discourage people from using public transport regularly (Soza-Parra et al., 2019). One of the main threats to reliability are disruptions, which can lead to increased travel times, missed connections, and overcrowding. These disruptions not only affect overall service quality but also influence passengers' transit route choices, forcing them to adapt their travel behavior in response to service changes due to the disruption.

Research has already been conducted on passengers' public transport route choices during disruptions, often using discrete choice models (Dixit et al., 2023). These models help reveal transit riders' preferences by assessing the relative importance of various travel attributes, offering policymakers valuable insights into how passengers make route choices in response to disruptions.

Many earlier studies on public transport route choice analysis were based on stated preference (SP) data (Shires et al., 2019, Li et al., 2020, Zhu et al., 2017), where passengers respond to hypothetical situations. However, SP surveys may not fully capture actual passenger behavior, potentially leading to biased estimates (Mo et al., 2022a). Later studies therefore turned to revealed preference (RP) data from travel surveys to estimate actual route choice preferences (Marra and Corman, 2023). Although RP surveys offer more accurate data, they are often costly and limited in sample size (Ingvardson et al., 2024). With the rise of large-scale disaggregate datasets, such as automatic fare collection (AFC) data, RP data can now be collected more efficiently, enabling a more accurate and large scale analysis of route choice behavior (Berggren et al., 2022).

AFC data has already been used in analyzing public transport users' choices during disruptions. For instance, Marra and Corman (Marra and Corman, 2023) used AFC data to examine how network disturbances affect passengers' route choices, while Eltved et al. (Eltved et al., 2021) analyzed the impacts of long-term service disruptions on travel behavior using smart card data. Whilst AFC data has been used to analyze passenger behavior during and after disruptions, it has rarely been applied to study route choice preferences specifically in the context of disruptions. Most existing studies rely on a combination of RP and SP surveys (Rahimi et al., 2020, Li et al., 2020), making it unclear whether AFC data alone can sufficiently capture route preferences. One exception is the study by Mo et al. (Mo et al., 2022b), which uses AFC data to examine mode choice preferences during disruptions rather than route choices. This means there is a gap in studies using only AFC data to look at route preferences during and after a disruption. Next to that there are also very little studies that only use AFC data to reveal preferences.

Another gap in the literature is that most research focuses on route choice during the disruption itself, rather than what happens after the disruption ends. While some studies, such as Eltved et al. (Eltved et al., 2021), have examined route choice before and after a disruption, their dataset only covered a period of three weeks before and after, making it difficult to determine whether changes in preferences persist over time.

This study aims to fill these gaps by extending the analysis period from weeks to several months before and after a disruption, allowing for a better understanding of whether changes in route preferences persist over time. Additionally, this research relies solely on AFC data to analyze changes in route preferences, a methodology that has not yet been applied to the Washington DC metro system. These contributions provide valuable scientific insights into how public transport users adapt to planned disruptions over extended periods and demonstrate the feasibility of using AFC data alone to infer route choice preferences.

Beyond its academic relevance, this research also has important practical implications. The Washington Metropolitan Area Transit Authority (WMATA), a key stakeholder in this study, plays a crucial role in promoting public transport as a sustainable alternative to private vehicles. Unlike unexpected disruptions, planned disruptions provide transit agencies with the opportunity to proactively manage service changes, as they can adjust

III. Literature review

Public transport disruptions, whether planned (e.g., scheduled maintenance) or unplanned (e.g., technical failures, accidents), affect passenger travel behavior in significant ways. Traditionally, studies on disruptions have focused on operational performance and minimizing service delays, but more recent research has shifted towards understanding passenger responses, particularly changes in route and mode choices (Marra and Corman, 2023, Eltved et al., 2021). Many studies have examined passenger decision-making during disruptions, but relatively few have explored how passengers adjust their behavior after the disruption ends (Wang et al., 2024).

The type of disruption influences how passengers respond. Unplanned disruptions typically cause more uncertainty, forcing passengers to make quick decisions with limited information (Li et al., 2020, Mo et al., 2022b). In contrast, planned disruptions allow passengers to anticipate changes in advance, but they often last longer, meaning the cumulative impact can be greater (Yap and Cats, 2021a). While several studies have examined planned disruptions (Eltved et al., 2021, Zhu et al., 2017), few have investigated whether passengers return to their pre-disruption travel habits or adopt new long-term behaviors.

A. Data Sources in Disruption Studies

Research on public transport disruptions relies on various data sources, each with strengths and limitations. Stated Preference (SP) surveys (Shires et al., 2019, Li et al., 2020) are commonly used to understand hypothetical choices, but responses may not always reflect actual behavior. Revealed Preference (RP) surveys (Marra and Corman, 2023) collect data on real travel choices, reducing hypothetical bias, but these surveys tend to be expensive and limited in sample size (Ingvardson et al., 2024).

With advances in data collection, Automatic Fare Collection (AFC) data (e.g., smart card records) has become an increasingly valuable tool for studying passenger behavior in disruptions (Yap et al., 2018). AFC data captures large-scale travel patterns, allowing for detailed analysis of how passengers adapt their route choices in real-world conditions. However, one major limitation of AFC data is the lack of demographic and perception-based information—it does not capture factors like comfort, safety concerns, or personal preferences (Mo et al., 2022a). As a result, AFC-based studies may overlook some of the subjective factors influencing route choice.

Despite its limitations, AFC data is particularly useful for disruption studies because it provides detailed information on how passengers adapt their behavior over time, which is crucial when examining post-disruption effects. However, while AFC data has been widely used to study mode choice and demand forecasting, relatively few studies have applied it to route choice analysis after disruptions (Mo et al., 2022b).

B. Study Methods in Disruption Research

Several methodological approaches have been used to study passenger behavior during disruptions. Data analysis methods (e.g., regression, clustering) have been applied to identify general trends and categorize different passenger responses (Liu et al., 2021, Nazem et al., 2018). Some studies use simulation models to predict how disruptions impact ridership (Deng et al., 2022, Wang et al., 2024), while others apply machine learning to forecast disruption impacts or passenger demand (Yap and Cats, 2022, Yap and Cats, 2021a).

One of the most widely used approaches for analyzing individual passenger behavior is discrete choice modeling. Multinomial Logit (MNL) models have been commonly used to estimate how passengers weigh different travel attributes when choosing routes (Marra and Corman, 2023, Shires et al., 2019). However, MNL models assume that all passengers have identical preferences, which may not always be realistic. Mixed Logit (ML) models (Yap et al., 2020) improve on this by allowing for individual heterogeneity. Another approach is the Path Size Logit (PSL) model, which accounts for route overlap by penalizing alternatives that share links or stations(Yap and Cats, 2021b).

C. Research Gaps and Contribution

Despite growing interest in passenger behavior during disruptions, several key gaps remain. Most studies focus on passenger responses during disruptions, with limited research on how route choice preferences evolve after disruptions (Eltved et al., 2021). It remains unclear whether passengers revert to their pre-disruption route preferences or if they adopt new, lasting travel behaviors.

Additionally, while Automatic Fare Collection (AFC) data has been widely used in disruption studies, it has rarely been applied to analyze route choice preferences after disruptions. Most AFC-based research focuses on mode choice or overall demand changes, rather than understanding how passengers adjust their specific route preferences over time (Mo et al., 2022b). This leaves an important gap in understanding long-term shifts in route selection following planned disruptions.

Furthermore, limited research has been conducted on the Washington, D.C. metro system in the context of disruptions. While AFC data is available for this network, most studies on transit disruptions focus on European or Asian systems. The only known study using AFC data for Washington, D.C. analyzed delay prediction, rather than how disruptions influence route choice preferences (Yap and Cats, 2021a).

To address these gaps, this study examines route choice preferences before and after a planned disruption using AFC data from the Washington, D.C. metro system. Unlike previous studies that primarily rely on RP and SP surveys, this research investigates whether AFC data alone can provide meaningful insights into route choice preferences. By analyzing an extended post-disruption period, this study offers a deeper understanding of whether travelers revert to pre-disruption route preferences or if their choices remain permanently altered following planned disruptions.

IV. Methodology and Case Study

This study follows a structured approach to identifying and analyzing planned disruptions in the Washington D.C. metro. To ensure meaningful analysis, disruptions must significantly impact passenger behavior and meet specific criteria before being selected for further study. Given the scale and importance of the Washington D.C. metro network, disruptions in this system provide an opportunity to examine how passengers adjust their route choices in response to disruptions.

The Washington D.C. metro, operated by the Washington Metropolitan Area Transit Authority (WMATA), serves over 600,000 passengers daily and is one of the busiest metro networks in the United States. The system consists of six color-coded lines—Red, Orange, Silver, Blue, Yellow, and Green—spanning 98 stations and more than 129 miles of track (Washington Metropolitan Area Transit Authority, 2024). A schematic of the network can be found in 1 Given its extensive network and high ridership, analyzing route choice behavior in response to disruptions provides valuable insights for both academic research and transit planning.

To conduct this analysis, the study utilizes automated fare collection (AFC) data provided by WMATA through the Smart Public Transport Lab at TU Delft. The dataset covers the period from August 2019 to December 2022, capturing various operational conditions, including normal service, planned maintenance, and disruptions. AFC data records individual passenger journeys, including tap-in and tap-out times, allowing for a detailed examination of how travelers adapt to disruptions. By systematically selecting and analyzing these disruptions, this methodology and case study allow for the assessment of how passengers' route preferences change after a disruption.

A. Disruption Identification

To ensure that selected disruptions provide meaningful insights into passenger behavior, the following criteria are applied:

- *The disruption should occur both on weekdays and weekends.*
- *The disruption should last several days in a station or a set of adjacent stations.*
- *There must be at least one month prior to the disruption as a pre-disruption period, and a minimum of three months following the disruption for post-disruption analysis.*
- *There should be no other disruptions lasting more than several hours in the affected area during both the pre- and post-disruption periods.*
- *There should be alternative route options available during the disruption.*

Each criterion ensures that the disruption has a measurable impact on passenger behavior and that changes in route choice can be attributed specifically to the disruption itself.

The requirement for a disruption to span both weekdays and weekends ensures that the analysis captures different travel patterns, as weekday travel is typically dominated by commuting behavior, while weekends often involve more discretionary trips. Additionally, focusing on disruptions lasting several days provides sufficient data to observe behavioral adjustments and allows passengers time to settle into new route choices. A clearly defined pre- and post-disruption period ensures a robust comparison of travel behavior before and after the disruption. The exclusion of overlapping disruptions ensures that observed behavioral changes are not influenced by other service interruptions. Finally, the availability of alternative routes ensures that passengers have viable choices, making the study of route preferences meaningful.

The selection process begins by identifying disruptions that meet the established criteria. This is done by analyzing the planned disruptions file, which contains details such as the start date, affected stations, and a description of each disruption. To filter for disruptions lasting multiple days, entries containing the word "Thru" in the message column are selected. These disruptions are then further examined to ensure they meet all criteria. After applying this process, an initial selection of 15 disruptions was made. These were further screened to confirm they occurred on both weekdays and weekends, had a clearly defined pre- and post-disruption period, and included viable alternative routes. This resulted in a final selection of four disruptions that met all criteria. These disruptions can be found in Table I.

Since all disruptions impacted the same section of the Green and Red Lines, they were combined into a single, continuous disruption lasting from June 21 to July 25, 2020. This approach provided a more comprehensive view of how passengers adapted their routes over an extended period. The disruption, caused by track work and platform repairs, offered

Fig. 1. Washington DC Metro Network

TABLE I
FINAL DISRUPTIONS

Date	Line	Affected Stations
21/06/2020 - 27/06/2020	Green	Between L'Enfant Plaza and Shaw Howard University
28/06/2020 - 02/07/2020	Green	Between Mt. Vernon Square and U street
06/07/2020 - 18/07/2020	Green	Between U street and Fort Totten
19/07/2020 - 25/07/2020	Red	Between Judiciary Square and Rhode Island Avenue

an opportunity to analyze passenger behavior after planned disruptions.

B. Selection of Pre- and Post-Disruption Periods

Defining pre- and post-disruption periods is essential to analyze changes in passenger behavior while minimizing external influences.

The pre-disruption period serves as a baseline for typical travel behavior. A period of one to two months is chosen to ensure stability in observed patterns while accounting for variations in daily and weekly travel. For this case study, April 21, 2020 – June 21, 2020 is selected, as no major disruptions occurred on the affected metro lines during this time.

To capture both short-term adjustments and longer-term behavioral changes, a three- to five-month post-disruption period is preferred. However, due to additional disruptions affecting the Green Line from October 31, 2020 – November 22, 2020, the post-disruption period is limited to three months to avoid interference.

To examine how passenger behavior evolves over time, the post-disruption period is further divided into three one-month intervals. This allows for a more detailed assessment of whether passengers immediately revert to pre-disruption travel patterns or if changes persist over time. Table II summarizes the final pre- and post-disruption periods.

TABLE II
PRE- AND POST-DISRUPTION PERIODS

Period	Start Date	End Date	Duration
Pre-disruption	21/04/2020	21/06/2020	2 months
Post-disruption total	25/07/2020	25/10/2020	3 months
Post-disruption 1	25/07/2020	25/08/2020	1 month
Post-disruption 2	26/08/2020	25/09/2020	1 month
Post-disruption 3	26/09/2020	25/10/2020	1 month

This structured selection ensures a clear comparison of passenger behavior before and after the disruption while allowing for an in-depth analysis of behavioral adaptation over time.

C. Selection of Affected OD Pair

The selection of the affected Origin-Destination (OD) pair is a crucial step in understanding how disruptions influence passenger route choices. This step ensures that the chosen

OD pair provides a meaningful basis for analyzing how passengers adjust their travel behavior when their usual routes are affected. The OD pair must have multiple viable routes, at least one of which is impacted by the disruption, and exhibit variation in the number of transfers required. The process of identifying the affected OD pair follows these steps:

- **Identification of all OD pairs**: All possible OD pairs in the network are first identified. Each OD pair represents a journey from a specific origin station to a destination station.
- **Identification of routes for each OD pair**: The next step involves mapping out the different routes available for each OD pair, including both direct routes and those involving transfers between lines.
- **Selection of OD pairs with multiple route options**: OD pairs with only a single route are excluded, as they do not allow for an analysis of route choice behavior. Only OD pairs with multiple viable routes are retained.
- **Selection of OD pairs with transfer variability**: OD pairs where different routes involve a varying number of transfers are prioritized. This ensures the dataset includes passengers who must decide between direct routes and those requiring transfers, making the analysis more meaningful.
- **Final OD pair selection**: From the remaining candidates, the OD pair with the highest number of passenger trips over the disrupted section is chosen, ensuring sufficient data for robust statistical analysis.

Applying this methodology to the Washington D.C. Metro, over 4,000 OD pairs were initially identified. After filtering for OD pairs with multiple route options and transfer variability, the dataset was reduced to 169 OD pairs. The selection was further refined to OD pairs where at least one route passes through the disrupted section on the Green Line between Gallery Place and Fort Totten.

Ordering the remaining OD pairs by passenger volume revealed that the first two candidates did not exhibit meaningful route variation, and the third included an incorrectly inferred route. The fourth OD pair, between Foggy Bottom and Greenbelt, was found to meet all criteria, offering multiple viable routes with different numbers of transfers. Thus, this OD pair was selected for further analysis.

After selecting the OD pair, one adjustment was made: all instances of the Yellow Line (YL) were reassigned to the Green Line (GR). This decision was based on the fact that the Yellow Line does not serve Greenbelt directly, and using it would require an impractical 8-mile walk to Greenbelt station. Changing YL into GR results in the following six realistic routes between *Foggy Bottom* and *Greenbelt*, which will be used throughout the rest of this study:

- Route 1: BL $\rightarrow$ GR
- Route 2: OR $\rightarrow$ GR
- Route 3: SV $\rightarrow$ GR
- Route 4: OR $\rightarrow$ RD $\rightarrow$ GR
- Route 5: BL $\rightarrow$ RD $\rightarrow$ GR
- Route 6: SV $\rightarrow$ RD $\rightarrow$ GR

This selection ensures a well-balanced dataset for analyzing how passengers respond to disruptions and how route choices evolve over time.

D. Attribute Identification

To generate the choice set, relevant attributes must first be identified. Literature indicates that time and cost are key factors influencing public transport route choice (Nielsen et al., 2021, Jánošíková et al., 2014). However, the inclusion of attributes in this study is constrained by data availability.

Based on the WMATA dataset, the following attributes are selected for the discrete choice models:

- **Veh_sec**: Total in-vehicle time (seconds).
- **Wait_sec**: Cumulative platform wait time (seconds).
- **Transfers**: Number of transfers inferred from the route taken.

Other attributes were considered but ultimately excluded. For instance, *Fare* was not selected due to minimal variation between routes, and *Reliability* was excluded as the delay data contained many inconsistencies. Given these limitations, the selected attributes provide the most reliable basis for analyzing route choice behavior.

E. Choice Set Generation

The choice set defines the available route alternatives for a given Origin-Destination (OD) pair. In this study, separate choice sets are constructed for the pre-disruption and post-disruption periods to assess changes in passenger route preferences. A balance is maintained between including realistic alternatives and avoiding excessively large choice sets that could distort model estimates.

To ensure accuracy, the dataset is first filtered and cleaned, removing invalid entries such as incomplete records, unrealistic journey durations, and duplicate transactions. The final dataset consists of metro trips with complete tap-in and tap-out data, focusing on passengers present in both time periods to allow for direct comparison. Observed route choices from smart card data serve as the foundation for constructing the choice set. To maintain realism, only routes chosen at least 1% of the time are included, and non-chosen alternatives are derived based on historically observed travel conditions rather than hypothetical assumptions.

Route attributes such as travel time, wait time, and the number of transfers are assigned based on historical data for specific time periods to reflect variations in travel conditions. The following time frames are used: morning peak (6 AM – 9 AM), midday (10 AM – 3 PM), evening rush (4 PM – 7 PM), and evening (8 PM – 12 AM). If data is unavailable for a given time frame, values are assigned using the closest available period to ensure realistic estimates. This structured approach ensures that the choice set accurately reflects passenger decision-making conditions while maintaining computational feasibility.

F. Estimate Discrete Choice model

To analyze passenger route choice behavior, this study employs discrete choice models, which estimate the probability of selecting a route based on its attributes. The models assume that passengers choose the alternative with the highest perceived utility, which consists of an observed component (based on route attributes like travel time and number of transfers) and an unobserved random component.

Given the complexity of route choice behavior and the need to account for different sources of variation, three models are estimated and compared: the Multinomial Logit (MNL) model, the Mixed Logit (ML) model with panel data, and the Path Size Logit (PSL) model.

The MNL model serves as a baseline and assumes that all passengers have identical preferences and that choices are independent of irrelevant alternatives (IIA). The ML model relaxes these assumptions by allowing individual-specific preference variation, capturing differences in how passengers value travel time and transfers. The PSL model further improves route choice estimation by correcting for the correlation between overlapping routes, introducing a path size term that penalizes highly similar alternatives.

To compare model performance, standard goodness-of-fit measures are used, including log-likelihood, Rho-square, and information criteria such as AIC and BIC. The final model is selected based on its ability to balance explanatory power and complexity while providing the most realistic representation of passenger route choice behavior.

V. RESULTS

This section examines the impact of the disruption on route choice behavior by analyzing key travel attributes and shifts in route choices.

A. Route choice adjustments

The disruption led to a noticeable redistribution of passengers across routes. Figure 2 presents the total trip counts per route across different periods, highlighting changes in ridership patterns.

Before the disruption, the most frequently used routes were $BL \rightarrow GR$ and $OR \rightarrow GR$, while multi-transfer routes such as $BL \rightarrow RD \rightarrow GR$, $SV \rightarrow RD \rightarrow GR$, and $OR \rightarrow RD \rightarrow GR$ had lower ridership. However, after the disruption, there was a substantial increase in the use of transfer-heavy routes. This indicates that passengers adjusted their route choices in response to the disruption, potentially due to disruptions affecting their usual travel patterns.

Despite this shift, travel attributes such as travel time and wait time remained relatively stable across periods. The disruption did not significantly alter the travel conditions on the alternative routes. However, passenger route choices still changed, suggesting that the disruption influenced how passengers perceived these attributes rather than the attributes themselves. This makes it particularly interesting to estimate a discrete choice model, as it allows us to examine whether the relative importance of travel time, wait time, and transfers shifted after the disruption. By analyzing changes in perception rather than just attribute values, we can better understand how disruptions shape passenger decision-making.

B. MNL results

Before estimating the MNL model, correlation analysis revealed a strong negative relationship between in-vehicle time and the number of transfers, suggesting that longer travel times are associated with fewer transfers. Due to this correlation, only in-vehicle time was included in the model, as it exhibits greater variation across trips and is a more informative predictor of route choice.

The MNL model was estimated using in-vehicle time and wait time as explanatory variables. The results, which can be found in table III and IV, indicate that the model provides a better fit than a random choice model, as evidenced by an improvement in log-likelihood and a Rho-square (ρ^2) value of 0.115. However, several estimated parameters exhibit counter-intuitive signs. Specifically, the coefficient for in-vehicle time is positive in some periods, which contradicts expectations that passengers prefer shorter travel times. This is likely due to the omitted transfer variable, as its correlation with in-vehicle time may have led to a misrepresentation of effects. Similarly, wait time coefficients were found to be unexpectedly positive in certain cases, suggesting potential inconsistencies in the choice set structure.

Further examination of the constructed choice set revealed dominated alternatives, meaning that some frequently chosen routes performed worse across all attributes compared to other available options. This suggests that observed route choices may be influenced by unobserved factors, distorting parameter estimation. To address this, an alternative model was tested where routes with the same number of transfers were merged. While this approach mitigated some inconsistencies, it limited the ability to capture more nuanced variations in route preferences, highlighting the need for a more sophisticated modeling approach.

C. Mixed Logit Model

Given the panel nature of the dataset, a Mixed Logit (ML) model was estimated to account for preference heterogeneity among travelers. Unlike the MNL model, which assumes homogeneous preferences across individuals, the ML model introduces random coefficients to capture individual-specific variation. This allows for a more flexible representation of route choice behavior, as different travelers may perceive in-vehicle time and wait time differently.

The model results, summarized in Table V, show that the final log-likelihood improved to -2393.276, compared to -2416.265 in the MNL model (see Table III). This results in a Rho-square (ρ^2) value of 0.141, which is higher than in the MNL model (0.114), indicating a better model fit. The AIC value also improved slightly (4826.552 vs. 4852.530 in the MNL model), although the BIC value increased (4933.832 vs. 4906.163), suggesting that the added model complexity does not necessarily justify the additional parameters.

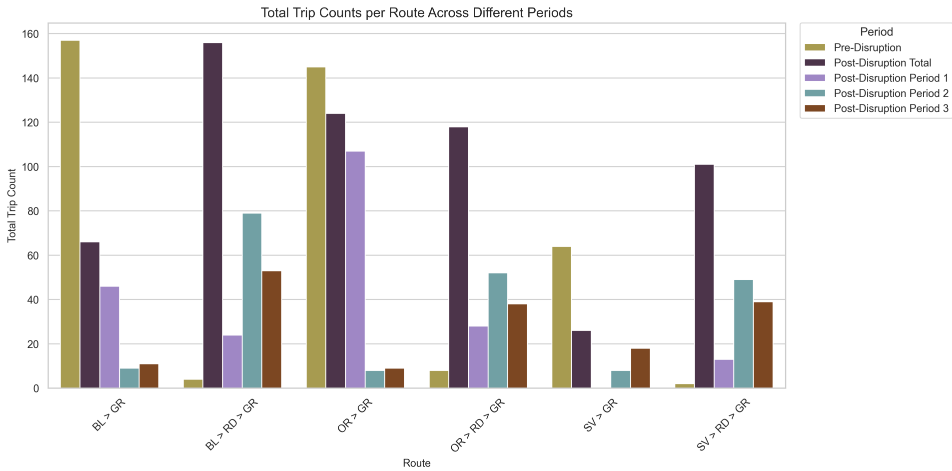


Fig. 2. Trip counts per route

TABLE III
MODEL ESTIMATION RESULTS MNL

Statistic	Value
Number of estimated parameters	10
Sample size	1577
Null log-likelihood	-2728.461
Final log-likelihood	-2416.265
Rho-square	0.115
Rho-square-bar	0.111
Akaike Information Criterion (AIC)	4852.530
Bayesian Information Criterion (BIC)	4906.163

TABLE IV
ESTIMATED PARAMETERS MNL

Parameter	Value	Rob. Std Err	Rob. t-test	Rob. p-value
BETA_VEH_POST_1	0.170982	0.019461	8.785678	0.000000
BETA_VEH_POST_2	-0.246755	0.026767	-9.218534	0.000000
BETA_VEH_POST_3	-0.171042	0.026250	-6.515931	0.000000
BETA_VEH_POST_TOTAL	-0.054247	0.011350	-4.779524	0.000002
BETA_VEH_PRE	0.601684	0.045771	13.145656	0.000000
BETA_WAIT_POST_1	0.051765	0.038428	1.347084	0.177953
BETA_WAIT_POST_2	-0.009663	0.053035	-0.182209	0.855419
BETA_WAIT_POST_3	-0.041786	0.035506	-1.176862	0.239251
BETA_WAIT_POST_TOTAL	0.093753	0.017437	5.376732	0.000000
BETA_WAIT_PRE	0.043243	0.021738	1.989264	0.046672

TABLE V
ESTIMATION REPORT MIXED LOGIT

Statistic	Value
Number of estimated parameters	20
Number of respondents	42
Sample size	1577
Initial log-likelihood	-2784.947
Final log-likelihood	-2393.276
Rho-square (initial model)	0.141
Rho-square-bar (initial model)	0.133
Akaike Information Criterion (AIC)	4826.552
Bayesian Information Criterion (BIC)	4933.832

The estimated parameters, presented in Table VI, show inconsistencies in in-vehicle time coefficients. BETA_VEH_PRE is unexpectedly positive, while post-disruption values vary, with BETA_VEH_POST_2 aligning with expectations. Wait time coefficients are mostly insignificant, contrasting with the MNL model results where some were unexpectedly positive.

The sigma values suggest limited variation in individual

TABLE VI
ESTIMATED PARAMETERS MIXED LOGIT

Parameter	Value	Std Err	z-test	p-value
VEH_SEC_PRE	0.670900	0.125000	5.386000	0.000000
VEH_SEC_POST_TOTAL	-0.025700	0.076000	-0.340000	0.734000
VEH_SEC_POST_1	0.201400	0.088000	2.284000	0.022000
VEH_SEC_POST_2	-0.272600	0.146000	-1.862000	0.063000
VEH_SEC_POST_3	-0.161400	0.087000	-1.859000	0.063000
WAIT_SEC_PRE	0.107100	0.080000	1.345000	0.179000
WAIT_SEC_POST_TOTAL	0.163200	0.221000	0.739000	0.460000
WAIT_SEC_POST_1	0.140800	0.257000	0.547000	0.584000
WAIT_SEC_POST_2	0.009600	0.184000	0.052000	0.958000
WAIT_SEC_POST_3	-0.053000	0.233000	-0.228000	0.820000
Sigma VEH_SEC_PRE	-0.067700	0.326000	-0.207000	0.836000
Sigma VEH_SEC_POST_TOTAL	0.073600	0.043000	1.699000	0.089000
Sigma VEH_SEC_POST_1	0.110700	0.093000	1.195000	0.232000
Sigma VEH_SEC_POST_2	-0.151200	0.171000	-0.883000	0.377000
Sigma VEH_SEC_POST_3	0.125600	0.136000	0.923000	0.356000
Sigma WAIT_SEC_PRE	0.146200	0.179000	0.819000	0.413000
Sigma WAIT_SEC_POST_TOTAL	-0.135200	0.120000	-1.124000	0.261000
Sigma WAIT_SEC_POST_1	0.254600	0.266000	0.959000	0.338000
Sigma WAIT_SEC_POST_2	-0.019500	0.537000	-0.036000	0.971000
Sigma WAIT_SEC_POST_3	0.025300	0.350000	0.072000	0.942000

preferences. None of the sigma values are significant, indicating that travelers perceive travel attributes similarly.

To formally compare the MNL and ML models, a Likelihood Ratio (LR) test was conducted, confirming that the ML model provides a statistically significant improvement over the MNL model. However, while the ML model improves log-likelihood, the practical benefits remain limited. The AIC and BIC values suggest only marginal gains, and unexpected coefficient signs persist. The estimated sigma values, which capture preference heterogeneity, indicate that none of the attributes show significant variation among individuals. This suggests that most attributes are viewed consistently across the sample. This limited heterogeneity, combined with the persistence of counterintuitive coefficient signs, suggests that the added complexity of the ML model does not yield meaningful insights over the simpler MNL model. Given the small improvements in fit and the continued presence of estimation issues, the ML model does not appear to justify its additional complexity.

D. Path Size Logit

Given the overlapping nature of routes in the dataset, a Path Size Logit (PSL) model was considered to correct for correlation among similar alternatives. However, calculations of the path size factors based on overlapping links and stations revealed nearly identical values across all routes. This suggests that the measure does not effectively differentiate between alternatives, making it unlikely to add value to the model. As a result, the PSL model was not estimated.

E. Summary

The results indicate that while the ML model offers a statistically improved fit over the MNL model, it does not substantially enhance explanatory power. The persistence of counterintuitive coefficient signs suggests that additional unobserved factors influence passenger route choice. While merging dominated alternatives improved model consistency, it also reduced the ability to capture variations in route preferences. The findings highlight the complexity of modeling route choices in disrupted transit networks and suggest the need for further refinements, potentially through alternative model specifications or additional data sources.

VI. Discussion and Conclusion

This study examined how public transport users' route choice preferences changed in response to a planned disruption, using AFC data from the Washington DC metro system. It addressed gaps in the literature by investigating whether route preferences shifted after a disruption, whether these changes persisted over time, and whether AFC data alone is sufficient to capture such behavior.

The descriptive analysis showed that although measured travel times and wait times remained relatively stable across different periods, passengers' route choices changed significantly. Before the disruption, travelers tended to prefer direct routes, even when these had longer travel times. After the disruption, however, passengers increasingly chose routes with more transfers but shorter in-vehicle times. This shift suggests a change in preferences, with travelers placing greater importance on minimizing in-vehicle time and demonstrating a greater willingness to accept transfers.

Discrete choice models were estimated to further explore these changes, but the results revealed unexpected findings: some models produced positive coefficients for travel time and wait time, contrary to theoretical expectations. These counterintuitive results are likely due to the strong negative correlation between in-vehicle time and the number of transfers, which made it difficult to isolate their individual effects. In addition, the presence of dominated alternatives—routes that were objectively worse across all measured attributes yet still chosen by passengers—indicates that unobserved factors also influenced decision-making.

Beyond the basic Multinomial Logit (MNL) model, a Mixed Logit (ML) model was estimated to account for preference heterogeneity. While the ML model provided a statistically better fit, practical improvements were limited, and the problem of unexpected coefficient signs persisted. Similarly, a Path Size Logit (PSL) model was estimated to account for route overlap, but the path size factors were nearly identical across routes, indicating that overlap did not meaningfully differentiate the alternatives.

Despite these modeling challenges, the broader patterns observed provide valuable insights. A lasting shift in route choice behavior was evident following the disruption, even if the underlying drivers of this change were difficult to pinpoint precisely.

Reflecting on the research questions, several conclusions can be drawn. Identifying suitable disruptions for analysis proved challenging, as many disruptions were too short, lacked viable alternatives, or overlapped with other events. Only one disruption provided the necessary conditions for clear analysis. Regarding the factors influencing route choice, the study focused on in-vehicle time, transfers, and wait time, but strong correlations between attributes limited the ability to isolate their effects. Still, a behavioral shift was detected: after the disruption, passengers placed greater emphasis on minimizing in-vehicle time, although it remains unclear whether this change was driven by increased time sensitivity, a reduced reluctance to transfer, or both. The evolution of preferences over time also indicated that passengers did not revert to their previous habits once the disruption ended. Finally, AFC data was highly valuable for detecting real-world behavioral changes, but its limitations in explaining underlying motivations highlight the need for complementary data sources in future research.

This study makes several scientific and societal contributions. Scientifically, it demonstrates that AFC data can reveal changes in behavior following a disruption, but researchers must be cautious when relying solely on AFC data, especially when key variables are strongly correlated or when perceptual factors are likely to influence choices. Societally, the findings suggest that planned disruptions can be opportunities, not just challenges. If passengers become more sensitive to travel time after disruptions, transit agencies must ensure that fast, convenient travel options remain available to maintain satisfaction and ridership. Proactively managing travel alternatives during disruptions could lead to lasting improvements in travel behavior.

Several limitations should be acknowledged. First, the use of AFC data alone limited the ability to capture perceptual factors such as comfort, reliability, or familiarity. Second, multicollinearity between in-vehicle time and transfers complicated model estimation. Third, the study focused on a single disruption event in one network, limiting generalizability. Fourth, the high degree of overlap between available routes restricted the variation necessary for robust model estimation.

Future research should address these limitations. Combining AFC data with stated preference surveys or qualitative passenger feedback would allow researchers to better capture unobserved influences. Conducting similar studies in networks with longer routes, greater variation between alternatives, or different structural characteristics would enhance generalizability. Additionally, future work could explore whether disruption duration affects the persistence of changes in preferences by analyzing networks with multiple disruptions of varying lengths.

In conclusion, this study shows that route choice preferences are adaptable and can shift permanently following planned disruptions. A disruption can prompt passengers to reconsider and adjust their travel behavior in ways that persist even after normal service is restored. For transit agencies, understanding and anticipating these shifts is essential for designing effective service recovery strategies that support passenger satisfaction and long-term system resilience.

References

Berggren, U., D'Agostino, C., Svensson, H., & Brundell-Freij, K. (2022). Intrapersonal variability in public transport path choice due to changes in service reliability. *Transportation*, *49*(6), 1517–1547. https://doi.org/10.1007/s11116-021-10218-z

Deng, Y., Liu, W., Zhang, Q., & Zhang, Z. (2022). Implementation and Application of a Stochastic Agent-based Model for Passenger Flow Distribution Predicting in Planned Metro Station Service Disruption Scenario. *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*, 2252–2257. https://doi.org/10.1109/ITSC55140.2022.9922254

Dixit, M., Cats, O., van Oort, N., Brands, T., & Hoogendoorn, S. (2023). Validation of a multi-modal transit route choice model using smartcard data. *Transportation*, *51*, 1809–1829. https://doi.org/10.1007/s11116-023-10387-z

Eltved, M., Breyer, N., Ingvardson, J., & Nielsen, O. (2021). Impacts of long-term service disruptions on passenger travel behaviour: A smart card analysis from the Greater Copenhagen area. *Transportation Research Part C: Emerging Technologies*, *131*. https://doi.org/10.1016/j.trc.2021.103198

Ingvardson, J. B., Thorhauge, M., Nielsen, O. A., & Eltved, M. (2024). Public transport route choice modelling: Reducing estimation bias when using smart card data. *Transportation Research Part A: Policy and Practice*, *179*, 103929. https://doi.org/10.1016/j.tra.2023.103929

Jánošíková, L., Slavík, J., & Koháni, M. (2014). Estimation of a route choice model for urban public transport using smart card data. *Transportation Planning and Technology*, *37*(7), 638–648. https://doi.org/10.1080/03081060.2014.935570

Lako, A., & Gjevori, S. (2023). Urban public transport as a basis for sustainable mobility development in the transition from private vehicles to urban public transport: A case study of elbasan. *International Journal* *of Sustainable Development and Planning*, *18*(11), 3609–3615. https://doi.org/10.18280/ijsdp.181126

Li, B., Yao, E., Yamamoto, T., Tang, Y., & Liu, S. (2020). Exploring behavioral heterogeneities of metro passenger's travel plan choice under unplanned service disruption with uncertainty. *Transportation Research Part A: Policy and Practice*, *141*, 294–306. https://doi.org/10.1016/j.tra.2020.09.009

Liu, T., Ma, Z., & Koutsopoulos, H. N. (2021). Unplanned Disruption Analysis in Urban Railway Systems Using Smart Card Data. *Urban Rail Transit*, *7*(3), 177–190. https://doi.org/10.1007/s40864-021-00150-x

Marra, A. D., & Corman, F. (2023). How different network disturbances affect route choice of public transport passengers. A descriptive study based on tracking. *Expert Systems with Applications*, *213*. https://doi.org/10.1016/j.eswa.2022.119083

Mo, B., von Franque, M., Koutsopoulos, H. N., Attanucci, J., & Zhao, J. (2022b). Impact of unplanned long-term service disruptions on public transit systems. *IEEE Open Journal of Intelligent Transportation Systems*, *3*, 551–569. https://doi.org/10.1109/OJITS.2022.3199108

Mo, B., Koutsopoulos, H. N., & Zhao, J. (2022a). Inferring passenger responses to urban rail disruptions using smart card data: A probabilistic framework. *Transportation Research Part E: Logistics and Transportation Review*, *159*. https://doi.org/10.1016/j.tre.2022.102628

Nazem, M., Lomone, A., Chu, A., & Spurr, T. (2018). Analysis of travel pattern changes due to a medium-term disruption on public transit networks using smart card data. *Transportation Research Procedia*, *32*, 585–596. https://doi.org/10.1016/j.trpro.2018.10.019

Nielsen, O. A., Eltved, M., Anderson, M. K., & Prato, C. G. (2021). Relevance of detailed transfer attributes in large-scale multimodal route choice models for metropolitan public transport passengers. *Transportation Research Part A: Policy and Practice*, *147*, 76–92. https://doi.org/https://doi.org/10.1016/j.tra.2021.02.010

Rahimi, E., Shamshiripour, A., Shabanpour, R., Mohammadian, A., & Auld, J. (2020). Analysis of Transit Users' Response Behavior in Case of Unplanned Service Disruptions. *Transportation Research Record*, *2674*(3), 258–271. https://doi.org/10.1177/0361198120911921

Shires, J. D., Ojeda-Cabral, M., & Wardman, M. (2019). The impact of planned disruptions on rail passenger demand. *Transportation*, *46*(5), 1807–1837. https://doi.org/10.1007/s11116-018-9889-0

Soza-Parra, J., Raveau, S., Muñoz, J. C., & Cats, O. (2019). The underlying effect of public transport reliability on users' satisfaction. *Transportation Research Part A: Policy and Practice*, *126*, 83–93. https://doi.org/https://doi.org/10.1016/j.tra.2019.06.004

Tuan, V. A., Van Truong, N., Tetsuo, S., & An, N. N. (2022). Public transport service quality: Policy prioritization strategy in the importance-performance analysis and the three-factor theory frameworks. *Transportation Research Part A: Policy and Practice*, *166*, 118–134. https://doi.org/https://doi.org/10.1016/j.tra.2022.10.006

Wang, Y., Zhao, O., & Zhang, L. (2024). Modeling urban rail transit system resilience under natural disasters: A two-layer network framework based on link flow. *Reliability Engineering & System Safety*, *241*. https://doi.org/10.1016/j.ress.2023.109619

Washington Metropolitan Area Transit Authority. (2024). Metrorail service information. https://www.wmata.com/service/rail/

Yap, M., & Cats, O. (2022). Analysis and prediction of ridership impacts during planned public transport disruptions. *Journal of Public Transportation*, *24*. https://doi.org/10.1016/j.jpubtr.2022.100036

Yap, M., & Cats, O. (2021a). Predicting disruptions and their passenger delay impacts for public transport stops. *Transportation*, *48*(4), 1703–1731. https://doi.org/10.1007/s11116-020-10109-9

Yap, M., & Cats, O. (2021b). Taking the path less travelled: Valuation of denied boarding in crowded public transport systems. *Transportation Research Part A: Policy and Practice*, *147*, 1–13. https://doi.org/10.1016/j.tra.2021.02.007

Yap, M., Nijënstein, S., & van Oort, N. (2018). Improving predictions of public transport usage during disturbances based on smart card data. *Transport Policy*, *61*, 84–95. https://doi.org/10.1016/j.tranpol.2017.10.010

Yap, M., Cats, O., & van Arem, B. (2020). Crowding valuation in urban tram and bus transportation based on smart card data. *Transportmetrica A: Transport Science*, *16*(1), 23–42. https://doi.org/10.1080/23249935.2018.1537319

Zhu, S., Masud, H., Xiong, C., Yang, Z., Pan, Y., & Zhang, L. (2017). Travel Behavior Reactions to Transit Service Disruptions: Study of Metro SafeTrack Projects in Washington, D.C. *Transportation Research Record*, *2649*(1), 79–88. https://doi.org/10.3141/2649-09