Master Thesis

Identifying Crowding Impact on Ridership in Public Transport: A Quasi-Experimental Study

Submitted by:
Alexis Ziegler
Chemin du Coteau 18, 1009 Pully VD
Switzerland
a.ziegler@student.vu.nl
2819982

Supervisor:
Eric Paul Kroes
School of Business and Economics, Faculty of Spatial Economics

Master's Spatial, Transport and Environmental Economics
(STREEM)

10th of August 2024

Abstract

This paper explores the impact of congestion on public transport ridership by analyzing passenger count data from six bus lines over the period from 2017 and 2023. The study finds that doubling the seating capacity on one line leads to a significant increase in ridership. The results show the increase in the number of passengers during morning peak hours being four times greater than during off-peak hours. These findings suggest that strategically increasing seating capacity during peak times is an effective method to enhance ridership.

Acknowledgements

I would like to express my gratitude to my thesis supervisor, Eric Paul Kroes, for sharing his expertise and for all the advice he provided throughout the process. Also, I would like to thank him for skillfully reminding me to remain result oriented when I got lost in trivial methodological details.

My thanks extend to Antonin Danalet for his orientation advice based on his reading of my research project, and to Joschka Bischoff for sharing the data and answering my questions. Although I regret that I was unable to use the SBB data to answer my research questions, the advice given were very useful for the rest of the analysis.

I am also thankful to Frédéric Bründler for sharing his extensive knowledge of the TL network and for recommending the analysis of the service capacity increase on line 60. Of course, I'd also like to thank Charlotte Fine for agreeing to send me the data and for gathering all the necessary acceptances, as well as Ilian Topalov for his work in extracting the necessary data.

Lastly, I wish to express my profound gratitude to my family for their support throughout my studies.

Table of content

List of figures

List of tables

List of abbreviations

A	Outward (direction of a ride)
ATT	Average treatment effect on the Treated
approx.	approximatively
av.	average
DiD	Differences-in-Differences
DiDiD	Differences-in-Differences-in-Differences
e. g.	exempli gratia
ep	evening peak
etc.	et cetera
charge	patronage, ridership, number of passengers
coef.	coefficient
i.aug19	Dummy var. to for the period after treatment in August 2019
i.line60	Dummy var. to for the treatment group
i.mp	Dummy var. to for the morning peak
i. e.	id est (that is)
L	Line
MP	morning peak
n°	number
ns	not statistically significant
OLS	Ordinary Least Squares
OP	off-peak
PT	Public Transport
oct18	October 2018
R	Return (direction of a ride)
RP	Revealed Preference
SBB	Swiss Federal Railways
SP	Stated Preferences
SE	Standard Errors
TL	Public transport Provider of Lausanne
TT	Travel time
R	Return (direction of a ride)
stat. sig.	statistically significant
var	variable
VIF	Variation Inflation Factor
w. r. t.	with respect to
%	percent
#	interaction

1 Introduction

1.1 Problem Statement and Relevance

Though less widely studied than road congestion, crowding externalities within public transport systems also significantly affects ridership. The logic behind this is that negative experiences from crowding can deter passengers from traveling, thereby lowering overall ridership. This understanding forms the basis of our investigation.

From an idealistic vision aimed at accelerating the modal shift towards public transport, the focus on crowding phenomena is grounded in compelling findings: Research such as a recent study by the Swiss territorial development office (ARE, 2022) suggests that enhancing agglomeration transit—rather than intercity transport—holds the greatest potential for increasing public transport's modal split. At the same time, the existing literature about crowding externalities is applied to urban centers such as London (Wardman & Whelan, 2011), Sydney/Melbourne (Li & Hensher, 2013) and Île-de-France (Kroes et al., 2014), indicating that congestion tends mostly to apply in high density areas. With the view to substantially increasing ridership in public transport, a congestion investigation seems consequently particularly promising.

1.2 Research Questions and Objectives

The research strategy is divided into two stages. The first involves studying the raw effect of a service capacity increase on ridership. More precisely, the aim is to gauge the responsiveness of ridership to increases in service capacity. The term "service capacity" can refer to both vehicle capacity and frequency increases. Accordingly, the first research question is:

"How do changes in public transport service capacity affect ridership?"

In the second stage, we focus on quantifying how much of the observed increase (or decrease) in ridership after the service capacity enhancements is due to reduced discomfort from crowding. While the positive impacts on ridership are partly influenced by congestion reductions, it is essential to consider that increases in service capacity might also improve other aspects of service quality, such as the perceived travel time, especially if it involves a frequency increase. Therefore, the second research question is:

“To what extent can the increase in ridership be attributed to a decrease in crowding externalities?”

Consequently, the objective of this study is to provide new empirical evidence on the impact of crowding on public transport ridership and to explore how enhancements in public transport services mitigate these effects. The findings aim to contribute both to theoretical knowledge and, albeit modestly, to the practical implementation in transportation supply planning for crowded routes.

1.3 Hypotheses

Based on the problem statement and the research questions, the following hypotheses are formulated:

Hypothesis 1: An increase in public transport service capacity results in an increased ridership.

Hypothesis 2: The impact of increased public transport serviced capacity on ridership is positively correlated with the level of crowding observed prior to the increase.

These two hypotheses will be tested through a detailed analysis of public transport data, focusing on how changes in capacity influence passenger numbers and how these effects vary depending on the crowding pattern observed.

1.4 Structure of the Thesis

First, the essential literature about public transport determinants and crowding disutility implementation in utility models is reviewed. Secondly the methodology chapter builds the backbone of this study. It explains the research design, provides important definitions, details the data collection, cleaning and aggregation process, calculates some summary statistics, explicates the triple difference estimation and finally sets up the model specification used in the analysis. Thereafter, the results are presented and interpreted.

2 Literature Review

This chapter reviews the essential literature for the underlying analysis. First, Chapter 2.1. provides a general overview of the explanatory variables of ridership. Chapter 2.2 explores the direction of causality between the service capacity supplied by the operator and ridership and proposes solutions to mitigate the potential issue of endogeneity. Chapter 2.3 presents two models, considered convincing, on how a crowding can be implemented. Please note that the terms “ridership”, “patronage” and “charge” all refer to the number of passengers. There are used interchangeably throughout this study

2.1 Determinants of PT Ridership

There are internal and external factors of ridership in Public Transport (PT). The internal factors relate to decisions, policies and conditions determined by the transit operator or the authority providing subsidies. They must be differentiated from external factors that usually equate wider economic influences, such as unemployment rate, GDP or density around stations as summarized by Boisjoly et al. (2018). This differentiation is necessary as solely the internal factors are considered in the underlying study. The first common internal factor is the fare and consensus dominates about its statistically significant negative relationship with ridership (Chen et al., 2011; Taylor et al., 2009). The second common internal factor is the service quality. There is however a high variety about the specification of service. It can be considered as e.g. the number of vehicles operated (McLeod et al., 1991) as network density (Currie & Wallis, 2008) or as service reliability (Paudel, 2021). Independently of the specification, consensus dominates about the positive impact of service quality on transit ridership (Kain & Liu, 1999; Taylor et al., 2009).

Also, it must be noted that the coefficients and elasticities of internal factors show high variations depending on the duration period they are calculated for (Voith, 1991). In the context of rail demand, Voith, (1991) resumed that commuter ridership depends on the share of potential riders choosing the train among the pool of these potential riders, this pool being fixed in the short-term. These potential riders are to distinguish from non-PT users, for whom, despite favorable changes during the parameters of PT explanatory choice variables, are unlikely to be swayed in the short-term. This resistance is evidenced by a slow evolution of the modal splits, even in the face of favorable policies. The introduction of free PT in Luxembourg in 2020 serves as an example. Indeed, despite a 10% modal shift of from car usage towards PT usage, car use and related congestion problems remain predominant (Bigi et al., 2023). Reasons for this rigidity are probably to be found partly justified by the nature of mobility decisions. Bubenhofer et al. (2018) evoked in the evaluation of a Micro census SP-study done by the Swiss government about mobility behavior, that mobility tools are a prerequisite for being mobile; their purchase being therefore often a long-term decision. To provide a thorough estimation, a study comparing the attributes of the other modes of transportation competing with PT would be required.

To model the competition between different discrete alternatives such as PT and car, a common approach for travel demand research is the use of disaggregate/behavioral travel-demand modeling, also called discrete-choice models, as these models usually analyze choices among discrete rather than continuous alternatives (Small & Verhoef, 2007). The most widely used theoretical framework is the additive random-utility model by (McFadden, 1974). Discrete choice models necessitate data about each alternative. However, the underlying analysis only considers PT data, therefore ruling out the utilization of these models. Moreover, discrete choices are often analyzed statically, as McFadden's model typically asserts fixed populations and therefore solely provides short-turn estimations (Voith, 1991). Although Xiong et al. (2015) proposed a new breakthrough with a hidden Markov modeling approach to consider the dynamic nature of travel mode choice, there is still a lack of theoretical frameworks for dynamic travel mode choice. As the underlying analysis focuses on the development over time of a supply shock on ridership, which is a second the use of discrete choice models seems inappropriate for a second reason.

2.2 Service Capacity & Ridership: Where is the Causality?

The direction of the causality between service capacity and ridership seems to be a typical case of simultaneous causality. As the literature regarding increases in vehicle capacity is scarce, the explanation of the so called "Mohring effect" in the context of frequency increases is used to highlight this duality. Thereafter, concrete solutions to avoid endogeneity and ensure the internal validity of the model are described in Section 2.2.3.

2.2.1 Ridership dependent on Service Capacity

The total service capacity provided by the operator is dependent on two parameters: the frequency of service and the capacity of the vehicles employed. While several studies provide evidence about the effect of frequency on ridership (Kain & Liu, 1995; Voith, 1991), the literature rarely examines the pure and simple increases in vehicle capacity. Concerning frequency increases, Voith found for instance considerable impacts of service attributes of ridership. Analyzing the effect of adding one train per hour, he found out coefficients of 4.17 in the peak and 4.19 in the off-peak, therefore indicating a similar positive absolute effect on ridership across the day. Translated in elasticities, Voith calculated for the peak: 0.14 in the short run and 0.36 in the long run; for the off-peak: 0.74 in the short run and of 1.89 in the long run in the off-peak. The off-peak elasticities are higher than peak elasticities because the ridership per off-peak train is lower on average (av.) than that for peak hour trains. However, an elasticity of 1.89 is interesting as it suggests a more than proportional increase of ridership with respect to frequency (Voith). The author justifies this higher frequency-elasticity in the off-peak with the argument that off-peak trips wouldn't be constrained to a regular schedule like work trips (Voith). However, this higher sensitivity of demand during off-peak periods may also be attributed to the fact that an increase in frequency has a more significant effect when the number of trips proposed is very low compared to when it is already high, as can be inferred from Mohring's model.

More specifically, the so-called Mohring-effect defines that the passenger waiting time at stops equals half the frequency, therefore asserting that a doubling of frequency reduces by half the waiting costs at stops (Mohring, 1972). Small & Verhoef (2007) resumed the following simplified version of Mohring's cost of waiting:

$$C_w = \frac{\alpha^w * q}{2V},$$

where	C_w	is the cost of waiting,
	α^w	is the value of time per hour,
	q	are the passenger per hour.

This equation shows how frequency can affect waiting costs and therefore ridership by reducing the waiting time of passengers. For example, doubling an hourly frequency to a semi-hourly frequency will reduce the av. waiting time from thirty minutes to fifteen minutes. Doubling a semi-hourly frequency to a quarter-hourly frequency will reduce the av. waiting time from fifteen minutes to seven and a half minutes. This theoretical framework is in line with more recent findings showing that routes with high headways are more sensitive to frequency increases (Evans IV, 2004; Verbas et al., 2015). However, the model also has limitations, as it omits the scheduling behavior of passengers who can minimize waiting time at stops.

However, this critics of Mohring's model can be attenuated by the schedule delay utility specification provided by Small, (1982). Assuming that passengers have a preferred arrival time at destination, Small defined a linear schedule delay early penalty for arriving too early than this preferred arrival time and a schedule delay late penalty for arriving too late. The delay late penalty is typically higher than the delay early penalty. Although this model was developed in the context of road congestion, it can be used to highlight how reducing headways can also contribute to lower schedule delays at arrival in PT.

2.2.2 Service Capacity dependent of Ridership

This direction of the causality suggests that the transit agency responds to increasing ridership or crowding variables with an increase in frequency. This suggests the presence of endogeneity, as this suggests that an explanatory variable is correlated with the error term (Stock & Watson, 2002). Paradoxically, Mohring's model primarily serves as a framework to establish a frequency that minimizes the operator's costs (Mohring, 1972). Thus, we can infer that according to Mohring, frequency is also fundamentally a response to demand. In fact, Mohring's model proposed to add the user waiting time defined above as cost input of the operator (Mohring). This modification induces the existence of increasing returns to scale and has a major influence on the optimal frequency of service. In fact, Mohring's finding supports the idea of an optimal frequency that is higher than a PT supplier would offer under a basic marginal pricing scheme. Summarized, Mohring recognized the existence of a virtuous circle in PT due to the positive simultaneous causality between frequency and demand, therefore posing a major challenge for the analyzing the impact of frequency increases on ridership.

2.2.3 Solutions to Simultaneous Causality Bias

There are two ways to mitigate simultaneous causality bias. The first relies on to use of instrumental variables catching the biasness and inconsistency polluting the regressor of interest (Stock & Watson, 2002). The second relies in designing and implementing a randomized controlled experiment in which the reverse causality channel is nullified (Stock & Watson). Unfortunately, it was neither possible to have an available variable able at catching the potential endogeneity of the treatment, nor was it possible to randomize the chosen links and this in a sufficient high sample size. It is therefore necessary finding another argument proving the introduction of an ameliorated supply to be exogenous to ridership.

To solve this problem before starting the analysis, it is therefore useful to elucidate now the reasons which brought the operator to increase the capacity of service on the lines which it communicated to us. In our case, the essential breakthrough to this potential endogeneity dilemma was provided by activity report of the operator, which provides information on the reasons for the treatment. Although the specificities of the lines under analysis will be reported in greater detail later in the analysis, it is useful already noting that service capacity increase on one of lines (Line 60) was made possible by the introduction of thirteen new double-decker buses (TL, 2020). The introduction occurred in August 2019. We use this information to contradict the argument of endogeneity of the treatment in our case, based on the following arguments:

First is assumed that the overall process between the decision and effective commissioning in 2019, including tendering and production, most certainly exceeded the three years preceding introduction considered in this study. Therefore, it is assumable that the decision of introducing new buses on this line was not related to the ridership previously observed. However, this argument is not very convincing, as it can be argued that operators are used to forecasting long-term tendencies, so they already had an estimate of the development of ridership in advance. However, L60 provides interesting insights from a second perspective: The purchased buses are specifically designed for use on the L60 and not on the other lines, most of which have overhead contact lines for electrical operation and are designed for single deck rolling stock. This helps in undermining the assertion that the operator could reallocate these buses to other routes, in response to varying ridership or crowding. On the contrary, the operator's flexibility is significantly constrained, being largely limited to adjustments on L60. Although these arguments may not completely dispel the assertion that the operator has the flexibility to respond to demand through service adjustments, they provide nuance and prevent premature termination of the analysis. Furthermore, additional elements in the subsequent analysis will introduce new quantitative elements concerning this issue.

2.3 Implementing a crowding variable

Regarding the implementation of crowding as an explanatory variable for ridership, it is important to note the diversity in approaches. This section revisits two models that have been highlighted for their compelling contributions in a prior research project. The first model has been developed by Kraus (1991) and introduces several different facets of crowding externalities. Then, another approach proposed by Tirachini et al., (2013) that proposes to introduce an interaction with the travel time variable is described.

2.3.1 Crowding in different facets

Kraus proposed an extension of Mohring's framework to consider crowding disutility (1991). Kraus distinguished three types of crowding externalities: A unloading externality (1) referred as the delay cost of a marginal passenger's unloading, an loading externality (2) referred as the delay cost of marginal passenger's loading, and a discomfort externality (3) referred as the cost that passengers who board at stop $J > i_0$ impose on passengers who boars at stop i_0. These three terms (1), (2) and (3) shall be considered as independent and are to be added up to assert the crowding effect.

(1)

$$F_j = \sum_{i=1}^{n} hy_i * v_i * \frac{v_2}{2}$$

where F_j is the marginal cost fare from a given stop J,

n is the number of stops i,

h	is the headway (h corresponds to $\frac{1}{V}$ in Mohring (1972)),
y_i	are the passengers per hour,
v_i	is the value of time in transit,
v_2	is the time for up/unloading.

(2)

$$\sum_{i=1}^{n} hy_i * v_i * v_2 +$$

(3)

$$\pi * T_{i0}$$

where	π	is the added value to transit time while standing instead of sitting,
	T_{i0}	is the transit time (with i_0 being the number of the highest number stop where not all boarding passengers find seats (i_0 is the largest value of i for which $h\sum_{j=i}^{n} y_i > \phi$, where ϕ is the sitting capacity of a train))

This method of Kraus (1991) is convincing, as it specificizes the interaction of crowding with the delay that passengers impose on each other. Kraus is also very precise, as he divides the weight of the unloading time by two, as seen in (1), in comparison with the loading time. According to Kraus (1991), the disutility due to boarding at a stop continues until the last passenger has boarded the vehicle. However, this disutility is assumed to disappear once the passenger has alighted, which is expected to occur halfway through the unloading process. All in all, this model is therefore interesting for two reasons: First, because it demonstrates several facets of crowding disutility. Second, because it suggests congestion also correlates with other explanatory variables such as the value of travel time.

2.3.2 The interaction with Travel Time

The study of Tirachini et al. (2013) is central for the methodology and the research set-up of this work, as it is the only study specifically referring to the "load factor" widely used in the subsequent analysis. More specifically, Tirachini et al. proposed incorporating a correlation between crowding and the valuation of travel time. Their analysis had two dimensions: operational effects of crowding on travel time and other variables on the one side and the impact of excluding crowding disutility on demand on the other side. Included are standee density, seat occupancy proportion, and the "load factor" to represent disutility, with a minimum "load factor" threshold set at 60% (2):

(1) $$U_m = \alpha_m + \beta_a t_{am} + \beta_h h_m + \beta_v t_{vm} + \beta_e t_{em} + \beta_c c_m$$

(2) $$U_m = \alpha_m + \beta_a t_{am} + \beta_h h_m + \beta_v t_{vm} + \beta_e t_{em} + \beta_c c_m + \beta_{lf90m} max(lf_m - 0.6, 0) t_{vm}$$

(3)

where	U_m	is the utility of mode m,
	t_{am} and t_{em}	are the access and egress times,
	h_m	is the headway between two vehicles, (h_m corresponds to $\frac{1}{V}$ in Mohring (1972)),
	t_{vm}	is the in-vehicle time,
	c_m	is the fare,
	l_{fm}	is the load factor,
	α_m	is the mode specific constant,
	β_k	are the parameters associated with the different variables.

With stated commuter preferences, Tirachini et al. (2013) identify a positive correlation between "load factor" and the value of travel time, with a threshold "load factor" of 60%. From this analysis, they derive the concept of "crowding multiplier", a factor that adjusts the value of in-vehicle time found under uncrowded conditions. Their findings indicate that the crowding multiplier increases as the "load factor" rises.

3 Data and Methodology

This chapter is dedicated to the methodology, establishing the foundation for the empirical results presented in Chapter 4. Chapter 3.1 discusses the reasons why this study settles on a quasi-experimental design and uses Revealed Preferences (RP). Chapter 3.2 defines the key terms utilized throughout the study. Chapter 3.3 outlies the data collection process. Based on essential summary statistics about each bus line, Chapter 3.4 assigns each line to either the treatment group or the control group. Chapter 3.4 also elucidates the reasons why only the specific case of the line 60 during the morning peak in the return direction will be analyzed. Chapter 3.5. details the reasons for performing a data aggregation process and explains each the steps that resulted in the reduction of the dataset' size from approx. 17 million observations to approx. 11 thousand. Chapter 3.6 specifies the differences-in-differences-in-differences method employed in the analysis. Finally, Chapter 3.7. justifies the exclusion of three potential control variables from the analysis.

3.1 Overview of Research Design

3.1.1 A Quasi-experimental Study

This study is designed as a quasi-experiment. Randomness is introduced by variations in individual circumstances that make it appear "as if" the treatment is randomly assigned (Stock & Watson, 2002). Indeed, the variations in the individual circumstances of the transportation links gathered for the analysis arose due to their specific location and the timing of the treatments. A pure random selection of links for treatment and control was impossible due to a major constraint: The analysis required detailed data collection for each transportation link, necessitating substantial preparation work by the transit companies that accepted communicating data. Consequently, links were carefully targeted based on precise criteria outlined in Figure 1, and variation was introduced by incorporating multiple control links. These control links differed in terms of their radial or transversal configurations, geographical locations, capacities, and frequencies.

3.1.2 Type of Data

The underlying analysis uses revealed preferences (RP) from secondary data. The reason for drawing on RP is twofold: First, the existing literature reported evidence regarding the effect of crowding in PT on ridership and mode choice decisions mostly using stated preferences (SP). In reality, only three major sources used RP: the first study found is of LT Marketing (1998), the second of Batarce et al. (2015) and the third of Tirachini et al. (2016). Opting for an RP approach enables therefore to address the relative scarcity of studies in this field. Secondly, SP studies are subjected to strong biases. For instance, Kroes et al. (2014) discovered that the revealed effective amount deciding to wait for a next, less crowded vehicle, to be lower than stated in survey's responses. This suggests the presence of bias, possibly the hypothetical bias if the respondent was not able to fully understand the question due to the stress of congestion, or the strategical bias, if the respondent hopes to influence the operator, as already suggested from the oldest study in the field of PT crowding by (Fowkes & Wardman, 1987). Choosing for RP therefore permits avoiding some biases SP studies are subjected to.

3.2 Definitions

- Charge: Term used by the operator to describe patronage, ridership or number of passengers. All these descriptions are used interchangeably.
- Evening peak: The evening peak starts at 16.00 and ends at 19.00[1].
- Frequency: Number of vehicles serviced operated during a defined time range in a direction. I no time range is specified; it is one hour.
- Morning peak: The morning peak starts at 07.00 and ends at 09.00[1].
- Off-peak: The off-peak includes all periods not included in the morning and evening peak.
- Passengers: Charge, patronage, ridership. All these descriptions are used interchangeably.

[1] No precise delimitation describing the start and end of morning and evening peak hours exists in the literature. It is context dependent. Accordingly, the delimitation applied by transportation planners from SBB is used.

- Patronage: Charge, ridership, number of passengers. All these descriptions are used interchangeably.
- Load factor: Percental Ratio of the patronage observed divided by the number of seats supplied.
- Ridership: charge, patronage, number of passengers. All these descriptions are used interchangeably.
- Seating capacity: Number of seats per vehicle.
- Timetable year: In Switzerland, PT operates based on timetable years. This usually lasts one year and begins on the second weekend in December of the previous year.
- Workday: Every day, weekends and vacations excluded in the Canton of Vaud in Switzerland.

3.3 Data Collection

The data collection process constituted the most significant effort of this work. It is summarized in the sections below.

3.3.1 Sources of Data

The two data provider are the Swiss federal railways (SBB) and the PT operator of the Lausanne Region (TL), located in western Switzerland. Although the datasets from both operators differ in the amount and format of the specified variables, they are both constructed based on a mixture of scheduled data and real data collected from the vehicles, such as the counting systems at the gates. The data collection was carried out at the explicit request at the address of the two transport companies in question, SBB and TL. For SBB, acceptance of the terms of use was required. For TL, signing a confidentiality contract, providing all the codes and models were prerequisites for data delivery.

3.3.2 Rail Data ruled out

SBB provided data for the five suburban train lines S5, S7, S9, S14, and S15 of the Zurich S-Bahn network for years since 2005. Regrettably, this data had to be excluded from the analysis.

Although the dataset was very rich, with each observation representing a train between two stops and including forty-nine columns of detailed information such as the train number, the seating capacity, the number of boardings and alightings, the patronage, etc., it lacked a variable describing the schedule of the rides prior to 2018. This limitation prevents us from addressing the second research question, as the observations cannot be grouped into peak and off-peak periods. Despite the possibility of constructing this information based on the train numbers included in the dataset and available in the historical timetables, this task was deemed too time-consuming. As a result, the data generously provided by SBB are unfortunately not included.

3.3.3 Data Details

TL provided data for 7 of its 38 operated links. This covers six bus lines (L2, L6, L13, L54, L60, L64) and one metro line (M2). All lines are located in the agglomeration of Lausanne counting a population of 442’036 in 2022 (Ville de Lausanne, 2024). The period covers the years 2017, 2018, 2019 and 2023. To reduce a hypothetical unobserved heterogeneity and to specifically catch commuter passengers, weekends and vacations are excluded. Each dataset provides information for one bus link over the whole period. Each observation represents a bus between two stops and provides information about: year, date, vehicle number, direction, stop number, stop name, scheduled departure time, scheduled arrival time, patronage. The number of observations ranges from 890,560 observations for L13 to 4,290,730 observations for L2. As the vehicle capacity is not specified, the number of seats is gathered manually for the 351 vehicle numbers based on manufacturer's brochures for the various models and stored in a separate file.

3.3.4 Data sampling techniques on the Network level

The first sampling stage took place before the data were even available. This was due to the considerable amount of work involved in extracting data for each line by the operator. It was consequently first necessary targeting the appropriate links to be analyzed. The choice was therefore based on 4 criteria, as depicted in Figure 1. No weighting is applied. Rather, each criterion shall be satisfied. The choice was made under the consultancy of the data provider.

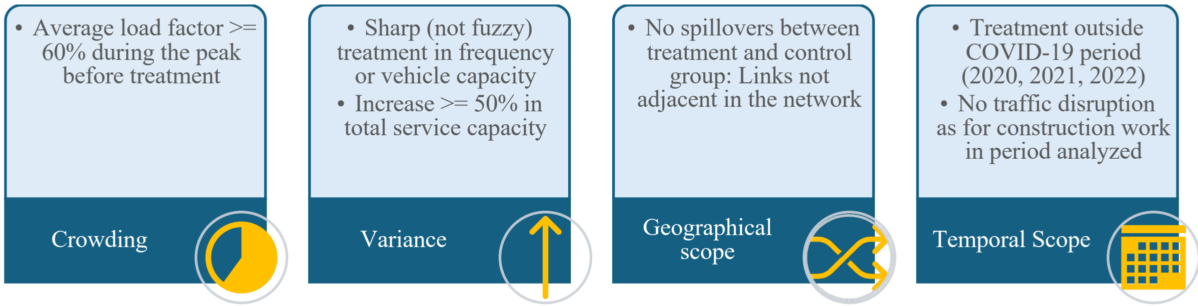

Figure 1: Selection Criteria on the network level

The first criterium stated that crowding must be present on the link under treatment. Therefore, a minimum threshold is set for the "load factor": It should reach at least 60% during peak hours on average (av.). This relatively low threshold offers some flexibility while aligning with existing literature which indicates that the negative perception of the passenger due to crowding start to negatively impact ridership once a "load factor" of 60% is reached (Fletcher & El-Geneidy, 2013; Tirachini et al., 2016). Since the data is not yet available, the fulfillment of this threshold is based on the operator's raw estimate. However, this criterion undermines the as-if randomization by suggesting a targeted selection of the analyzed links. Despite this, it is essential to ensure that a link subjected to crowding can be identified and retrieved. Second, the increase in supply shall substantial and sharp to provide sufficient variance pre-/posttreatment. Third, there must be no spillovers between the links under treatment and those under control as this a key identifying assumption of the Difference-in-Differences method, which will detailed in Chapter 3.6.3. Fourth, the treatment must have occurred outside the period affected by COVID-19[1], as this period substantially disrupted PT usage. Consequently, six bus lines and one metro line are retained for the analysis.

3.3.5 Data cleaning techniques on the link-specific level[2]

Once the data are available, several data cleaning operations at the link-specific level are conducted in preparation for analysis. Since the descriptive statistics specific to this analysis, such as the number of available seats or frequency, require retaining the raw number of

[1] COVID-19: CoronaVirus Disease provoked by SARS-CoV-2 arrived in Switzerland on 25 February 2020. It dramatically decreased PT usage. On 1 April 2022, all federal measures related to pandemics were dropped (Hintermann et al., 2023).

[2] For more detailed information, please refer to the Appendix X for the entire code, including comments.

observations, the data aggregation process that finalizes the preparation of the data for analysis is dealt separately in Chapter 3.5.

Data cleaning is performed separately for each line. The code is generally similar for each line, although the documents to be imported differ. Figure 2 provides a visual summary of the cleaning process for a given line X. In the first step, the documents from different years are merged. In the second step, a separate document providing information on the seating capacity and total capacity (seating + standing) of the vehicles is joined, using the variable "no parc" as the key. From this point on, each observation is also assigned capacity variables. Subsequently, the variable "load factor" = charge/seating capacity is created. Additionally, the format of several variables is modified to enable further operations. Several operations are also performed on the time variables to facilitate analysis and determine the travel time for each run and at each stop.

Despite this, the data provided by TL is characterized by a very high degree of quality. Missing observations had already been removed from the raw database. Additionally, for rides where the total number of boardings did not equal the total number alightings due to inaccuracies in the precision of the counting systems at the gates, the data was adapted. Specifically, the variable "charge" was slightly adjusted to a more accurate value that meets the condition of equal number of boardings and alightings per ride. Unfortunately, the calculation model for this preliminary cleaning is not known.

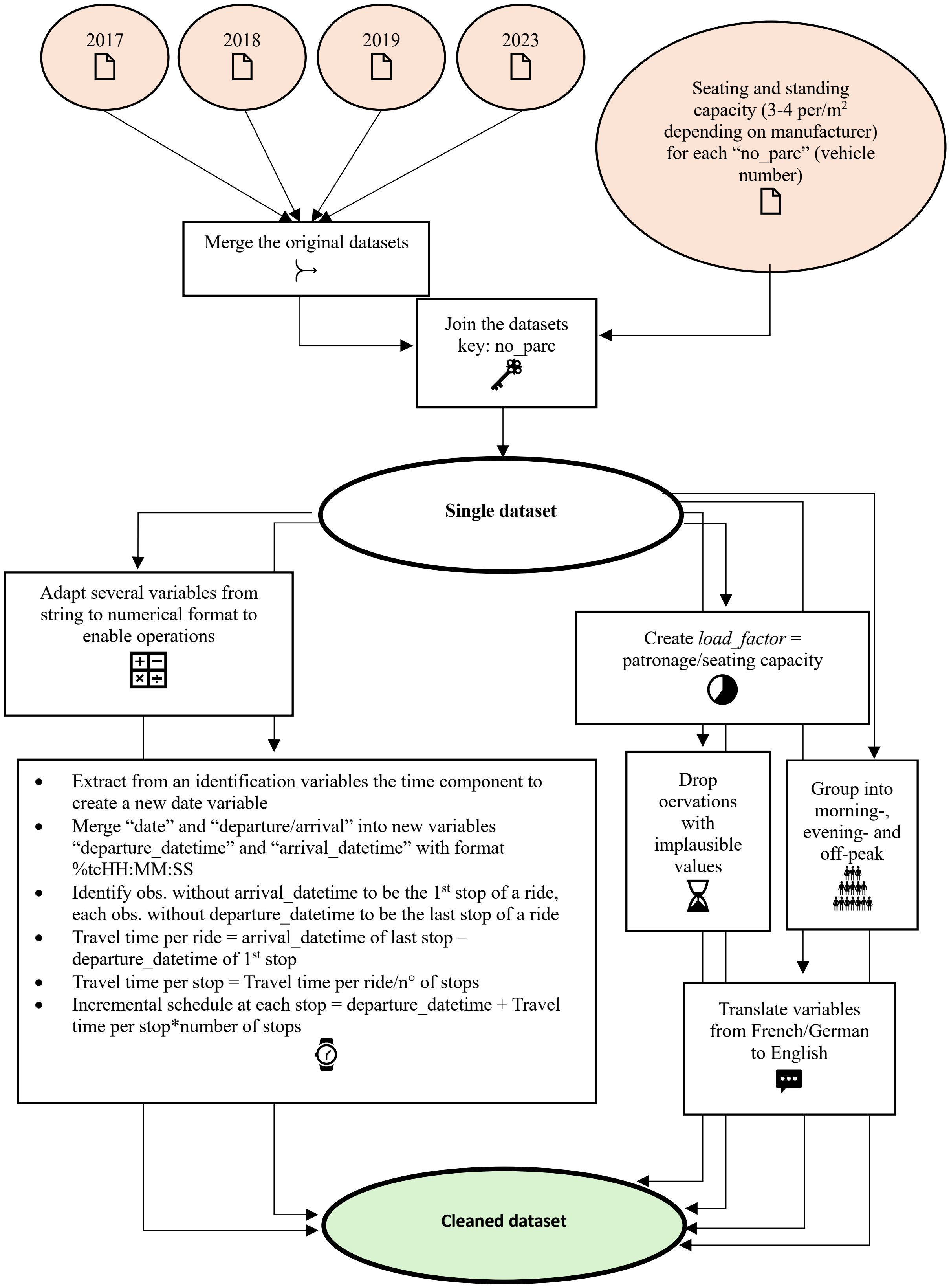

Figure 2: Heuristic scheme of data cleaning on the line-specific level.

3.4 Control and Treatment group

After the dataset is cleaned, several summary statistics about the supply variables of the bus lines are performed. The results are summarized in Table 1, Table 2 and Table 3 and present the evolution between Timetable year 2017 and Timetable year 2023. Consequently, the bus lines can be categorized into the control or treatment group based on the evolution in the number of buses and of seats supplied per day. The bus lines that do not sufficiently meet the identifying assumptions of the econometric model that will be explained in Chapter 3.6. are already excluded here from the analysis.

Figure 3: Classification of the bus lines to the Control or Treatment Group

3.4.1 Control Group

L2, L6 and L64 are retained for the control group due to their relative stability in the supply variables, as displayed in Table 1. Indeed, the number of buses and of seats per day are relatively stable, especially on L2 and L6. Regarding L64, it should be noted that both the number of buses and the number of seats slightly increased. However, this increase is not considered to be sufficient to be incorporated into the treatment group. Also, it can be observed that the number of stops on L6 increased from 24 to 26, which transgresses the stable composition condition of the econometric method that will be used for analysis (see Chapter 3.6.3). However, as this is not an extensive extension but rather a densification of the stop policy, the potential for endogeneity is considered as low.

Table 1: Overview of the Control Group

Line		2		6		64	
Year		2017	2023	2017	2023	2017	2023
At stop		Grey		Vallonnette		Planches	
Direction		R		R		R	
Number of buses per day	Per m-p (7am - 9am)	13	13	12	12	6	7
	Per e-p (4pm - 7pm)	18	18	18	18	10	13
	Per o-p	79	80	80	80	40	40
	Total	110	111	110	110	56	60
Number of seats per day	Per m-p (7am - 9am)	550	550	530	530	340	400
	Per e-p (4pm - 7pm)	700	750	800	800	210	250
	Per o-p	3450	3500	3370	3370	1250	1350
	Total	4700	4800	4700	4700	1800	2000
Charge per day	Per m-p (7am - 9am)	310	280	500	400	230	270
	Per e-p (4pm - 7pm)	190	220	400	380	120	200
	Per o-p	700	700	1300	1100	250	300
	Total	1200	1200	2200	1880	600	770
Travel time	Per m-p (7am - 9am)	34	35	35	37	15	15
	Per e-p (4pm - 7pm)	38	38	40	41	13	14
	Per o-p (10pm - 11pm)	33	35	26	28	10	11
Number of stops		26	26	24	26	12	12

3.4.2 Treatment Group Candidates

L13 and L60 have been selected as potential "candidates" for the treatment group, reflecting their significant increases in seating capacity as documented in Table 2. These lines remain candidates pending further analysis of crowding patterns on each, which will be detailed in Section 3.4.4.

Foremost, Table 2 provides major information regarding the type of treatment occurred on L13 and L60. On L13, we can depict a decrease in frequency coupled with an increase in seating capacity during both the morning peak (MP) and the evening peak (MP). On L60, we can depict a stable frequency coupled with an increase in seating capacity in the MP; and an increase in frequency coupled with an increase in seating capacity in the EP. The dimensions are notable, as both lines experienced an approximate doubling in service capacity, if not more.

However, the fact that we observe two treatments on L13 and L60 during the EP, both in term of frequency and vehicle capacity, might confound the estimations. Before taking a final decision on the potential erasure of the EP, it is however necessary also observing the crowding pattern applying before the treatment, as explained in Section 3.4.4.

Moreover, we can also infer that both L13 and L60 exhibit advantages and disadvantages: L13 is characterized by a stable composition over the periods as the number of stops did not change. However, L13 also shows very low dimensions for “charge per day” compared to the control links and a small number of stops which subjects it to stronger bias. L60, on the other hand, shows dimensions for “charge per day” that are comparable with those displayed for the control lines. However, L60 has an unstable composition across the periods as the number of stops decreased from 33 to 17 and “Travel time” also slightly decreased by a few minutes.

Table 2: Overview of Candidates for the Treatment Group

Line		13		60					
Year		2017	2023	2017			2023		
At stop		St-Francois		Froid. Croisee (north of the line)	Coppoz-Poste (middle of the line)	Bellevaux (south of the line)	Froid. Croisee (north of the line)	Coppoz-Poste (middle of the line)	Bellevaux (south of the line)
Direction		R		R					
Number of buses per day	Per m-p (7am - 9am)	10	8	8			8		
	Per e-p (4pm - 7pm)	15	12	6			10		
	Per o-p	51	53	25			27-29		
	Total	76	73	39			45-47		
Number of seats per day	Per m-p (7am - 9am)	110	237	260	350	350	600	600	600
	Per e-p (4pm - 7pm)	170	355	260	260	260	750	750	750
	Per o-p	570	1608	1180	1090	1090	2250	2250	2250
	Total	850	2200	1700			3600		
Charge per day	Per m-p (7am - 9am)	70	80	60	300	380	150	350	310
	Per e-p (4pm - 7pm)	70	80	20	180	180	40	220	200
	Per o-p	130	180	170	400	440	200	420	390
	Total	270	340	250	880	1000	390	990	900
Travel time	Per m-p (7am - 9am)	10	10	45			41		
	Per e-p (4pm - 7pm)	10	9	43			41		
	Per o-p (10pm - 11pm)	7	6	39			37		
Number of stops		8	8	33			17		

3.4.3 Eliminated group

L54 and M2 are excluded from the analysis. L54 has undergone considerable expansion as the number of stops increased from 10 to 20. To avoid clearly violating the identifying assumption of stable pre- and post-treatment composition (see Chapter 3.6.3), it is excluded from the analysis. Secondly, M2 is not retained for comparison, as it is considered a different mode of transport than buses. We may note that there is a debate about the existence of a rail bonus (Axhausen et al., 2001), which describes a preference for the rail option over the other alternatives.

Table 3: Overview of non-selected Lines

Line		54		M2	
Year		2017	2023	2017	2023
At stop		Renens-Gare nord		Lausanne-Flon	
Direction		R		A	
Number of buses per day	Per m-p (7am - 9am)			50	56
	Per e-p (4pm - 7pm)			68	82
	Per o-p			177	227
	Total			295	365
Number of seats per day	Per m-p (7am - 9am)			3000	3500
	Per e-p (4pm - 7pm)			4000	5000
	Per o-p			10500	13500
	Total			17500	22000
Charge per day	Per m-p (7am - 9am)			5500	6000
	Per e-p (4pm - 7pm)			6000	7500
	Per o-p			12000	16000
	Total			3950	5150
Travel time	Per m-p (7am - 9am)			19	19
	Per e-p (4pm - 7pm)			19	19
	Per o-p (10pm - 11pm)			18	18
Number of stops		10	20	14	14

3.4.4 Choice of peak period(s) and direction(s)

After categorizing the bus lines into the control and treatment group (candidates), the next step involves selecting the specific peak periods for detailed analysis. There are two approaches to consider: the first is to focus exclusively on either the morning peak (MP) or the evening peak (EP) in each regression. The second approach involves incorporating interactions for both the MP and EP, which necessitates simultaneous analysis of both directions of travel.

However, this dual approach introduces complexities due to the directional nature of the routes. For instance, L13 and L64 both operate from the outskirts to the city center, leading to differences in crowding levels between the inbound and outbound journeys. Figures 4 and 5 illustrate this asymmetry, suggesting that crowding phenomena may not mirror each other in both directions. Based on this, the analysis will be confined to a single peak period and one direction to ensure clarity and precision in the regression results.

In a second step, a choice between the MP and the EP must be made. One disadvantage of the MP is that passenger flows on a workday would potentially demonstrate more evident spatiotemporal heterogeneity, as observed by Lu et al. (2024), compared to the EP. Since this study focuses on internal factors and ignores the external factors of ridership (see Chapter 2.1), focusing on the evening might help in reducing the potential bias of these unobserved factors.

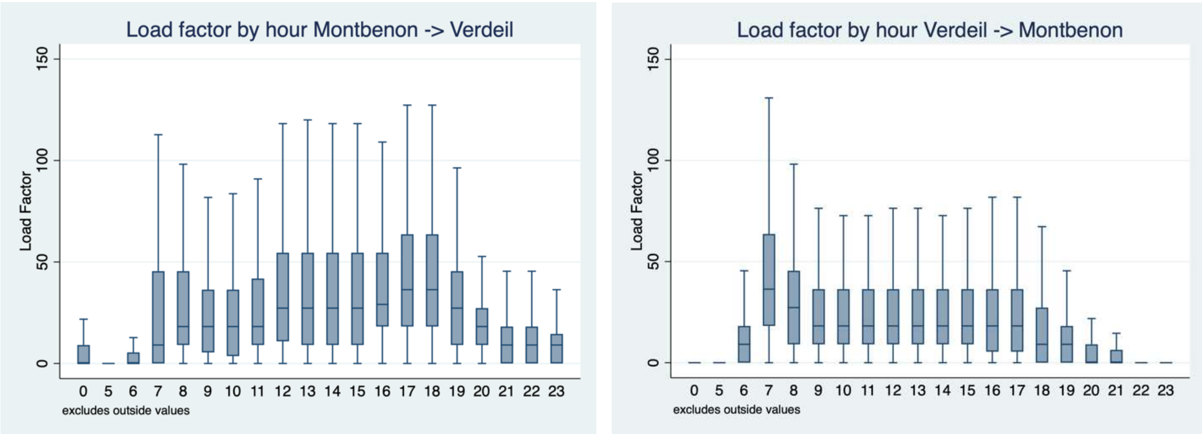

Figure 4: L13: Load factor per hour and direction before treatment (aggregated over entire link).

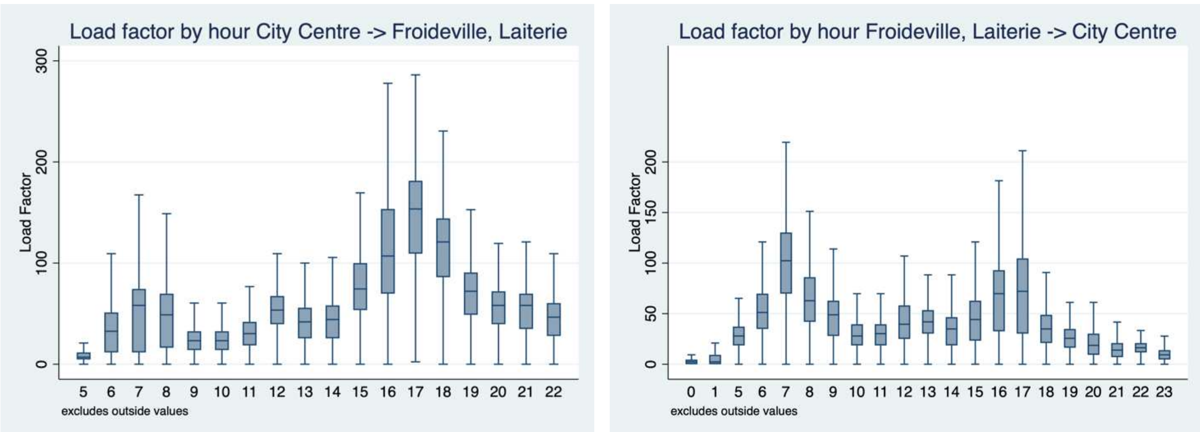

Figure 5: L60: Load factor per hour and direction before treatment (aggregated over entire link)[1]*.*

Figure 4 and Figure 5 provide boxplots representing the distribution of the “load factor” observed on L13 and L60 before treatment in both directions. Accordingly, the following points are noteworthy:

[1] The first, respectively the last stop of L60 is stated as City Centre as it changed from *Lausanne-Flon* to *Riponne-Maurice Béjart* on the 12.12.2023.

- First, the “load factor” varies greatly according to direction and hour. This confirms the previously mentioned argument that a one-way analysis of both peaks would be irrelevant.

- Second, the distribution of the “load factor” is slightly more compact during the MP than during the EP. This supports the definition provided in Chapter 3.2 that the MP would last for two hours, while the EP would last for three hours.

- Concerning L13, no substantial crowding is observed. Hence, the selection criterion set in Figure 1 of a minimum 60% threshold to consider any crowding disutility is not filled.

- Concerning L13 moreover, it is worth noting that the data shows equal averages as well as 25%th- and 75%th- percentile values over several hours across the workday. It is assumed that the vehicles used prior to the treatment did not have automatic counters at gates and that counting was carried out by hand at sporadic intervals, with values subsequently extrapolated for other periods. This calls into question the quality of the data and the accuracy of the subsequent analysis with L13.

- Concerning L60, crowding varies strongly depending on the direction. In the downward direction from the upper outskirts to the city center, over 50% of the observations between 07:00 and 07:59 show a “load factor” exceeding 100%, whereas solely slightly more than a quarter of the observations reach this level between 17:00 and 17:59. In the upward direction from the city center to the upper outskirts, no substantial crowding pattern is observed in the morning hours. However, the av. “load factor” exceeds 100% between 16.00 and 18.59, and even 150% between 17.00 and 17.59.

Consequently, it has become necessary to exclude L13 from the analysis. Several factors contribute to this decision: L13 exhibits low passenger counts, experiences two simultaneous interventions, shows minimal evidence of crowding, and the data quality is questionable. The analysis will instead focus on L60 during the morning peak (MP) in the direction from Froideville, Laiterie towards the City Centre, also referred to as the return direction (R). Despite lower crowding levels compared to the evening peak (EP) in the opposite direction, the singular nature of the treatment—namely the increase in capacity without concurrent frequency changes—offers a clear advantage. This distinct setup eliminates some potential confounding effects, thereby strengthening the validity of analyzing this specific scenario.

3.5 Data aggregation

After cleaning and describing the datasets, several data aggregation operations are conducted to eliminate confounding effects and to condense the dataset size, facilitating more efficient regression analyses within a manageable timeframe. After each aggregation step, a descriptive table for “charge” was calculated in the code to ensure that the operations do not cause any major distortion of the data. Figure 6 provides a visual summary of the aggregation process. Before starting the aggregation process, the cleaned and described datasets of L2, L6, L60 and L64 are merged. As each line rides’ ID have conflicting numbers, the reallocation of a unique “id” for each ride is necessary. Firstly, the direction is chosen: “R” for the MP. By accident, it turns out that “R” describes the best the MP experienced for all lines. As a result, half of the observations are erased. Secondly, 17 stops occurring on L60 are erased. The detailed explanation of the reasons for this choice is provided in Section 4.1.2. Thirdly, each ride’s last observation is erased. This step is justified by the fact that these instances always have value "0" for “charge+ and “load factor.” Indeed, each observation corresponds to a stop. But the value for charge and “load factor” captures the value between the specified stop and the next stop. And as the last stop per ride has no subsequent stop, “charge” and “load factor” is always display “0”. This adjustment has the effect of increasing the av. charge and “load factor” per ride, especially for the lines with few stops. Fourthly, only five months are kept for analysis. This operation is performed to make the model more efficient errors. Fifthly, the dataset is collapsed. This is the major operation of the aggregation process: Data are aggregated by the ID of each ride. As a result, each observation therefore no longer corresponds to a stop, but to a ride. Therefore, the values displayed for “charge” and “load factor” represent from now on the av. over the trip. Sixthly, a proxy is set for the MP (07:00 – 07:59) and the OP (10:00 – 10:59). Only the rides inside these time ranges are kept in[1]. Seventhly, random sampling is applied to test the robustness of our estimates as we increase the sample size. As a result, the

[1] The retained time per ride is the departure time at the first stop. E.g. a 30-minutes ride departing at 07.59 with arrival time is 08.29 is kept in, whereas a 30-minutes ride departing at 06.59 with arrival time is 07.29 is erased. This measure is essential to avoid breaking down the structure of the rides and keep a longitudinal consistency.

number of observations is reduced from approx. 17 million observations to approx. 11 thousand observations.

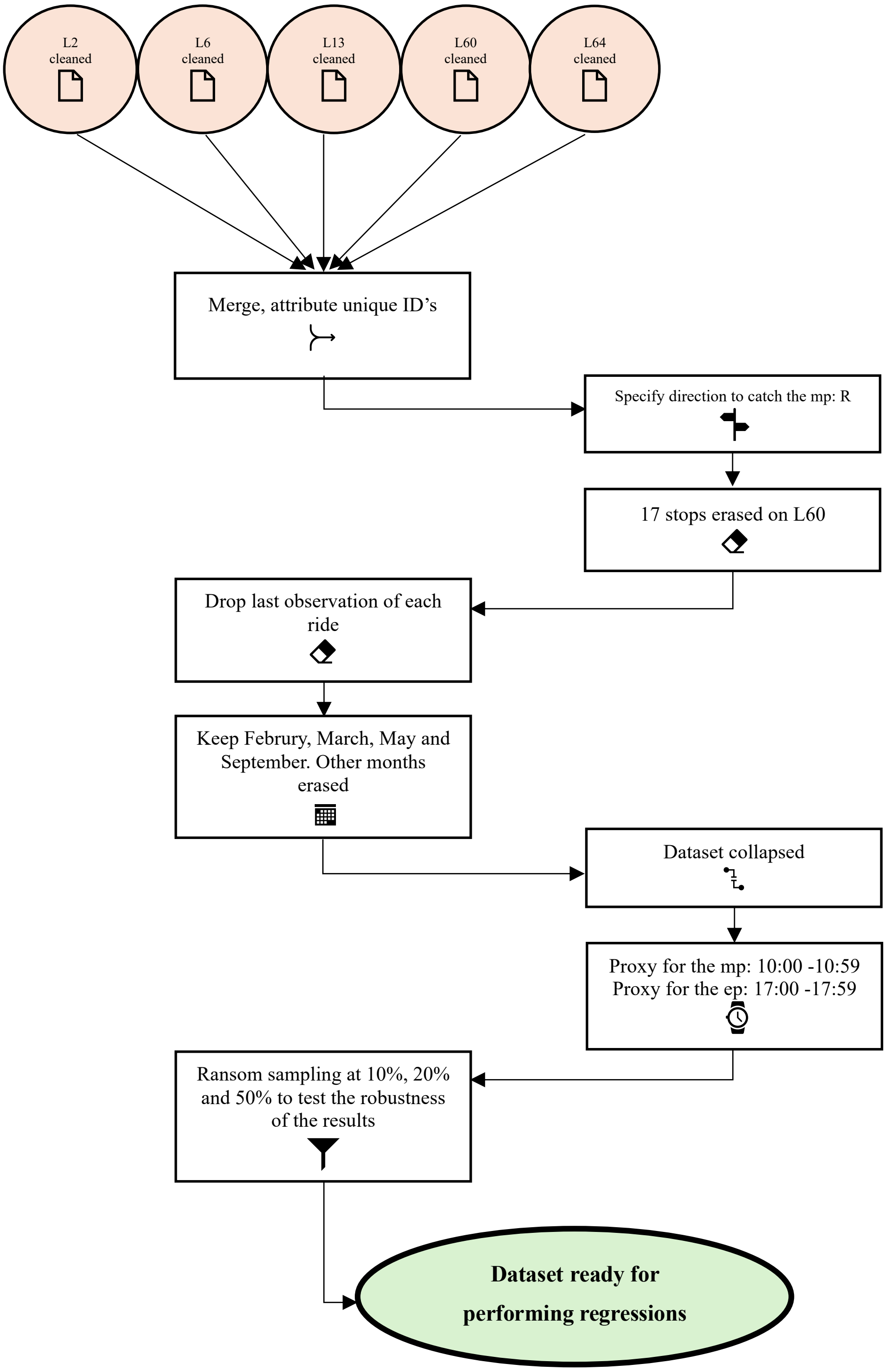

Figure 6: Heuristic scheme of data aggregation

3.6 Model

The analysis relies on differences-in-differences-in-differences estimations, also called triple difference estimations (DiDiD).

3.6.1 Rationale for selecting this Method

The choice of difference estimations for answering the research questions is founded on three key reasons: Firstly, the double difference method is prevalently applied in quasi-experimental research designs and is recognized as a fundamental analytical approach in Econometrics (Stock & Watson 2002). Secondly, difference estimations are particularly useful when randomization is infeasible, which is the case in this study (refer to Chapter 3.3 for details). Thirdly, this method is ideally suited to address both research questions within a unified analysis framework. The first objective is to determine how changes in PT service capacity affect ridership, achievable through a two-dimensional difference estimation (time, treatment). The second objective is to assess the extent to which the increase in ridership can be attributed to a decrease in crowding externalities. This is facilitated by a three-dimensional difference estimation (time, treatment, peak), allowing for additional differentiation based on whether the period experienced crowding prior to the capacity increase.

3.6.2 The Double Difference Estimator

Basically, a one-dimensional regression considering the effect of a treatment on a dependent variable, could be described by an ordinary least squares regression. In contrast, the classical differences-in-differences estimation does not compare the post-treatment evolution of the dependent variable in isolation. Instead, it adds a second temporal dimension comparing this evolution to that of another group of control. Consequently, the unidimensional effect is either attenuated or reinforced by incorporating this second dimension. Consequently, this dual-dimensional approach allows for the calculation of the double difference estimator as follows:

$$\hat{\beta}^{diffs-in-diffs} = \left(\overline{Y}^{L60=1,\,after=1} - \overline{Y}^{L60=1,\,after=0}\right) - \left(\overline{Y}^{L60=0,\,after=1} - \overline{Y}^{L60=0,\,after=0}\right)$$

If the treatment is randomly assigned, $\beta^{diffs-in-diffs}$ is an unbiased and consistent estimator of the causal effect (Stock & Watson, 2002).

3.6.3 The Triple Difference Estimator

The triple difference method extends the classical difference-in-differences approach by incorporating an additional third dimension. Introduced by Gruber in 1994, this methodology addresses potential biases that arise when significant disparities between the treatment and control groups persist, which classical difference-in-differences estimations may not fully account for (Olden & Møen, 2022). This third dimension justifies the enrichment of the model, allowing for a more nuanced accounting of these disparities. Specifically, the triple difference approach involves comparing the treatment group not only with the initial control group but also with an alternative control group. This second control group is selected to more closely match the characteristics of the treatment group, thereby providing a more precise measure of the treatment's impact (Olden & Møen). In our study, this third dimension is represented by a dummy variable for the morning peak (MP), enhancing the robustness of our model. Consequently, the triple difference estimator can be calculated as follows:

$$\begin{aligned}
\hat{\beta}^{triple-in-diffs} = & \left[\left(\overline{Y}^{L60=1,MP=1,After=1} - \overline{Y}^{L60=1,MP=1,After=0}\right)\right. \\
& \left. - \left(\overline{Y}^{L60=0,MP=1,After=1} - \overline{Y}^{L60=0,MP=1,After=0}\right)\right] \\
& - \left[\left(\overline{Y}^{L60=1,MP=0,After=1} - \overline{Y}^{L60=1,MP=0,After=0}\right)\right. \\
& \left. - \left(\overline{Y}^{L60=0,MP=0,After=1} - \overline{Y}^{L60=0,MP=0,After=0}\right)\right]
\end{aligned}$$

Summarized, the triple difference estimator calculates the difference between two difference-in-differences estimators (Olden & Møen, 2022).

3.6.4 Identifying assumptions

For the triple difference to be an unbiased and consistent estimator of the causal effect, several identifying assumptions need to be satisfied. Firstly, the intervention must not be related to the outcome at the baseline, ensuring the exogeneity of the model. This criterion is discussed in

Chapter 2.2.3, where both the theoretical framework and the practical claim for the treatment's exogeneity on L60 are examined. Second, for difference estimations to be considered reliable, it’s important that the treatment and control groups follow similar trends before the treatment. This helps ensure that the results we see are actually caused by the treatment (Lechner, 2011). Before we do any regression analysis, we will check this by using a method from Riveros-Gavilanes (2023). This method looks at how significant the trends were before the treatment was applied. The main purpose of checking these trends is to depict the potential presence of anticipation on the dependent variable before the treatment, causing the estimates to be biased (Callaway & Sant’Anna, 2021). Fortunately, when we use triple difference estimations, we don’t have to check this for every single group separately (Olden & Møen, 2022). This requirement is the same as what we use for double difference estimations. Third, the composition of the groups must remain consistent throughout the observation periods, as changes in composition can complicate the interpretation of the results (Callaway & Sant’Anna, 2021). For this reason, L54 was excluded from the analysis. Unfortunately, this assumption is not met for L60, but strategies to mitigate this issue are discussed in Section 4.1.1 and Section 4.1.2. Fourth, there should be no spillover effects between the groups (Butts, 2021). Although all lines are located within the same network, they are considered independent apart from Stop “Bessières”, where both L6 and L60 intersect. However, this assumption of no spillovers does not hold for the third dimension, as passengers who traveled during the off-peak period before the treatment may be the same ones traveling during the peak period afterward. This limitation will be discussed in detail in Chapter 5.3.

3.7 Control variables

Price, travel time, reliability: Although the definition of the travel time variable can vary from taking into account only the time spent in the vehicle (Parry & Small, 2007), to the addition of access and egress times (Tirachini et al., 2013), consensus dominates in the literature that these variables are major explanatory variables of PT-Utility. The reason for assessing the addition of control variables is twofold: First, control variables with substantial explanatory power reduces the residual variance and increases the precision of the causal effect of interest. Second, including more regressors might account for differences between groups and make the parallel trend identifying assumption needed for identification more credible (Olden & Møen, 2022).

3.7.1 Price

No price variable can be retrieved or constructed from the available datasets. As Swiss PT systems do not use a card identification system for each passenger as for example practiced in the Netherlands, it is impossible to infer a relation between a boarding or alighting observed and the ticket/subscription purchased by the passenger concerned. Hence, it is also impracticable including this variable. However, it's worth noting that nominal PT prices have remained stable from Timetable year 2016 to Timetable year 2023 (Alliance SwissPass, 2023).

3.7.2 Travel time

Travel time is established for each ride and at each stop thanks to the data cleaning process performed in Section 3.3.5. Moreover, this variable demonstrates an interesting variation between the peak and off-peak periods, as displayed Table 1 and Table 2. This is the result of the increased passenger in- and outflow during the peak and, above all, of the scheduled increased travel time because of the congestion imposed by road traffic during the peak. However, this variable resulted to be highly multicollinear with several other variables, as will be shown in Chapter 4.4. It's contribution in increasing the fit will be proven to be low, while considerably inflating the variance in the pallet of specifications tested. Unfortunately, the inclusion of travel time as control variable is therefore rejected.

3.7.3 Service reliability

Implementing a control variable considering reliability or service quality appeared primarily difficult as the datasets provided by TL do not communicate about the real effective departure and arrival time of the rides. Nevertheless, it provides the vehicle number, whereas this information was used to construct the "load factor" variable (see Section 3.3.5).

The interesting element is that a non negligeable number of the observations for "vehicle number" have value 0. The data provider explained this identifier to be used for trips where the data was not correctly retrieved by the vehicle's on-board systems, or where the retrieved data was of poor quality. In such cases, the ride data would be imputed with high-quality data from a similar period, including the same section, time of year, day of the week, and time of day. The operator therefore assumes the ride to have occurred. At the same time, it could be retrieved from code-related descriptives, that the pattern of frequency demonstrated an implausibly stable and equal number of trips observed across each workday, suggesting a perfect bus ride occurrence of 100%, as according to the schedule. This state of matter calls for a more detailed study.

To test whether bus numbers equaling 0 might represent rides that did not occur as scheduled, a logical reasoning would suggest that a not-occurred ride would transfer the patronage, if not entirely then at least partially, to the next ride. Consequently, the following null hypothesis is tested:

"A vehicle number equaling 0 leads to a statistically significant higher number of passengers on the next ride w. r. t. other rides. "

This involves creating a variable to capture the patronage for the subsequent trip, aiming to analyze data reliability and imputation practices. Accordingly, instances following suspected rides, and not equaling zero themselves, are flagged. This flagging allows for a comparison of the observed patronage with trips following those without data issues. Using these flagged variables, t-tests are performed to determine if there is a significant difference in patronage between the groups.

Table 4: T-Tests Comparing Trip Patronage Following Suppressed Rides[1]

1st Two-sample t test with equal variances on L60, direction R

```
------------------------------------------------------------------------------
  Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
      0 |  28,927    16.27751    .0712584    12.11959    16.13784    16.41718
      1 |   1,323    17.36153    .4247214    15.44842    16.52833    18.19473
---------+--------------------------------------------------------------------
        |  30,250    16.32492     .070638    12.28574    16.18647    16.46338
---------+--------------------------------------------------------------------
   diff |           -1.084015    .3453574                -1.76093   -.4071002
------------------------------------------------------------------------------
   diff = mean(0) - mean(1)                                      t =  -3.1388
: diff = 0                                      Degrees of freedom =    30248
: diff < 0                 Ha: diff != 0                 Ha: diff > 0
(T < t) = 0.0008         Pr(|T| > |t|) = 0.0017          Pr(T > t) = 0.9992
```

2nd Two-sample t test with equal variances on all 5 lines, direction R

```
------------------------------------------------------------------------------
  Group |     Obs        Mean    Std. err.   Std. dev.   [95% conf. interval]
---------+--------------------------------------------------------------------
      0 | 239,362    12.36807    .0180257     8.81899    12.33274     12.4034
      1 |   9,288    12.50046    .0983341    9.476873    12.30771    12.69322
---------+--------------------------------------------------------------------
        | 248,650    12.37301    .0177369     8.84446    12.33825    12.40778
---------+--------------------------------------------------------------------
   diff |           -.1323943    .0935354               -.3157212    .0509326
------------------------------------------------------------------------------
   diff = mean(0) - mean(1)                                      t =  -1.4154
: diff = 0                                      Degrees of freedom =   248648
: diff < 0                 Ha: diff != 0                 Ha: diff > 0
(T < t) = 0.0785         Pr(|T| > |t|) = 0.1569          Pr(T > t) = 0.9215
```

Table 4 presents the results for L60 specifically in direction "R", as well as for all five lines, also in direction "R". Although the 1st t-test demonstrates a statistically significant difference between both groups at the 0.001 level, it is necessary to acknowledge that the effect is too modest to support the construction of a convincing proxy variable representing the service reliability. Furthermore, when all five lines are considered as in the 2nd t-test, the null hypothesis–that rides following those suspected to have been suppressed would exhibit higher patronage–is rejected at the 0.1 level. Consequently, the option of incorporating a proxy control variable for describing the service reliability is also discarded.

[1] Please excuse the poor visual quality of this table. All automatic means tried all failed in displaying the results with better visual quality.

4 Empirical results

This chapter applies the methods and models outlined in Chapter 3. Section 4.1 details the treatment applied to Line 60, including the necessary adjustments, along with the correlation matrices and the testing of parallel trends prior to treatment. Section 4.2 presents the regression results and explains the implications of the coefficients. Section 4.3 offers a methodological interpretation of the findings. Section 4.4 evaluates the efficiency of the estimated coefficients.

4.1 Treatment on L60

The capacity expansion on Line 60 (L60) was achieved by introducing new double-decker buses. As a result, the number of seats available is roughly doubled (see Table 2). This upgrade, effective from August 2019, consists of the introduction of "Enviro 500" double-decker buses. Nonetheless, the deployment of single-decker buses like the "Lion's City GL" and the "Citaro C2 G" continued to some extent. Despite the augmentation in seating capacity, the frequency of service during both the morning peak (MP) and off-peak (OP) periods remained constant.

4.1.1 Evolution of Ridership during the Morning Peak

The observed value for the dependent variable “charge” in Figure 8, Figure 9 and Figure 10 demonstrate a different pattern depending on stop and period analyzed. Indeed, the trend indicates an increase in charge for all observed periods at the stops “Froideville, Croisée” and “Coppoz, Poste” following the increase in capacity, whereas it tends to stagnate or decrease at stop “Bellevaux”.

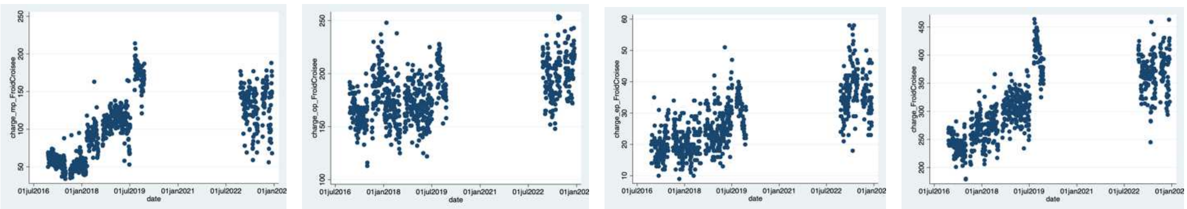

Figure 7:Charge per day during the MP, OP, EP and in total at Froideville, Croisée

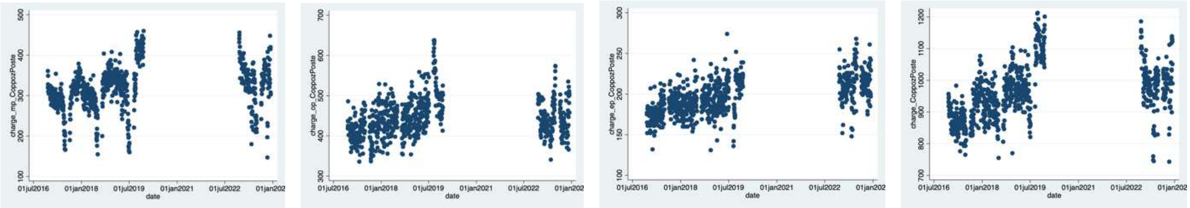

Figure 8: Charge per day during the MP, OP, EP and in total at Coppoz, Poste

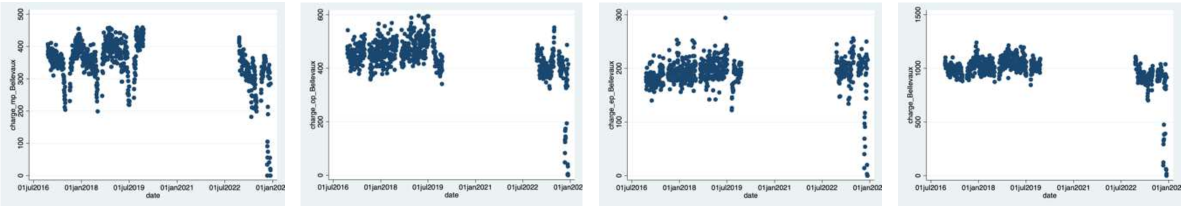

Figure 9: Charge per day during the MP, OP, EP and in total at Bellevaux

4.1.2 Necessary Adjustments

The evolution depicted at stop "Bellevaux" in Figure 9 is explained by the changes in the stop policy on this line, as shown in Figure 10. Many stops between "Coppoz-Poste" and "Lausanne-Flon"/"Riponne-M.Béjard" have been removed. This change has had the positive effect of reducing travel time by four minutes, but it has also diminished service coverage. It is assumed that passengers from the outskirts located at the northern stops of the line, such as "Froideville, Croisée" were minimally impacted by this modification. It is assumed these passengers mostly travel directly to the city center and not to the erased stops. Unfortunately, this hypothesis cannot be tested as the origin-destination matrix of passengers is unknown. However, the effect seems being particularly strong on the southern part of the line, such as at "Bellevaux". Consequently, it was necessary to make a constraining choice to limit these effects, which might complicate the further steps of the analysis. Therefore, all stops between "Coppoz-Poste" and "Bellevaux" are excluded from the analysis due to the significant changes they underwent, which clearly violate the stable composition assumption of the difference estimation (see Chapter 3.6.3). While the assumption is also transgressed in the retained area, it is assumed to a lesser extent.

Stop/Period	12.12.16-07.12.18	10.12.18-05.07.19	26.08.19-13.12.19	12.12.22-13.12.23
1. Froideville-Lait	●	●	●	●
2. Froid. Croisee	●	●	●	●
3. Froid. Village	●	●	●	●
4. Rusteriaz	●	●	●	●
5. Bottens, Croisee	●	●	●	●
6. Chalet-Coucou	●			

7. Bretigny/Morrens	●	●	●	●
8. Cugy VD, Moulin	●	●	●	●
9. Cugy VD, Village	●	●	●	●
10. Cugy VD, Poste	●	●	●	●
11. Cugy, Cavenette	●	●	●	●
12. Le Mont, Budron	●	●	●	●
13. Le Mont Fougeres	●	●	●	●
14. Le Mont, Etavez	●	●	●	●
15. Grand-Mont	●	●	●	●
16. Coppoz	●	●		
17. Coppoz-Poste	●	●	●	●
18. Petit-Mont	●	●		
19. Cotes	●	●		
20. Martines	●	●		
21. Rionzi	●	●		
22. Grangette	●	●		
23. Bellevaux	●	●	●	●
24. Foret	●	●		
25. Motte	●	●		
26. Vieux-Moulin	●	●		
27. Grande-Borde	●	●		
28. Memise	●	●		
29. Tunnel	●	●	●	●
30. Place du Nord	●	●		
31. Bessieres	●	●	●	
32. Rotillon	●	●		
33. Lausanne-Flon	●	●	●	
34. Riponne-M.Bej.				●

Figure 10: Evolution of the stop policy of L60 in the direction R. Described stops in bold. Erased stops in red.

4.1.3 Correlation Matrix

If multiple regressors are strongly multicollinear, the coefficient on one or more of these regressors will be imprecisely estimated and the standard errors will increase (Stock & Watson, 2002). Consequently, two correlation matrixes are presented in Table 5 and Table 6 below. The aim of a correlation matrix is to observe if potential predictors are multicollinear with one another. The reason for drawing two tables is justified by the fact that the underlying method uses dummy variables. Without diving into the details, it is generally assumed that tetrachoric correlations produce more precise estimates for binary variables (Ethington, 1987). However, continuous variables as "load factor" and "travel time" were also widely thematized in the upstream analysis. Therefore, the more widely used Pearson's correlation is used to measure linear association between these continuous variables (Liu, 2019) and to justify the exclusion of "load factor" from the analysis.

	i.aug19	i.line60	i.mp
i.aug19	1		
i.line60	0.0815***	1	
i.mp	-0.1397***	0.1764***	1

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

Table 5: Tetrachoric Correlation matrix for binar3y variables

	TT	load factor	i.MP	bus per hour
TT	1			
load factor	-0.00278	1		
i.mp	0.147***	0.694***	1	
bus per hour	0.432***	0.0650***	0.347***	1

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

Table 6: Pearson Correlation matrix

Two major elements can be deducted from Table 5 and Table 6. First, no problematic multicollinearity is to be inferred between the three binary variables in Table 3. This is a major milestone as it allows to test and perform regressions based on these variables and to hope for consistent and efficient estimates. Second, the correlation between "load factor" and "i.mp" is 0.694 and stat. sig at 0.001 level, indicating a strong linear relationship between the crowding observed inside vehicles and the morning peak[1]. This result makes a lot of sense. However, it also means that a choice must be made between both variables. As the underlying differences estimations utilizes a dummy variable for both the treatment dimension and the time dimension, it is also necessary selecting the dummy variable "i.mp" and thus erasing the "load factor" from the analysis.

[1] Although "i.mp" is a dummy, it was necessary also including it in Table 4 to demonstrate that a choice must be made between the "load factor" and the "i.mp" because of the obvious multicollinearity between these two variables

4.1.4 Parallel Trends

A natural way to check for the validity of difference estimation design is to examine pre-treatment outcomes, whereas similar pre-treatment trends lends credibility to the fulfillment of the parallel trends assumption (Angrist & Pischke, J-S., 2009). As proposed by Bertrand et al. (2004), we test the presence of differential pre-trends by running a regression of the outcome on a time trend interacted with the treatment indicator using pre-treatment data. More specifically, we run the same DiD-regression as will be run in the second model (M2) in the regression results.

Table 7: Test Parallel Trends Pre- Treatment

	(1)	
VARIABLES	DiD charge	p-value -----
time_trend_before	0.00157 (0.0336)	0.963
1.line60	7.978*** (2.682)	0.003
1.line60#c.time_trend_before	0.188 (0.170)	0.270
1.line6	7.809*** (0.664)	0.000
1.line64	0.515 (1.022)	0.614
Constant	15.25*** (0.622)	
Observations	970	
R-squared	0.144	

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

The results in Table 7 provide limited evidence supporting the validation of the parallel trends assumption. Although the pre-treatment difference-in-differences interaction term is ns at the 0.270 level, indicating that any anticipation effects can be excluded, the R-squared is very low and complicates the interpretation of the table. This result only slightly strengthens the argument that the operator's intervention is not related to the baseline outcome. Also, the sample size is relatively small, which also impacts the interpretation of the results and tend to increase the p-values, thereby reducing the significance of the estimates. Moreover, L60 has a relatively high standard error, which also contributes to augmenting the p-value and thus leading to ns values. Moreover, it must be recalled the period after treatment was evidently influenced by the

COVID-19 pandemic as clearly visible in Figure 13, which perturbed the post-treatment data. As a result, caution should be exercised when interpreting these findings.

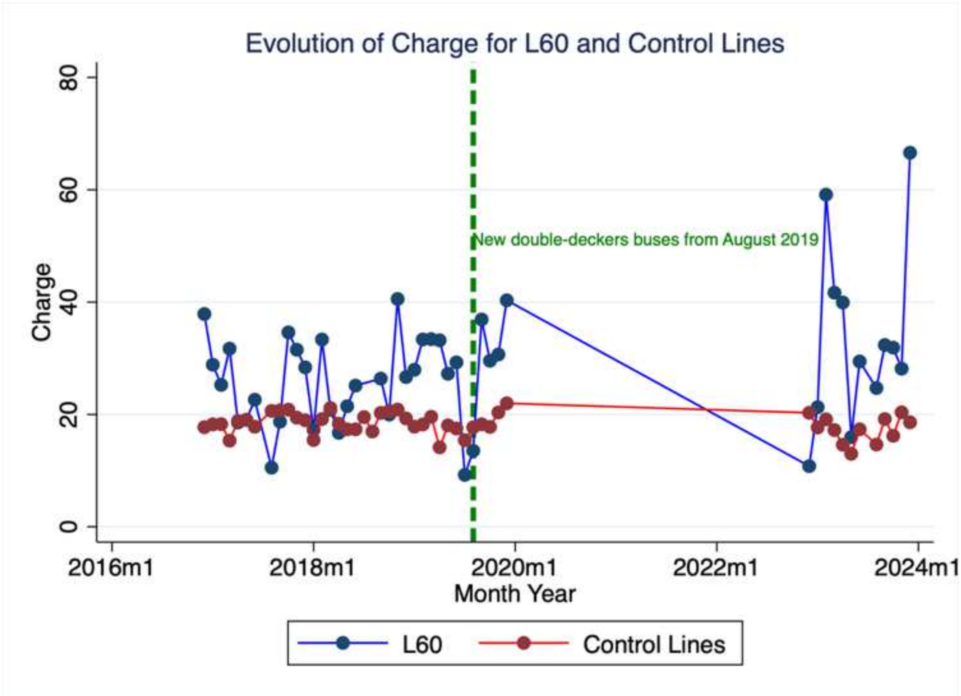

Figure 11: Pre-/Post-Evolution of the average charge per month in both L60 and Control Lines

4.2 Final Results

The results are presented in Table 8. Five regressions are conducted, each one incorporating an incremental addition of variable interactions. The objective is to analyze the Average Treatment Effect on the Treated (ATT), which is represented by the introduction of double-decker buses, on the dependent variable "charge".

4.2.1 Regression Results

All variables used in these regressions are dummy variables and depicted with a "1." in front in Table 7. Note that the coefficients do not indicate the slope coefficient, but the average difference in the dependent variable across all the binary combinations specified. From now on, the variables are displayed in **bold** to facilitate the understanding of the results' explanations

Table 8: Regression Results

	(M1)	(M2)	(M3)	(M4)	(M5)
	OLS	DiD	DiDiD	DiDiD	DiDiD with line-specific constants
VARIABLES	charge	charge	charge	charge	charge
1.aug19	6.243***	-1.547***	-1.547***	-0.112	-0.112
	(1.016)	(0.203)	(0.203)	(0.132)	(0.117)
		5.469***	5.469***	-1.864***	1.926***
		(0.435)	(0.435)	(0.149)	(0.155)
1.aug19#1.line60		7.791***	-12.07***	2.138***	2.138***
		(1.036)	(0.548)	(0.345)	(0.339)
1.mp				15.23***	15.31***
				(0.160)	(0.124)
1.aug19#1.mp				-2.366***	-2.174***
				(0.271)	(0.219)
1.line60#1.mp				8.231***	8.156***
				(0.459)	(0.448)
1.aug19#1.line60#1.mp			29.79***	8.693***	8.502***
			(1.113)	(1.224)	(1.213)
1.line6					8.937***
					(0.117)
1.line64					1.076***
					(0.147)
Constant	24.34***	18.88***	18.88***	10.57***	6.777***
	(0.417)	(0.124)	(0.124)	(0.0804)	(0.0913)
Observations	1,854	11,973	11,973	11,973	11,973
R-squared	0.027	0.071	0.139	0.544	0.648

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

First, M1 develops a simplistic one-dimensional approach using an Ordinary Least Squares (OLS) regression with a single dimension. Given that only L60 is considered in this model, **1.aug.19** represents the average difference in charge pre-/post-treatment of L60. Although the coef. is stat. sig. at the 0.01 level, the R-squared value indicates that only 2.7% of the variation is explained by the model, as the dummy variable **1.aug.19** is the sole variable capable of explaining the variation. In conclusion, no useful insights can be derived from this regression. The contribution of M1 lies in elucidating the methodology for performing M2.

Second, M2 represents a classical DiD estimation. Therefore, an additional dimension is added and the basic model in M1 has now a second dimension. Accordingly, the dummy variable **1.line60** is added to represent L60, as opposed to the control lines. It is clearly visible that **1.aug19#1.line60** corresponds to the subtraction of **1.aug19** in M1 with **1.aug19** in M2. Therefore, after adding a second dimension, the result of the DiD interaction is reinforced as the overall tendency represented by the control lines is negative. Note that **1.line60** represents the average difference in charge of L60 compared to the control lines pre-treatment. The coef. is stat. sig. at the 0.01 level, indicating that L60 has a higher baseline charge compared to the control group of 7.791 passengers. However, the R-squared value shows that solely 0.071% of the variation is explained by the model. Although the decision to enhance a model with additional variables always involves a bias-variance trade-off (Stock & Watson, 2002), this value is considered as too low, necessitating further analysis to achieve a better model fit. As no control variables are available, as discussed in Chapter 3.6.4, the strategy of triple difference estimations is pursued in M3-M5.

Third, the purpose of M3 is purely explanatory, as it introduces a biased triple difference (DiDiD) regression by erasing all the additional interactions underlying the triple difference approach. The aim of this model is to illustrate that a DiDiD-estimation can be understood as the decomposition of a classical single interaction (DiD) into two double interaction estimations. Specifically, the estimator of the ATT is sectioned into the MP and the OP. It is crucial to note that the interaction term **1.aug19#1.line60** differs between M2 and M3. While **1.aug19#1.line60** captures the overall ATT in M2, it describes the ATT in the OP specifically in M3, in contrast to **1.aug19#1.line60#mp**, which specifies the ATT in the MP specifically. Although these distinctions can be confusing, they are necessary to avoid displaying numerous additional interaction variables and to ensure a concise presentation of the regression results, as shown in Table 5.

Fourth, M4 presents a triple difference regression while retaining all the underlying interactions of the method. M4 demonstrates a substantially better fit as the R-squared reaches 54.4%. The three dimensions considered are henceforward **aug19**, **line60** and **mp**. Consequently:

- **1.aug19** (1 0 0): The coef. suggests that the av. change in charge post-treatment on the control lines during the OP is ns.

- **1.line60** (0 1 0): The coef. suggests that the av. charge pre-treatment on L60 during the OP is lower compared to the control lines by approx. 1.864 passengers. The coef. is stat. sig. at the 0.01 level.

- **1.aug19#1.line60** (1 1 0): The coef. suggests that the ATT on L60 during the OP is a slight increase by approx. 2.138 passengers. The coef. is marginally stat. sig. at the 0.01 level.

- **1.mp** (0 0 1): The coef. suggests that the av. charge pre-treatment on the control lines is higher during the MP than during the OP by approx. 15.23 passengers. The coef. is stat. sig at the 0.01 level.

- **1.aug19#1.mp** (1 0 1): The coef. suggests that the av. change in charge post-treatment on the control lines during the MP is a decrease by approx. 2.366 passengers. The coef. is stat. sig at the 0.01 level.

- **1.line60#1.mp** (0 1 1): The coef. suggests that the av. difference in charge between the MP and the OP is higher on Line 60 by approximately 8.156 units compared to 1.mp. The coef. is stat. sig. at the 0.01 level.

- **1.aug19#1.line60#1.mp** (1 1 1): The coef. suggests that the ATT on L60 during the MP is an increase by approx. 8.693 passengers. The coef. is marginally stat. sig., at the 0.01 level.

- **Constant** (0 0 0): The coef. suggests that the av. charge pre-treatment on the control lines during the OP was approx. 10.57 passengers. The coef. is stat. sig. at the 0.01 level.

Fifth, M5 enhances M4 by incorporating line-specific constants for two of the control lines: **1.line6** and . Therefore, the new Constant is henceforth a synonym for a hypothetical **1.line2**, which is obviously omitted to avoid the dummy variable trap. The choice for the L2 as Constant is based on the bias-variance generated relative to what L6 and L64 would have exhibited. As a result, M5 demonstrates a better fit as the R-squared increases again to 64.8% without having any substantial impact on the variance. Consequently:

- **1.aug19** (1 0 0): The coef. suggests that the av. change in charge post-treatment on L2 during the OP is ns, as in M4.

- **1.line60** (0 1 0): The coef. suggests that the av. charge pre-treatment on L60 during the OP is higher compared to L2 during the OP by approx. 1.926 passengers. The coef. is stat. sig. at the 0.01 level. This result increased by 3.79 passengers compared to M4 because Constant henceforth represents L2 uniquely, without considering the other control lines.

- **1.aug19#1.line60** (1 1 0): The coef. suggests that the ATT on L60 during the OP is an increase by approx. 2.138 passengers. The coef. is stat. sig. at the 0.01 level and stable compared to M4.

- **1.mp** (0 0 1): The coef. suggests that the av. charge pre-treatment on L2 is higher during the MP than the OP by approx. 15.31 passengers. The coef. is stat. sig at the 0.01 level and stable compared to M4.

- **1.aug19#1.mp** (1 0 1): The coef. suggests that the av. change in charge post-treatment on L2 during the MP is a decrease by approx. 2.174 passengers. The coef. is stat. sig. at the 0.01 level and stable compared to M4.

- **1.line60#1.mp** (0 1 1): The coef. suggests that the av. difference in charge between the MP and the OP is higher on Line 60 by approximately 8.156 units compared to 1.mp. The coef. is stat. sig. at the 0.01 level and stable compared to M4.

- **1.aug19#line60#1.mp** (1 1 1): The coef. suggests that the ATT on L60 during the MP is an increase by approx. 8.502 passengers. The coef. is stat. sig. at the 0.01 level and slightly lower compared to M4.

- **1.line6** (0 0 0): The coef. suggests that the av. charge pre-treatment on L6 during the OP was higher compared to L2 by approx. 8.937 passengers. The coef. is stat. sig. at the 0.01 level.

- **1.line64** (0 0 0): The coef. suggests that the av. charge pre-treatment on L6 during the OP was higher compared to L2 by approx. 1.076 passengers. The coef. is stat. sig. at the 0.01 level.

- **Constant** (0 0 0): The coef. suggest that the av. charge pre-treatment on L2 during the OP was approx. 6.777 passengers. This is lower of 3.79 passengers compared to M4. Logically, this decrease is to the same extend as the increase observed in 1.line60. The coef. is stat. sig. at the 0.01 level.

4.3 Methodological Interpretation

First, it appears necessary to find a way to enrich the difference-in-differences (DiD) regression in M2. The low R-squared suggests a severe inability of the calculated coefficients to explain the variation of the data. Given the impossibility of introducing convincing control variables as thematized in Chapter 3.6.4, employing a triple-difference (DiDiD) regression model seems to have been an imperative strategy.

Second, introducing the triple difference estimation to account for the differential ATT resulted in strongly attenuating the ATT in the OP and slightly reinforcing the ATT in the MP. On the basis of the single interaction **1.aug19#1.line60** of 7.791 passengers calculated in M2, the double interactions **1.aug19#1.line60**(#0mp) and **1.aug19#1.line60#1.mp** decreased to 2.138 and increased to 8.693 respectively in M4. This non-symmetrical divergence is explained by the fact that part of the variation was absorbed from both estimators by the other interactions variables created by the DiDiD method in M4. Introducing the line-specific introduced in M5 also lowered the estimator for **1.aug19#1.line60#1.mp** from 8.693 to 8.502 in M5.

Third, as the Table 8 do not provide the overview about the effective evolution of the dependent variable, the below calculations must be read to understand how the value of "charge" can be calculated based on the estimates displayed in Table 8:

Pre-treatment:

- Av. charge on L2 during the OP: **Constant:** <u>6.777 passengers</u>
- Av. charge on L2 during the MP: **Constant + 1.mp**: 6.777 + 15.31 = <u>22.087 passengers</u>

- Av. charge on L60 during the OP: **Constant + 1.line60**: 6.777 + 1.926 = 8.703 passengers
- Av. charge on L60 during the MP: **Constant** + **1.line60 + 1.mp + 1.line60#1.mp**: 6.777 + 15.31 + 8.156 + 1.926 = 32.169 passengers

Post-treatment:

- Av. charge on L2 during the OP: **Constant** + **1.aug19:** 6.777 +(– 0.112) = 6.665 passengers.
- Av. charge on L2 during the MP: **Constant** + **1.aug19** + **1.mp** + **1.aug19#1.mp**: 6.777 +(– 0.112) + 15.31 +(– 2.174) = 20.521 passengers.
- Av. charge on L60 during the OP: **Constant + 1.aug19** + **1.line60** + **1.aug19#1.line60**: 6.777 +(– 0.112) + 1.926 + 2.138 = 10.729 passengers.
- Av. charge on L60 during the MP: **Constant + 1.aug19** + **1.line60** + **1.aug19#1.line60** + **1.mp** + **1.aug19#1.mp** + **1.line60#1.mp** + **1.aug19#1.line60#1.mp** : 6.777 +(– 0.112) + 1.926 + 2.138 + 15.31 +(– 2.174) + 8.156 + 8.502 = 40.523 passengers.

Note that it is necessary adding all the variables' estimates to calculate the av. charge post treatment on L60 during the MP. Consequently, the coefficients must always be interpreted with great caution, as it is easy to become confused if one forgets that each coefficient is binary categorized across all three dimensions.

4.4 Efficiency

The efficiency refers to the standard errors, and therefore to the precision of coefficients (Stock & Watson, 2002). In Table 8, M4 and M5 display smaller standard errors compared to M1 and M2. This implies a better efficiency of the estimates.

We also employed the Variation Inflation Factor (VIF) to evaluate the precision of our model's estimates. Typically utilized to detect multicollinearity among predictors, the VIF is utilized here with the objective of underscoring why a further addition of variables to M5 in Table 8 was halted. Typically, high VIF values indicate increased variance in the regression coefficients due to collinearity among predictors, compared to when predictors are orthogonal. Such collinearity diminishes estimator precision, validating the use of VIF not only to check for multicollinearity but also to assess the efficiency of our model estimates. (Murray et al., 2012)

Specifically, the VIF for the i^{th} predictor variable can be expressed by:

$$VIF_i = r^{ii} = \frac{1}{1 - R_i^2}, i, \ldots, p,$$

where R_i^2 is the multiple correlation coefficient of the regression between X_i and the remaining $p - 1$ predictors. (Murray et al., 2012)

Although dissensus dominates in the literature about cutoff values above which a regressor must be excluded, the rule of thumb of a VIF's not exceeding 5 is retained (Murray et al., 2012).

Table 9: Variation Inflation Factors of the Regressors

	(M1)	(M2)	(M3)	(M4)	(M5)	(MX)
	OLS	DiD	DiDiD	DiDiD	DiDiD with line-specific constants	DiDiD with line-specific constants and Travel Time
VARIABLES	VIF	VIF	VIF	VIF	VIF	VIF
1.aug19	1	1.18	1.18	2.56	2.56	2.62
1.line60		1.46	1.46	4.21	4.44	5.91
1.aug19#1.line60		1.65	3.55	4.73	4.73	4.77
1.mp				1.71	1.71	2.48
1.aug19#1.mp				3.18	3.18	3.20
1.line60#1.mp				4.58	4.58	4.64
1.aug19#1.line60#1.mp			2.90	4.95	4.95	5.02
1.line6					1.34	1.53
1.line64					1.28	28.79
Travel Time						34.69
MEAN VIF	1	1.43	2.28	3.70	3.20	9.37

Based on the Variance Inflation Factor (VIF) results presented in Table 9, a nuanced perspective with respect to the model fit is provided. Firstly, it is evident that the introduction of additional interaction variables leads to an inflation of the variance. Secondly, M5 not only provides a better fit than M4 as previously shown in Table 8, but also demonstrates lower variance, highlighting its superior performance. However, it is to note that this reduction in variance is primarily attributable to the addition of line-specific constants, while the inflation behavior of most other variables remained unchanged. Thirdly, the decision to exclude Travel Time from the model is justified due to its significant multicollinearity with the other variables, and **i.line64** in particular, as shown in MX. An alternative model that included Travel Time but excluded **i.line64** was also tested but showed inferior performance on both R-squared and variance compared to M5, leading to its exclusion from the analysis.

5 Discussion

This chapter presents the major interpretations that can be drawn from the results, as well as possible implications for policy and practice.

5.1 Interpretation of Results

First and foremost, the diverging estimation for the morning peak (MP) and off-peak (OP) implies that the average treatment effect on the treated (ATT) is an increase of 2.138 passengers during the OP as well as an increase of 8.502 passengers during the MP. These coef. are stat. sig. at the 0.001 level. These results are highly encouraging as they seem to confirm the predominant role of crowding on ridership: When buses are empty in the OP, there is only limited reason for the introduction of vehicles with a higher seating capacity to have a substantial positive impact on ridership. As a result, ridership increases by a smaller av. number of 2.138 passengers. On the other hand, when buses are full in the MP, there is a valid reason for the introduction of vehicles with a higher seating capacity to have a substantial impact on ridership. As a result, ridership increases by a larger av. number of 8.502 passengers. Consequently, the estimate for the ATT in the MP suggests that the increase is roughly four times stronger compared to the OP.

Second, these findings corroborate the initial hypothesis proposed in the introduction: an increase in service capacity leads to higher ridership. Although the "load factor" was excluded to prevent multicollinearity with **i.mp**, the ATT allows us to deduce the extent to which the rise in ridership can be attributed to reduced crowding externalities. Consequently, these outcomes also validate the second hypothesis.

The results presented above are not without their potential counterarguments. For example, one might argue that the observed increase in ridership during the MP relative to the OP could simply reflect a broader trend of growing ridership during the peak hours rather than an effect specific to the intervention. However, this argument can be countered by the fact that the model already accounts this tendency through the inclusion of **1.aug19#1.mp** and **1.aug**(0.mp).

Specifically, this estimate indicates a decrease in av. ridership of 2.174 passengers in the control group during the MP, while the evolution during the OP remains ns.

Interestingly, this result for the control group also points to a broader trend: the narrowing gap between peak and off-peak travel patterns. This phenomenon is consistent with larger trends observed both in Switzerland and internationally: In Switzerland, there is noticeable stagnation in commuter traffic using PT, while other types of traffic as for leisure destinations are on the rise (Lustenberger et al., 2021). Globally, similar shifts in transportation behavior are being forecasted, particularly in urban centers, as remote work becomes more prevalent (Currie et al., 2021).

5.2 Implications for Policy and Practice of the Operator

The results have two major implications for the operator, which may seem contradictory. First, a targeted capacity increase can have a substantially higher impact when a link is experiencing congestion. Although the fourfold impact observed in this very specific case of L60 in direction R during the MP is expected to vary if the same model is applied to other scenarios, a differential effect in the MP as opposed to the OP appears to be undeniable. However, it must be remained that the dependent variable calculated represents here the number of passengers per vehicle averaged over each ride. However, as most passengers are not assumed to travel over the entire length of the line, the true number of additional PT users per ride on L60 might be substantially higher than the estimates for ATT displayed in Table 8. This calculation would however require the identification of the passengers to assess the stop at which they board and alight.

On the other hand, the trend observed on the control lines during the MP also suggests a slight decrease of 2.174 passengers, while no significant trend is apparent in the OP. If this tendency persists, it could also indicate that the negative impact of crowding on ridership during commuter peaks might diminish in the middle and long run. Given that improvements in infrastructure or rolling stock require substantial financial investments, coupled with the long duration from decision-making to actual implementation, which can span several years, these results should also be interpreted as warning not to disregard the observed general trend towards a more evenly distributed ridership throughout the day.

5.3 Limitations

This chapter outlines significant limitations inherent in the data and the methods used.

5.3.1 Limitations related to the Data

First, it's crucial to note that the dataset was dramatically reduced from over 17.6 million observations to just 11,973. This significant reduction was not due to random sampling but resulted from multiple aggregation processes designed to facilitate the performance of regression and to eliminate confounding effects. This aggregation restricted the analysis to specific lines, months, days and hours. The aggregation also involved collapsing the dataset such that each observation represents an entire ride rather than individual stops. Consequently, our approach fails to account for potential variations in crowding along the route, instead assuming uniform crowding throughout each ride. This is a considerable limitation, as stop-specific crowding patterns are overlooked. As Wardman & Whelan (2011) suggest, assuming constant crowding levels over a specified time can lead to distortions, since crowding can fluctuate significantly over time and along different segments of a route.

Second, the data necessitated using seating capacities as the primary metric for determining bus capacity, as the definition of standing capacities were inconsistent across the bus manufacturers. Some manufacturers formulate a maximal capacity of 3 persons per square meter, other a maximal capacity of 4 persons per square meter. Consequently, seating capacity served as the denominator for calculating crowding, as defined in Chapter 3.2. However, this metric might still not fully reflect the actual experience of congestion for passengers. As the dataset spans approx. 15 different bus models, it therefore inherently presupposes that a "load factor" of, e. g., 80% uniformly affects passenger perception of crowding independently of the vehicle. However, variation in the size of standing areas across the bus models could further influence perceptions of crowding. Moreover, the data is not able differentiating between seating passengers versus those who are standing, even though studies, such as the meta-analysis by Wardman & Whelan (2011), suggest significant differences in travel time multipliers between these two groups.

Thirdly, as the bus data did not allow for the inclusion of additional control variables, such as “travel time”, these could not be utilized due to their tendency to inflate the variance of the model. Furthermore, although the bus data includes both departure and arrival times for each ride, it lacks the level of detail found in rail data, which, starting from 2019 on, provided precise second-by-second departure and arrival times at each stop. Consequently, this limitation in data granularity significantly constrained the generation of additional control variables, such as travel time or service reliability.

Fourthly, another significant limitation is that the analysis was confined to specific links available in the dataset, ignoring other potential connections. However, the origin-destination matrixes of passengers often extend beyond the analyzed links. As a result, the travel choice is also influenced by the factors of adjacent links and beyond.

5.3.2 Limitations related to the Methodology

The differences-in-differences-in-differences methodology, along with our exclusive use of dummy variables, introduces a significant limitation. This method substantially simplifies the analysis by assuming that the Average Treatment Effect (ATT) will remain constant from the day after the new buses are introduced to the end of the period for which data is available. However, a logical reasoning would suggest that improvements in services more likely have a gradual impact, as passengers might progressively adjust and integrate these changes into their decision-making. This simplification could bias the results, as extending the treatment period analysis to the first half of 2024 could potentially reveal a more pronounced effect on the ATT, or also not.

Secondly, the decision-making process in this study did not incorporate a probabilistic approach to evaluating crowding impacts, which, according to Wardman & Whelan (2011), could be more influential on behavior than actual crowding levels. On the contrary, Wardman & Whelan argued that drawing on the “load factor” as a decision variable for travel choice might be problematic, as passengers do not necessarily know the “load factor” they will face or whether they will obtain a seat. Even though, real-time crowding information has enhanced passenger predictions through user interfaces (Hoppe et al., 2023), there's no guarantee passengers will utilize this information. Accordingly, Wardman & Whelan argued that it's the perceived likelihood of having to stand for various durations that shapes passenger behavior, rather than the certainty of standing for a set amount of time. For forecasting purposes, the focus should therefore ideally be on the probability of having to stand and how passengers respond to this likelihood rather than merely assessing the impacts of different definite standing durations. Possibly, this probabilistic approach could be better approached using a SP-approach, as by Kroes et al., (2014). However, we use in our analysis the "load factor" and then the morning peak (MP) as a proxy for crowding's influence on ridership. This simplification might seem to downplay the complexity of real-world interactions. Nevertheless, the meta-analysis provided by Wardman & Whelan (2011), also did not reveal any significant differences in outcomes between using a probabilistic approach and quantifying specific levels of standing time.

Thirdly, not all identifying assumptions outlined in Chapter 3.6.3 could be fully adhered to. Specifically, two assumptions present challenges: the assumption of stable composition and the assumption of no spillovers between units. The stable composition assumption was compromised on L60 due to changes in its stop policy. This issue is partially mitigated by focusing on only part of the line, as detailed in Chapter 4.1.2. However, considering the transportation line as a cohesive unit means this limitation cannot be completely overcome. As for the stable composition, careful measures were taken to ensure to respect the assumption of absence of spillovers between the treatment and control lines.

However, guaranteeing no spillover effects between the morning peak (MP) and off-peak (OP) periods is not feasible and represents an important downside of our study. Indeed, it is impossible to verify that a passenger observed on L60 traveling during the OP before the treatment is not the same individual observed traveling during the MP after the treatment. On the contrary, if crowding significantly affected L60 before the treatment, it's plausible that some passengers altered their travel schedules to the OP. With the reduction in crowding during the MP post-treatment, it's even logical for passengers to shift their travel back to their preferred MP, demonstrating a potential behavioral adaptation to the changed conditions. Therefore, this element significantly undermines the methodology's robustness.

5.4 Knowledge Gaps and Future Research

In the context of Switzerland, it would be insightful to explore congestion phenomena and their potential mitigation through capacity increases aimed at leisure travel. Specifically, tourist train lines experience significant crowding during peak winter days for ski destinations. Unlike the stagnation or decline observed in traditional commuter ridership that we observed, leisure traffic, according to regional train operators like RhB (2023) is burgeoning. This presents an opportunity to examine the impacts of alleviating congestion on routes designated primarily for other travel purposes.

Further Research Question: “To what extent does leisure ridership in Public Transport respond to reductions in crowding externalities?”

Unlike our study, an analysis focusing on leisure traffic could offer several benefits:

1. Leisure travel during peak times could show more spatiotemporal uniformity compared to commuter traffic, which may facilitate a more homogeneous analysis of travel purposes, even if passenger identities remain unknown.
2. Addressing leisure traffic crowding would tap into an under-researched area of public transportation studies, aligning with broader trends observed in ridership.

Regarding the research design, the methodological framework utilized in this study, such as difference estimations, could be adapted while ideally incorporating a broader range of control variables to capture the nuances of leisure travel behaviors and their responses to changes in service conditions.

6 Conclusion

This research aimed to identify the impact of congestion externalities on public transport ridership, guided by two primary research questions: "*How do changes in the service capacity of public transport affect ridership?*" and *"To what extent can the increase in ridership be attributed to a decrease in crowding externalities?"*. The analysis of revealed preferences data provided by public transport operators demonstrated that a doubling of the seating capacity significantly increased ridership, thereby addressing the first question. Notably, the impact was fourfold during the morning peak compared to the off-peak, addressing the second question.

Despite initial concerns about the potential endogeneity of our research design, the data supported the treatment's exogeneity. However, the possible presence of spillover effects between the morning peak and off-peak periods, suggesting dynamic passenger behavior adjustments, could compromise both the internal and external validity of the findings.

Innovatively, this study managed to isolate the effect of increased vehicle capacity without concurrent changes in frequency—an unintended and beneficial circumstance that eliminated the confounding effects induced from frequency enhancements, such as the effect of reducing waiting times at stops. This aspect is particularly noteworthy, given that most of existing literature on the subject tends to focus on frequency increases.

It would therefore be advantageous for future research to continue exploring congestion phenomena using revealed preferences data. Although our model simplifies the analysis by primarily using average values based on binary categories, future studies could enhance our proposed triple difference model by incorporating continuous data variables. This inclusion would allow for a more nuanced understanding of the data granularity.

7 References

Alliance SwissPass. (2023, April 4). öV-Preise steigen per Dezember 2023 erstmals seit sieben Jahren an. Alliance SwissPass. http://www.allianceswisspass.ch/de/asp/News/Newsmeldung?newsid=567

Angrist, J. D., & Pischke, J-S. (2009). Mostly Harmless Econometrics: An Empiricistís Companion.

ARE, O. fédéral du développement territorial. (2022). Perspectives d'évolution du transport 2050.

Axhausen, K. W., Haupt, T., Fell, B., & Heidl, U. (2001). Searching for the rail bonus: Results from a panel SP/RP study. European Journal of Transport and Infrastructure Research, 1(4), 353–369. https://doi.org/10.3929/ethz-a-006098486

Batarce, M., Muñoz, J. C., de Dios Ortúzar, J., Raveau, S., Mojica, C., & Ríos, R. A. (2015). Use of Mixed Stated and Revealed Preference Data for Crowding Valuation on Public Transport in Santiago, Chile. Transportation Research Record, 2535(1), 73–78. https://doi.org/10.3141/2535-08

Bertrand, M., Duflo, E., & Mullainathan, S. (2004). How Much Should We Trust Differences-In-Differences Estimates?*. The Quarterly Journal of Economics, 119(1), 249–275. https://doi.org/10.1162/003355304772839588

Bigi, F., Schwemmle, N., & Viti, F. (2023). Evaluating the impact of Free Public Transport using agent-based modeling: The case-study of Luxembourg.

Boisjoly, G., Grisé, E., Maguire, M., Veillette, M.-P., Deboosere, R., Berrebi, E., & El-Geneidy, A. (2018). Invest in the ride: A 14 year longitudinal analysis of the determinants of public transport ridership in 25 North American cities. Transportation Research Part A: Policy and Practice, 116, 434–445. https://doi.org/10.1016/j.tra.2018.07.005

Bubenhofer, J., Hool, A., Naef, C., & Hess, J. (2018). Dichte und Mobilitätsverhalten. https://www.are.admin.ch/are/de/home/medien-und-publikationen/publikationen/grundlagen/dichte-und-mobilitatsverhalten.html

Butts, K. (2021). Difference-in-Differences Estimation with Spatial Spillovers.

Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-Differences with multiple time periods. Journal of Econometrics, 225(2), 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Chen, C., Varley, D., & Chen, J. (2011). What Affects Transit Ridership? A Dynamic Analysis involving Multiple Factors, Lags and Asymmetric Behaviour. Urban Studies, 48(9), 1893–1908. https://doi.org/10.1177/0042098010379280

Currie, G., Jain, T., & Aston, L. (2021). Evidence of a post-COVID change in travel behaviour – Self-reported expectations of commuting in Melbourne. Transportation Research Part A: Policy and Practice, 153, 218–234. https://doi.org/10.1016/j.tra.2021.09.009

Currie, G., & Wallis, I. (2008). Effective ways to grow urban bus markets – a synthesis of evidence. Journal of Transport Geography, 16(6), 419–429. https://doi.org/10.1016/j.jtrangeo.2008.04.007

Ethington, C. A. (1987). The Robustness of LISREL Estimates in Structural Equation Models with Categorical Variables. The Journal of Experimental Education, 55(2), 80–88.

Evans IV, J. (2004). Transit Scheduling and Frequency—Traveler Response to Transportation System

Changes. World Transit Research. https://www.worldtransitresearch.info/research/3051

Fletcher, G., & El-Geneidy, A. (2013). Effects of Fare Payment Types and Crowding on Dwell Time: Fine-Grained Analysis. Transportation Research Record: Journal of the Transportation Research Board, 2351, 124–132. https://doi.org/10.3141/2351-14

Fowkes, A. S., & Wardman, M. (1987). The Values of Overcrowding and Departure Time Variations for Inter City Rail Travellers.

Hintermann, B., Schoeman, B., Molloy, J., Schatzmann, T., Tchervenkov, C., & Axhausen, K. W. (2023). The impact of COVID-19 on mobility choices in Switzerland. Transportation Research Part A: Policy and Practice, 169, 103582. https://doi.org/10.1016/j.tra.2023.103582

Hoppe, J., Schwinger, F., Haeger, H., Wernz, J., & Jarke, M. (2023). Improving the Prediction of Passenger Numbers in Public Transit Networks by Combining Short-Term Forecasts With Real-Time Occupancy Data. IEEE Open Journal of Intelligent Transportation Systems, 4, 153–174. https://doi.org/10.1109/OJITS.2023.3251564

Kain, J. F., & Liu, Z. (1995). Secrets of Success: How Houston and San Diego Transit Providers Achieved Large Increases in Transit Ridership. (PB98103922). Texas Univ. at Dallas, Richardson.; Harvard Univ., Cambridge, MA.; Federal Transit Administration, Washington, DC. Office of Planning. https://ntrl.ntis.gov/NTRL/dashboard/searchResults/titleDetail/PB98103922.xhtml

Kain, J. F., & Liu, Z. (1999). Secrets of success: Assessing the large increases in transit ridership achieved by Houston and San Diego transit providers. Transportation Research Part A: Policy and Practice, 33(7), 601–624. https://doi.org/10.1016/S0965-8564(99)00009-9

Kraus, M. (1991). Discomfort externalities and marginal cost transit fares. Journal of Urban Economics, 29(2), 249–259. https://doi.org/10.1016/0094-1190(91)90018-3

Kroes, E., Kouwenhoven, M., Debrincat, L., & Pauget, N. (2014). Value of Crowding on Public Transport in île-de-France, France. Transportation Research Record: Journal of the Transportation Research Board, 2417(1), 37–45. https://doi.org/10.3141/2417-05

Lechner, M. (2011). The Estimation of Causal Effects by Difference-in-Difference MethodsEstimation of Spatial Panels. Foundations and Trends® in Econometrics, 4, 165–224. https://doi.org/10.1561/0800000014

Liu, X. (2019). A probabilistic explanation of Pearson's correlation. Teaching Statistics, 41. https://doi.org/10.1111/test.12204

Lu, Q.-C., Li, J., Xu, P.-C., Zhang, L., & Cui, X. (2024). Modeling cascading failures of urban rail transit network based on passenger spatiotemporal heterogeneity. Reliability Engineering & System Safety, 242, 109726. https://doi.org/10.1016/j.ress.2023.109726

Lustenberger, N., Becker, F., Hintermann, B., & Axhausen, K. W. (2021). Änderung des Verkehrsverhaltens während der COVID-19 Pandemie [Report]. ETH Zurich. https://doi.org/10.3929/ethz-b-000519119

McFadden, D. (1974). Conditional Logit Analysis of Qualitative Choice Behavior. In FRONTIERS IN ECONOMETRICS (Vol. 8, pp. 105–142).

McLeod, MS, Flannelly, J., Flannelly, K., Behnke, L., & Davidson, R. (1991). Multivariate Time-Series Model of Transit Ridership based on historical, aggregate data: The past, present and future of Honolulu. SWOV. https://swov.nl/en/publicatie/multivariate-time-series-model-transit-ridership-based-historical-aggregate-data-past

Mohring, H. (1972). Optimization and Scale Economies in Urban Bus Transportation. American Economic Review, 62, 591–604.

Murray, L., Nguyen, H., Lee, Y.-F., Remmenga, M., & Smith, D. (2012). VARIANCE INFLATION FACTORS IN REGRESSION MODELS WITH DUMMY VARIABLES. Conference on Applied Statistics in Agriculture. https://doi.org/10.4148/2475-7772.1034

Olden, A., & Møen, J. (2022). The triple difference estimator. The Econometrics Journal, 25(3), 531–553. https://doi.org/10.1093/ectj/utac010

Parry, I., & Small, K. (2007). Should Urban Transit Subsidies Be Reduced? American Economic Review, 99. https://doi.org/10.1257/aer.99.3.700

Paudel, J. (2021). Bus ridership and service reliability: The case of public transportation in Western Massachusetts. Transport Policy, 100, 98–107. https://doi.org/10.1016/j.tranpol.2020.10.010

RhB. (2023). Geschäftsbericht 2023.

Riveros-Gavilanes, J. M. (2023). A simple test of parallel pre-trends for Differences-in-Differences.

Small, K. A. (1982). The Scheduling of Consumer Activities: Work Trips. The American Economic Review, 72(3), 467–479.

Small, K. A., & Verhoef, E. T. (2007). The Economics of Urban Tranportation.

Stock, J. H., & Watson, M. W. (2002). Introduction to Econometrics. Pearson.

Taylor, B. D., Miller, D., Iseki, H., & Fink, C. (2009). Nature and/or nurture? Analyzing the determinants of transit ridership across US urbanized areas. Transportation Research Part A: Policy and Practice, 43(1), 60–77. https://doi.org/10.1016/j.tra.2008.06.007

Tirachini, A., Hensher, D. A., & Rose, J. M. (2013). Crowding in public transport systems: Effects on users, operation and implications for the estimation of demand. Transportation Research Part A: Policy and Practice, 53, 36–52. https://doi.org/10.1016/j.tra.2013.06.005

Tirachini, A., Sun, L., Erath, A., & Chakirov, A. (2016). Valuation of sitting and standing in metro trains using revealed preferences. Transport Policy, 47, 94–104. https://doi.org/10.1016/j.tranpol.2015.12.004

TL. (2020). RAPPORT D’ACTIVITÉ 2019.

Verbas, O., Frei, C., Mahmassani, H., & Chan, R. (2015). Stretching resources: Sensitivity of optimal bus frequency allocation to stop-level demand elasticities. Public Transport, 7, 1–20. https://doi.org/10.1007/s12469-013-0084-6

Ville de Lausanne. (2024). Portrait statistique 2024.

Voith, R. (1991). The long-run elasticity of demand for commuter rail transportation. Journal of Urban Economics, 30(3), 360–372. https://doi.org/10.1016/0094-1190(91)90055-C

Wardman, M., & Whelan, G. (2011). Twenty Years of Rail Crowding Valuation Studies: Evidence and Lessons from British Experience. Transport Reviews, 31, 379–398. https://doi.org/10.1080/01441647.2010.519127

Xiong, C., Chen, X., He, X., Guo, W., & Zhang, L. (2015). The analysis of dynamic travel mode choice: A heterogeneous hidden Markov approach. Transportation, 42(6), 985–1002. https://doi.org/10.1007/s11116-015-9658-2