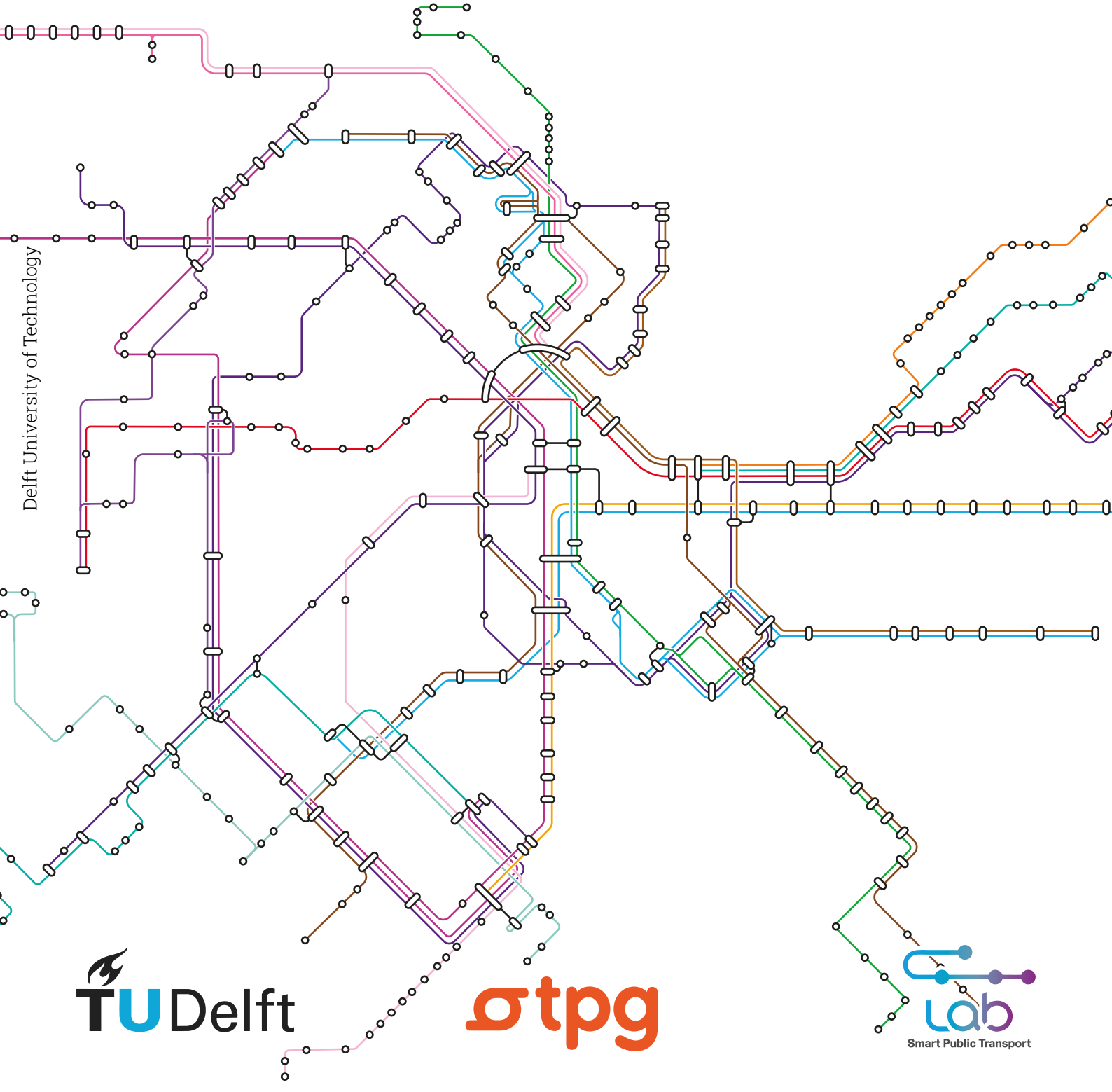
When Schedules Meet the Streets

Public transport reliability, local traffic exposure, and mitigating measures in urban bus and tram segments

CIEM0500: Master Thesis

Rona Jak



Delft University of Technology

TU Delft

tpg

Lab Smart Public Transport


Grand-Saconnex, place
Grand-Saconnex, Le Pommier
Genève, Vieusseux
Genève, Servette
Genève, Maison de la Paix
Genève, Butini
Genève, Môle
Genève, gare Cornavin
Genève, Chantepoulet
Genève, Mont-Blanc
Genève, Métropole
Genève, 31-Décembre
Genève, Place des Eaux-Vives
Chêne-Bourg, Place Favre
Chêne-Bougeries, Grange-Falquet
Vernier, Avenue de l'Ain
Petit-Lancy, Chemin du Bac
Genève, Mercier
Genève, Bel-Air
Genève, Stand
Genève, Palladium
Genève, Jonction
Genève, Plainpalais
Genève, Uni-Mail
Lancy-Pont-Rouge, gare/Etoile
Grand-Lancy, Mairie de Lancy
Grand-Lancy, Palettes
Grand-Lancy, Pontets

When Schedules Meet the Streets

Public transport reliability, local traffic exposure, and mitigating measures in urban bus and tram segments

by

Rona Jak

to obtain the degree of Master of Science in Civil Engineering

at the Delft University of Technology,

to be defended publicly on August 28, 2026 at 2:00 PM local time.

Student number:	5024234
Programme:	MSc Civil Engineering
Track:	Traffic & Transport Engineering
Faculty:	Faculty of Civil Engineering and Geosciences, TU Delft
Project duration:	February 2026 - August 2026
Graduation committee:	Prof. dr. O. (Oded) Cats, Dr. ir. A. (Alexandra) Gavriilidou
External collaboration:	Dr. D.J. (Daniel) Reck, transports publics genevois (tpg)
Data provided by:	transports publics genevois (tpg)
Cover:	Own work
Style:	TU Delft Report Style

A digital version of this thesis is available in the collection of TU Delft Repository.

TU Delft

tpg

Lab Smart Public Transport

[*This page intentionally left blank*]

Acknowledgements

This thesis marks the end of my Master's in Civil Engineering at Delft University of Technology. Looking back, this project has been both challenging and rewarding. It required me to work with large datasets, make many methodological choices, question my own results more than once, and slowly turn a broad interest in public transport reliability into a complete thesis. I am grateful to everyone who supported me throughout this process.

First, I would like to thank my graduation committee. I am grateful to Prof. dr. Oded Cats for his guidance, critical view, and ability to bring the research back to the bigger picture. I would also like to thank Dr. ir. Alexandra Gavriilidou for her valuable feedback, personal supervision, and encouragement throughout the project. Your comments helped me improve the academic quality of this thesis.

I would like to express my sincere thanks to Dr. Daniel Reck and transports publics genevois for the collaboration and for providing the data that made this research possible. Being able to work with a real public transport case made the thesis both more complex and more interesting; sometimes I almost forgot that every data point ultimately represents a bus or tram trying to make its way through an actual city.

I am also thankful to the people who helped me over the past half year, whether by discussing modelling choices, reading parts of my work, or simply listening when I needed to explain for the tenth time why one coefficient did not behave the way I expected. These conversations made the process less overwhelming and often helped me see the next step more clearly.

A special thank you goes to all the friends I met during my Bachelor's, board year, and Master's; you made the past years in Delft so much more enjoyable. Thank you for the study sessions, coffee breaks, shared deadlines, and necessary distractions. Writing a thesis can feel quite lonely at times, but having people around who understand the process made a big difference.

Finally, I would like to thank my family and girlfriend for their support, patience, and encouragement throughout my studies. Thank you for believing in me (and sorry that you had to learn far more about buses, trams, traffic counters, and regression models than you ever asked for). Your support, occasional unexpected insights, and endless willingness to listen gave me the space and confidence to finish this thesis.

With this thesis, my time as a student at TU Delft comes to an end. It has been a period of learning, growth, and lots of challenges, but also one that I look back on with pride. It feels right to finish my time in Delft with a project that combines three things I genuinely enjoy: public transport operations, far too much data, and trying to understand why reality does not always behave the way you would like it to.

Rona Jak
Delft, The Netherlands, August 2026

[*This page intentionally left blank*]

Abstract

Public transport reliability is an important component of service quality, particularly in dense urban environments where buses and trams operate in close interaction with surrounding traffic. In such settings, realised operations may differ substantially from planned operations because vehicles are exposed to signalised intersections, limited road space, frequent stops, and varying levels of interaction with general traffic. At the same time, reliability is multidimensional. Schedule adherence, travel time variability, and headway regularity describe different aspects of service performance and may respond differently to the same operating conditions. Existing studies provide important insights into these individual dimensions, but fewer combine planned operations, realised operations, local traffic exposure, and mitigating measures within one detailed segment-level framework. This study therefore examines how operational public transport reliability in dense urban bus and tram segments is associated with local traffic exposure and mitigating measures.

A multidimensional segment-hour framework is developed and applied to 15 selected directed bus and tram segments in Geneva, Switzerland. Geneva is used as an information-rich case because it combines dense surface public transport operations, variation in right-of-way conditions, and the availability of scheduled, realised, and road traffic data within one coherent network. GTFS Schedule data, GTFS Realtime data, and hourly road traffic counter data are combined into a dataset of 228,094 valid segment-hour observations covering the period from 2019 to 2024. Reliability is represented through three complementary indicators: OTP for schedule adherence, CV_{TT} for travel time variability, and CV_h for headway regularity. Local traffic exposure is operationalised through an intensity/capacity ratio derived from hourly traffic counts and estimated reference capacities. The empirical analysis combines descriptive and within-segment analyses with a hypothesis-guided blockwise OLS regression strategy, parsimonious model specifications, and outcome-specific robustness checks.

The results show that the relationship between local traffic exposure and public transport reliability differs substantially across the three reliability dimensions. The clearest and most consistent relationship is observed for CV_{TT}. Higher local traffic exposure is associated with greater travel time variability, indicating that realised stop-to-stop travel times become less stable under more intensive traffic conditions. This relationship remains statistically significant in both the OLS and log-transformed specifications and provides the strongest support for the hypothesis that higher local traffic exposure is associated with lower reliability. The result also illustrates the close correspondence between the spatial scale of CV_{TT} and the analysed stop-to-stop segment, as the indicator is derived directly from realised movement through the selected segment.

The findings for OTP and CV_h are less consistent with the expected traffic-related relationships. For OTP, local traffic exposure is positively associated with the outcome after controlling for the broader operating context, contrary to the hypothesised direction. This positive association persists in the fractional logit specification and therefore cannot be attributed solely to the use of OLS for a bounded outcome. Downstream punctuality is instead strongly associated with the operational state with which vehicles enter the analysed segment, particularly upstream departure delay. For CV_h, the local traffic coefficient is weak and contrary to the expected direction in the OLS specification and loses statistical significance after log transformation. These findings indicate that punctuality and headway regularity reflect broader upstream and route-level operational processes in addition to conditions within a single stop-to-stop segment.

Dedicated public transport infrastructure and traffic signal priority are represented as a mitigating-measure package because they are not observed as fully independent conditions in the selected sample. The presence of this package is not consistently associated with improved reliability: its main effect is statistically insignificant for OTP and CV_{TT} and is associated with higher CV_h. The results therefore do not support a general direct reliability advantage of the mitigating-measure package. For CV_{TT}, however, the interaction between local traffic exposure and the mitigating-measure package is negative and remains weakly significant across the OLS and log-transformed specifications. This provides

limited evidence that, within the selected sample, the positive association between traffic exposure and travel time variability is weaker when the package is present. Because the measures are not randomly distributed across the selected segments, this interaction should be interpreted as a conditional difference in traffic sensitivity rather than as a causal moderating effect.

The findings reinforce the conceptualisation of operational public transport reliability as multidimensional. OTP, CV_{TT}, and CV_h should not be treated as interchangeable indicators, because they capture different operational mechanisms and operate at different effective spatial scales. The segment-hour framework is particularly informative for studying local travel time stability, whereas punctuality and headway regularity require greater consideration of upstream and route-level conditions. The analysis also highlights important methodological limitations. The intensity/capacity ratio provides a consistent proxy for local traffic exposure, but does not directly observe traffic speed, queue formation, spillback, signal phase, or the congestion experienced by an individual public transport vehicle. In addition, the selected segments form a purposive sample and the mitigating measures are not randomly distributed, so the observed relationships should be interpreted as conditional associations rather than network-wide or causal effects.

The main scientific contribution of this study is the development and application of a transferable segment-level framework that integrates planned operations, realised operations, local traffic exposure, and multiple dimensions of reliability for both bus and tram services. The findings show that the interpretation of reliability depends not only on the operating conditions considered, but also on the spatial scale and operational meaning of the selected indicator. For practice, reliability monitoring should therefore combine multiple indicators and distinguish local traffic-related travel time instability from punctuality and headway problems that may originate or propagate elsewhere in the network. Future research should strengthen the causal assessment of mitigating measures through longitudinal before–after designs with comparable control segments, incorporate more detailed traffic, signal, passenger-demand, dwell-time, and upstream operational data, and apply the framework to larger and more diverse samples of segments or other cities.

Contents

Acknowledgements iii

Abstract v

1 Introduction 1
- 1.1 Background and motivation . . . 1
- 1.2 Research problem and relevance . . . 2
- 1.3 Research objective and research questions . . . 2
- 1.4 Scope and case study . . . 3
- 1.5 Structure of the thesis . . . 3

2 Literature Review 5
- 2.1 Literature search and selection . . . 5
- 2.2 Reliability and indicators in public transport . . . 5
  - 2.2.1 Travel time variability indicators . . . 6
  - 2.2.2 Schedule adherence indicators . . . 7
  - 2.2.3 Headway regularity indicators . . . 7
- 2.3 Suitability of reliability indicators across modes and service contexts . . . 8
  - 2.3.1 Bus and tram operations . . . 8
  - 2.3.2 Low-frequency and high-frequency services . . . 8
  - 2.3.3 Operational and behavioural perspectives in urban contexts . . . 9
- 2.4 Literature gap . . . 9
- 2.5 Indicator selection for this study . . . 11

3 Methodology 13
- 3.1 Operationalisation of public transport reliability . . . 13
- 3.2 Construction of the analysis dataset . . . 14
- 3.3 Reliability indicators . . . 15
- 3.4 Segment selection . . . 15
- 3.5 Descriptive and correlation analysis . . . 16
  - 3.5.1 Descriptive analysis . . . 16
  - 3.5.2 Within-segment binned associations . . . 16
  - 3.5.3 Correlation analysis . . . 17
- 3.6 Hypothesis-guided regression strategy . . . 17
  - 3.6.1 Blockwise model construction . . . 19
  - 3.6.2 Manual backward elimination with hierarchy constraints . . . 22
- 3.7 Model diagnostics and robustness checks . . . 23
  - 3.7.1 Multicollinearity . . . 23
  - 3.7.2 Residual diagnostics . . . 23
  - 3.7.3 Log-transformed variability models . . . 24
  - 3.7.4 Bounded nature of on-time performance variable . . . 24

4 Case Study Application 25
- 4.1 Case study area . . . 25
- 4.2 Public transport context . . . 26
- 4.3 Temporal scope and data sources . . . 27
  - 4.3.1 Road traffic sensor data . . . 28
  - 4.3.2 GTFS Schedule . . . 29
  - 4.3.3 GTFS Realtime . . . 29
  - 4.3.4 SITG . . . 29
- 4.4 Segment selection and counter matching . . . 29
- 4.5 Operationalisation of case-specific variables . . . 30

4.5.1 Local traffic exposure . . . 30
4.5.2 Mitigating measures . . . 31
4.5.3 Infrastructure and service context variables . . . 31
4.5.4 Punctuality definition . . . 31
4.6 Selected segments and classification . . . 31
4.7 Construction and coverage of the final analysis dataset . . . 32

5 Results **35**
5.1 Traffic exposure and reliability outcomes . . . 35
5.1.1 Traffic exposure . . . 35
5.1.2 Reliability outcomes . . . 36
5.2 Correlation structure and model implications . . . 36
5.2.1 Within-segment associations between traffic exposure and reliability . . . 36
5.2.2 Correlations between explanatory variables . . . 37
5.3 Regression results . . . 38
5.3.1 Local traffic exposure . . . 38
5.3.2 Mitigating measures . . . 39
5.3.3 Moderation of traffic sensitivity . . . 39
5.3.4 Other contextual associations . . . 40
5.4 Outcome-specific model specifications and robustness . . . 40
5.4.1 Residual diagnostics of the OLS baseline models . . . 40
5.4.2 Log-transformed variability models . . . 42
5.4.3 Fractional logit specification for on-time performance . . . 42
5.5 Summary of empirical findings . . . 43

6 Discussion **45**
6.1 Discussion of the research questions . . . 45
6.1.1 Reliability dimensions and indicator choice . . . 45
6.1.2 Local traffic exposure and reliability . . . 46
6.1.3 Mitigating measures and traffic sensitivity . . . 46
6.2 Methodological reflection . . . 47
6.2.1 Segment-hour unit of analysis . . . 47
6.2.2 Operationalisation of local traffic exposure . . . 48
6.2.3 Temporal aggregation . . . 48
6.2.4 Selection and operationalisation of mitigating measures . . . 48
6.2.5 Model specification and outcome-specific robustness . . . 48
6.3 Implications for reliability evaluations and practice . . . 49
6.4 Limitations . . . 50

7 Conclusions **51**
7.1 Answers to the research questions . . . 51
7.2 Answer to the main research question . . . 52
7.3 Scientific contribution . . . 52
7.4 Practical recommendations and future research . . . 52
7.5 Final conclusion . . . 53

References **55**

A GTFS Structure **59**

B Information on Selected Segments **61**
B.1 Location within Geneva . . . 61
B.2 Segments and traffic-counter information . . . 61
B.3 Street-level view . . . 62

C Traffic Data from Counting Station **65**

D Descriptive Distributions of Reliability Outcomes **71**
D.1 Distribution of segment-hour OTP by segment . . . 71
D.2 Distribution of segment-hour CV_{TT} by segment . . . 72

D.3 Distribution of segment-hour CV_h by segment 72

E Regression Model Development **75**
E.1 On-time performance model development 75
E.2 Travel time variability model development 77
E.3 Headway regularity model development 79
E.4 Multicollinearity during model development 80

F Additional Outcome-Specific Model Diagnostics **83**
F.1 Distributional characteristics of the variability indicators 83
F.2 Fitted-value diagnostics for on-time performance 84

G Full Regression Output **85**
G.1 Full on-time performance regression output 85
G.2 Full travel time variability regression output 89
G.3 Full headway regularity regression output 94
G.4 Full log-transformed variability model output 97
G.5 Full fractional logit output for on-time performance 98

H Statement on the Use of Generative AI **99**

List of Figures

3.1 Methodological framework with the conceptualisation of public transport reliability. . . . 13
3.2 Aggregation structure used to construct the segment-time analysis dataset. 14
3.3 Hierarchical blockwise regression sequence and corresponding hypotheses. 18

4.1 Administrative boundaries of the Geneva metropolitan area, canton, and city. 25
4.2 Overview of the Geneva public transport network. Click on the figures to enlarge. 26
4.3 Spatial coverage of the main bus and tram services in the Geneva urban area. 27
4.4 Fixed road traffic counters associated with the bus and tram network during the morning peak (07:00-09:00) and evening peak (16:00-18:00). 28
4.5 Number of candidate segments retained after each selection step. 30

5.1 Distribution of segment-hour IC ratios by selected segment. 35
5.2 Within-segment binned associations between deviations in local traffic exposure and deviations in reliability outcomes. 36
5.3 Pairwise correlations between explanatory variables used in the regression analysis. . . 37
5.4 Residuals versus fitted values and Q-Q plot for the final parsimonious OTP model. . . . 40
5.5 Residuals versus fitted values and Q-Q plot for the final parsimonious CV_{TT} model. . . 41
5.6 Residuals versus fitted values and Q-Q plot for the final parsimonious CV_h model. . . . 41

A.1 Structure of the GTFS dataset; the main tables and the key columns linking them. . . . 59

B.1 Locations of the selected stop-to-stop public transport segments and associated road traffic counters in the Geneva urban area. 61
B.2 Illustrative street-level view of the selected segment environments (Apple Inc., 2026). . 64

C.1 Daily average vehicle counts from counting stations at all segment locations, 2017–2024. 70

D.1 Distribution of segment-hour OTP values by selected segment. 71
D.2 Distribution of segment-hour CV_{TT} values by selected segment. 72
D.3 Distribution of segment-hour CV_h values by selected segment. 72

F.1 Original and log-transformed distributions of CV_{TT}. 83
F.2 Original and log-transformed distributions of CV_h. 84
F.3 Distribution of fitted values from the final parsimonious OLS model for OTP. 84

List of Tables

2.1 Overview of the main studies reviewed and their relevance to the present study. 10

3.1 Overview of dependent variables, explanatory variables, and level of variation used in the regression analysis. 19

4.1 Classification of selected segments. 32
4.2 Observation coverage of the final analysis dataset by segment. 33

5.1 Final parsimonious regression models. 38
5.2 Comparison of coefficient estimates between the original and log-transformed final CV_{TT} model. 42
5.3 Comparison of coefficient estimates between the original and log-transformed final CV_h model. 42
5.4 Comparison of coefficient estimates between the final parsimonious OLS and fractional logit OTP models. 43

B.1 Selected stop-to-stop public transport segments and associated road traffic counters. . 62

E.1 Blockwise OLS regression results for On-Time Performance (OTP). 75
E.2 Backward Elimination OLS regression results for On-Time Performance (OTP). 76
E.3 Blockwise OLS regression results for Coefficient of Variation of Travel Time (CV_{TT}). . . 77
E.4 Backward Elimination OLS regression results for Coefficient of Variation of Travel Time (CV_{TT}). 78
E.5 Blockwise OLS regression results for Coefficient of Variation of Headway (CV_h). 79
E.6 Backward Elimination OLS regression results for Coefficient of Variation of Headway (CV_h). 80
E.7 Variance inflation factors for the final blockwise model specifications. 81
E.8 Variance inflation factors for the final parsimonious model specifications. 81

F.1 Distributional summary of CV_{TT} and CV_h in the final model sample. 83
F.2 Summary statistics of fitted values from the final OLS and fractional logit specifications for OTP. 84

[*This page intentionally left blank*]

1 Introduction

1.1. Background and motivation

Public transport reliability is widely recognised as an important component of service quality. When services do not operate consistently, passengers may experience longer waiting times, missed connections, and less predictable journeys. As a result, unreliable service can reduce passenger satisfaction and weaken the attractiveness of public transport as a competitive mode of transport (Soza-Parra et al., 2019). For operators, reliability is equally important, since delays and irregular service patterns can affect the efficiency of operations and the overall performance of the network (Buran, 2025; Soza-Parra et al., 2019).

In dense urban environments, maintaining reliable public transport services is particularly challenging. Surface public transport modes such as buses and trams operate in streets where performance may be affected by signalised intersections, limited street space, frequent stops, and interactions with surrounding traffic. Under such conditions, realised operations may differ substantially from planned operations. Vehicles may experience delays, irregular spacing, or variation in realised travel times. Reliability in urban segments is therefore shaped not only by the timetable itself, but also by the operating environment in which the service is executed (Anderson & Geroliminis, 2020; Van Oort, 2011).

This makes the distinction between planned and realised operations particularly relevant. Planned operations describe the intended service pattern through, for example, scheduled arrival times, travel times, and headways, whereas realised operations reflect how services are actually performed in practice. Comparing these two layers can reveal deviations from the timetable as well as instability in realised travel times and vehicle spacing. Analysing the relationship between schedule and reality is therefore essential for understanding public transport reliability in operational terms (Jenelius, 2018).

One important factor that may influence realised service performance is the surrounding traffic environment. Especially for surface public transport modes, local traffic conditions may affect the consistency of operations through congestion, intersection delay, and broader mixed-traffic interactions (Anderson & Geroliminis, 2020). At the same time, operators and planners implement measures intended to reduce these interactions. Dedicated public transport infrastructure can separate buses and trams from general traffic, while traffic signal priority can reduce interference at signalised intersections. Such measures are generally intended to support more reliable operations, although their relevance may depend on the local operating conditions in which they are implemented (Fadaei & Cats, 2016; Federal Highway Administration, 2017).

Developments in data availability create new opportunities to study these relationships in greater detail. Scheduled timetable data can describe the intended service, realised operational data can reveal how the service performs in practice, and road traffic data can provide information on the surrounding operating environment. Combining these different perspectives makes it possible to study public transport reliability at a detailed operational level and provides the motivation for examining how local traffic conditions and mitigating measures relate to reliability in dense urban bus and tram operations.

1.2. Research problem and relevance

Although public transport reliability has been studied extensively, the existing literature remains fragmented in the way service performance is conceptualised and analysed. Previous studies often focus on specific indicators, individual data sources, or particular operational interventions. In addition, different studies emphasise different dimensions of reliability, including schedule adherence, travel time variability, headway regularity, and passenger-oriented outcomes. Consequently, the interpretation of reliability depends strongly on the indicator and operational perspective considered (Gittens & Shalaby, 2015; Tirachini et al., 2022; Trompet et al., 2011).

This fragmentation is especially relevant for dense urban bus and tram operations. Reliability problems may arise from the interaction between the planned service, the realised operation, local traffic conditions, infrastructure characteristics, and operational measures. These mechanisms are not necessarily reflected equally by different reliability indicators. A service may adhere relatively well to the timetable while still showing substantial variation in realised travel times or irregular vehicle spacing. Conversely, unstable timetable performance may partly originate upstream rather than within the local segment being observed. Evaluating reliability therefore requires consideration of multiple operational dimensions rather than reliance on a single performance indicator.

As discussed in Chapter 2, relatively few studies explicitly combine planned public transport operations, realised operations, and local road traffic conditions within one detailed segment-level analytical framework. Existing research provides important insights into individual reliability indicators, operational data, traffic interactions, and mitigating interventions, but these elements are less frequently examined jointly for both bus and tram operations. This limits the ability to assess how different dimensions of reliability respond to the same local operating conditions (De Brito-Filho & Oliveira-Neto, 2025; Fadaei & Cats, 2016).

The role of mitigating measures creates an additional evaluation challenge. Dedicated public transport infrastructure and traffic signal priority are intended to reduce the influence of surrounding traffic, but such measures are generally not distributed randomly across an urban network. They may be implemented at locations where interactions with traffic are already particularly relevant. Comparing locations with and without such measures therefore requires consideration of the operating context in which they are observed and limits the extent to which observed differences can be interpreted as direct effects of the measures.

Addressing these issues is scientifically relevant because it provides a more integrated understanding of how operational reliability relates to local traffic conditions and mitigating measures across different reliability dimensions. It is also practically relevant for public transport operators and planners, since identifying the type and spatial scale of a reliability problem is an important step in determining which operational or infrastructural responses may be appropriate.

This study therefore focuses on the relationship between operational public transport reliability, local traffic exposure, and mitigating measures in dense urban bus and tram segments. By considering multiple dimensions of reliability within the same analytical setting, the study aims to provide a framework for distinguishing reliability problems that are closely related to local traffic interactions from those that reflect broader operational processes.

1.3. Research objective and research questions

The objective of this study is to develop and apply a multidimensional framework for analysing operational public transport reliability in dense urban bus and tram segments, with particular attention to local traffic exposure and mitigating measures. To achieve this objective, the following main research question is addressed:

➜ **How is public transport reliability in dense urban bus and tram segments associated with local traffic exposure and mitigating measures?**

This main question is supported by the following sub-questions:

1. What conceptualisation and measurement framework is suitable for analysing operational reliability in dense urban bus and tram segments?

2. To what extent is local traffic exposure associated with different dimensions of reliability across the selected segments?
3. How are mitigating measures, consisting of dedicated public transport infrastructure and traffic signal priority, associated with the different dimensions of reliability across the selected segments?
4. How does the relationship between local traffic exposure and reliability differ between segments with and without the mitigating-measure package?
5. What operational and infrastructural recommendations follow from the empirical findings for the analysed segments?

1.4. Scope and case study

This study focuses on operational public transport reliability in selected bus and tram segments in the Geneva urban area. Geneva provides a relevant empirical setting because it combines dense surface public transport operations with variation in traffic exposure and right-of-way conditions within one coherent urban network. The study considers selected segments operated by transports publics genevois (tpg) and uses Geneva as an information-rich case rather than as a statistically representative case of urban public transport more generally. A detailed description of the case study, data sources, and selected segments is provided in Chapter 4.

The empirical scope is limited to directed stop-to-stop segments for which scheduled operations, realised operations, and local traffic exposure can be combined consistently. The selected segments are therefore intended to represent analytically meaningful operating contexts rather than the Geneva network as a whole. The findings should consequently be interpreted for the analysed segment-time observations and should not be understood as network-wide average effects.

The study focuses on operational reliability derived from scheduled, realised, and traffic-related data. It does not analyse passenger perception, customer satisfaction, route choice behaviour, crowding, or network-wide accessibility. These aspects are important for understanding the broader consequences of reliability, but they are outside the scope of this study. The focus is instead on the operational manifestation of reliability through schedule adherence, travel time variability, and headway regularity.

Finally, the study examines mitigating measures through their observed presence within the selected segment contexts and does not aim to estimate causal treatment effects of individual interventions. Dedicated public transport infrastructure and traffic signal priority are not observed as fully independent conditions within the selected sample, and the empirical analysis therefore considers them as a mitigating-measure package. The resulting associations describe differences between observed operating contexts and should not be interpreted as isolated causal effects of a specific measure.

Although the empirical findings are case-specific, the methodological framework is intended to be transferable to other public transport systems where scheduled operations, realised operations, and local traffic data can be integrated at a comparable spatial and temporal scale.

1.5. Structure of the thesis

The remainder of this study is structured as follows. Chapter 2 reviews the literature on public transport reliability, reliability indicators, traffic exposure, and mitigating measures. It identifies the main research gap and selects the reliability indicators used in the empirical analysis. Chapter 3 presents the methodology, including the operationalisation of reliability, the construction of the segment-time dataset, the segment selection logic, and the descriptive, correlation, and regression analyses. Chapter 4 introduces the Geneva case study, describes the public transport context and data sources, and explains how the selected segments and case-specific variables are operationalised.

Chapter 5 presents the empirical results, including descriptive patterns, within-segment associations, regression results, and outcome-specific robustness checks. Chapter 6 discusses the findings in relation to the research questions, reflects on the methodological choices and their implications, and considers the broader meaning and limitations of the results. Chapter 7 concludes the thesis by answering the research questions, summarising the scientific contribution, and presenting recommendations for practice and future research.

The appendices provide supporting material for the empirical analysis. Appendix A summarises the GTFS data structure, while Appendices B and C provide additional information on the selected segments and traffic counting stations. Appendix D presents supplementary descriptive distributions, and Appendices E, F, and G contain the detailed regression model development, additional diagnostics, and full model output. Finally, Appendix H contains the statement on the use of generative AI.

2 Literature Review

This chapter addresses the first sub-question of this study: "What conceptualisation and measurement framework is suitable for analysing operational reliability in dense urban bus and tram segments". To do so, it reviews how reliability is defined in the literature, which indicator groups are commonly used, and how their suitability differs across operational contexts. The chapter then identifies the main gaps in the existing literature and uses these insights to position the present study and to develop the theoretical framework for the empirical analysis.

2.1. Literature search and selection

A structured literature review was conducted to identify relevant conceptualisations and operational indicators of public transport reliability, as well as studies examining the relationship between reliability, traffic conditions, and mitigating measures.

Relevant literature was identified through searches in *Google Scholar*, *ScienceDirect*, and *ScienceGate*. The search combined terms related to public transport reliability, including *public transport reliability*, *bus reliability*, *tram reliability*, *schedule adherence*, *on-time performance*, *travel time variability*, *headway regularity* and *service regularity*, with terms related to the operational context, such as *urban traffic*, *mixed traffic*, *dedicated lane*, *bus lane* and *traffic signal priority*. Variations and combinations of these terms were used to account for differences in terminology across studies.

Publications were selected based on their relevance to operational reliability, their applicability to urban bus or tram services, and their contribution to the conceptualisation, measurement, or explanation of reliability. Studies focusing primarily on passenger perceptions or behavioural responses were included where they contributed to the conceptual distinction between operational and passenger-oriented reliability but were not central to the indicator selection. Additional studies were identified through snowballing from key publications. The resulting literature base was used to classify reliability indicators, assess their suitability across service contexts, and identify limitations in existing empirical frameworks.

2.2. Reliability and indicators in public transport

Public transport reliability is a widely used concept in the literature, but its exact meaning is not always defined in the same way. A common aspect in many definitions is that reliability is defined as the ability of the service to provide a consistent service over time. For example, Gittens and Shalaby (2015) relate reliability to the variability of service, and Polus (1978) and Lomax et al. (2003) similarly emphasise service consistency. These definitions suggest that public transport reliability is not only about average performance, but mainly about the extent to which service outcomes remain predictable.

At the same time, the literature shows that different aspects of reliability can be emphasised. Bowman and Turnquist (1981), for example, relate reliability primarily to schedule adherence, whereas other studies place more emphasis on service regularity, particularly in the context of frequent urban services (Tirachini et al., 2022; Trompet et al., 2011). In addition, Bates et al. (2001) and Kimpel et al. (2000) suggest that reliability cannot be reduced to only one of these aspects, since a service should be both consistent and reasonably close to its planned operation.

This implies that reliability should not be treated as a single directly measurable property. Instead, it

is better understood as a broader operational concept that can be observed through different types of indicators, each capturing a different dimension of service performance. For the purpose of this study, this distinction is important, because the evaluation of mitigating measures requires indicators that reflect not only whether service is delayed, but also how stable and predictable realised operations are under varying traffic conditions.

Based on the reviewed literature, operational reliability indicators can be broadly organised into three complementary categories: travel time variability, which captures the dispersion of realised travel times (Gittens & Shalaby, 2015; Kieu et al., 2014), schedule adherence, which evaluates realised operations relative to the timetable (Trompet et al., 2011; Van Oort, 2011), and headway regularity, which describes the spacing between consecutive vehicles (Tirachini et al., 2022; Trompet et al., 2011).

The literature also shows that not every indicator is equally useful for every purpose. Studies that compare reliability measures in practice argue that indicator selection should depend on criteria such as interpretability, comparability, objectivity, customer relevance, and data requirements (Currie et al., 2011; Strathman et al., 1999; Trompet et al., 2011). These considerations are relevant for answering the first sub-question, because the aim is not simply to list available measures, but to identify which indicators are suitable for analysing reliability in dense urban bus and tram segments and for evaluating the role of mitigating measures.

The literature suggests that indicator suitability depends not only on mode and frequency but also on the analytical perspective of the study. Passenger-oriented indicators are appropriate when the objective is to understand perceived service quality or behavioural response. Operational indicators are more appropriate when the objective is to evaluate realised service performance using vehicle-level data. Since this study focuses on operational reliability at segment level, the selected indicators must be calculable from scheduled and realised public transport observations.

2.2.1. Travel time variability indicators

The first category consists of travel time variability indicators. These indicators capture the extent to which realised travel times vary across observations and therefore reflect how predictable service performance is in operation. Their main function is to describe not only the average travel time but also the spread of the realised performance.

Within this category, the literature includes several indicators that approach travel time uncertainty in different ways:

- **Reliability buffer time** represents the additional travel time required, beyond the average travel time, to achieve a specified probability of arriving on time. It translates travel time variability into an intuitive margin that can inform both passenger journey planning and operator timetable design. However, its focus lies primarily on the practical consequences of variability rather than on variability as an operational phenomenon itself (Lomax et al., 2003; Schrank & Lomax, 2002).
- **Reliability buffer index** expresses the additional time relative to the average travel time. This supports comparison across contexts, since the indicator standardises the required buffer with respect to the typical duration of the trip. However, like reliability buffer time, it remains strongly linked to traveller risk and journey uncertainty, rather than to operational variability itself (Lomax et al., 2003; Schrank & Lomax, 2002).
- **Travel time window** reflects the range within which travel times are likely to fall. This makes it useful for describing the overall breadth of travel time uncertainty and for illustrating how wide the distribution of realised outcomes may be. However, it provides less direct insight into relative dispersion across services, since the width of the window may still be strongly influenced by the absolute scale of the trip (Uniman et al., 2010).
- **Coefficient of variation of travel time** expresses variability relative to the mean travel time. This makes it particularly suitable for comparing services or segments with different average travel times, because it captures dispersion in a normalised way rather than in absolute terms. At the same time, its interpretation may be less intuitive than measures framed more directly in terms of additional travel time, especially for non-technical audiences (Gittens & Shalaby, 2015).

These studies show that travel time variability indicators are especially useful when the aim is to assess

how stable realised operations are under varying conditions. However, they also differ in orientation: some are more closely connected to the passenger perspective and journey planning, while others are more directly suited to operational comparison. This distinction is important when assessing which travel time variability indicators are most appropriate in a given analytical context.

2.2.2. Schedule adherence indicators

The second category consists of schedule adherence indicators, which compare realised arrival or departure times with the timetable. These indicators are used to assess the extent to which services operate according to schedule and are therefore closely related to punctuality. Their main value is that they directly connect realised operations to the planned timetable, making them relevant in service contexts where passengers rely on scheduled departure or arrival times.

Within this category, the literature includes several indicators that reflect timetable deviation in different ways:

- **On-time performance** expresses the share of trips or stop events that fall within a predefined tolerance window around the scheduled time. This makes it one of the most widely used schedule adherence indicators, as it is intuitive, easy to communicate, and straightforward to calculate (Jenks et al., 2013). At the same time, its threshold-based nature means that it does not show the magnitude of deviations outside the accepted range.
- **Earliness-based indicators** focus specifically on vehicles departing or arriving too early. This is relevant because early departures may be particularly problematic for passengers, as they can lead to missed services even when the deviation from schedule is small in absolute terms. However, such indicators capture only one side of schedule deviation and therefore provide a more partial view of adherence (Saberi et al., 2013).
- **Weighted delay measures** place more emphasis on the severity of timetable deviations by incorporating the magnitude of delay into the indicator itself. This makes them more informative than purely threshold-based measures when the aim is to distinguish between small and substantial deviations. At the same time, their interpretation is often less immediate than that of simpler measures such as on-time performance (Camus et al., 2005).

These studies show that schedule adherence indicators are useful because they relate realised performance directly to the timetable and are therefore closely aligned with conventional notions of punctuality. However, they also differ in analytical depth: while some indicators offer a simple and communicable summary of whether service is on time, others provide more insight into the direction or magnitude of timetable deviations.

2.2.3. Headway regularity indicators

The third category consists of headway regularity indicators, which focus on the spacing between consecutive vehicles rather than on deviations from scheduled times. These indicators are especially relevant in frequent services, where passengers are often less dependent on exact timetable adherence and more affected by the regularity with which vehicles arrive. Their main value is that they capture how evenly service is distributed in operation, which is important in urban contexts where irregular spacing may lead to bunching, uneven waiting times, and crowding.

Within this category, the literature includes several indicators that describe headway irregularity in different ways:

- **Coefficient of variation of headway** expresses variability relative to the mean headway. This makes it particularly useful for comparing the regularity of services with different average frequencies, since it captures dispersion in a normalised and continuous way (Trompet et al., 2011). At the same time, its interpretation may be less direct than measures expressed more explicitly in terms of service irregularity.
- **Headway regularity** measures are designed more explicitly around the extent to which vehicles remain evenly spaced in operation. Their main strength is that they reflect the operational phenomenon that matters most in frequent services, namely whether passengers experience a balanced stream of vehicles rather than long gaps followed by bunching. However, the exact for-

mulation of such measures may differ between studies, which can make direct comparison less straightforward (Saberi et al., 2013; Trompet et al., 2011).

- **Distribution-based measures**, such as width indices, describe the spread of observed headways within a service. This makes them useful for illustrating how broad the range of realised spacing is and for identifying unstable service patterns. At the same time, such measures may provide less direct insight into relative comparability across services when the absolute scale of the headway differs substantially (Cramer et al., 2009).

These studies show that headway regularity indicators are especially useful in frequent urban services, where reliability is often experienced less through strict timetable adherence and more through the regular arrival of vehicles. However, these indicators also differ in emphasis: some are more suitable for continuous comparison across services, while others are better at describing the operational manifestation of irregular spacing itself.

2.3. Suitability of reliability indicators across modes and service contexts

The previous section reviewed the main categories of reliability indicators and showed that each captures a different dimension of service performance. However, the literature also makes clear that no single indicator is equally informative in every context. The suitability of an indicator depends on the type of service being analysed, the operational environment in which that service runs, and the analytical purpose for which reliability is assessed. This section examines how the relevance of different indicator groups varies across modes and service contexts, as a basis for the indicator selection in the present study.

2.3.1. Bus and tram operations

A first distinction concerns differences between public transport modes. Reliability may differ between modes because buses, trams, metro services, and other rail-based services operate under different operational conditions. Soza-Parra et al. (2021) show that reliability varies across modes and that both headway variability and travel time variability can play an important role in explaining passenger preferences. Similarly, Dixit et al. (2019) apply a reliability-based approach to multimodal public transport journeys in a network that includes bus, tram, and metro services, showing that travel time reliability can be analysed across modes within one system. These studies suggest that modal differences should be considered when assessing reliability, but they do not imply that each mode necessarily requires a separate set of indicators.

For the present study, the relevance of mode lies mainly in the interpretation of reliability indicators. Van Oort (2011) discusses service reliability in urban bus, tram, and metro operations and shows that reliability can be understood at different levels, including planning, operations, and passenger effects. From an operational perspective, this means that schedule adherence, travel time variability, and headway regularity can all be meaningful for both bus and tram services, but the mechanisms behind these outcomes may differ. Bus services often operate more directly in mixed traffic and may therefore be more sensitive to congestion and delays caused by surrounding road traffic. Tram services may be more protected through dedicated infrastructure, but can still be affected by intersections, shared right-of-way sections, and local traffic interactions.

2.3.2. Low-frequency and high-frequency services

A second distinction concerns the service context itself, especially the difference between high-frequency and low-frequency services. The literature consistently suggests that schedule adherence indicators are more suitable where passengers rely on a timetable, whereas headway-based indicators become more relevant when vehicles operate frequently and passengers are less concerned with exact departure times. This distinction is also reflected in the review by Tirachini et al. (2022), which identifies headway variability as a central reliability issue in frequent urban public transport services. In such contexts, irregular spacing between vehicles can lead to bunching, uneven waiting times, and crowding, even when average punctuality remains acceptable.

At the same time, punctuality alone is often too narrow to capture the full extent of operational un-

reliability. Kieu et al. (2014) emphasise that public transport travel time variability should be treated as a distinct concept and monitored explicitly rather than being reduced to simple timetable deviation measures. This is particularly relevant in dense urban segments, where interactions with road traffic, intersections, and operational disturbances may lead not only to delay, but also to increased variation in travel time between observations. Travel time variability indicators are therefore especially useful when the aim is to capture the spread and predictability of realised performance, rather than only whether a service falls within a given threshold.

2.3.3. Operational and behavioural perspectives in urban contexts

The literature on urban public transport reliability also includes approaches that move beyond vehicle operations alone. Jenelius (2018), for example, integrates travel time and travel conditions into a broader concept of experienced service reliability, while Shelat et al. (2023) show that waiting time reliability can affect route choice in urban public transport networks. These studies are important because they demonstrate that reliability has consequences beyond the operation of vehicles itself and can influence how passengers experience and respond to service quality.

At the same time, these broader perspectives are not equally suitable for every research purpose. Indicators related to perceived reliability, waiting time reliability, or behavioural response are especially relevant when the aim is to understand passenger experience or travel behaviour. By contrast, studies that focus on operational service performance require indicators that can be derived directly from realised and planned service data. The distinction between behavioural and operational perspectives is therefore not only conceptual but also methodological, since it determines which forms of reliability can be measured with the available data.

The literature thus suggests that the suitability of reliability indicators depends on both mode and service context. Schedule adherence indicators are especially useful when timetable compliance is central, headway regularity indicators are particularly relevant in frequent urban services, and travel time variability indicators are valuable when the aim is to understand the spread of operational performance under varying traffic conditions. At the same time, broader passenger-oriented indicators may be important in other research settings, but are less directly suited to studies centred on operational performance. These differences will provide the basis for selecting indicators that are appropriate for the analytical focus of the present study.

Overall, the literature demonstrates that travel time variability, schedule adherence, and headway regularity provide complementary perspectives on operational reliability. However, the reviewed studies differ in the extent to which these dimensions are considered jointly. As summarised in Table 2.1, many studies emphasise a specific reliability dimension or analytical context, whereas fewer combine multiple dimensions within the same operational assessment.

2.4. Literature gap

The reviewed literature provides a broad basis for understanding and measuring public transport reliability, and several studies already offer multidimensional or integrated assessment approaches. However, these approaches do not fully address the analytical setting of the present study. Existing frameworks often focus on a specific reliability dimension, mode, service context, or intervention type, and they do not consistently combine schedule adherence, travel time variability, and headway regularity with segment-level traffic exposure and mitigating measures. For example, Fadaei and Cats (2016) provide a strong basis for evaluating operational measures, but do not explicitly integrate external road traffic exposure as a central explanatory component. De Brito-Filho and Oliveira-Neto (2025) combine multiple data sources in an intervention-oriented analysis, but focus exclusively on bus operations. Multimodal studies such as Soza-Parra et al. (2021) and Dixit et al. (2019) consider differences between modes, but are less focused on segment-level traffic interactions and the joint evaluation of mitigating measures. These studies therefore provide important building blocks, but no directly transferable framework was identified that jointly relates multiple operational reliability dimensions to planned and realised operations, local traffic conditions, mode, and mitigating measures in dense urban bus and tram segments.

A second limitation concerns the perspective from which reliability is analysed. A substantial part of

the literature focuses on passenger experience, for example through perceived reliability, waiting time reliability, route choice, or customer satisfaction. These approaches are valuable for understanding how unreliability is experienced by travellers, but they are less directly suited to a segment-level operational analysis based on planned and realised vehicle data. For studies that aim to evaluate operational performance, indicators are needed that can be derived consistently from service observations and linked to local operating conditions.

A further limitation concerns the integration of operational and contextual data. Existing studies combine planned or realised public transport data with traffic conditions and intervention characteristics in different ways, but the reviewed literature does not provide a directly transferable approach that jointly links planned service characteristics, realised operations, local time-varying road traffic exposure, and mitigating measures at the stop-to-stop segment level. In particular, local traffic conditions are not consistently incorporated as a central explanatory variable alongside multiple operational reliability indicators. As a result, existing approaches provide limited support for distinguishing how reliability is associated with the planned service, the realised traffic environment, and the presence of protective infrastructure within one analysis.

Finally, the literature provides limited evidence on how mitigating measures should be evaluated in relation to traffic exposure and mode within the same segment-level framework. Measures such as dedicated public transport infrastructure and traffic signal priority are often intended to reduce the influence of surrounding traffic on public transport operations, but their effects are difficult to isolate from the broader operational and infrastructural context in which they are implemented. It therefore remains insufficiently understood how traffic conditions, right-of-way characteristics, mode, and mitigating measures jointly relate to reliability outcomes when planned and realised operations are analysed together.

Table 2.1: Overview of the main studies reviewed and their relevance to the present study.

Study	Mode[a]	Context[b]	Reliability indicators[c]					Data used[d]	Traffic[e]	Measures[f]	Relevance / limitation
			TTV	SA	HR	PO	MO				
Gittens and Shalaby, 2015	B	–	✔	✔	✔	✔	✘	review	✘	✘	Useful for indicator classification; no link to traffic conditions or segment measures.
Camus et al., 2005	B	–	✘	✔	✘	✘	✘	AVL	✔	✘	Useful beyond OTP; not focused on segment context or interventions.
Trompet et al., 2011	B	U, HF	✘	✘	✔	✘	✘	operator performance	✘	✘	Strong on frequent-service regularity; limited to bus benchmarking.
Kieu et al., 2014	PT	–	✔	✘	✘	✘	✘	operational	✘	✘	Important for travel time variability; not focused on modes or measures.
Soza-Parra et al., 2021	B, T, M	U	✔	✘	✔	✔	✘	service, passenger	✔	✘	Strong on modal differences; less focused on operational traffic interactions.
Dixit et al., 2019	B, T, M	U	✔	✘	✘	✘	✘	journey	✘	✘	Useful for multimodal comparison; less focused on segment-level operations.
Van Oort, 2011	B, T, M	U	✔	✔	✘	✔	✘	conceptual	✔	✘	Core source on urban PT reliability; limited integration of traffic data.

Table continues on next page.

Table continued from previous page.

Study	Mode[a]	Context[b]	Reliability indicators[c] TTV	SA	HR	PO	MO	Data used[d]	Traffic[e]	Measures[f]	Relevance / limitation
Tirachini et al., 2022	PT	U, HF	✘	✘	✔	✘	✘	review	✔	✘	Strong on metric choice in frequent urban services; not an empirical segment study.
Jenelius, 2018	PT	U	✘	✘	✘	✔	✘	AVL, APC	✔	✘	Useful for broader reliability concept; more passenger-oriented than this study.
Fadaei and Cats, 2016	B	U	✘	✘	✘	✔	✔	AVL, APC	✔	✔	Strong for evaluating measures; limited explicit use of external traffic data.
De Brito-Filho and Oliveira-Neto, 2025	B	C	✘	✘	✘	✘	✔	AVL, APC, GTFS	✔	✔	Strong multi-data intervention study; limited to bus segments.
This study	**B, T**	**U, C**	✔	✔	✔	✘	✘	**GTFS, GTFS-RT, traffic counts**	✔	✔	**Combines planned and realised operations with traffic conditions for bus and tram segments.**

[a] Mode: **B** = bus, **T** = tram, **M** = metro, **PT** = general public transport.
[b] Context: **U** = urban, **HF** = high-frequency, **C** = corridor/segment.
[c] Reliability indicator: **TTV** = travel time variability, **SA** = schedule adherence, **HR** = headway regularity, **PO** = passenger-oriented, **MO** = measure-oriented.
[d] Data: **AVL** = automated vehicle location, **APC** = automatic passenger counting, **GTFS** = General Transit Feed Specification.
[e] Road traffic included: ✔ = explicitly included, ✔ = included indirectly or not as a main variable, ✘ = not included.
[f] Evaluation interventions/measures: ✔ = explicit evaluation of interventions or mitigating measures, ✘ = no explicit evaluation.

2.5. Indicator selection for this study

For dense urban bus and tram services, contextual dependence is important. In such settings, public transport operations are often exposed to interactions with general traffic, intersections, and varying degrees of right-of-way protection. This makes reliability not only a matter of timetable adherence but also of service regularity and variation in realised travel times. The literature suggests that an operational assessment of reliability in urban segments should not rely on a single indicator alone but should recognise the complementary value of different indicator dimensions.

At the same time, the review indicates that not all indicators are equally suitable. Since this study adopts an operational perspective and focuses on segment-level performance under varying traffic conditions, the selected indicators must be both analytically meaningful and compatible with the available planned and realised service data. In addition, they should allow comparison across different segment types and support the evaluation of mitigating measures in a consistent way.

Based on these considerations, three indicators are selected: on-time performance, the coefficient of variation of headway, and the coefficient of variation of travel time. These indicators reflect the three main dimensions of reliability identified in the literature, while also fitting the analytical focus of the study.

- On-time performance is selected to represent the schedule adherence dimension of reliability. It is one of the most widely used and most intuitive indicators in public transport analysis, because it directly expresses the extent to which realised operations remain close to the timetable. This makes it useful for assessing whether services continue to operate according to plan and whether mitigating measures help reduce deviations from the schedule. At the same time, on-time performance is easy to interpret and communicate, which strengthens its usefulness as a

complementary indicator in the empirical analysis.

- The coefficient of variation of headway is selected to represent the headway regularity dimension of reliability. In frequent urban bus and tram services, timetable adherence alone is not sufficient, because passengers are often affected more by irregular spacing between vehicles than by strict deviations from scheduled times. Uneven headways may lead to longer and less predictable waiting times, bunching, and uneven passenger loads. The coefficient of variation of headway is particularly suitable in this context because it expresses variability relative to the mean headway, thereby enabling comparison across services and segments with different average frequencies. This makes it a useful and consistent operational indicator for the present study.
- The coefficient of variation of travel time is selected to represent the travel time variability dimension of reliability. This indicator is particularly suitable for examining local segment-level variability under changing traffic conditions, because it captures the stability of realised stop-to-stop travel times. By expressing variability relative to the mean travel time, it supports comparison across segments with different average travel times and provides a normalised measure of operational instability. This makes it particularly appropriate for a study in which differences between bus and tram segments, and between more and less protected operating conditions, are compared within one framework.

Based on the reviewed literature, public transport reliability is most appropriately conceptualised as a multidimensional operational concept rather than as a single measurable outcome. For the purpose of evaluating mitigating measures in dense urban bus and tram segments, the framework therefore combines three complementary dimensions: schedule adherence, travel time variability, and headway regularity. These dimensions are operationalised using on-time performance, the coefficient of variation of travel time, and the coefficient of variation of headway, respectively. Together, these indicators provide a measurement framework that captures both adherence to planned operations and the stability and regularity of realised service performance. This answers the first sub-question by defining reliability as a multidimensional operational concept and by selecting a complementary set of indicators that can be applied consistently to dense urban bus and tram segments.

3 Methodology

After answering the first sub-question in Chapter 2, this chapter develops the methodological approach used to answer the main research question: “How is public transport reliability in dense urban bus and tram segments associated with local traffic exposure and mitigating measures?”

The chapter is structured as follows. Section 3.1 operationalises public transport reliability and defines the role of the selected indicator groups. Section 3.2 explains how the scheduled, realised, and traffic datasets are transformed into a segment-time analysis dataset, and describes the validity criteria applied during this process. Section 3.3 defines the reliability indicators used in the empirical analysis. Section 3.4 presents the criteria used to select the network segments. Section 3.5 and Section 3.6 describe the descriptive, correlation, and regression analyses used to examine the relationships between reliability outcomes and explanatory variables. Finally, Section 3.7 discusses model diagnostics and robustness checks, including multicollinearity, residual diagnostics, log-transformed variability models, and the bounded nature of the on-time performance outcome.

3.1. Operationalisation of public transport reliability

Chapter 2 conceptualised public transport reliability as a multidimensional operational concept and identified three complementary dimensions: schedule adherence, travel time variability, and headway regularity. This section translates that conceptualisation into the methodological framework used in the empirical analysis.

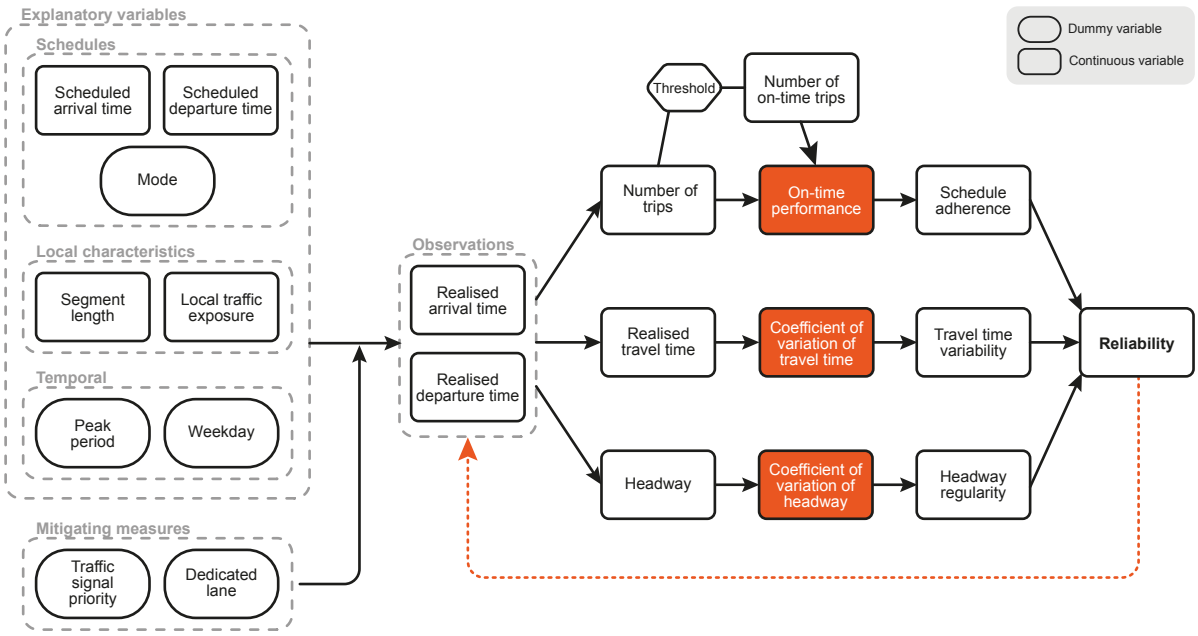


Figure 3.1: Methodological framework with the conceptualisation of public transport reliability.

Figure 3.1 summarises how these reliability dimensions are operationalised from scheduled and realised public transport observations. Schedule adherence is represented by on-time performance (OTP), which is derived by comparing realised and scheduled arrival times. Travel time variability is represented by the coefficient of variation of travel time (CV_{TT}), calculated from realised stop-to-stop travel times. Headway regularity is represented by the coefficient of variation of headway (CV_h), based on the spacing between consecutive vehicle observations. The exact definitions and calculation of these indicators are provided in Section 3.3.

The framework also distinguishes between the reliability outcomes and the characteristics of the operating environment that are used to explain variation in these outcomes. These explanatory variables include local characteristics, temporal characteristics, mode, and local traffic exposure. Mitigating measures are represented separately by the presence of dedicated public transport infrastructure and traffic signal priority, as these measures are of specific interest in assessing whether a more protected operating environment is associated with differences in reliability.

The framework therefore links the available scheduled and realised operational data to three complementary reliability outcomes and the explanatory factors examined in the empirical analysis. The remainder of this chapter describes how the underlying segment-time dataset is constructed, how the reliability indicators are calculated, and how their associations with traffic conditions and mitigating measures are analysed.

3.2. Construction of the analysis dataset

The empirical analysis is based on three main datasets: GTFS Schedule, GTFS Realtime, and road traffic sensor data. Aggregation is necessary to create a common unit of analysis across these sources. The scheduled and realised public transport data are initially observed at stop level, while the traffic data are observed at a different spatial and temporal resolution. Directly combining these records would therefore mix different observation units. To avoid this, the stop-level public transport records are first converted into directed stop-to-stop movements and then aggregated to segment-time observations.

The window must be long enough to contain multiple realised vehicle movements, so that variability-based reliability indicators can be calculated in a stable way. At the same time, it must remain short enough to preserve relevant temporal variation in traffic exposure and public transport performance.

For each stop-to-stop movement, realised travel time is calculated as the time difference between the realised departure time at the upstream stop and the realised arrival time at the downstream stop. Scheduled travel time is calculated in the same way using the scheduled departure and scheduled arrival times. Delay is calculated at the downstream stop as the difference between the realised and scheduled arrival time. Headways are calculated between consecutive vehicles serving the same line, direction, and segment within the same segment-time observation.

Although the analysis is referred to as segment-time level, observations are separated by service where needed. For segments used by both bus and tram services, bus and tram movements are not combined into a single mixed-mode observation. Instead, reliability indicators are calculated separately for each mode within the same directed segment and time window. The mode variable is therefore assigned at observation level rather than at segment level.

This aggregation structure can be summarised as follows:

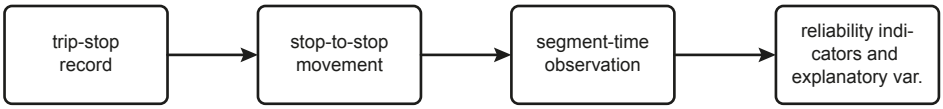


Figure 3.2: Aggregation structure used to construct the segment-time analysis dataset.

After aggregation, the resulting segment-time records are checked for validity and data sufficiency. First, a segment-time observation is retained only if a valid matched traffic observation is available and if the observation falls within the selected period of the day. Traffic observations with a value of zero are treated as invalid measurements, because in this context they are more likely to indicate missing

or faulty traffic counter records than actual zero traffic. Second, the segment-time observation must contain a sufficient number of realised public transport movements to calculate the reliability indicators. Segment-time observations with a realised service frequency below one vehicle every 10 minutes are excluded, because reliability indicators based on too few movements are highly sensitive to individual observations and may not represent recurring segment-time conditions. The 10-minute threshold is used to retain the focus on high-frequency urban public transport services (Harmann et al., 2025). Records that do not meet these criteria are excluded from the analysis dataset. No interpolation is applied, because the analysis aims to relate public transport reliability only to directly observed traffic conditions.

3.3. Reliability indicators

The three reliability dimensions identified in Chapter 2 are operationalised using on-time performance (OTP), the coefficient of variation of travel time (CV_{TT}), and the coefficient of variation of headway (CV_h). These indicators are treated as complementary outcomes, each representing a distinct aspect of operational reliability. None of the three indicators is treated as the primary reliability measure; instead, their combined analysis is used to examine whether traffic exposure and mitigating measures are associated differently with different dimensions of reliability.

On-time performance measures the share of trips that arrive within a predefined punctuality window at the downstream stop. A trip is classified as on time when its realised arrival time falls within the case-specific accepted range relative to the scheduled arrival time. The punctuality threshold applied for this study is specified in Subsection 4.5.4.

$$\mathrm{OTP}_{i,t} = \frac{n_{OT,i,t}}{n_{\mathrm{total},i,t}} \tag{3.1}$$

where $n_{OT,i,t}$ is the number of valid on-time trips for segment i during time t, and $n_{\mathrm{total},i,t}$ is the total number of valid trips for that segment-time. A higher value indicates better schedule adherence.

The coefficient of variation of travel time is used to represent the travel time stability dimension of operational reliability. While OTP captures adherence to the timetable, travel time variability reflects the predictability of vehicle movements between consecutive stops. This indicator is particularly relevant for understanding how local traffic exposure and mitigating measures are associated with the consistency of realised operations, even when timetable adherence remains relatively stable. The coefficient of variation of travel time is defined as:

$$CV_{TT,i,t} = \frac{\sigma_{TT,i,t}}{\mu_{TT,i,t}} \tag{3.2}$$

where $CV_{TT,i,t}$ is the coefficient of variation of travel time for segment i during time t, $\sigma_{TT,i,t}$ is the standard deviation of realised travel times within that segment-time, and $\mu_{TT,i,t}$ is the mean realised travel time within that segment-time. A higher value indicates greater travel time variability and therefore lower reliability.

The coefficient of variation of headway is used to represent the headway regularity dimension of reliability. It measures the variability in realised headways relative to the mean headway. This indicator is particularly relevant for the frequent urban services considered, where irregular spacing between consecutive vehicles can indicate unstable service operation even when schedule adherence remains relatively stable. It is defined as:

$$CV_{h,i,t} = \frac{\sigma_{h,i,t}}{\mu_{h,i,t}} \tag{3.3}$$

where $CV_{h,i,t}$ is the coefficient of variation of headway for segment i during time t, $\sigma_{h,i,t}$ is the standard deviation of realised headways within the segment-time, and $\mu_{h,i,t}$ is the mean realised headway. A higher value indicates more irregular headways and therefore lower reliability.

3.4. Segment selection

The unit of analysis is a directed stop-to-stop segment. This level of analysis allows local operating conditions to be linked more precisely to reliability outcomes than route-level averages. The interpretation of the indicators nevertheless differs across this spatial unit. On-time performance and headway

regularity may partly reflect delay or irregularity accumulated upstream, whereas stop-to-stop travel time is measured directly between the upstream departure and downstream arrival within the selected segment. Travel time variability can therefore be interpreted more directly in relation to local segment conditions.

The segment selection is guided by four considerations. First, segments must be identifiable in both the GTFS Schedule and GTFS Realtime data. Second, segments must be located within the study area and form part of the analysed surface public transport network. Third, segments must be located near a fixed road traffic counter or another suitable source of traffic data, so that local traffic exposure can be linked to public transport operations. Fourth, the selected set should contain variation in mode, traffic exposure, infrastructure characteristics, and the presence of mitigating measures. To ensure variation in transport mode across the sample, the selected segments include bus-only, tram-only, and segments used by both bus and tram services.

The selected segments form a purposive sample rather than a statistically representative sample of the full network. The purpose of the selection is not to estimate network-wide average effects, but to construct a focused analytical sample with meaningful variation in operating conditions. The findings should therefore be interpreted as evidence from the selected urban segments within the study area, while the analytical framework itself is intended to be transferable to other public transport networks.

3.5. Descriptive and correlation analysis

Before estimating the regression models, a descriptive and correlation analysis is conducted. This step is used as a pre-modelling diagnostic to understand the empirical structure of the segment-time dataset. This is important because the explanatory variables vary at different levels and may be strongly related to each other.

3.5.1. Descriptive analysis

The descriptive analysis first examines the distribution of the three reliability indicators across segments and time periods. For each selected segment, the analysis reports the number of valid segment-time observations, the number of underlying public transport movements, the distribution of local traffic exposure, and the distribution of reliability outcomes. This makes it possible to identify whether some segments contribute substantially more observations than others, whether the variability-based indicators are based on sufficient data, and whether the selected sample contains meaningful variation in the outcomes of interest.

The analysis then examines differences between relevant segment and time categories. Reliability outcomes and local traffic exposure are compared across bus and tram observations, peak and off-peak periods, weekdays and weekends, and segments with and without a mitigating-measure package. These comparisons are descriptive and are not interpreted as causal effects. Their purpose is to describe the empirical context in which the regression models are estimated and to identify patterns that may affect the interpretation of the model results.

3.5.2. Within-segment binned associations

In addition to the descriptive comparisons, within-segment binned associations are used to explore how reliability outcomes vary with local traffic exposure within the same segment. This step is included because absolute differences between segments may reflect fixed segment characteristics, such as length, right-of-way conditions, or surrounding urban form, rather than the short-term relationship between traffic exposure and reliability. To focus on within-segment variation, both the traffic exposure variable and the reliability outcomes are expressed as deviations from their segment-specific mean. For each segment-time observation, the deviation is calculated as the observed value minus the average value for the same segment.

The resulting deviations are then grouped into bins according to the deviation in local traffic exposure. For each bin, the mean deviation in the reliability outcome is calculated. This produces a visual diagnostic of whether time periods with relatively higher traffic exposure than usual on the same segment are also associated with better or worse reliability outcomes than usual on that segment. The binned associations are not used as causal evidence and do not replace the regression models. Instead, they

provide an exploratory indication of the direction and shape of the within-segment relationship between local traffic exposure and reliability, while reducing the influence of fixed differences between segments.

3.5.3. Correlation analysis

A correlation analysis is performed for the explanatory variables before model estimation. This is necessary because several variables may capture partly overlapping aspects of the operating environment. For example, traffic exposure may be related to the moment of day, scheduled travel time may be related to segment length, and mitigating measures may be associated with specific infrastructure characteristics. Pairwise correlations are inspected to identify possible dependencies between explanatory variables and to assess whether some variables may lead to unstable coefficient estimates when included in the same model.

The correlation analysis is treated as an exploratory diagnostic rather than as an automatic variable selection procedure. High pairwise correlations do not directly lead to variable removal, because the inclusion of explanatory variables is guided by the theoretical and operational framework of the study. Instead, the correlation results are used to inform the regression strategy and to support the interpretation of the estimated coefficients. Multicollinearity is assessed more formally in the model diagnostics through variance inflation factors, as described in Section 3.7.

The Pearson correlation coefficient is used to quantify the pairwise linear association between explanatory variables. The coefficient between explanatory variables x_j and x_k is calculated as:

$$\rho_{jk} = \frac{\sum_{n=1}^{N}(x_{j,n} - \bar{x}_j)(x_{k,n} - \bar{x}_k)}{\sqrt{\sum_{n=1}^{N}(x_{j,n} - \bar{x}_j)^2} \cdot \sqrt{\sum_{n=1}^{N}(x_{k,n} - \bar{x}_k)^2}} \tag{3.4}$$

where ρ_{jk} is the correlation between variables x_j and x_k, N is the number of observations, and $\bar{x}_j$ and $\bar{x}_k$ are their sample means. Values close to 1 or -1 indicate strong linear relationships, while values close to 0 indicate weak linear relationships (Stewart, 2026).

Several explanatory variables in this study are binary dummy variables, such as the weekday, peak-period, mode, upstream delay, and mitigating-measure indicators. For these variables, the interpretation of the correlation coefficient depends on the measurement level of the paired variables. When one variable is continuous and the other is binary, the coefficient corresponds to a point-biserial correlation. When both variables are binary, the coefficient corresponds to the ϕ-coefficient. Both measures are mathematically equivalent to Pearson's correlation coefficient when binary variables are coded as 0 and 1 (Ekström, 2011; Kornbrot, 2014; Schober & Schwarte, 2018). Therefore, the same pairwise correlation matrix can be used for the explanatory variables, while the interpretation of individual coefficients accounts for whether the variables are continuous or binary.

The resulting correlation matrix is used as an exploratory diagnostic because it shows which specific pairs of variables may be closely related. The correlations will be interpreted as indicators of possible overlap between explanatory variables before the regression models are estimated.

3.6. Hypothesis-guided regression strategy

The regression analysis is designed to assess conditional associations between reliability outcomes and explanatory variables rather than to estimate causal effects. The models are not intended to be purely predictive, but are used to examine whether reliability differs systematically with local traffic exposure, temporal patterns, service characteristics, infrastructure characteristics, and mitigating measures. In particular, mitigating measures are not randomly assigned across the network. Their presence may depend on operational, infrastructural, or traffic conditions that are themselves associated with reliability, creating a potential endogeneity concern. Estimated coefficients for the mitigating measures are therefore interpreted as conditional associations after accounting for the included explanatory variables, rather than as isolated causal effects.

The analysis applies a hierarchical blockwise Ordinary Least Squares (OLS) regression strategy. Hierarchical regression progressively adds theoretically defined groups of explanatory variables, allowing

changes in model fit and estimated associations to be examined as additional characteristics of the operating context are accounted for (Aivaz & Șerbănescu, 2024).

The hypotheses are grounded in two mechanisms identified in previous research. First, surface public transport reliability is influenced by characteristics of the operating environment, including surrounding traffic conditions (Hu & Shalaby, 2017; Van Oort, 2011). Second, public transport design and priority measures are intended to reduce operational disturbances and improve service performance (De Brito-Filho & Oliveira-Neto, 2025; Fadaei & Cats, 2016). Based on these mechanisms, the analysis tests whether higher local traffic exposure is associated with lower reliability, whether mitigating measures are associated with better reliability, and whether such measures reduce the sensitivity of reliability outcomes to traffic exposure.

- **H1**: Higher local traffic exposure is associated with lower reliability. For the coefficient of variation of travel time and the coefficient of variation of headway, this corresponds to a positive association between traffic exposure and variability. For on-time performance, it corresponds to a negative association between traffic exposure and punctuality.
- **H2**: Segments with mitigating measures are associated with better reliability than segments without such measures, after accounting for infrastructure, service, temporal, and traffic characteristics.
- **H3**: Mitigating measures reduce the sensitivity of reliability outcomes to local traffic exposure. The association between increasing traffic exposure and deteriorating reliability is therefore expected to be weaker on protected or prioritised segments. For the coefficient of variation outcomes, this corresponds to a negative interaction coefficient, whereas for on-time performance a positive interaction coefficient is expected.

Mode is included as a contextual service characteristic to account for systematic differences between bus and tram observations and is therefore not treated as a separate hypothesis.

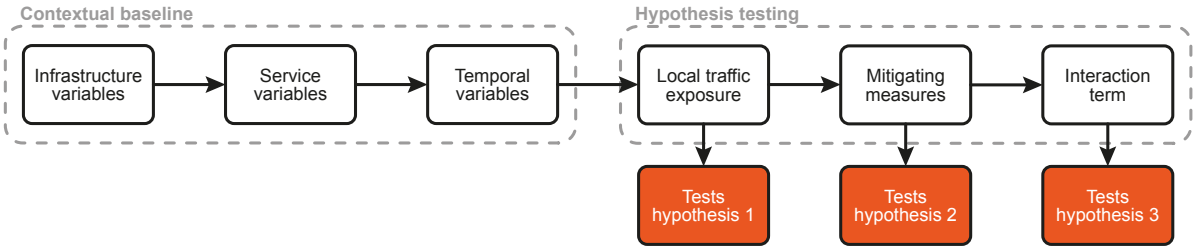


Figure 3.3: Hierarchical blockwise regression sequence and corresponding hypotheses.

In the study, the blockwise structure distinguishes between contextual variables and variables directly related to the hypotheses, as can be seen in Figure 3.3. Infrastructure, service, and temporal characteristics are introduced first to establish a contextual baseline. Local traffic exposure is subsequently added to assess H1, followed by the specification of mitigating measures to assess H2. Finally, the interaction between traffic exposure and mitigating measures is introduced to assess H3. The sequence is predetermined according to the analytical role of the variables rather than their statistical significance and is therefore not used as an automatic variable-selection procedure.

The full hypothesis-testing specification is defined as:

$$RI_{i,t} = \beta_0 + \beta_I I_i + \beta_S \boldsymbol{S}_{i,t} + \beta_T \boldsymbol{T}_t + \beta_F F_{i,t} + \beta_M M_i + \beta_{FM}(F_{i,t} \times M_i) + \varepsilon_{i,t} \tag{3.5}$$

Here, $RI_{i,t}$ represents one of the reliability indicators for segment i during time t. I_i represents infrastructure characteristics, $\boldsymbol{S}_{i,t}$ contains service and timetable characteristics, $\boldsymbol{T}_t$ contains time-related variables, $F_{i,t}$ represents local traffic exposure, and M_i the mitigating measure specification. The coefficients β_F, β_M, and β_{FM} correspond to the associations examined under H1, H2, and H3, respectively. The interaction term allows the traffic association to differ according to the presence of mitigating measures.

The model is estimated separately for on-time performance, the coefficient of variation of travel time, and the coefficient of variation of headway. The same model structure is applied for all three outcomes, but the interpretation of coefficient signs differs across indicators. For $CV_{TT,i,t}$ and $CV_{h,i,t}$, positive coefficients indicate higher variability and therefore worse reliability. For $\mathrm{OTP}_{i,t}$, positive coefficients indicate better schedule adherence and therefore better reliability.

Ordinary Least Squares (OLS) is used as the common baseline specification because it provides directly interpretable associations and allows changes across the blockwise model sequence to be compared transparently. Alternative specifications are subsequently used as robustness checks to address characteristics of the individual dependent variables, including the bounded nature of OTP and the skewed distributions of the coefficient-of-variation outcomes. The OLS estimator minimises the sum of squared differences between observed and predicted values:

$$\hat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} \sum_{i,t} \left(RI_{i,t} - \hat{Y}_{i,t} \right)^2 \tag{3.6}$$

where $RI_{i,t}$ is the observed reliability indicator and $\hat{Y}_{i,t}$ is the value predicted by the regression model.

3.6.1. Blockwise model construction

The model is developed in hierarchical blocks. Each block adds a theoretically meaningful group of variables. The order of the blocks follows the operational logic of the segment-time analysis. Infrastructure characteristics are introduced first because they describe the fixed physical context of each selected segment. Service and timetable characteristics are then added to account for the planned public transport operation within that segment context. Time-related variables are added next to account for recurring temporal patterns in reliability. Local traffic exposure is subsequently introduced to test whether traffic conditions are associated with reliability outcomes. Finally, mitigating measures are added as the main variables of interest, allowing their conditional association with reliability to be assessed after accounting for infrastructure, service, temporal, and traffic characteristics.

This hierarchical structure is used to make the modelling process transparent. It shows whether associations related to local traffic exposure and mitigating measures are already visible in simpler specifications, and whether they remain after additional layers of the operating context are included. The blockwise structure therefore supports the interpretation of the estimated coefficients, rather than serving as an automatic variable selection procedure.

Table 3.1 summarises the dependent variables and explanatory variables used in the regression analysis. The table also indicates the level at which each variable varies, because some variables change between segment-time observations, while others are fixed at segment level. This distinction is important for interpreting the estimated associations.

Table 3.1: Overview of dependent variables, explanatory variables, and level of variation used in the regression analysis.

Variable	Description	Model		
		OTP	CV_{TT}	CV_h
Dependent variables				
On-time performance	Share of trips arriving within the predefined punctuality window around scheduled arrival time at the downstream stop.	✦	✘	✘
Coefficient of variation of travel time	Standard deviation of realised travel time divided by mean realised travel time within a segment-time window.[a]	✘	✦	✘
Coefficient of variation of headway	Standard deviation of realised headway divided by mean realised headway within a segment-time window.[a]	✘	✘	✦
Infrastructure variables				
Segment length	Length of the analysed directed stop-to-stop segment.	✔	✔	✔
Service variables				
Scheduled travel time	Planned travel time between the upstream and downstream stop of the segment.	✔	✘[b]	✔

Table continues on next page.

Table continued from previous page.

Scheduled headway	Planned time interval between consecutive services on the same line, direction, segment.	✔	✔	✘[b]
Mode	Dummy variable = 1 if the observation concerns a bus movement, 0 if it concerns a tram movement.	✔	✔	✔
Departure at upstream stop with delay[c]	Dummy variable = 1 if the vehicle is already delayed at the upstream stop, 0 otherwise.	✔	✔	✔
Arrival at upstream stop with delay[c]	Dummy variable = 1 if the vehicle arrives delayed at the upstream stop, 0 otherwise.	✔	✔	✔
Time variables				
Peak period	Dummy variable = 1 during the morning or afternoon peak period, 0 otherwise.	✔	✔	✔
Weekday	Dummy variable = 1 for Monday to Friday, 0 for Saturday or Sunday.	✔	✔	✔
Traffic variables				
Intensity/Capacity ratio	Ratio between the observed traffic intensity and the estimated road capacity at the matched traffic counter, used as a proxy for local traffic exposure.	✔	✔	✔
Mitigating measure variables				
Dedicated public transport infrastructure[d]	Dummy variable = 1 if dedicated public transport infrastructure is present on the segment, 0 otherwise.	✔	✔	✔
Traffic signal priority[d]	Dummy variable = 1 if traffic signal priority is present on the segment, 0 otherwise.	✔	✔	✔

✦ = dependent variable, ✔ = included in model, ✘ = not included in model.

[a] Each observation represents a segment-time, or a mode-specific segment-time where bus and tram services use the same directed segment.

[b] Scheduled travel time and scheduled headway are excluded from the models in which they would be mechanically related to the dependent variable.

[c] Upstream delay variables indicate whether the vehicle is already delayed at the upstream stop, using the same punctuality threshold as the on-time performance definition.

[d] Dedicated public transport infrastructure and traffic signal priority are conceptually distinct variables, but are represented jointly in the main regression models as a mitigating-measure package because they are not observed as fully independent conditions in the selected sample.

The selection of explanatory variables also accounts for mechanical relationships between dependent and explanatory variables. Variables that are direct components of a dependent variable are not included as explanatory variables in the corresponding model. For example, the mean and standard deviation of realised travel time are not included in the CV_{TT} model, because they directly define the dependent variable. Similarly, the mean and standard deviation of realised headway are not included in the CV_h model. For the OTP model, variables directly derived from the same downstream arrival delay used to classify trips as on time are excluded from the main explanatory specification. This avoids circular interpretation, where a dependent variable would be explained by variables that are mathematically part of its own definition.

Step 1: infrastructure characteristics (part of contextual baseline)
The first model includes only infrastructure characteristics:

$$RI_{i,t} = \beta_0 + \beta_I I_i + \varepsilon_{i,t} \tag{3.7}$$

I_i represents segment length. This model provides the segment-context baseline, because reliability outcomes may differ systematically between segments with different physical and operational characteristics. These variables are introduced first because they are fixed at segment level and describe the environment in which public transport operations take place.

Step 2: service and timetable characteristics (part of contextual baseline)
The second model adds planned service characteristics:

$$RI_{i,t} = \beta_0 + \beta_I I_i + \beta_S \boldsymbol{S}_{i,t} + \varepsilon_{i,t} \tag{3.8}$$

The service vector $\boldsymbol{S}_{i,t}$ includes scheduled travel time, scheduled headway, mode, and upstream delay variables where applicable. These variables account for the planned public transport operation within the physical segment context. Realised reliability cannot be interpreted separately from the planned service pattern, because scheduled travel times and scheduled headways influence how delay, variability, and irregularity are observed. Mode is included because bus and tram observations may differ in their exposure to surrounding traffic and in their degree of operational protection.

Step 3: time characteristics (part of contextual baseline)
The third model adds time-related variables:

$$RI_{i,t} = \beta_0 + \beta_I I_i + \beta_S \boldsymbol{S}_{i,t} + \beta_T \boldsymbol{T}_t + \varepsilon_{i,t} \tag{3.9}$$

The time vector $\boldsymbol{T}_t$ includes peak-period and weekday variables. These variables account for recurring temporal patterns in reliability and help distinguish general time-of-day effects from traffic-exposure effects.

Step 4: local traffic exposure (testing hypothesis 1)
After establishing the contextual baseline, local traffic exposure is introduced as the first hypothesis-related variable. This step directly addresses H1 and tests whether traffic exposure is associated with reliability after accounting for infrastructure, service, and temporal characteristics. Introducing traffic exposure after the contextual baseline allows its association with reliability to be evaluated conditional on relevant characteristics of the operating environment.

The fourth model adds local traffic exposure:

$$RI_{i,t} = \beta_0 + \beta_I I_i + \beta_S \boldsymbol{S}_{i,t} + \beta_T \boldsymbol{T}_t + \beta_F F_{i,t} + \varepsilon_{i,t} \tag{3.10}$$

$F_{i,t}$ represents the local traffic exposure variable. In the case study, this is operationalised as the intensity/capacity ratio at the matched road traffic counter. The intensity/capacity ratio is interpreted as a proxy for local traffic exposure, not as a direct measure of congestion experienced by individual public transport vehicles.

Before inclusion in the regression models, the intensity/capacity ratio is mean-centred. The centred variable is calculated as the observed intensity/capacity ratio minus the sample mean intensity/capacity ratio.

$$IC_{i,t}^{\text{centred}} = IC_{i,t} - \overline{IC}_{i,t} \tag{3.11}$$

This transformation does not change the underlying variation in local traffic exposure but improves the interpretation of models that include interaction terms. In particular, when the interaction between local traffic exposure and the mitigating-measure package is included, the coefficient of the mitigating measure variable can be interpreted at average traffic exposure rather than at an intensity/capacity ratio of zero (Dalal & Zickar, 2012).

Step 5: mitigating-measure package (testing hypothesis 2)
The fifth model adds the mitigating measure specification:

$$RI_{i,t} = \beta_0 + \beta_I I_i + \beta_S \boldsymbol{S}_{i,t} + \beta_T \boldsymbol{T}_t + \beta_F F_{i,t} + \beta_M M_i + \varepsilon_{i,t} \tag{3.12}$$

M_i represents the presence of measures intended to reduce the influence of surrounding traffic on public transport operations. In this study, two measures are considered: dedicated public transport infrastructure and traffic signal priority. Dedicated infrastructure is expected to reduce direct interaction with general traffic, while traffic signal priority is expected to support more stable progression through signalised intersections.

Although these measures are conceptually distinct, they are not observed as fully independent conditions in the selected sample. In the main regression specification, they are therefore represented as a mitigating-measure package:

$$M_i = \begin{cases} 1, & \text{if the mitigating-measure package is present on segment } i \\ 0, & \text{otherwise} \end{cases} \tag{3.13}$$

This specification reflects the empirical structure of the case study and avoids attempting to distinguish between effects that cannot be separately identified in the available data. The coefficient associated with M_i is interpreted as the conditional association between the presence of the mitigating-measure package and the reliability outcome, after controlling for the included service, time, traffic, and infrastructure variables. It is not interpreted as a direct causal effect of either dedicated infrastructure or traffic signal priority separately. This step is directly related to H2.

Step 6: interaction specification (testing hypothesis 3)
The final step introduces an interaction between local traffic exposure and the mitigating-measure package to address H3. This specification tests whether the association between local traffic exposure and reliability differs between segments with and without the mitigating-measure package:

$$RI_{i,t} = \beta_0 + \beta_I I_i + \beta_S \boldsymbol{S}_{i,t} + \beta_T \boldsymbol{T}_t + \beta_F F_{i,t} + \beta_M M_i + \beta_{FM}(F_{i,t} \times M_i) + \varepsilon_{i,t} \tag{3.14}$$

The interaction coefficient β_{FM} represents the difference in the association between local traffic exposure and reliability between segments with and without the mitigating-measure package. For CV_{TT} and CV_h, a negative interaction coefficient is consistent with H3, as it indicates that the increase in variability associated with higher local traffic exposure is weaker when the mitigating-measure package is present. For OTP, a positive interaction coefficient is consistent with H3, as it indicates that the expected decrease in punctuality associated with higher local traffic exposure is weaker when the mitigating-measure package is present.

The interaction specification is interpreted together with the corresponding main effects of local traffic exposure and the mitigating-measure package. Because the measures are not randomly distributed across the selected segments, the interaction coefficient is interpreted as a conditional difference in traffic sensitivity rather than as a causal moderating effect. Evidence for H3 is assessed based on the direction, statistical significance, and consistency of the interaction coefficient across the relevant model specifications.

3.6.2. Manual backward elimination with hierarchy constraints

After the full blockwise model specifications are estimated, a manual backward elimination procedure with hierarchy constraints is applied to obtain more parsimonious final model specifications. The purpose of this step is not to replace the hypothesis-guided blockwise modelling strategy, but to assess which associations remain in a reduced model after variables with limited explanatory contribution are removed.

The backward elimination procedure starts from the full interaction specification developed in the blockwise model sequence. Variables are then removed step by step when they do not contribute meaningfully to the model, based on statistical significance, model fit, and theoretical relevance. A removal threshold of $p > 0.10$ is used as a guiding criterion. This threshold is chosen because the regression analysis is explanatory and exploratory rather than purely confirmatory, and because potentially relevant operational associations may otherwise be removed too aggressively. The threshold is therefore used as a decision aid rather than as a fully automatic rule (Faraway, 2002; Hu & Shalaby, 2017).

The elimination procedure is subject to model hierarchy constraints. When an interaction term is retained, the corresponding lower-order main effects are retained as well, irrespective of their individual statistical significance. This follows the hierarchy principle for models containing interactions: removing a constituent main effect would impose an additional restriction on the model and complicate interpretation of the interaction term (Burnham & Anderson, 2002). Accordingly, if the interaction between local traffic exposure and the mitigating-measure package is retained, both main effects remain in the model.

Retention for reasons of model hierarchy should not be interpreted as evidence in support of a hypothesis. Hypotheses are evaluated based on the sign, magnitude, statistical significance, and consistency of the relevant coefficients across model specifications. A non-significant coefficient is therefore interpreted as insufficient evidence for the hypothesised association in that specification, even if the variable is retained because of theoretical relevance or model hierarchy.

This approach follows the principle of model parsimony, where simpler models are preferred when they provide sufficient explanatory value, while avoiding the removal of terms that are required for the interpretation of hierarchical model structures (Field, 2009; Hair et al., 2010).

The resulting parsimonious models are used to support the interpretation of the main findings and to assess whether the associations observed in the full blockwise models remain visible after reducing the model complexity. They are not interpreted as purely data-driven optimal models. Instead, they provide a complementary specification that balances statistical parsimony with the theoretical structure of the research design.

3.7. Model diagnostics and robustness checks

After estimating the baseline OLS models, several diagnostics and outcome-specific model checks are performed. The OLS specifications provide a common and interpretable baseline across the three reliability indicators, allowing coefficient direction and model development to be compared consistently. However, the statistical properties of the dependent variables differ. The checks in this section are therefore used not only to assess whether the baseline results can be interpreted reliably, but also to evaluate whether the substantive conclusions remain supported under specifications that better reflect the characteristics of each outcome. The diagnostics focus on multicollinearity and residual patterns, while alternative specifications address the skewed nature of the coefficient of variation indicators and the bounded nature of on-time performance.

3.7.1. Multicollinearity

Multicollinearity is assessed using the correlation analysis and variance inflation factors. Multicollinearity occurs when explanatory variables are strongly related to each other. This is relevant because several variables in the analysis may describe overlapping aspects of the operating environment. For example, scheduled travel time may be related to segment length, local traffic exposure may be related to peak-period operation, and mitigating measures may be associated with specific infrastructure characteristics.

Multicollinearity does not necessarily reduce the overall explanatory power of a model, but it can make individual coefficient estimates unstable. This means that the sign, magnitude, or statistical significance of a coefficient may change when additional variables are added to the model (Kim, 2019). Since the interpretation of local traffic exposure and mitigating measures is central to this study, it is important to assess whether these coefficients are affected by strong dependencies between explanatory variables.

The variance inflation factor for explanatory variable x_j is defined as:

$$\mathrm{VIF}_j = \frac{1}{1 - R_j^2} \tag{3.15}$$

where R_j^2 is obtained by regressing x_j on all other explanatory variables in the model. A higher VIF indicates that a larger share of the variation in x_j can be explained by the other regressors. High VIF values do not automatically lead to variable removal, because some correlation between theoretically relevant variables is expected in an urban public transport context. Instead, VIF values are used to interpret coefficient stability and to assess whether a more parsimonious specification is preferable (CFI Team, 2020).

3.7.2. Residual diagnostics

Residual diagnostics are used to assess whether the OLS model specification is broadly appropriate for the observed data. For each observation, the residual is defined as:

$$\hat{\varepsilon}_{i,t} = RI_{i,t} - \hat{Y}_{i,t} \tag{3.16}$$

where $RI_{i,t}$ is the observed reliability outcome for segment i during time t, and $\hat{Y}_{i,t}$ is the fitted value predicted by the model.

Residuals are inspected using residual-versus-fitted plots and Q-Q plots (Chatelain & Ralf, 2014; StatsTest, 2026). Residual-versus-fitted plots are used to assess whether the residuals are randomly distributed around zero. A random spread suggests that the linear specification is broadly appropriate,

while systematic patterns may indicate non-linearity or omitted structure. Funnel-shaped patterns may indicate heteroscedasticity, meaning that the variance of the residuals changes across the range of fitted values. Q-Q plots are used to inspect whether the residuals approximately follow a normal distribution. These diagnostics are not used as strict pass-or-fail tests, but as tools to support the interpretation of the regression results.

3.7.3. Log-transformed variability models

The coefficient of variation of travel time and the coefficient of variation of headway are non-negative indicators. In practice, these variables may be right-skewed, because most segment-time observations may show relatively limited variability, while a smaller number of observations show substantially higher variability (Mullahy & Norton, 2022). This distributional pattern is relevant for OLS regression, because strong skewness can affect residual behaviour and the influence of high-value observations.

To assess whether the main findings are sensitive to this distributional characteristic, log-transformed robustness checks are estimated for the two coefficient of variation outcomes. The transformed dependent variable is defined as:

$$\ln\left(CV_{TT,i,t}\right) = \beta_0 + \beta_I \cdot I_i + \beta_S \cdot \boldsymbol{S}_{i,t} + \beta_T \cdot \boldsymbol{T}_t + \beta_F \cdot \boldsymbol{F}_{i,t} + \beta_M \cdot \boldsymbol{M}_i + \varepsilon_{i,t} \tag{3.17}$$

$$\ln\left(CV_{h,i,t}\right) = \beta_0 + \beta_I \cdot I_i + \beta_S \cdot \boldsymbol{S}_{i,t} + \beta_T \cdot \boldsymbol{T}_t + \beta_F \cdot \boldsymbol{F}_{i,t} + \beta_M \cdot \boldsymbol{M}_i + \varepsilon_{i,t} \tag{3.18}$$

The log-transformed models are estimated as outcome-specific alternatives to the baseline OLS specifications. The original-scale OLS models remain useful because their coefficients are directly interpretable and comparable across reliability outcomes, while the log-transformed models better account for the skewed distribution of the variability indicators. The substantive conclusions for CV_{TT} and CV_h are therefore evaluated based on the consistency of coefficient direction, statistical significance, and interpretation across both specifications (West, 2021).

3.7.4. Bounded nature of on-time performance variable

On-time performance is a proportional outcome. It represents the share of valid trips within a segment-time observation that arrive within the predefined punctuality window. As a result, the indicator is bounded between 0 and 1. This bounded nature is relevant because OLS models do not explicitly restrict fitted values to this interval (Fonsberg, 2026).

The main on-time performance model is estimated using OLS to maintain consistency and interpretability across the three reliability indicators. However, a robustness check is performed to assess whether the bounded nature of the dependent variable affects the conclusions. For this purpose, a fractional logit specification is estimated for the on-time performance outcome. The OLS and fractional logit results are considered jointly when evaluating the robustness of the substantive OTP findings (University of Wisconsin-Madison, 2024).

The fractional logit model can be written as:

$$\mathbb{E}(\mathrm{OTP}_{i,t} \mid \boldsymbol{X}_{i,t}) = \frac{\exp(\boldsymbol{X}_{i,t}\boldsymbol{\beta})}{1 + \exp(\boldsymbol{X}_{i,t}\boldsymbol{\beta})} \tag{3.19}$$

where $\mathrm{OTP}_{i,t}$ is expressed as a proportion, $\boldsymbol{X}_{i,t}$ is the vector of explanatory variables, and $\boldsymbol{\beta}$ is the vector of estimated coefficients. The fractional logit results are compared with the OLS results to assess whether the main conclusions are sensitive to the linear specification. The robustness check focuses on the direction and relative importance of the estimated associations rather than on a direct comparison of coefficient magnitudes, because the coefficients from OLS and fractional logit models are not interpreted on the same scale (Papke & Wooldridge, 1996).

4

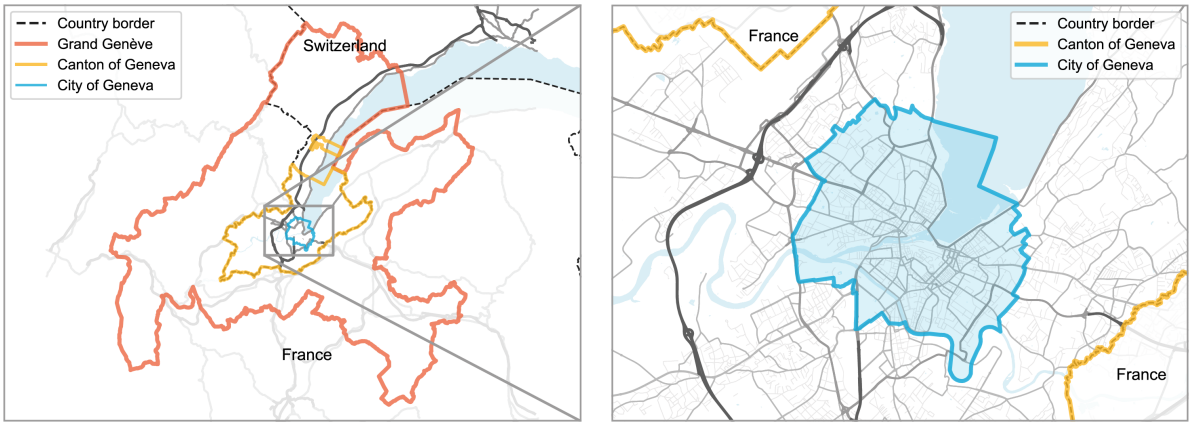
Case Study Application

Chapter 3 defined the general methodological framework for assessing public transport reliability at the stop-to-stop segment level. This chapter translates that framework into the Geneva case study by specifying the study area, the available data sources, the temporal scope of the analysis, the selected public transport segments, and the operationalisation of local traffic exposure and mitigating measures for the empirical analysis.

The chapter is structured as follows. Section 4.1 introduces the study area, followed by the public transport context in Section 4.2. Section 4.3 describes the temporal scope and data sources used in the analysis. Section 4.4 explains the selection of public transport segments and their matching to road traffic counters, while Section 4.5 defines the case-specific variables used in the empirical analysis. Section 4.6 presents the final set of selected segments and their classification, and Section 4.7 describes the construction and coverage of the final analysis dataset.

4.1. Case study area

The case study is situated in the Geneva urban area, located in the westernmost part of Switzerland near the French border. Geneva is the capital of the Republic and Canton of Geneva and forms the urban core of the cross-border metropolitan region of *Grand Genève*, as shown in Figure 4.1. While the City of Geneva forms the central part of the study area, the empirical analysis is not restricted strictly to the municipal boundary. Selected public transport segments may also include adjacent urban municipalities that are functionally connected to the same surface public transport network operated by transports publics genevois (hereafter tpg).



(a) Overview of the metropolitan area, canton, and city boundaries.

(b) Detailed view of the canton and city boundaries.

Figure 4.1: Administrative boundaries of the Geneva metropolitan area, canton, and city.

The Geneva urban area is compact and highly urbanised. The City of Geneva covers approximately 16 km^2 and accommodates a little over 200,000 inhabitants, resulting in a high urban population density

within the Swiss context (Office Cantonal de la Statistique, 2025). This dense urban structure is relevant for the present study because surface public transport services in such environments frequently operate under constrained spatial conditions. Stops, signalised intersections, limited street space, and interactions with surrounding road traffic may all affect the consistency with which buses and trams can operate.

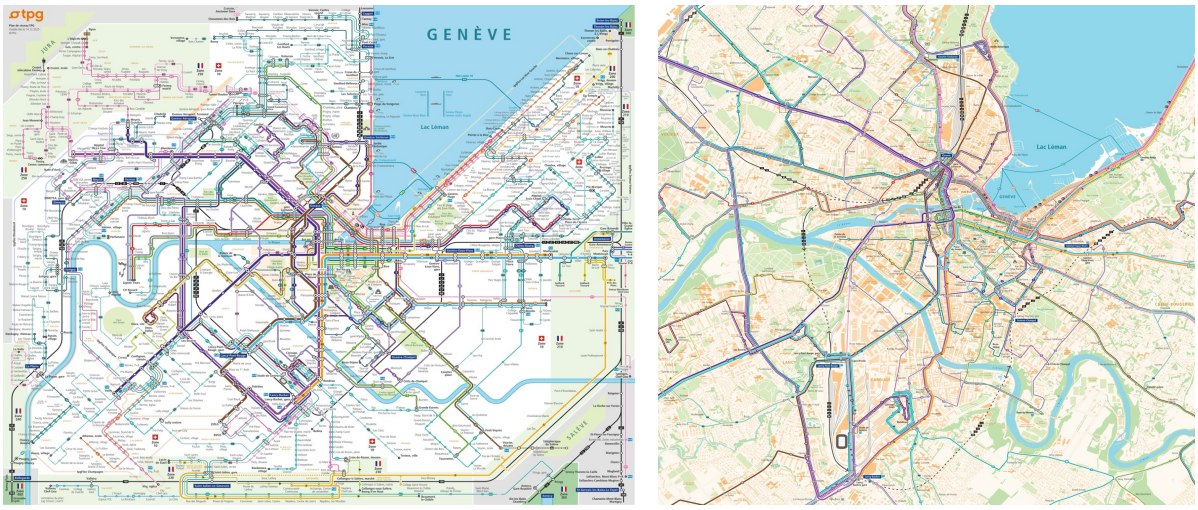
Geneva is used as an information-rich case rather than as a statistically representative case. Its relevance lies in the coexistence of dense surface public transport operations, variation in right-of-way protection, and the availability of scheduled, realised, and road traffic data within one coherent urban network. This makes it possible to compare reliability outcomes across different segment conditions while keeping the broader institutional and network context constant.

4.2. Public transport context

Public transport in Geneva is operated by tpg and consists of tram, trolleybus, and bus services. In recent years, network use has remained high. In March 2026, more than 20 million boardings were recorded on tpg vehicles (tpg, 2026a). This high level of use underlines the importance of public transport in daily mobility in the Geneva urban area and makes service reliability an important operational concern.

The present tpg network should also be understood against the background of Geneva's long history of surface public transport. At the beginning of the 20th century, Geneva already had an extensive tram system, which reached its historical peak in the 1920s before much of the tram network was gradually replaced by bus and trolleybus services. From the late 1980s onwards, public transport again entered a period of expansion, with renewed investment in the network, higher service levels, and later extensions of tram and trolleybus services (tpg, 2008). This historical development helps explain why the current network consists of a mixture of tram, trolleybus, and bus services operating under different infrastructure and traffic conditions.

The broader network context is relevant because Geneva relies strongly on surface public transport. However, the empirical analysis in this study focuses exclusively on conventional bus and tram services. Trolleybus services are therefore excluded from the analysis rather than merged with the bus category. This distinction is made because trolleybuses operate under a partly different technological and operational setting, while the study aims to maintain a clear comparison between conventional bus and tram operations under varying levels of traffic exposure and operational protection. Trolleybuses are nevertheless described here as part of the broader public transport context.



(a) Schematic overview (tpg, 2025). **(b)** Geographic overview (tpg, 2024).

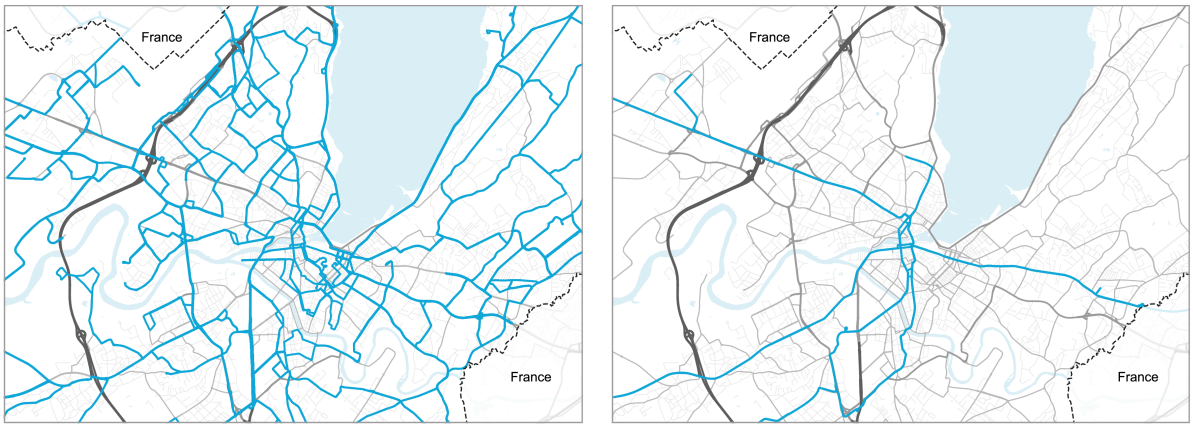
Figure 4.2: Overview of the Geneva public transport network. Click on the figures to enlarge.

The tpg network forms the backbone of surface public transport in the study area. It includes tram lines

with varying degrees of right-of-way protection and bus services that operate partly in mixed traffic. This variation is important for the present research because the reliability of surface public transport may be affected by both the level of exposure to surrounding traffic and the degree to which public transport vehicles are protected from that traffic.

The Geneva network is particularly relevant because it contains a range of operational environments within one coherent public transport system. Services operate in streets with frequent stops, signalised intersections, constrained road space, and different forms of interaction with general traffic (SITG, 2026). This makes it possible to examine segment-level reliability outcomes under different operational and infrastructural conditions while keeping the broader urban and institutional context constant.

The Geneva public transport network includes 59 regular bus lines, 5 tram lines, and 6 trolleybus lines. In addition, the system operates 2 express lines, 1 seasonal line, 8 school routes, and 5 cross-border intercity lines in cooperation with other operators (tpg, 2026b). In March 2026, the network comprised 866 stops. Figure 4.2 shows the schematic and geographic public transport network maps, while Figure 4.3 shows the spatial coverage of the main surface public transport modes in the study area. These figures provide the broader network context for the segment-level analysis conducted in this study.



(a) Spatial coverage of bus services in the study area.

(b) Spatial coverage of tram services in the study area.

Figure 4.3: Spatial coverage of the main bus and tram services in the Geneva urban area.

4.3. Temporal scope and data sources

The empirical analysis covers the period from 2019 to 2024. All available days within this period are considered, including weekdays and weekends. Weekends are not excluded, because the objective is to analyse reliability patterns over the available observation period rather than only under weekday commuter conditions. Weekday and weekend conditions are instead distinguished in the analysis through explanatory variables.

In the Geneva case study, the common segment-time unit introduced in Section 3.2 is operationalised as a segment-hour observation. This choice follows from the temporal resolution of the available road traffic sensor data, which are provided as hourly traffic counts. The one-hour window also provides a practical balance between preserving temporal variation in traffic exposure and public transport performance, while allowing multiple realised public transport movements to be observed within the same analysis window. If traffic data with a finer temporal resolution were available, the same methodological framework could in principle be applied using a shorter segment-time window, provided that each window contains at least six valid public transport movements, consistent with the minimum data sufficiency criterion applied in this study.

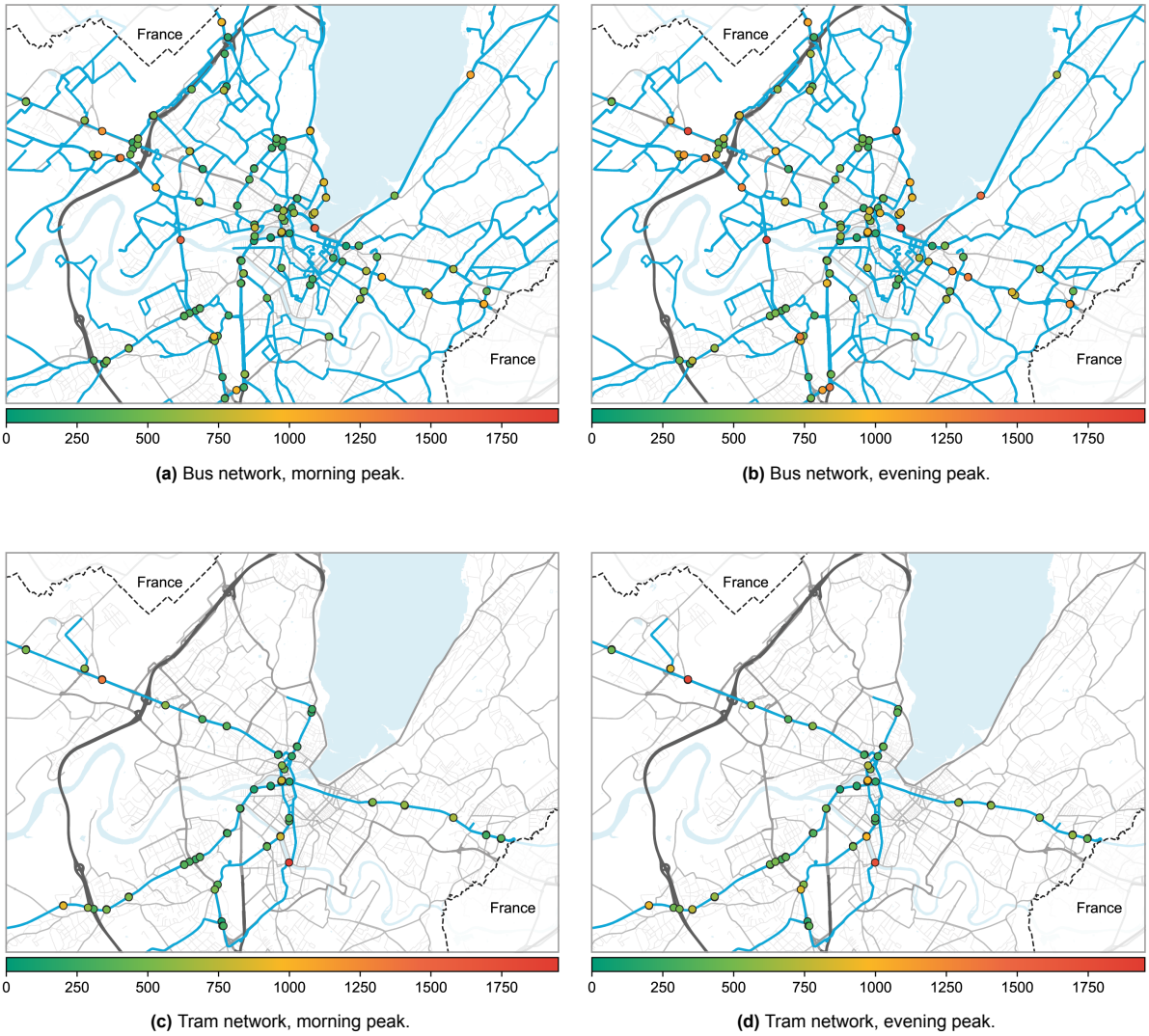
The analysis focuses on observations between 07:00 and 19:00, subject to data availability. This time window is used because it captures a substantial part of daytime public transport operations while excluding night-time periods, during which both public transport frequency and traffic counter observations may be less suitable for the intended analysis. Within this period, a peak-period variable is used

to distinguish busier daytime traffic conditions from other parts of the day. The morning peak period is defined as 07:00-09:00 and the evening peak period as 16:00-18:00. These peak periods are used to characterise traffic intensity patterns in the available traffic data.

The case study combines three main operational data sources: road traffic sensor data, GTFS Schedule data, and GTFS Realtime data. Together, these datasets make it possible to compare planned and realised public transport operations and to relate observed service performance to local road traffic exposure near selected network segments. In addition, geospatial reference data from SITG are used to support the spatial interpretation of the network, traffic counters, and infrastructure characteristics.

4.3.1. Road traffic sensor data

Road traffic sensor data are used to represent local road traffic intensity near the selected public transport segments. These data contain hourly traffic counts from fixed counters distributed across the Geneva urban area. For this case study, traffic counter data are available for the period 2019-2024. Observations between 07:00 and 19:00 are used where available.



(a) Bus network, morning peak.

(b) Bus network, evening peak.

(c) Tram network, morning peak.

(d) Tram network, evening peak.

Figure 4.4: Fixed road traffic counters associated with the bus and tram network during the morning peak (07:00-09:00) and evening peak (16:00-18:00).

The traffic counters provide a local indication of road traffic activity and are used as indicators of traffic exposure in the vicinity of the selected bus and tram segments. Morning and evening peak periods are used to compare traffic intensity across counter locations during busier parts of the day. This

comparison helps identify spatial differences in traffic exposure within the Geneva urban area and supports the selection of segments with varying levels of surrounding traffic intensity. Figure 4.4 shows the fixed road traffic counters associated with the selected bus and tram segments during the morning and evening peak periods. The legend indicates the observed vehicle counts at each traffic counter during the corresponding peak period.

It is important to note that the traffic counters measure the number of vehicles passing a fixed counter location per hour during the 2019-2024 analysis period. They do not capture queue formation, traffic speeds, turning movements, signal timing, or spillback effects. The counter data are therefore used as a consistent proxy for local traffic exposure near the selected segments, rather than as a direct measure of congestion experienced by public transport vehicles. This distinction is important for the interpretation of the results: high traffic volumes may indicate stronger exposure to surrounding traffic, but they do not necessarily imply that a public transport vehicle was delayed by congestion at that exact location or moment.

4.3.2. GTFS Schedule

Planned public transport operations are represented using GTFS Schedule data obtained from the Open Transport Data Switzerland archive (opentransportdata.swiss, 2026a). These data describe the scheduled structure of the public transport system, including routes, trips, stops, stop times, and service calendars. In this study, the GTFS Schedule data provide the planned reference against which realised operations are compared. Scheduled arrival and departure times are used to derive scheduled travel times, scheduled headways, and timetable-based reliability measures.

4.3.3. GTFS Realtime

Realised public transport operations are represented using GTFS Realtime data, also obtained from the Open Transport Data Switzerland archive (opentransportdata.swiss, 2026b). The archived real-time data contain observed stop-level operational information, including actual arrival and departure times. After processing, these data provide the realised operational records required to evaluate service performance at the segment level. The derivation of the reliability indicators from the scheduled and realised stop-level data is described in Chapter 3.

4.3.4. SITG

Geospatial reference data are obtained from the Geneva territorial information system, *SITG* (SITG, 2026). These data are used to support the spatial construction and interpretation of the case study. In particular, *SITG* provides spatial context for the public transport network, road environment, traffic counter locations, and infrastructure characteristics used to classify the selected segments. It is used to verify the spatial relationship between selected stop-to-stop segments and nearby road traffic counters, to interpret the surrounding road environment, and to support the classification of infrastructure characteristics such as road layout, signalised intersections, and the presence of dedicated public transport infrastructure.

4.4. Segment selection and counter matching

The general segment-selection criteria defined in Section 3.4 were applied to the Geneva network to construct the case-study sample. Candidate directed stop-to-stop segments were screened for data availability, spatial relevance, suitability of nearby road traffic measurements, and variation in operating conditions. The application of these criteria to the Geneva network is described below.

Starting from 4219 candidate directed stop-to-stop segments in the Geneva surface public transport network, segments were first screened for availability in both the GTFS Schedule and GTFS Realtime datasets. Segments outside the predefined urban study area were then removed. The remaining candidates were assessed for the availability of a suitable nearby road traffic counter. For each remaining candidate segment, nearby fixed road traffic counters were identified spatially. A counter was considered suitable when it satisfied two conditions: (1) it was located on the same road corridor as the public transport segment, or on an immediately adjacent road section that directly represented the traffic interacting with the segment; and (2) its measured traffic stream corresponded to the direction and road section relevant to the analysed stop-to-stop movement. No fixed distance threshold was applied, be-

cause the nearest counter was not necessarily the most representative of the traffic stream interacting with the public transport segment. The matches were verified using SITG road-network information and street-level inspection of the local road layout.

Finally, the sample was refined to retain sufficient variation in mode, traffic exposure, infrastructure characteristics, and mitigating measures. This procedure resulted in a final sample of 15 directed segments. Figure 4.5 summarises the number of candidate segments retained after each selection step.

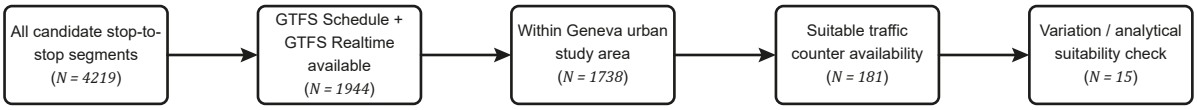


Figure 4.5: Number of candidate segments retained after each selection step.

The final sample therefore represents a purposive set of 15 directed segments for which scheduled and realised public transport data could be consistently linked to a locally representative road traffic counter while retaining variation in mode, infrastructure, traffic exposure, and mitigating measures.

4.5. Operationalisation of case-specific variables

Several explanatory variables used in the empirical analysis are defined for the selected segments and segment-hour observations. These include local traffic exposure, the presence of mitigating measures, segment length, number of lanes, speed limit, number of signalised intersections, and the bus or tram lines serving the segment. Together, these variables describe the local operating environment in which public transport services are observed.

4.5.1. Local traffic exposure

Local traffic exposure is operationalised using the hourly road traffic count measured at the matched road traffic counter. Because road traffic counters differ in their typical volume levels and associated road capacity, the raw hourly count is converted into an intensity/capacity ratio. This ratio expresses the observed hourly traffic intensity relative to the estimated capacity of the counted road section and is used as the main proxy for local traffic exposure in the empirical analysis.

The intensity/capacity ratio is calculated as:

$$IC_{i,t} = \frac{q_{i,t}}{C_i} \tag{4.1}$$

where $IC_{i,t}$ is the intensity/capacity ratio for segment i during hour t, $q_{i,t}$ is the observed hourly traffic count at the matched road traffic counter, and C_i is the estimated hourly capacity of the corresponding counted road section. The capacity value is assigned at counter level and is based on the local road characteristics of the counted section, including the number of relevant lanes and the type of road environment.

The capacity values should be interpreted as assumed effective capacities used for normalising traffic counts across counter locations, rather than as exact theoretical road capacities. This distinction is important because the available traffic counter data do not include signal timing, effective green time, turning movements, queue formation, spillback, or lane-specific traffic composition. In signalised urban environments, the effective capacity of a lane depends strongly on such local conditions. A single universal capacity value per lane would therefore not be appropriate for all selected counter locations.

For this reason, each matched traffic counter was assigned to one of three capacity classes. The classification is based on the local road environment and the expected interaction between general traffic, turning movements, signal control, and public transport priority. Low-capacity road sections represent locations where turning or combined traffic movements at signalised intersections coincide with public transport priority, for which an additional reduction in effective green time is assumed. Medium-capacity road sections represent signalised locations with turning or combined traffic movements, but without

this additional priority-related reduction. High-capacity road sections represent locations dominated by straight-ahead traffic movements, for which a higher effective saturation flow is assumed.

The assumed capacity values are derived from simplified effective-capacity calculations. For straight-ahead traffic, an effective saturation flow of 1700 passenger-car equivalents per hour of green time is assumed, while for combined or turning lanes an effective saturation flow of 1350 passenger-car equivalents per hour of green time is assumed. The effective lane capacity is then calculated by multiplying the saturation flow by an assumed green-time ratio. For high-capacity road sections, this gives approximately $1700 \cdot 0.45 = 765$, rounded to 800 veh/h/lane. For medium-capacity road sections, this gives approximately $1350 \cdot 0.40 = 540$, rounded to 500 veh/h/lane. For low-capacity road sections, an additional reduction factor is applied to account for more restrictive interaction with public transport priority, giving approximately $1350 \cdot 0.40 \cdot (1 - 0.30) = 378$, rounded to 400 veh/h/lane. These assumptions are based on simplified traffic engineering principles and Swiss design and traffic management references (Bundesamt f"ur Strassen ASTRA, 2013, 2018; Stadt Winterthur, 2022).

The resulting capacity classes are therefore defined as 400 veh/h/lane for low-capacity urban road sections, 500 veh/h/lane for medium-capacity urban road sections, and 800 veh/h/lane for high-capacity urban road sections. The total capacity assigned to a counter is calculated by multiplying the assumed capacity per lane by the number of relevant lanes at the counted road section. These classes will be assigned to the segments in Table 4.1.

4.5.2. Mitigating measures

In the Geneva case study, mitigating measures are represented by the presence of dedicated public transport infrastructure and traffic signal priority. These two measures are not observed as fully independent conditions within the selected sample, as described in Section 3.6.

This choice has consequences for interpretation. Associations between the mitigating measure variable and reliability outcomes should be interpreted as associations with the presence of a protected or prioritised operating context. The variable does not allow the separate contribution of dedicated infrastructure and traffic signal priority to be identified.

4.5.3. Infrastructure and service context variables

An infrastructure variable is included to describe differences between the selected segments. The segment length captures the physical distance between the upstream and downstream stop.

Service context is represented by the bus and tram lines serving each selected segment, as well as by scheduled travel time, scheduled headway, and mode in the segment-hour dataset. These variables are important because reliability outcomes cannot be interpreted only in relation to local traffic exposure or mitigating measures. Segments may differ in length, number of intersections, right-of-way conditions, service frequency, scheduled running time, and the mix of services using the segment.

4.5.4. Punctuality definition

For the Geneva case study, a public transport movement is classified as on time when it arrives at the downstream stop between 0 and 3 minutes after its scheduled arrival time. This threshold was adopted following consultation with tpg (D. Reck, personal communication, 2026). Arrivals before the scheduled arrival time and arrivals more than three minutes late are therefore classified as outside the on-time window. The same threshold is used for delay-related variables in the empirical analysis.

4.6. Selected segments and classification

The final selected sample consists of 15 directed stop-to-stop public transport segments. These segments include bus-only, tram-only, and bus-and-tram segments, and they differ in local traffic exposure, infrastructure characteristics, and the presence of mitigating measures. This variation forms the empirical basis for comparing reliability outcomes across segment types within the Geneva case study.

The segment selection focuses on regular urban bus and tram services. Express lines, school routes, seasonal services, and other services that do not form part of the regular analysed bus and tram network are excluded from the empirical analysis. In the selected dataset, this means that the express lines E+,

G+, and 5+, as well as the seasonal line 29, are not included.

Table 4.1: Classification of selected segments.

		DL[a]	TSP	Package	Length [km]	Cap. class[b]	Bus line(s)[c]	Tram line(s)[c]
●	**S1**	✘	✘	✘	1.300	high	22 23	
●	**S2**	✔44.4%	✘	✘	0.450	high	5 8 20 28 60 61 F[d]	
●	**S3**	✔37.1%	✔	✔	0.350	low	5 8 9 20 25	
●	**S4**	✔3.1%	✔	✘	0.650	low	9 33 A	
●	**S5**	✔50.9%	✔	✔	0.550	high	5 8 9 20 25	
●	**S6**	✔	✔	✔	0.500	medium		15
●	**S7**	✔	✔	✔	0.450	medium		12 14 15 17 18
●	**S8**	✔46.2%	✔	✔	0.650	high		12 17
●	**S9**	✔	✘	✘	0.450	medium		15
●	**S10**	✔	✔	✔	0.400	medium		12 14 15 17 18
●	**S11**	✘	✔	✘	0.275	low	80 D[e]	12 14 17 18
●	**S12**	✔	✔	✔	0.450	high	22 23 42	12 15 17 18
●	**S13**	✔71.4%	✔	✔	0.700	high	43	12 15 17
●	**S14**	✔37.5%	✔	✔	0.400	low	1 9	12 14 15 17 18
●	**S15**	✘	✔	✘	0.350	medium	80 D[e]	14

[a] ✔= the segment is completely covered with a dedicated lane, ✔% = the segment is covered $< 20\%$ with a dedicated lane, ✔% = the segment is covered $> 20\%$ with a dedicated lane. Coverage percentage is shown in superscript.
[b] Capacity class used to assign the assumed effective capacity per lane for the calculation of the intensity/capacity ratio: low = 400 veh/h/lane, medium = 500 veh/h/lane, high = 800 veh/h/lane.
[c] Within tpg, several changes were made to the naming of tram and bus lines. As a result, the line names reported here represent the lines that operated on a given segment at some point between 2019 and 2024. However, due to these naming changes, not every line name refers to the same service throughout the entire period.
[d] Bus line F has been renamed to 60.
[e] Bus line D has been renamed to 80.

Table 4.1 classifies the selected segments according to the presence of a mitigating-measure package, segment length, capacity class, and the bus and tram lines operating on the segment. This classification provides the link between the descriptive case study and the explanatory variables used in the regression analysis. It also shows that the selected sample contains variation in both operational and infrastructural conditions.

The spatial distribution of the 15 selected segments and their associated road traffic counters is shown in Figure B.1. The map illustrates how the selected segments are distributed across the Geneva urban area and provides the spatial context for the segment-level analysis. Additional information on the selected segments, including stop identifiers, matched traffic counters, and illustrative street-level views of the segment environments, is also provided in Appendix B.

4.7. Construction and coverage of the final analysis dataset

The methodology described in Chapter 3 was applied to the 15 selected Geneva segments to construct the final segment-hour analysis dataset. Each observation represents one directed stop-to-stop seg-

ment during a one-hour period and combines scheduled and realised public transport information with the local traffic exposure measured at the matched road traffic counter.

Observations were retained only when valid public transport and traffic data were simultaneously available for the corresponding segment-hour. Traffic counter observations identified as invalid were excluded and not interpolated. In addition, each segment-hour was required to contain at least six valid public transport movements to ensure sufficient observations for calculating the reliability indicators.

After these filtering steps, the final analysis dataset contains 228,094 valid segment-hour observations across the 15 selected segments. These observations are based on 2,662,191 valid public transport movements and cover the period from April 17, 2019 to December 31, 2024, subject to data availability. Table 4.2 summarises the observation coverage by segment. The table reports the number of valid segment-hour observations, the number of underlying public transport movements, the number of observed days, the mean local traffic exposure, and the mean values of the main reliability indicators.

Table 4.2: Observation coverage of the final analysis dataset by segment.

	Valid segment-hours	Valid PT movements	Days observed	Mean IC ratio [-]	Mean OTP [-]	Mean CV_{TT} [-]	Mean CV_h [-]
S1	12,686	105,352	1,226	0.706	0.580	0.155	0.232
S2	14,404	168,922	1,479	0.704	0.699	0.269	0.362
S3	17,975	305,294	1,728	0.808	0.608	0.412	0.324
S4	7,465	77,120	820	0.201	0.630	0.244	0.460
S5	10,310	172,018	988	0.697	0.585	0.215	0.340
S6	7,040	71,440	673	0.632	0.729	0.209	0.236
S7	17,774	251,663	1,674	0.318	0.660	0.252	0.468
S8	8,807	132,247	839	0.745	0.767	0.142	0.470
S9	19,495	201,205	1,876	0.657	0.677	0.259	0.268
S10	18,217	278,717	1,773	0.685	0.700	0.253	0.457
S11	21,630	211,576	1,526	0.409	0.697	0.210	0.348
S12	6,284	44,884	485	0.257	0.659	0.200	0.278
S13	14,101	146,453	1,402	0.594	0.593	0.155	0.417
S14	26,168	242,528	1,728	1.170	0.673	0.256	0.320
S15	25,738	252,772	1,782	0.247	0.605	0.205	0.362
Total / overall mean	**228,094**	**2,662,191**	-	**0.614**	**0.665**	**0.237**	**0.357**

The analysis period includes the COVID-19[1] pandemic, during which road traffic volumes differed substantially from typical conditions. These observations were retained because local traffic exposure is explicitly included as an explanatory variable in the analysis. Periods with unusually low traffic volumes therefore provide additional variation in the traffic conditions under which public transport reliability is observed, rather than constituting invalid observations by themselves.

The number of valid observations differs between segments. These differences are mainly caused by variation in service frequency, traffic counter availability, and the number of realised public transport movements available within each segment-hour. Segments served by multiple lines or higher-frequency services generally contribute more observations than segments with fewer services. In addition, some segments have lower observation coverage because the associated traffic counter contains missing or invalid measurements during parts of the analysis period.

The uneven distribution of observations across segments is important for the interpretation of the empirical results. The final dataset is not balanced across segments or years, because the availability

[1] COVID-19 refers to the global pandemic caused by the SARS-CoV-2 virus. From 2020 onwards, measures introduced to limit transmission substantially affected mobility patterns, travel demand, and public transport use in many countries.

of valid public transport records and traffic counter observations differs between locations and time periods. The results should therefore be interpreted as evidence from the selected segment-hour observations for which scheduled operations, realised operations, and local traffic exposure could be combined consistently, rather than as network-wide average reliability outcomes for Geneva.

5 Results

This chapter presents the empirical results of the Geneva case study and examines how public transport reliability in the selected bus and tram segments is associated with local traffic exposure, operating context, and the presence of mitigating measures. The analysis is structured around the three complementary reliability outcomes introduced in Chapter 3: OTP, CV_{TT}, and CV_h.

The chapter proceeds in five steps. Section 5.1 first describes the observed distributions of local traffic exposure and the reliability outcomes. Section 5.2 examines within-segment associations between traffic exposure and reliability and the correlation structure among explanatory variables. Section 5.3 presents the final parsimonious OLS models, which provide a common baseline for comparing associations across the three reliability dimensions. Section 5.4 subsequently evaluates these findings using outcome-specific model specifications, including log-transformed models for CV_{TT} and CV_h and a fractional logit specification for OTP. Finally, Section 5.5 synthesises the empirical findings across the three reliability dimensions. Detailed model-development steps, supplementary diagnostics, and complete regression output are provided in the appendices.

5.1. Traffic exposure and reliability outcomes

Before examining the regression results, this section provides a descriptive overview of local traffic exposure and the three reliability outcomes in the selected segment-hour sample. The purpose is to characterise the observed range and distribution of the variables and to provide context for the subsequent regression analysis.

5.1.1. Traffic exposure

Figure 5.1 shows the distribution of segment-hour intensity/capacity (IC) ratios for each selected segment. The IC ratio is used as the main proxy for local traffic exposure, with higher values indicating that observed hourly traffic volumes are closer to the estimated reference capacity of the corresponding road section. Each observation represents one directed segment during one hour.

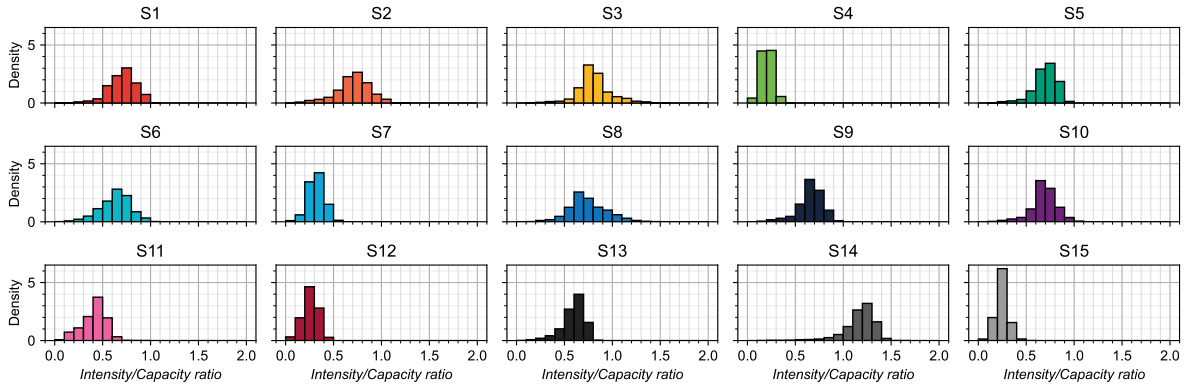


Figure 5.1: Distribution of segment-hour IC ratios by selected segment.

The distributions show substantial differences in local traffic exposure across the selected segments. Segments such as S4, S11, S12, and S15 are characterised by relatively low and narrowly distributed IC ratios, while segments such as S5 and S14 show higher values and broader distributions. The remaining segments generally fall between these extremes, indicating that the selected sample covers a wide range of observed traffic conditions.

IC ratios above 1 can occur because road capacity is represented by an estimated reference value rather than an observed physical maximum. Such values indicate that the recorded hourly traffic volume exceeds the assumed reference capacity for that road section and should therefore be interpreted as periods of high traffic exposure rather than as direct observations of physical capacity being exceeded.

5.1.2. Reliability outcomes

The three reliability outcomes also show clear variation across the selected segment-hour observations. On-time performance is generally concentrated in the middle to upper part of its feasible range, although both the central tendency and dispersion differ between segments. The coefficient of variation of travel time is concentrated at relatively low values but shows a pronounced right tail for several segments, while the coefficient of variation of headway generally displays a wider spread across observations. The full segment-level distributions of OTP, CV_{TT} and CV_h are provided in Appendix D.

These distributional characteristics are relevant for the subsequent modelling. OTP is bounded between 0 and 1, whereas both coefficients of variation are non-negative and show right-skewed distributions. The implications of these properties for model specification and interpretation are examined in the robustness checks presented in Section 5.4.

5.2. Correlation structure and model implications

Before estimating the regression models, two descriptive analyses are used to examine the association structure in the data. First, within-segment associations are examined to assess how deviations in local traffic exposure relate to deviations in reliability within the same segment. Second, correlations between explanatory variables are inspected to identify potential overlap between predictors. Both analyses are descriptive and are used to support the interpretation of the subsequent regression results.

5.2.1. Within-segment associations between traffic exposure and reliability

Figure 5.2 shows the within-segment binned associations between local traffic exposure and the three reliability outcomes. Both traffic exposure and reliability are expressed as deviations from their segment-specific means. The figure therefore captures variation over time within the same segment rather than differences between structurally high- and low-traffic segments.

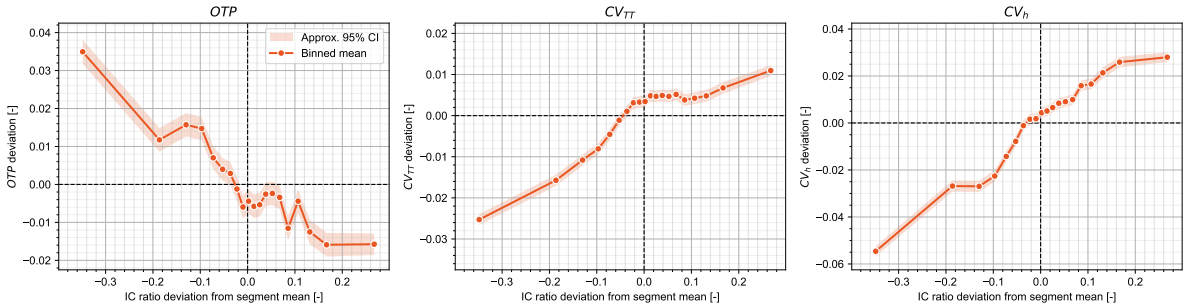


Figure 5.2: Within-segment binned associations between deviations in local traffic exposure and deviations in reliability outcomes.

A consistent descriptive pattern is visible across the three reliability indicators. Higher-than-average local traffic exposure within a segment is associated with lower-than-average OTP and higher-than-average CV_{TT} and CV_h. Conversely, segment-hours with below-average traffic exposure generally show better-than-average reliability. The association appears particularly clear for OTP and CV_{TT}, while the pattern for CV_h is less pronounced. These results are consistent with the expected direction

of H1, but do not account for other time-varying or segment-specific characteristics and should therefore not be interpreted as causal effects.

5.2.2. Correlations between explanatory variables

Figure 5.3 presents the pairwise correlations between the explanatory variables used in the regression analysis. As described in Subsection 3.5.3, the matrix includes correlations between continuous variables, binary dummy variables, and mixed variable pairs. The coefficients are therefore interpreted as Pearson, point-biserial, or ϕ-correlations depending on the measurement level of the paired variables.

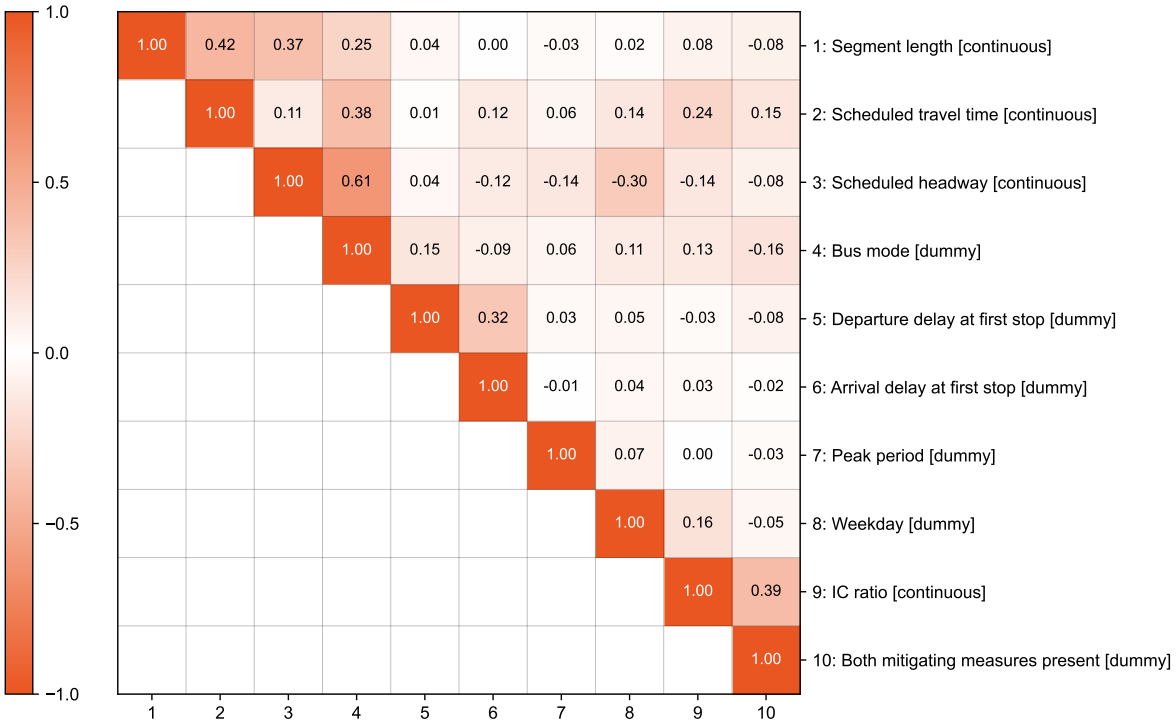


Figure 5.3: Pairwise correlations between explanatory variables used in the regression analysis.

The strongest observed association is between scheduled headway and bus mode ($r = 0.61$), indicating that scheduled service frequency differs systematically between bus and tram observations in the selected sample. Moderate positive correlations are also observed between segment length and scheduled travel time ($\rho = 0.42$), segment length and scheduled headway ($\rho = 0.37$), and between arrival and departure delay at the upstream stop ($\phi = 0.32$). These relationships indicate that several variables capture partly overlapping aspects of the service and operating context.

A further relevant association is observed between local traffic exposure and the mitigating-measure variable ($r = 0.39$). Segments with the mitigating-measure package therefore tend to be associated with higher traffic exposure in the selected sample. This is important for the subsequent regression analysis, because the mitigating measures are not observed independently of the operating context in which they are implemented. Their estimated coefficients should consequently be interpreted as conditional associations rather than as isolated effects of the measures.

Overall, the correlation structure does not preclude the joint inclusion of the explanatory variables, but it indicates that individual coefficients should be interpreted in the context of the other included predictors. Potential multicollinearity is assessed during model development using variance inflation factors; detailed results are reported in Appendix E.4.

5.3. Regression results

This section presents the final parsimonious OLS models for the three reliability outcomes: on-time performance (OTP), the coefficient of variation of travel time (CV_{TT}), and the coefficient of variation of headway (CV_h). These models provide a common baseline specification that allows associations to be compared consistently across OTP, CV_{TT}, and CV_h. The OLS models result from the hierarchical model-building and backward-elimination procedure described in Section 3.6. Full blockwise specifications and intermediate elimination steps are reported in Appendix E. Outcome-specific alternative specifications are subsequently used in Section 5.4 to assess whether the substantive conclusions remain supported when the statistical properties of each dependent variable are taken into account.

The interpretation of coefficient signs differs across the three reliability indicators. For OTP, higher values indicate better schedule adherence and therefore higher reliability. For CV_{TT} and CV_h, higher values indicate greater variability and therefore lower reliability. Table 5.1 presents the three final model specifications side by side to facilitate comparison across reliability dimensions.

Table 5.1: Final parsimonious regression models.

Variable		OTP	CV_{TT}	CV_h
`const`	*Model constant*	0.776 ***	0.304 ***	0.325 ***
Infrastructure variables				
`len`	*Segment length*	-0.085 ***	-0.188 ***	✘
Service variables				
`scheduled_tt_min`	*Scheduled travel time*	✘	✘	-0.026 **
`scheduled_hw_min`	*Scheduled headway*	✘	✘	✘
`mode_bus`	*Mode: bus*	✘	0.066 **	✘
`DT_with_delay`	*Departure delay upstream*	-0.230 ***	✘	0.092 ***
`AT_with_delay`	*Arrival delay upstream*	-0.052 ***	✘	-0.051 ***
Time variables				
`peak_period`	*Peak period*	✘	✘	0.026 ***
`weekday`	*Day: weekday*	-0.039 ***	✘	0.054 ***
Traffic variables				
`IC_ratio`	*IC ratio*	0.078 **	0.121 ***	-0.077 *
Mitigating measure variables				
`measure_package`	*mitigating-measure package*	-0.019	0.012	0.076 **
Interaction terms				
`traffic_x_measure`	`IC_ratio` × `measure_package`	-0.068 *	-0.101 *	✘
Model information				
No. Observations		228,094	228,094	228,094
R-squared		0.217	0.245	0.138
Adj. R-squared		0.217	0.245	0.138
F-statistic		265.200	3.948	34.184
Prob (F-statistic)		0.000	0.019	0.000
AIC		-203,300	-470,421	-270,519

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.3.1. Local traffic exposure

The association between local traffic exposure and reliability differs across the three indicators. For CV_{TT}, the `IC_ratio` coefficient is positive and statistically significant ($\beta = 0.121$, $p < 0.01$). Since higher CV_{TT} values indicate greater travel time variability, this result is consistent with H1 and indicates that higher local traffic exposure is associated with less stable realised stop-to-stop travel times.

For OTP, the `IC_ratio` coefficient is also statistically significant but positive ($\beta = 0.078$, $p < 0.05$). Since higher OTP represents better reliability, this direction is contrary to H1. The coefficient should therefore not be interpreted as evidence that higher traffic exposure improves punctuality. Rather, it represents a conditional association after accounting for upstream delay, weekday conditions, segment

length, the mitigating-measure variable, and the interaction between traffic exposure and mitigating measures.

For CV_h, the `IC_ratio` is weakly significant and negative ($\beta = -0.077$, $p < 0.10$). This direction is again contrary to H1, which would predict higher headway variability under higher traffic exposure. The result therefore provides no robust evidence that local traffic exposure is associated with deteriorating headway regularity in the final specification.

Taken together, H1 is supported for travel time variability but not for on-time performance or headway regularity. This difference suggests that local traffic exposure is most clearly reflected in the stability of realised stop-to-stop travel times, whereas schedule adherence and headway regularity are more strongly shaped by additional operational conditions.

5.3.2. Mitigating measures

The `measure_package` variable does not show a consistent association with improved reliability across the three models. For OTP, the coefficient is negative and statistically insignificant ($\beta = -0.019$, $p > 0.10$), providing no evidence that the presence of the mitigating-measure package is associated with higher punctuality after controlling for the other variables in the final model.

For CV_{TT}, the coefficient is positive but statistically insignificant ($\beta = 0.012$, $p > 0.10$). This likewise provides no evidence that the mitigating-measure package is associated with lower travel time variability in the final specification.

For CV_h, the coefficient is positive and statistically significant ($\beta = 0.076$, $p < 0.05$). Since higher CV_h indicates lower headway regularity, this association is contrary to H2.

These results do not support H2. However, the coefficients should not be interpreted as evidence that mitigating measures worsen reliability. The measures are not randomly distributed across the network and, in the selected sample, are associated with operating contexts characterised by higher local traffic exposure. The estimated coefficients do not reflect isolated causal effects of the measure, but they do reflect conditional differences between segment contexts with and without the mitigating-measure package.

5.3.3. Moderation of traffic sensitivity

The interaction between local traffic exposure and the mitigating-measure package is retained in the final OTP and CV_{TT} models, but not in the final CV_h specification.

For CV_{TT}, the `traffic_x_measure` coefficient is negative and weakly significant ($\beta = -0.101$, $p < 0.10$). Because `IC_ratio` is positive, the negative `traffic_x_measure` indicates that the increase in travel time variability associated with higher traffic exposure is smaller when the mitigating-measure package is present. This direction is consistent with H3 and provides weak evidence that mitigating measures may reduce the sensitivity of travel time variability to local traffic exposure.

For OTP, the `traffic_x_measure` coefficient is also negative and weakly significant ($\beta = -0.068$, $p < 0.10$). However, because higher OTP indicates better reliability, a mitigating effect would imply that the negative relationship between traffic exposure and OTP becomes weaker when the mitigating-measure package is present. This would correspond to a positive `traffic_x_measure` under the hypothesised relationship. The observed negative interaction therefore does not support H3 for OTP. In addition, the positive main traffic coefficient in the final OTP model complicates any interpretation of the interaction as a protective effect.

For CV_h, `traffic_x_measure` was not retained in the final parsimonious model, indicating no evidence of a systematic difference in traffic sensitivity between segments with and without the mitigating-measure package.

Overall, H3 receives only weak and partial support, with the clearest indication observed for travel time variability. The weak significance level and non-random placement of the mitigating measures mean that this interaction should be interpreted as an exploratory association rather than as evidence of a causal moderating effect.

5.3.4. Other contextual associations

Several contextual variables remain associated with reliability in the final models. Upstream operating conditions are particularly important for OTP and CV_h. `DT_with_delay` shows a strong negative association with OTP ($\beta = -0.230$, $p < 0.01$) and a positive association with CV_h ($\beta = 0.092$, $p < 0.01$), indicating that delayed departures are associated with both lower downstream punctuality and more irregular vehicle spacing. `AT_with_delay` is also negatively associated with OTP ($\beta = -0.052$, $p < 0.01$). For CV_h, `AT_with_delay` is negative ($\beta = -0.051$, $p < 0.01$); this conditional association should be interpreted cautiously because it does not imply that arrival delay improves headway regularity.

The `len` coefficient is negatively associated with OTP ($\beta = -0.085$, $p < 0.01$) and CV_{TT} ($\beta = -0.188$, $p < 0.01$). For CV_{TT}, this may partly reflect the relative nature of the coefficient of variation, as similar absolute disturbances produce smaller proportional variation when mean travel times are longer. Bus observations are also associated with higher travel time variability than tram observations in the final CV_{TT} model ($\beta = 0.066$, $p < 0.05$).

Temporal conditions remain particularly relevant for headway regularity. Peak-period observations are associated with higher CV_h ($\beta = 0.026$, $p < 0.01$), as are weekday observations ($\beta = 0.054$, $p < 0.01$). Weekday operation is also negatively associated with OTP ($\beta = -0.039$, $p < 0.01$). These differences indicate that the three reliability indicators respond differently to local and upstream operating conditions and should therefore be interpreted as complementary dimensions of public transport reliability rather than interchangeable measures.

5.4. Outcome-specific model specifications and robustness

The OLS models presented in Section 5.3 provide a common baseline across the three reliability outcomes and allow the estimated associations to be compared within a consistent modelling framework. However, the statistical properties of the dependent variables differ. OTP is bounded between 0 and 1, while CV_{TT} and CV_h are non-negative and show right-skewed distributions.

This section examines whether the substantive findings remain supported under model specifications that better reflect these outcome-specific characteristics. First, residual diagnostics are used to identify limitations of the baseline OLS specifications. Second, log-transformed models are estimated for CV_{TT} and CV_h, followed by a fractional logit specification for OTP. The outcome-specific specifications are considered alongside the OLS baseline when assessing the robustness and substantive interpretation of the main findings.

5.4.1. Residual diagnostics of the OLS baseline models

Residual diagnostics were inspected for the final parsimonious OLS models to assess where the common baseline specification does not fully reflect the statistical properties of the individual reliability outcomes. For each model, residuals were plotted against fitted values and a Q-Q plot was used to assess departures from normality. Figure 5.4, 5.5, and 5.6 present these diagnostics for OTP, CV_{TT}, and CV_h, respectively.

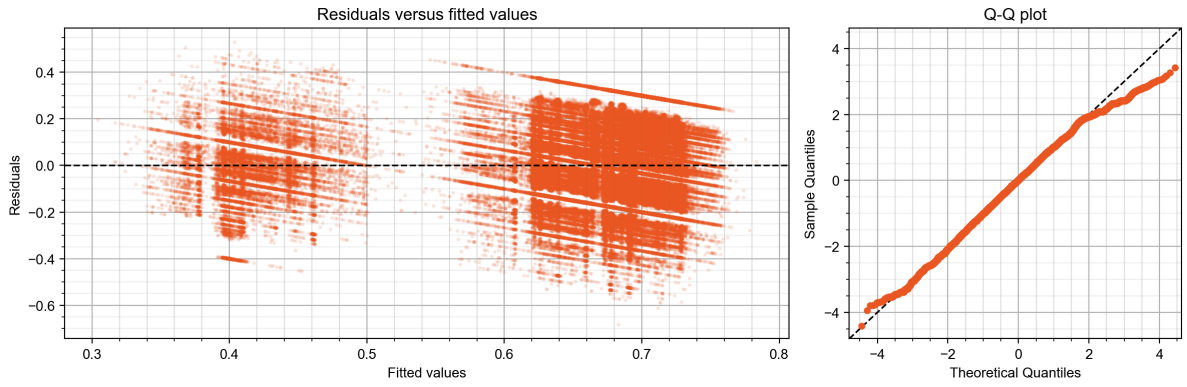


Figure 5.4: Residuals versus fitted values and Q-Q plot for the final parsimonious OTP model.

Figure 5.4 shows the residual diagnostics for the final parsimonious OTP model. The residual-versus-fitted plot shows a structured pattern in which the spread of the residuals varies across fitted values. This is consistent with the bounded nature of OTP: as fitted values approach the upper part of the $[0, 1]$ interval, positive residuals become increasingly constrained. The Q-Q plot additionally shows deviations from normality, particularly in the tails. Although the OLS fitted values remain within the feasible range, the residual structure indicates that a model explicitly accounting for the bounded response may provide a more appropriate representation of the outcome. This motivates the fractional logit specification examined in Subsection 5.4.3.

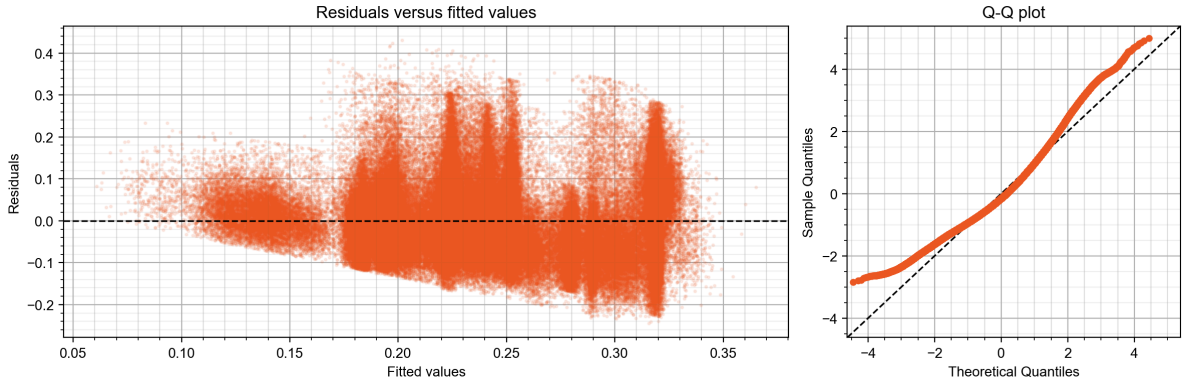


Figure 5.5: Residuals versus fitted values and Q-Q plot for the final parsimonious CV_{TT} model.

For CV_{TT}, Figure 5.5 shows non-constant residual variance and a pronounced departure from the Q-Q reference line, particularly in the upper tail. These patterns are consistent with the non-negative and right-skewed distribution of travel time variability observed in Subsection 5.1.2. The diagnostics therefore indicate that the original-scale OLS specification does not fully capture the distributional characteristics of CV_{TT}, motivating the log-transformed specification examined in Subsection 5.4.2.

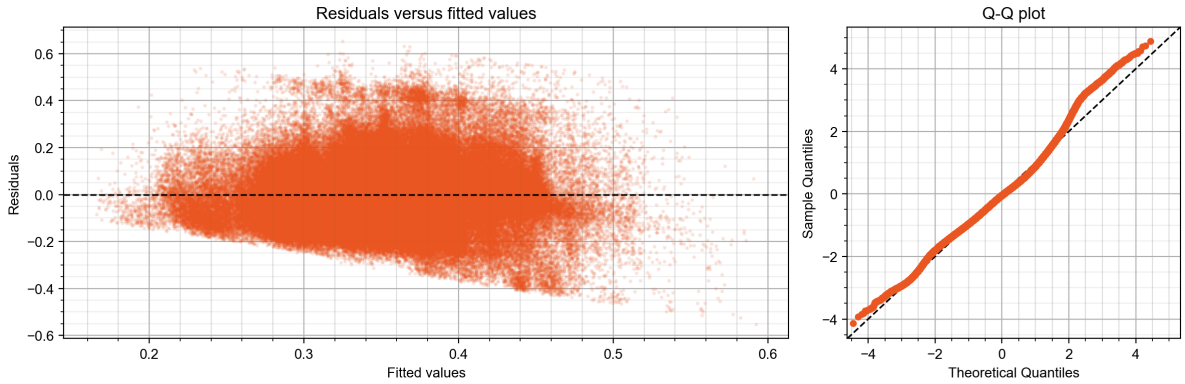


Figure 5.6: Residuals versus fitted values and Q-Q plot for the final parsimonious CV_h model.

The CV_h diagnostics in Figure 5.6 show a similar, although less pronounced, pattern. The residual variance changes across fitted values and the Q-Q plot deviates from normality, indicating that the original-scale OLS model does not fully capture the distributional characteristics of headway variability. This likewise motivates comparison with a log-transformed specification.

The diagnostics indicate that outcome-specific specifications are needed when assessing the strength of the substantive findings. The OLS models are retained as a common baseline for cross-indicator comparison. The following sections therefore compare the OLS results with log-transformed models for CV_{TT} and CV_h and a fractional logit model for OTP.

5.4.2. Log-transformed variability models

Because both CV_{TT} and CV_h are non-negative and right-skewed, log-transformed specifications were estimated to assess whether the substantive findings from the OLS baseline remain supported under a model form that better reflects the distributional characteristics of the outcomes. The comparison focuses on coefficient direction and statistical significance rather than on direct comparison of coefficient magnitudes, since the dependent-variable scale differs between the original and log-transformed models.

Table 5.2: Comparison of coefficient estimates between the original and log-transformed final CV_{TT} model.

Variable	Original coef.	Original p-value	Log coef.	Log p-value	Same sign?
`len`	-0.188	0.000	-0.822	0.000	✔
`mode_bus`	0.066	0.032	0.259	0.014	✔
`IC_ratio`	0.121	0.007	0.517	0.010	✔
`measure_package`	0.012	0.653	0.028	0.780	✔
`traffic_x_measure`	-0.101	0.053	-0.434	0.058	✔

Table 5.2 compares the final OLS and log-transformed specifications for CV_{TT}. The coefficient signs remain unchanged across the two models. `len` remains negatively associated with travel time variability, while `mode_bus` and `IC_ratio` remain positively associated with the outcome. In particular, `IC_ratio` remains positive and statistically significant in the log-transformed specification ($\beta = 0.517$, $p < 0.05$), confirming that the association between higher local traffic exposure and greater travel time variability is not dependent on the original-scale OLS specification.

The traffic-measure interaction (`traffic_x_measure`) also remains negative and weakly significant after log transformation ($\beta = -0.434$, $p < 0.10$). This is consistent with the pattern observed in the OLS baseline: the positive association between local traffic exposure and travel time variability is weaker when the mitigating-measure package is present. The mitigating-measure main effect remains statistically insignificant in both specifications. Overall, the main traffic-related findings for CV_{TT} are therefore robust to the log transformation.

Table 5.3: Comparison of coefficient estimates between the original and log-transformed final CV_h model.

Variable	Original coef.	Original p-value	Log coef.	Log p-value	Same sign?
`scheduled_tt_min`	-0.026	0.019	-0.094	0.040	✔
`DT_with_delay`	0.092	0.000	0.249	0.000	✔
`AT_with_delay`	-0.051	0.000	-0.167	0.000	✔
`peak_period`	0.026	0.002	0.083	0.004	✔
`weekday`	0.054	0.001	0.203	0.000	✔
`IC_ratio`	-0.077	0.077	-0.174	0.263	✔
`measure_package`	0.076	0.014	0.226	0.041	✔

For CV_h, the coefficient directions likewise remain unchanged across the OLS and log-transformed specifications. `scheduled_tt_min` and `AT_with_delay` remain negatively associated with headway variability, while `DT_with_delay`, `peak_period`, `weekday`, and `measure_package` remain positively associated with the outcome.

The main difference concerns `IC_ratio`. In the original OLS specification, `IC_ratio` is negative and weakly significant ($\beta = -0.077$, $p < 0.10$), whereas in the log-transformed model the coefficient remains negative but is no longer statistically significant ($\beta = -0.174$, $p > 0.10$). The traffic-related association observed in the OLS model should therefore not be considered robust for CV_h. This reinforces the interpretation that local traffic exposure is not consistently associated with deteriorating headway regularity in the selected sample.

5.4.3. Fractional logit specification for on-time performance

Because OTP is expressed as a proportion bounded between 0 and 1, a fractional logit specification was estimated as an outcome-specific alternative to the OLS baseline. Unlike OLS, fractional logit

explicitly accounts for the bounded response structure. The comparison is therefore used to assess whether the substantive OTP findings remain supported under a model form that is better suited to the distribution of this dependent variable. As the coefficients of the two models are expressed on different scales, the comparison focuses on coefficient direction and statistical significance rather than coefficient magnitude.

Table 5.4: Comparison of coefficient estimates between the final parsimonious OLS and fractional logit OTP models.

Variable	OLS coef.	OLS p-value	Frac. logit coef.	Frac. logit p-value	Same sign?
`len`	-0.085	0.003	-0.384	0.003	✔
`DT_with_delay`	-0.230	0.000	-0.947	0.000	✔
`AT_with_delay`	-0.052	0.000	-0.231	0.000	✔
`weekday`	-0.039	0.000	-0.179	0.000	✔
`IC_ratio`	0.078	0.023	0.356	0.023	✔
`measure_package`	-0.019	0.258	-0.088	0.246	✔
`traffic_x_measure`	-0.068	0.068	-0.313	0.067	✔

Table 5.4 shows a high degree of consistency between the final OLS and fractional logit specifications. All coefficient signs remain unchanged. `len`, `DT_with_delay`, `AT_with_delay`, and `weekday` remain negatively associated with OTP, while `IC_ratio` remains positive. The `measure_package` coefficient and the `traffic_x_measure` interaction also retain their negative signs.

The statistical significance patterns are similarly stable across the two specifications. The variables `len`, `DT_with_delay`, `AT_with_delay`, `weekday`, and `IC_ratio` remain statistically significant. The `measure_package` main effect remains statistically insignificant, while `traffic_x_measure` remains weakly significant in both models. The positive association between `IC_ratio` and OTP therefore persists under the fractional logit specification and cannot be attributed solely to the use of a linear model for a bounded outcome.

5.5. Summary of empirical findings

The baseline OLS models and outcome-specific specifications together show that the relationship between local traffic exposure and public transport reliability differs substantially across the three reliability dimensions. The alternative specifications refine but do not fundamentally overturn the conclusions from the OLS baseline models.

The most consistent traffic-related finding is observed for CV_{TT}. Higher local traffic exposure is associated with greater travel time variability in both the OLS baseline and the log-transformed specification. In addition, the negative traffic-measure interaction remains weakly significant across both specifications. This provides the strongest support for H1 and limited evidence in support of H3, suggesting that the mitigating-measure package may reduce the sensitivity of travel time variability to increasing traffic exposure.

The findings for OTP and CV_h are less consistent with the expected traffic-related relationships. For OTP, local traffic exposure is positively associated with the outcome in both the OLS and fractional logit specifications, contrary to H1. Schedule adherence is instead strongly associated with the operational state of vehicles before entering the analysed segment, particularly upstream departure delay. For CV_h, the traffic coefficient is weak and contrary to the expected direction in the OLS model and loses statistical significance after log transformation. H1 is therefore not supported for either OTP or CV_h.

The mitigating-measure variable does not show a consistent direct association with improved reliability. Its main effect is statistically insignificant for OTP and CV_{TT} and is associated with higher headway variability for CV_h, providing no support for H2. For H3, the only indication of the hypothesised protective effect is found for CV_{TT}; the interaction for OTP is in the opposite direction, while no interaction is retained for CV_h. These findings should not be interpreted as causal effects of the measures, because mitigating measures are not randomly distributed across the selected segments and are associated with the operating context in which they are implemented.

Overall, the results provide the strongest evidence for a relationship between local traffic exposure

and the stability of realised stop-to-stop travel times. Comparable evidence is weaker for schedule adherence and headway regularity, which appear to reflect a broader set of upstream, temporal, and operational conditions. The findings therefore reinforce the use of OTP, CV_{TT}, and CV_h as complementary rather than interchangeable dimensions of public transport reliability.

6 Discussion

This chapter discusses the empirical findings presented in Chapter 5 and relates them to the research questions and hypotheses of this study. The discussion focuses on the interpretation of the three reliability dimensions, the role of local traffic exposure and mitigating measures, and the methodological considerations and limitations that affect these relationships.

6.1. Discussion of the research questions

The empirical findings show that the relationship between local traffic exposure, mitigating measures, and public transport reliability depends strongly on how reliability is defined. This section discusses the main findings in relation to the research questions and the hypotheses formulated in respectively Chapter 1 and 3. The discussion focuses on the different operational mechanisms captured by each outcome and on what these differences imply for the interpretation of traffic exposure and mitigating measures.

6.1.1. Reliability dimensions and indicator choice

The first research question concerned the development of a suitable framework for evaluating reliability in dense urban bus and tram segments. The empirical results reinforce the decision to treat reliability as a multidimensional operational concept. Although OTP, CV_{TT}, and CV_h all describe aspects of a reliable service, they respond to different parts of the operating process and provide different information about the same segment-hour observations.

OTP primarily reflects schedule adherence at the downstream stop. Its strong association with the delay state at the upstream stop illustrates an important characteristic of this indicator at segment level: downstream punctuality depends not only on what happens within the analysed segment, but also on the operational state with which a vehicle enters it. A vehicle that is already delayed can remain outside the punctuality window even if its movement through the selected segment is relatively stable. Conversely, a vehicle entering the segment on time may tolerate some local disturbance without immediately being classified as late. The strong role of upstream delay should therefore not be interpreted as a contradiction of the segment-level approach, but as evidence that punctuality inherently carries information from earlier parts of the route.

CV_{TT} captures a more locally defined dimension of reliability because it is calculated directly from realised stop-to-stop travel times within the analysed segment-hour. This makes it less dependent on the accumulated delay state of vehicles before entering the segment. The stronger relationship between CV_{TT} and local traffic exposure is therefore consistent with the spatial scale of the indicator: variation in the conditions experienced while travelling through the selected road section is reflected directly in variation in realised travel time.

CV_h, in contrast, reflects the spacing between consecutive vehicles and therefore depends on a broader operational process. Even if travel times within a single segment are relatively stable, vehicles may already enter that segment with irregular spacing. Such irregularity may have developed through upstream delay propagation, uneven departures, bunching, interactions between services, or operational control elsewhere along the route. A local stop-to-stop segment can consequently contribute to headway irregularity without necessarily being the location where that irregularity originated.

The framework is particularly useful because it distinguishes reliability problems that are closely related to local vehicle movement from those that reflect broader route-level processes. At the same time, the results show that the spatial scale of an indicator should correspond to the operational mechanism being investigated. Segment-level analysis is especially informative for realised travel time stability, while schedule adherence and headway regularity require greater consideration of upstream conditions. This also implies that indicator suitability is context-dependent: an indicator that is highly informative for one analytical scale may be less suitable for attributing reliability changes to another. In the present study, OTP and CV_h remain valuable complementary outcomes, but their interpretation at stop-to-stop segment level requires more caution than that of CV_{TT}.

6.1.2. Local traffic exposure and reliability

The second research question examined how local traffic exposure is associated with reliability across the selected segments. The findings provide only partial support for H1, which expected higher traffic exposure to be associated with deteriorating reliability. The strongest and most consistent evidence is found for CV_{TT}, whereas the relationships with OTP and CV_h are weaker or contrary to the hypothesised direction.

The CV_{TT} results suggest that increasing traffic exposure is primarily reflected in the stability of vehicle movement through the analysed road section. Higher traffic volumes relative to the estimated road capacity may increase the variability of interactions with surrounding traffic, intersections, and queues. Importantly, this does not require every public transport movement to become systematically slower. Reliability can deteriorate because the realised travel time becomes less predictable, even when the average travel time changes only modestly. The persistence of this association in the log-transformed specification strengthens the interpretation that the observed relationship is not driven solely by the skewed distribution of CV_{TT}.

The results for OTP highlight a different mechanism. The descriptive within-segment analysis indicates that segment-hours with higher-than-usual traffic exposure tend to coincide with lower-than-usual punctuality. After controlling for the broader operating context, however, the regression coefficient for local traffic exposure becomes positive. This difference suggests that the descriptive relationship between traffic exposure and punctuality is intertwined with other operational characteristics. In particular, the strong role of upstream delay means that downstream punctuality cannot be attributed solely to local conditions within the selected segment. The consistency of the positive traffic coefficient across the OLS and fractional logit specifications indicates that this unexpected direction is not simply a consequence of modelling a bounded outcome with OLS.

For CV_h, the absence of a robust traffic relationship further illustrates the limits of interpreting headway regularity as a purely local outcome. The weak negative association in the OLS model disappears after log transformation and is contrary to H1. This does not demonstrate that road traffic has no influence on vehicle spacing. Rather, it indicates that traffic exposure measured at one local road section does not consistently explain the realised headway structure observed there once broader operational conditions are considered.

H1 should therefore be interpreted as outcome-dependent rather than universally supported or rejected. Local traffic exposure is clearly relevant to the stability of realised stop-to-stop travel times, but its relationship with schedule adherence and headway regularity is less direct. This distinction is important for both analysis and practice: the effect of traffic conditions may be underestimated or mischaracterised when reliability is evaluated using only a timetable-based or headway-based indicator.

6.1.3. Mitigating measures and traffic sensitivity

The third and fourth research questions concern the association between mitigating measures and reliability and the conditions under which these measures may be most effective. The results do not support H2, which expected the presence of the mitigating-measure package to be directly associated with better reliability. At the same time, the interaction results provide limited evidence for H3 in the CV_{TT} model.

The absence of a consistently beneficial main effect should be interpreted in light of how the measures are distributed across the selected segments. Dedicated public transport infrastructure and traffic signal

priority are not randomly assigned. In the selected sample, the mitigating-measure package is associated with operating contexts characterised by higher local traffic exposure. As a result, differences between segments with and without the package combine the potential contribution of the measures with pre-existing differences in the locations where those measures are implemented. The main measure coefficient therefore cannot be interpreted as the isolated effect of dedicated infrastructure or signal priority.

This selection problem makes the interaction with traffic exposure a useful perspective, although it does not resolve the underlying non-random placement of the measures. For CV_{TT}, the negative interaction indicates that the positive association between traffic exposure and travel time variability is weaker when the mitigating-measure package is present. The result is only weakly significant and remains observational, but its direction is consistent across the OLS and log-transformed specifications. This suggests that protected or prioritised segments may be less sensitive to additional traffic pressure than otherwise comparable operating contexts.

This distinction between an average difference and a difference in sensitivity is important when evaluating infrastructure measures. A segment with dedicated infrastructure or signal priority may still display relatively poor reliability because it operates in a more demanding environment. The relevant question is therefore not only whether protected segments perform better on average, but also whether their reliability deteriorates less as traffic exposure increases.

The evidence for such moderation is not observed consistently for OTP or CV_h. For OTP, the interaction is in the opposite direction to the hypothesised protective relationship, while no interaction is retained for CV_h. These results do not imply that signal priority or dedicated infrastructure cannot improve punctuality or headway regularity. Their effect depends partly on how such measures are designed and operated. For example, signal-control strategies can potentially respond to schedule deviation or vehicle spacing rather than simply minimising intersection travel time. The present analysis does not observe the specific control logic or activation of individual priority interventions and therefore cannot distinguish between these mechanisms.

6.2. Methodological reflection

The methodological choices made in this study shaped both the strengths and the limits of the empirical findings. The segment-hour framework enabled scheduled operations, realised public transport movements, local traffic exposure, and mitigating measures to be analysed on a common spatial and temporal scale. At the same time, the resulting associations depend on how the spatial unit, temporal aggregation, explanatory variables, and reliability indicators were operationalised. This section reflects on the implications of these choices for the interpretation of the results.

6.2.1. Segment-hour unit of analysis

The use of directed stop-to-stop segment-hours created a detailed operational unit that is well suited to studying local variation in realised travel times. This is especially relevant for CV_{TT}, which is calculated directly from movements through the selected segment and therefore corresponds closely to the spatial scale of the explanatory traffic variable. The relatively clear traffic-related findings for CV_{TT} are consistent with this alignment between outcome and exposure.

The same spatial unit is less complete for OTP and CV_h. Both indicators may reflect processes that developed before vehicles entered the analysed segment. Downstream punctuality inherits part of the upstream delay state, while headway irregularity can propagate along a route and may already be present at the upstream stop. The segment-hour framework can therefore observe these outcomes locally without fully identifying where the underlying reliability problem originated. This does not make the indicators unsuitable, but it limits how directly their variation can be attributed to conditions within the selected segment.

The results consequently suggest that the appropriateness of the segment-hour unit depends on the reliability mechanism being studied. It is particularly informative for local travel time stability, whereas analyses of punctuality and headway regularity benefit from additional route-level or upstream information.

6.2.2. Operationalisation of local traffic exposure

Local traffic exposure was represented using the segment-hour IC ratio derived from fixed road traffic counters and estimated reference capacities. This provided a consistent measure of the relative intensity of road traffic across segments and time periods and made it possible to relate changing traffic conditions to public transport reliability at the same temporal resolution.

However, the IC ratio represents exposure to road traffic rather than the congestion directly experienced by an individual public transport vehicle. It does not observe traffic speed, queue length, spillback, turning movements, signal phase, or the exact position of a bus or tram within surrounding traffic. Two segment-hours with similar traffic volumes may therefore generate different operational consequences depending on how traffic is distributed in space and time.

This distinction is important when interpreting the regression coefficients. A significant association with the IC ratio indicates that reliability changes systematically with local traffic exposure, but it cannot identify the specific traffic mechanism responsible for that change. Conversely, the absence of a strong association for OTP or CV_h does not imply that local congestion is irrelevant; it may also reflect the limited ability of hourly traffic volume to represent the operational disturbance experienced by individual vehicles.

6.2.3. Temporal aggregation

The one-hour aggregation interval was primarily determined by the resolution of the available road traffic data and by the need to observe multiple public transport movements within each analysis period. This choice improves comparability across datasets and provides sufficient observations for the calculation of variability-based reliability indicators.

At the same time, hourly aggregation smooths short-lived operational disturbances. Queue formation, signal delay, bunching, incident-related disruption, and temporary priority activation may occur on much shorter time scales. Their effects may therefore be diluted when averaged within a segment-hour. This is particularly relevant for traffic-related mechanisms that are highly dynamic and may partly explain why some expected relationships are weaker for OTP and CV_h than for CV_{TT}.

The results should consequently be interpreted as relationships between reliability and relatively broad local operating conditions within an hour, rather than as estimates of the effect of individual congestion or signal events.

6.2.4. Selection and operationalisation of mitigating measures

The treatment of mitigating measures is another important methodological consideration. Dedicated public transport infrastructure and traffic signal priority are represented as a combined measure package rather than as fully independent interventions. This reflects the structure of the selected sample, but it means that the analysis cannot distinguish the contribution of the individual measures or the way in which they are operated.

Moreover, the presence of the measure package is not randomly assigned across the network. Segments with mitigating measures are embedded in different operating contexts from segments without them, including differences in local traffic exposure. The resulting main-effect coefficient therefore combines potential measure effects with pre-existing differences between segment contexts.

The interaction between traffic exposure and the measure package partly addresses a different question by examining whether the traffic-reliability relationship changes when the package is present. However, even this interaction remains observational and cannot be interpreted as a causal treatment effect. The analysis identifies conditional differences in sensitivity rather than the effect of implementing a measure on an otherwise identical segment.

6.2.5. Model specification and outcome-specific robustness

The use of OLS as a common baseline provided a consistent framework for comparing model development and coefficient direction across OTP, CV_{TT}, and CV_h. This comparability is useful because the three indicators represent different reliability dimensions but are evaluated against the same set of explanatory concepts.

The diagnostic results also showed that a single model form does not fully reflect the statistical properties of all three outcomes. The right-skewed variability indicators and bounded OTP outcome motivated log-transformed and fractional logit specifications. The consistency of the central CV_{TT} and OTP findings across these alternative models strengthens confidence that these patterns are not solely artefacts of the baseline OLS specification. Conversely, the loss of significance of the traffic coefficient in the log-transformed CV_h model illustrates why robustness across model forms is important when evaluating weaker findings.

The alternative specifications therefore do more than confirm model assumptions: they help distinguish findings that are stable across reasonable model choices from those that depend more strongly on a particular specification.

6.3. Implications for reliability evaluations and practice

The findings have implications for how reliability problems and mitigating measures are evaluated in dense urban public transport networks. A central implication is that the interpretation of reliability depends on the operational mechanism that an indicator is expected to capture. The three indicators used in this study provide different perspectives on the same operating environment and therefore lead to different conclusions about the role of local traffic exposure.

For the evaluation of local traffic conditions and infrastructure interventions, CV_{TT} appears particularly informative because its spatial definition corresponds closely to the analysed stop-to-stop movement. Changes in the stability of realised travel times can therefore be related more directly to conditions experienced within the selected segment. This makes travel time variability especially relevant when the objective is to assess whether local road conditions or infrastructure affect the consistency with which public transport vehicles move through a corridor.

OTP provides a different perspective. Its direct relationship with the timetable makes it intuitive and operationally meaningful, but its interpretation at segment level requires consideration of the delay state with which a vehicle enters the segment. Poor downstream punctuality cannot automatically be attributed to the local segment, while acceptable punctuality does not necessarily imply that the segment itself operated without disturbance. For local evaluations, OTP therefore provides useful information about the realised passenger-facing schedule outcome, but it does not by itself identify where the underlying disturbance originated.

A similar distinction applies to CV_h. Irregular vehicle spacing observed within a segment may result from disturbances that occurred locally, but can also represent irregularity propagated from upstream parts of the route. The weaker relationship between local traffic exposure and CV_h in this study therefore does not reduce the operational importance of headway regularity. Instead, it indicates that its evaluation requires a broader spatial perspective than the local traffic conditions of a single stop-to-stop segment.

These distinctions are also relevant for the evaluation of mitigating measures. Dedicated infrastructure and traffic signal priority are designed to influence vehicle movement through locations where public transport interacts with surrounding traffic and signalised intersections. It is therefore plausible that their most direct effect is reflected in local travel time stability. However, the consequences for punctuality and headway regularity depend on how local improvements interact with the wider service-control process. Signal priority, for example, can in principle be operated not only to reduce intersection delay but also in response to schedule deviation or vehicle spacing. The present analysis does not observe the specific control logic or activation of individual priority interventions and therefore cannot distinguish between these operational strategies.

In the Geneva context, public transport operations and traffic signal control are part of a broader integrated traffic-management environment. Geneva's central traffic control system brings together several actors, including the cantonal transport authorities, the road police, and tpg, and enables real-time adaptation of traffic flows and signal timings (République et Canton de Genève, 2017). The empirical analysis isolates only a subset of these mechanisms through the observed infrastructure and traffic variables. The absence of a clear relationship between a local measure package and every reliability indicator should therefore not be interpreted as evidence that broader service-regulation practices are ineffective. Rather, the findings illustrate that local infrastructure measures and network-level opera-

tional control act at different spatial and operational scales.

The results suggest that the evaluation of public transport reliability should begin with the mechanism of interest rather than with a single preferred performance indicator. Local travel time variability is well suited to identifying instability associated with traffic exposure, timetable adherence reflects both local and accumulated delay, and headway regularity captures a broader service-regulation process. Evaluations that combine these perspectives are therefore better able to distinguish between local traffic-related reliability problems and reliability problems that require a wider operational explanation.

6.4. Limitations

Several limitations should be considered when interpreting the findings of this study. These limitations primarily concern the representativeness of the selected segments, the operationalisation of traffic exposure and mitigating measures, and the availability of explanatory information beyond the analysed segment.

First, the selected segments form a purposive sample rather than a statistically representative sample of the Geneva public transport network. The selection was designed to include variation in mode, infrastructure, traffic exposure, and mitigating measures, which supports the comparison of different operating contexts. However, the resulting coefficients should not be interpreted as network-wide average effects. The conclusions apply to the analysed segment contexts and should be generalised to the wider network with caution.

Second, the local traffic exposure variable represents traffic intensity relative to an estimated reference capacity rather than the congestion directly experienced by individual public transport vehicles. The IC ratio does not capture traffic speed, queue formation, turning movements, spillback, signal phase, or priority activation. As a result, the analysis can identify systematic relationships with traffic exposure but cannot determine which specific traffic mechanisms produce the observed reliability changes. This may also attenuate relationships where short-lived congestion events are not well represented by hourly traffic volumes.

Third, the mitigating measures are observed as a combined package and are not randomly distributed across the selected segments. Dedicated infrastructure and traffic signal priority therefore cannot be evaluated as independent interventions, and differences between segments with and without the package also reflect differences in the operating environments in which these measures are implemented. The estimated coefficients and interaction effects consequently represent conditional associations rather than causal effects of implementing a specific measure.

Finally, several operational processes that may influence reliability were outside the scope of the available data. These include passenger demand, dwell times, vehicle occupancy, incidents, weather conditions, detailed signal operations, and upstream service-control decisions. Their omission is particularly relevant for OTP and CV_h, which are influenced by processes that extend beyond the selected stop-to-stop segment. The hourly temporal aggregation further limits the ability to observe short-duration disturbances such as temporary queue formation or bunching events.

These limitations do not prevent the framework from identifying meaningful patterns in the observed data, but they constrain the interpretation of those patterns. The findings are therefore most appropriately understood as evidence of conditional relationships between operating context and reliability rather than as estimates of network-wide or causal intervention effects.

7 Conclusions

This chapter concludes the thesis by answering the research questions based on the empirical findings and their interpretation in Chapter 6. It then summarises the scientific contribution of the study, presents practical recommendations and directions for future research, and closes with the main conclusion.

7.1. Answers to the research questions

This thesis examined how public transport reliability in dense urban bus and tram segments is associated with local traffic exposure and mitigating measures. To answer the main research question, five sub-questions were formulated. The answers to these sub-questions are summarised below.

➜ **SQ1. What conceptualisation and measurement framework is suitable for analysing operational reliability in dense urban bus and tram segments**

A suitable framework treats public transport reliability as a multidimensional operational concept and combines planned operations, realised operations, and local operating conditions at a consistent spatial and temporal scale. In this study, OTP, CV_{TT}, and CV_h provide complementary perspectives on schedule adherence, travel time variability, and headway regularity. The directed segment-hour is particularly suitable for analysing local travel time stability, while punctuality and headway regularity require additional consideration of upstream and route-level conditions.

➜ **SQ2. To what extent is local traffic exposure associated with different dimensions of reliability across the selected segments?**

Local traffic exposure is most consistently associated with travel time variability. Higher traffic exposure is associated with higher CV_{TT} in both the OLS and log-transformed specifications, indicating less stable realised stop-to-stop travel times under more intensive traffic conditions. No comparable robust evidence is found for OTP or CV_h. The relationship between local traffic exposure and reliability therefore depends on the reliability dimension considered.

➜ **SQ3. How are mitigating measures, consisting of dedicated public transport infrastructure and traffic signal priority, associated with the different dimensions of reliability across the selected segments?**

The mitigating measure package is not consistently associated with improved reliability across the selected segments. Its main effect is statistically insignificant for OTP and CV_{TT} and is associated with higher CV_h in the final model. H2 is therefore not supported. Because the measures are not randomly distributed across the selected segments, these associations should be interpreted as conditional differences between operating contexts rather than as causal effects of the measures.

➜ **SQ4. How does the relationship between local traffic exposure and reliability differ between segments with and without the mitigating-measure package?**

The clearest indication of a difference in traffic sensitivity is found for CV_{TT}. The negative interaction between local traffic exposure and the mitigating measure package indicates that the positive association between traffic exposure and travel time variability is weaker when the package is present. This interaction is only weakly significant, but its direction remains consistent across the OLS and log-transformed specifications. No comparable evidence is found for OTP or CV_h.

➜ **SQ5. What operational and infrastructural recommendations follow from the empirical findings for the analysed segments?**

The findings indicate that reliability problems should be diagnosed according to the operational mechanism involved. Local traffic-related instability is most directly reflected in travel time variability, while punctuality should be interpreted together with upstream delay information and headway regularity in relation to broader route-level processes. The evaluation of mitigating measures should account for the traffic and operating context in which they are implemented rather than relying only on average differences between segments with and without the measure package.

7.2. Answer to the main research question

The main research question of this thesis is:

➜ **How is public transport reliability in dense urban bus and tram segments associated with local traffic exposure and mitigating measures?**

The results show that the relationship between local traffic conditions and public transport reliability is strongly dependent on the reliability dimension considered. Local traffic exposure is most clearly associated with the stability of realised stop-to-stop travel times: higher traffic exposure is linked to greater travel time variability. Its relationship with schedule adherence and headway regularity is less direct, as these outcomes are more strongly influenced by upstream delay and broader route-level operational processes.

Mitigating measures do not show a consistent direct association with improved reliability across the selected segments. However, there is limited evidence that the mitigating-measure package reduces the sensitivity of travel time variability to increasing traffic exposure. The role of these measures should therefore be understood in relation to the operating context in which they are implemented rather than through average comparisons between protected and unprotected segments.

Overall, public transport reliability in dense urban bus and tram operations is shaped by an interaction between local traffic conditions, upstream operational states, and infrastructure or control measures. A multidimensional evaluation is therefore required to distinguish locally generated travel time instability from punctuality and headway problems that originate or propagate elsewhere in the network.

7.3. Scientific contribution

The scientific contribution of this thesis lies in the development and application of a segment-hour framework that integrates planned public transport operations, realised operations, and local road traffic exposure. By combining GTFS Schedule, GTFS Realtime, and traffic counter data at a common spatial and temporal scale, the framework enables multiple dimensions of operational reliability to be analysed simultaneously.

The findings also contribute to the understanding of public transport reliability as a multidimensional concept. OTP, CV_{TT}, and CV_h capture different operational mechanisms and respond differently to local traffic exposure and mitigating measures. In particular, the results show that local traffic exposure is most clearly reflected in travel time variability, while schedule adherence and headway regularity are more strongly influenced by upstream and route-level operational conditions.

The study therefore demonstrates that the interpretation of reliability depends not only on the explanatory variables considered, but also on the spatial scale and operational meaning of the reliability indicator itself. This provides a methodological basis for more targeted analyses of reliability in dense urban surface public transport networks.

7.4. Practical recommendations and future research

The findings lead to several practical recommendations for public transport operators and planners. Reliability monitoring should combine multiple indicators rather than rely on punctuality alone. For diagnosing local traffic-related instability, travel time variability provides particularly relevant information because it corresponds closely to realised movement through the analysed segment. Punctuality indicators should be interpreted together with upstream delay information, while headway regularity should

be evaluated in relation to broader route-level service conditions and operational control.

The evaluation of mitigating measures should also account for the operating context in which these measures are implemented. Simple comparisons between segments with and without dedicated infrastructure or traffic signal priority may confound measure effects with pre-existing differences between locations. Where possible, future evaluations should therefore use longitudinal before-after designs, ideally combined with comparable control segments, to better distinguish intervention effects from pre-existing differences between locations.

Future research could further improve the traffic exposure measure by incorporating traffic speed, queue length, signal timing, priority activation, and other indicators of the conditions directly experienced by public transport vehicles. Higher temporal resolution would also make it possible to analyse short-lived congestion, signal delays, and bunching events that may be smoothed within hourly observations.

Additional operational variables such as passenger demand, dwell times, vehicle occupancy, incidents, and upstream service-control decisions would be especially valuable for explaining OTP and CV_h. Future studies could also extend the spatial unit from individual stop-to-stop segments to short corridors or sequences of consecutive segments, allowing delay propagation and headway instability to be studied over a larger part of the route while retaining a detailed spatial scale.

Finally, applying the framework to a larger and more diverse sample of segments or to other cities would help assess the extent to which the relationships observed in Geneva are transferable to other dense urban public transport networks.

7.5. Final conclusion

This thesis shows that public transport reliability in dense urban bus and tram operations cannot be understood through a single indicator or a single source of disruption. Local traffic exposure is most clearly reflected in the variability of realised stop-to-stop travel times, while schedule adherence and headway regularity are shaped more strongly by upstream and route-level operational processes. Mitigating measures do not show a consistent direct association with improved reliability, but there is limited evidence that they may reduce the sensitivity of travel time variability to increasing traffic exposure.

A detailed segment-level analysis using multiple reliability indicators therefore provides a more nuanced understanding of how local traffic conditions interact with public transport operations. Such an approach helps distinguish locally generated travel time instability from reliability problems that originate or propagate elsewhere in the network, and provides a stronger basis for selecting appropriate operational and infrastructural responses.

[*This page intentionally left blank*]

References

Aivaz, K. A., & Șerbănescu, M. (2024). “Ecosystem Services Evaluation of the Danube Delta: An Analysis Using Hierarchical Multifactor Regression”. *Studies in Business and Economics*, *(vol. 19)*(1), 5–21. https://doi.org/10.2478/SBE-2024-0001

Anderson, P., & Geroliminis, N. (2020). “Dynamic lane restrictions on congested arterials”. *Transportation Research Part A: Policy and Practice*, *(vol. 135)*(2124), 224–243. https://doi.org/10.1016/j.tra.2020.03.009

Apple Inc. (2026). Apple Maps Look Around [Street-level imagery captured from Apple Maps Look Around for the selected segment environments. Accessed: 2026-06-01]. https://maps.apple.com

Bates, J., Polak, J., Jones, P., & Cook, A. (2001). “The valuation of reliability for personal travel”. *Transportation Research Part E: Logistics and Transportation Review*, *(vol. 37)*(2-3), 191–229. https://doi.org/10.1016/S1366-5545(00)00011-9

Bowman, L. A., & Turnquist, M. A. (1981). “Service frequency, schedule reliability and passenger wait times at transit stops”. *Transportation Research Part A: General*, *(vol. 15)*(6), 465–471. https://doi.org/10.1016/0191-2607(81)90114-X

Bundesamt f”ur Strassen ASTRA. (2013). *Gestaltung von Ausr”ustungen der Anschl”usse an das Nationalstrassennetz* (Ausgabe 2013, V2.00, tech. rep. No. ASTRA 85006) (Accessed: 2026-06-23). Bundesamt f”ur Strassen ASTRA. https://www.astra.admin.ch/dam/astra/de/dokumente/standards_fuer_nationalstrassen/astra_85006_gestaltungvonausruestungenderanschluesseandasnationa.pdf.download.pdf/astra_85006d.pdf

Bundesamt f”ur Strassen ASTRA. (2018). *Anzuwendende Standards auf Nationalstrassen 3. Klasse* (Ausgabe 2018, V1.00, tech. rep. No. ASTRA 8A009) (Accessed: 2026-06-23). Bundesamt f”ur Strassen ASTRA. https://www.astra.admin.ch/dam/astra/de/dokumente/standards_fuer_nationalstrassen/astra%208A009%20anzuwendendestandardsaufnationalstrassen3klasse.pdf.download.pdf/astra%208A009%20anzuwendendestandardsaufnationalstrassen3klasse2018v100.pdf

Buran, B. (2025). “Passenger satisfaction modeling in public bus transportation based on business model approach: Ten city case studies”. *Case Studies on Transport Policy*, *(vol. 21)*(3), 101472. https://doi.org/10.1016/j.cstp.2025.101472

Burnham, K. P., & Anderson, D. R. (2002, July). *Model Selection and Multimodel Inference* (2nd ed.). Springer New York. https://doi.org/10.1007/B97636

Camus, R., Longo, G., & Macorini, C. (2005). “Estimation of transit reliability level-of-service based on automatic vehicle location data”. *Transportation Research Record*, (1927), 277–286. https://doi.org/10.3141/1927-31

CFI Team. (2020, June). Variance Inflation Factor (VIF). https://corporatefinanceinstitute.com/resources/data-science/variance-inflation-factor-vif/

Chatelain, J. B., & Ralf, K. (2014). “Spurious regressions and near-multicollinearity, with an application to aid, policies and growth”. *Journal of Macroeconomics*, *(vol. 39)*(PA), 85–96. https://doi.org/10.1016/J.JMACRO.2013.11.003

Cramer, A., Cucarese, J., Tran, M., Lu, A., & Reddy, A. (2009). “Performance Measurements on Mass Transit: Case Study of New York City Transit Authority”. *Transportation Research Record: Journal of the Transportation Research Board*, *(vol. 2111)*, 125–138. https://www.researchgate.net/publication/287103609_Performance_measurements_on_mass_transit_-_New_York_city_transit_authority_case_study

Currie, G., Douglas, N. J., & Kearns, I. (2011). “An Assessment of Alternative Bus Reliability Indicators”. *Australasian Transport Research Forum*. https://australasiantransportresearchforum.org.au/wp-content/uploads/2022/03/2012_Currie_Douglas_Kearns.pdf

Dalal, D. K., & Zickar, M. J. (2012). “Some Common Myths About Centering Predictor Variables in Moderated Multiple Regression and Polynomial Regression”. *Organizational Research Methods*, *(vol. 15)*(3), 339–362. https://doi.org/10.1177/1094428111430540

De Brito-Filho, F. N., & Oliveira-Neto, F. M. (2025). “Analysis of the causal effects of exclusive bus lanes on performance indicators of public transport corridors”. *Transportation Research Part A: Policy and Practice*, *(vol. 196)*, 104485. https://doi.org/10.1016/J.TRA.2025.104485

Dixit, M., Brands, T., van Oort, N., Cats, O., & Hoogendoorn, S. (2019). “Passenger Travel Time Reliability for Multimodal Public Transport Journeys”. *Transportation Research Record*, *(vol. 2673)*(2), 149–160. https://doi.org/10.1177/0361198118825459

Ekström, J. (2011, October). *The Phi-coefficient, the Tetrachoric Correlation Coefficient, and the Pearson-Yule Debate* (tech. rep.). University of California. https://escholarship.org/uc/item/7qp4604r

Fadaei, M., & Cats, O. (2016). “Evaluating the impacts and benefits of public transport design and operational measures”. *Transport Policy*, *(vol. 48)*, 105–116. https://doi.org/10.1016/J.TRANPOL.2016.02.015

Faraway, J. J. (2002). *Practical Regression and Anova using R*. University of Michigan.

Federal Highway Administration. (2017, February). *Traffic Congestion and Reliability: Linking Solutions to Problems* (tech. rep.). U.S. Department of Transportation. https://ops.fhwa.dot.gov/congestion_report_04/chapter2.htm

Field, A. (2009). *Discovering statistics using SPSS* (3rd ed.). SAGE Publications.

Fonsberg, O. (2026, March). The Issue of Boundedness. https://stats.libretexts.org/Courses/Knox_College/Linear_Models_and_Rurita_Kralovstvi/08%3A_Fixing_the_Violations/8.01%3A_The_Issue_of_Boundedness

Gittens, A., & Shalaby, A. (2015). “Evaluation of bus reliability measures and development of a new composite indicator”. *Transportation Research Record*, *(vol. 2533)*, 91–99. https://doi.org/10.3141/2533-10

Hair, J. F., Black, W., Babin, B. J., & Anderson, R. E. (2010, March). *Multivariate Data Analysis: A Global Perspective* (7th ed.). Pearson.

Harmann, D., Burmeister, L., Häbel, R., Bienzeisler, L., & Friedrich, B. (2025). “Optimizing Public Transit: A Critical Analysis of Increased Bus Frequency and its Effects on Traffic Dynamics”. *Transportation Research Procedia*, *(vol. 86)*, 282–289. https://doi.org/10.1016/J.TRPRO.2025.04.036

Hu, W. X., & Shalaby, A. (2017). “Use of automated vehicle location data for route- and segment-level analyses of bus route reliability and speed”. *Transportation Research Record*, *(vol. 2649)*, 9–19. https://doi.org/10.3141/2649-02

Jenelius, E. (2018). “Public transport experienced service reliability: Integrating travel time and travel conditions”. *Transportation Research Part A: Policy and Practice*, *(vol. 117)*, 275–291. https://doi.org/10.1016/j.tra.2018.08.026

Jenks, C. W., Jencks, C. F., Schwager, D. S., Delaney, E. P., & Hitchcock, S. E. (2013). *TCRP Report 165: Transit Capacity and Quality of Service Manual* (tech. rep.). Transportation Research Board of the National Academies. Washington DC. https://onlinepubs.trb.org/onlinepubs/tcrp/tcrp_rpt_165fm.pdf

Kieu, L. M., Bhaskar, A., & Chung, E. (2014). “Public Transport Travel-Time Variability Definitions and Monitoring”. *Journal of Transportation Engineering*, *(vol. 141)*(1), 04014068. https://doi.org/10.1061/(ASCE)TE.1943-5436.0000724

Kim, J. H. (2019). “Multicollinearity and misleading statistical results”. *Korean Journal of Anesthesiology*, *(vol. 72)*(6), 558. https://doi.org/10.4097/KJA.19087

Kimpel, T. J., Strathman, J. G., Dueker, K., Kimpel, T. J., Strathman, J. G., & Dueker, K. (2000). “Time Point-Level Analysis of Passenger Demand and Transit Service Reliability”. *Center for Urban Studies Publications and Reports*, *(vol. 7)*. https://pdxscholar.library.pdx.edu/cus_pubs/7

Kornbrot, D. (2014). “Point Biserial Correlation”. *Statistics Reference Online*. https://doi.org/10.1002/9781118445112.STAT06227

Lomax, T., Schrank, D., Turner, S., Margiotta, R., Johnson, C., Paniati, J., & Lindley, J. (2003, May). *Selecting Travel Reliability Measures* (tech. rep.). Texas Transportation Institute. https://scispace.com/pdf/selecting-travel-reliability-measures-1sciijnek2.pdf

Mullahy, J., & Norton, E. C. (2022). “Why Transform Y? A Critical Assessment of Dependent-Variable Transformations in Regression Models for Skewed and Sometimes-Zero Outcomes”. https://doi.org/10.3386/W30735

Office Cantonal de la Statistique. (2025). Les 45 communes genevoises: Genève-ville [Accessed: 2026-06-19]. https://statistique.ge.ch/communes/apercu.asp?commune=21

opentransportdata.swiss. (2026a). GTFS | Fahrplandaten-Archiv. https://archive.opentransportdata.swiss/timetable_gtfs_archive.htm

opentransportdata.swiss. (2026b). GTFS | Realtime. https://data.opentransportdata.swiss/en/dataset/gtfsrt

Papke, L. E., & Wooldridge, J. M. (1996). “Econometric methods for fractional response variables with an application to 401(k) plan participation rates”. *Journal of Applied Econometrics*, *(vol. 11)*(6), 619–632.

Polus, A. (1978). “Modeling and measurements of bus service reliability”. *Transportation Research*, *(vol. 12)*(4), 253–256. https://doi.org/10.1016/0041-1647(78)90067-9

République et Canton de Genève. (2017, December). Une centrale de régulation du trafic pour plus de sécurité et de fluidité [Communiqué de presse]. https://www.tpg.ch/sites/default/files/2019-11/2017121%20Centrale%20de%20r%C3%A9gulation%20du%20trafic.pdf

Saberi, M., Ali Zockaie, K., Feng, W., & El-Geneidy, A. (2013). “Definition and Properties of Alternative Bus Service Reliability Measures at the Stop Level”. *Journal of Public Transportation*, *(vol. 16)*(1), 97–122. https://doi.org/10.5038/2375-0901.16.1.6

Schober, P., & Schwarte, L. A. (2018). “Correlation Coefficients: Appropriate Use and Interpretation”. *Anesthesia and Analgesia*, *(vol. 126)*(5), 1763–1768. https://doi.org/10.1213/ANE.0000000000002864

Schrank, D., & Lomax, T. (2002, June). *The 2002 Urban Mobility Report* (tech. rep.). Texas Transportation Institute. https://doi.org/10.21949/1503647

Shelat, S., Cats, O., van Oort, N., & van Lint, J. W. (2023). “Evaluating the impact of waiting time reliability on route choice using smart card data”. *Transportmetrica A: Transport Science*, *(vol. 19)*(2), 2028929. https://doi.org/10.1080/23249935.2022.2028929

SITG. (2026). Carte interactive | SITG Le Territoire Genevois à la Carte. https://map.sitg.ge.ch/app/

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

Soza-Parra, J., Raveau, S., & Muñoz, J. C. (2021). “Public transport reliability across preferences, modes, and space”. *Transportation*, *(vol. 49)*(2), 621–640. https://doi.org/10.1007/s11116-021-10188-2

Stadt Winterthur. (2022, August). *Steuerungs- und Dosierungskonzept Frauenfelderstrasse* (tech. rep.). Stadt Winterthur. https://stadt.winterthur.ch/themen/leben-in-winterthur/verkehr-mobilitaet/strategien-konzepte/ov-hochleistungskorridor/ergaenzungsstudien/frauenfelderstrasse/

StatsTest. (2026, January). Linear Regression Assumptions and Diagnostics in Practice. https://www.statstest.com/linear-regression-assumptions-diagnostics-practice

Stewart, K. (2026, March). Pearson's correlation coefficient. https://www.britannica.com/topic/Pearsons-correlation-coefficient

Strathman, J. G., Dueker, K., Kimpel, T. J., Gerhart, R., Turner, K., Taylor, P., Callas, S., & Griffin, D. (1999). "Automated Bus Dispatching, Operations Control, and Service Reliability: The Initial Tri-Met Experience". *Center for Urban Studies Publications and Reports*, *(vol. 86)*. https://pdxscholar.library.pdx.edu/cus_pubs/86

Tirachini, A., Godachevich, J., Cats, O., Muñoz, J. C., & Soza-Parra, J. (2022). "Headway variability in public transport: a review of metrics, determinants, effects for quality of service and control strategies". *Transport Reviews*, *(vol. 42)*(3), 337–361. https://doi.org/10.1080/01441647.2021.1977415

tpg. (2008). Histoire des TPG - Les TPG ont plus de 100 ans. Un grand passé. Et un immense avenir. https://web.archive.org/web/20081207145840/http://www.tpg.ch/fr/au-sujet-des-tpg/histoire/index.php

tpg. (2024, December). Plans du réseau | Plan tpg géographique 2025. https://www.tpg.ch/sites/default/files/2024-11/Geneve%20TPG%20Plan%20Geographique_15122024.pdf

tpg. (2025, December). Plans du réseau | Plan tpg diurne 2026. https://www.tpg.ch/sites/default/files/2025-12/Geneve%20TPG%20Plan%20Schematique%202025-10-02_20251214_low.pdf

tpg. (2026a, April). Accueil - tpg opendata. https://opendata.tpg.ch/pages/accueil/

tpg. (2026b). Horaires par arrêt pour la ligne. https://www.tpg.ch/fr/lignes

Trompet, M., Liu, X., & Graham, D. J. (2011). "Development of key performance indicator to compare regularity of service between urban bus operators". *Transportation Research Record: Journal of the Transportation Research Board*, (2216), 33–41. https://doi.org/10.3141/2216-04

Uniman, D., Attanucci, J., Mishalani, R., & Wilson, N. (2010). "Service reliability measurement using automated fare card data". *Transportation Research Record*, (2143), 92–99. https://doi.org/10.3141/2143-12

University of Wisconsin-Madison. (2024). Linearity. https://sscc.wisc.edu/sscc/pubs/RegDiag-R/linearity.html

Van Oort, N. (2011). *Service Reliability and Urban Public Transport Design* (tech. rep.). Technische Universiteit Delft. https://repository.tudelft.nl/file/File_243bbe9b-0d7c-4bc5-8ec0-9b9aaeb6afba

West, R. M. (2021). "Best practice in statistics: The use of log transformation". *Annals of Clinical Biochemistry*, *(vol. 59)*(3), 162. https://doi.org/10.1177/00045632211050531

A

GTFS Structure

All the GTFS datasets are from the Open Transport Data Switzerland archive (opentransportdata.swiss, 2026a). A GTFS dataset includes several linked tables with information about different transit services, see Figure A.1.

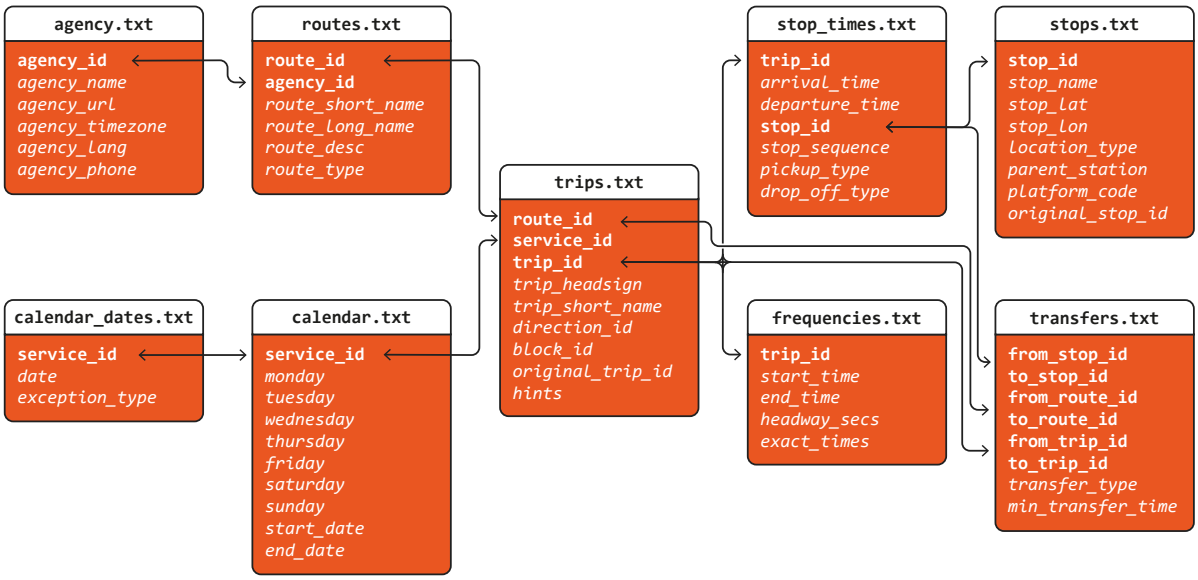


Figure A.1: Structure of the GTFS dataset; the main tables and the key columns linking them.

[*This page intentionally left blank*]

B

Information on Selected Segments

This appendix provides additional information on the 15 selected directed stop-to-stop segments used in the Geneva case study. It complements the classification presented in Chapter 4 by documenting their spatial location, the corresponding stop and traffic-counter identifiers, and the local street environments in which the analysed public transport movements take place.

B.1. Location within Geneva

Figure B.1 shows the spatial distribution of the selected segments and their associated road traffic counters within the Geneva urban area.

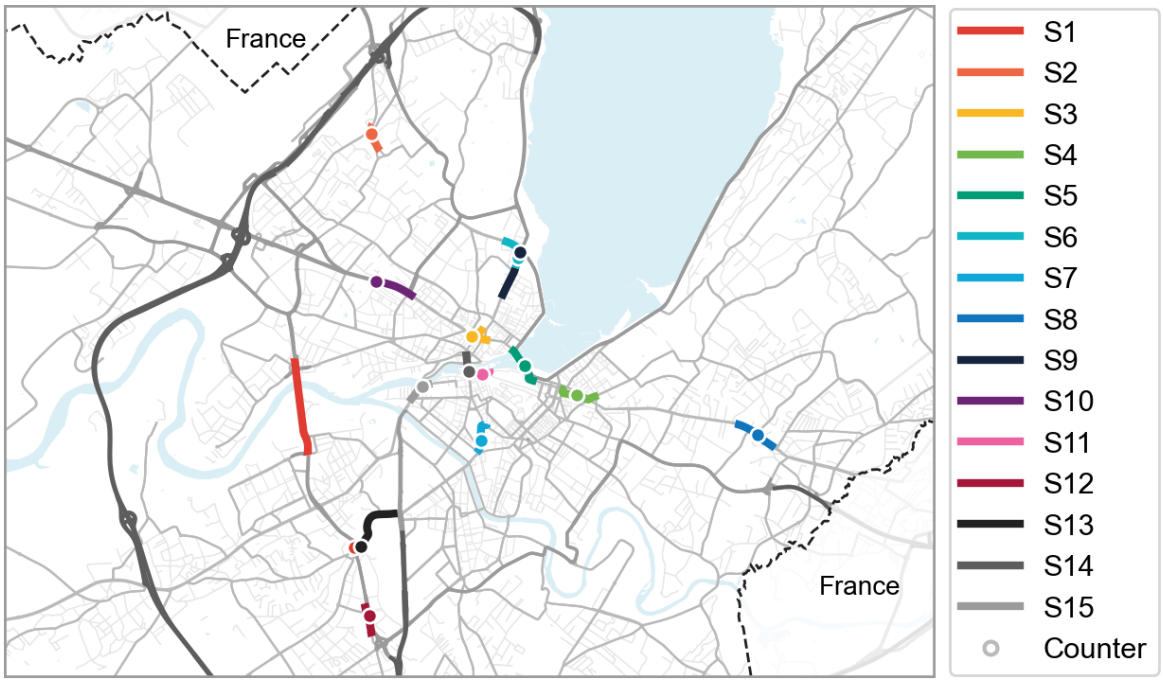


Figure B.1: Locations of the selected stop-to-stop public transport segments and associated road traffic counters in the Geneva urban area.

B.2. Segments and traffic-counter information

Table B.1 identifies the upstream and downstream stops defining each directed segment, together with the corresponding GTFS stop identifiers and the road traffic counter matched to that segment. The stop identifiers provide the link to the scheduled and realised public transport data, while the counter identifier represents the local traffic observation used to construct the segment-hour traffic exposure variable. Several stop names changed during the 2019-2024 study period; these alternative names

are reported in the table notes where applicable.

Table B.1: Selected stop-to-stop public transport segments and associated road traffic counters.

	First stop name	Stop ID	Second stop name	Stop ID	Counter
S1	Vernier, Avenue de l'Ain	8593138	Petit-Lancy, Chemin du Bac	8592978	6058.2
S2	Grand-Saconnex, place	8587914	Grand-Saconnex, Le Pommier	8592945	6041.2
S3	Genève, gare Cornavin[a]	8587057	Genève, Chantepoulet	8592801	6079.2
S4	Genève, Place des Eaux-Vives	8592883	Genève, 31-Décembre[b]	8592773	82.1
S5	Genève, Métropole	8592856	Genève, Mont-Blanc	8592864	2005.1
S6	Genève, Maison de la Paix	8592825	Genève, Butini	8592795	6021.1
S7	Genève, Uni-Mail	8592915	Genève, Plainpalais	8587907	6007.2
S8	Chêne-Bourg, Place Favre	8592713	Chêne-Bougeries,Grange-Falquet	8592706	80.2
S9	Genève, Butini	8592795	Genève, Môle	8592863	6021.2
S10	Genève, Vieusseux	8592922	Genève, Servette	8592903	6076.2
S11	Genève, Bel-Air	8587387	Genève, Stand	8592906	39.1
S12	Grand-Lancy, Palettes	8592988	Grand-Lancy, Pontets	8592992	258.1
S13	Lancy-Pont-Rouge, gare/Etoile[c]	8592679	Grand-Lancy, Mairie de Lancy[d]	8587902	6062.1
S14	Genève, Mercier[e]	8592841	Genève, Stand	8592906	2017.2
S15	Genève, Jonction	8587062	Genève, Palladium	8592874	6124.2

[a] *Genève, gare Cornavin* was also named: *Genève, Cornavin*.
[b] *Genève, 31-Décembre* was also named: *Genève, 31 Décembre*.
[c] *Lancy-Pont-Rouge, gare/Etoile* was also named: *Carouge GE, P+R Etoile*.
[d] *Grand-Lancy, Mairie de Lancy* was also named: *Grand-Lancy, mairie*.
[e] *Genève, Mercier* was also named: *Genève, Isaac-Mercier*.

B.3. Street-level view

Figure B.2 provides an illustrative street-level view of the local environment of each selected segment. The images are included to support the interpretation of differences in road layout, public transport infrastructure, and interaction with surrounding traffic that are difficult to capture fully through the numerical segment characteristics alone. Each image is clickable and links to the corresponding map location.



(a) Segment 1.



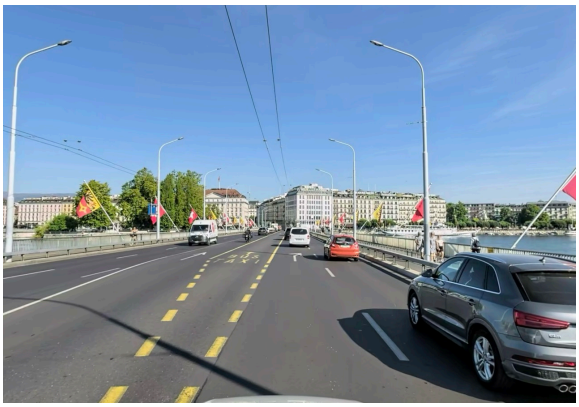
(b) Segment 2.



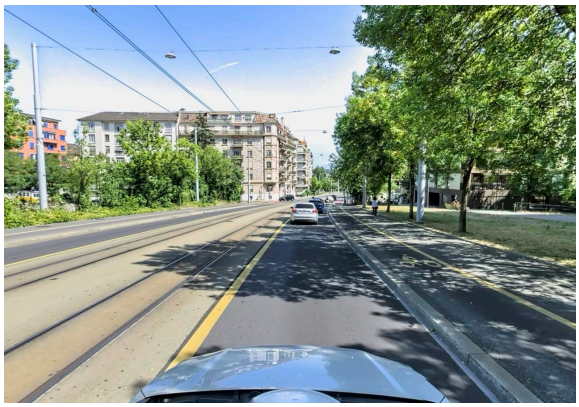
(c) Segment 3.



(d) Segment 4.



(e) Segment 5.



(f) Segment 6.



(g) Segment 7.



(h) Segment 8.



(i) Segment 9.



(j) Segment 10.



(k) Segment 11.



(l) Segment 12.



(m) Segment 13.



(n) Segment 14.

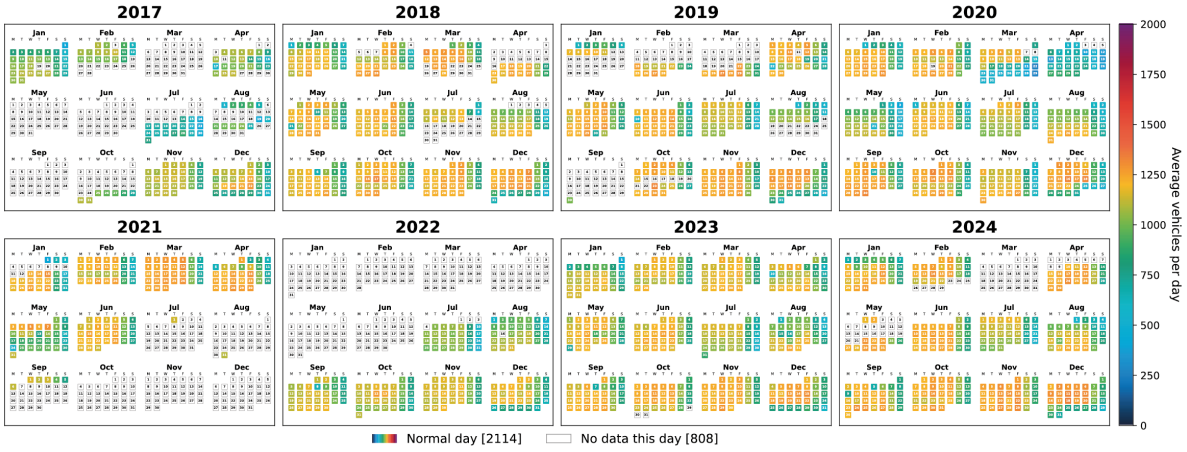


(o) Segment 15.

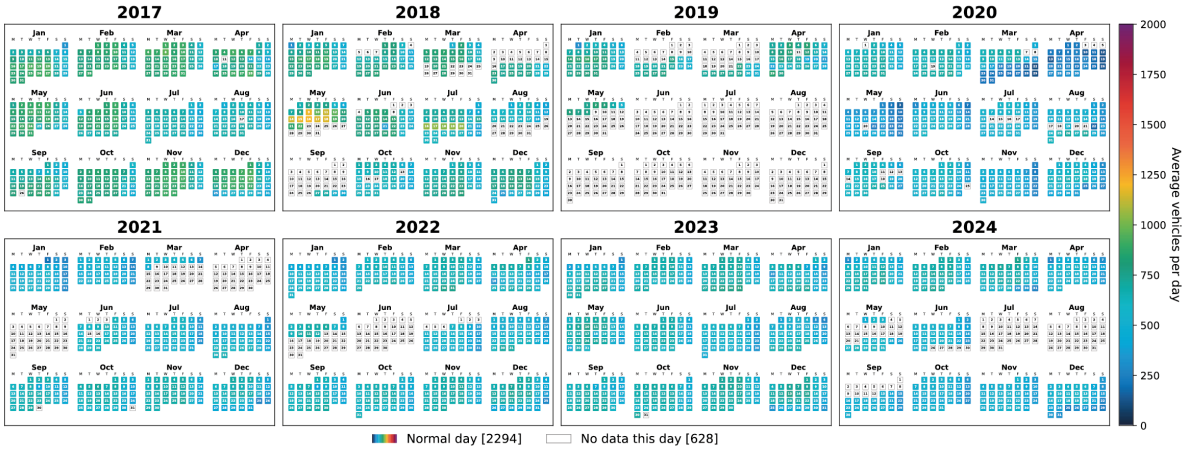
Figure B.2: Illustrative street-level view of the selected segment environments (Apple Inc., 2026).

Traffic Data from Counting Station

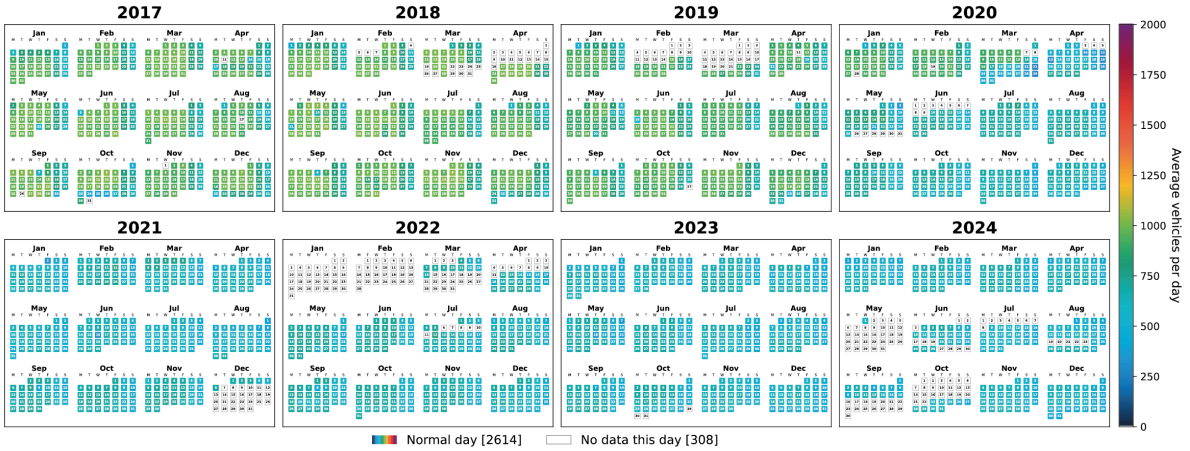
The figures in this appendix show calendar plots of the daily average vehicle counts per counting station. Each cell represents one day. The colour scale indicates the average number of vehicles per hour, while white highlighted (☐) cells indicate missing data, outliers, or suspiciously constant values.



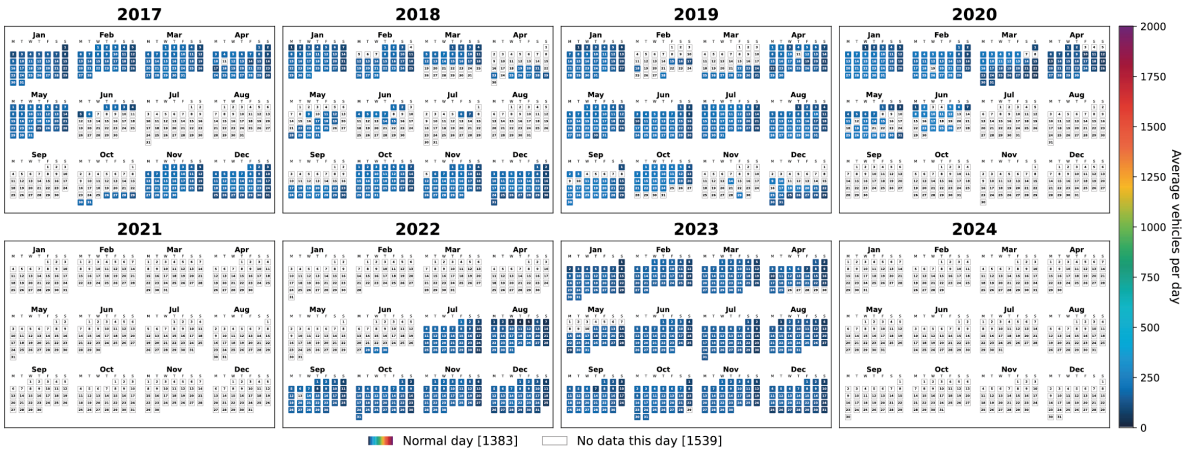
(a) Segment 1, counting station 6058.2.



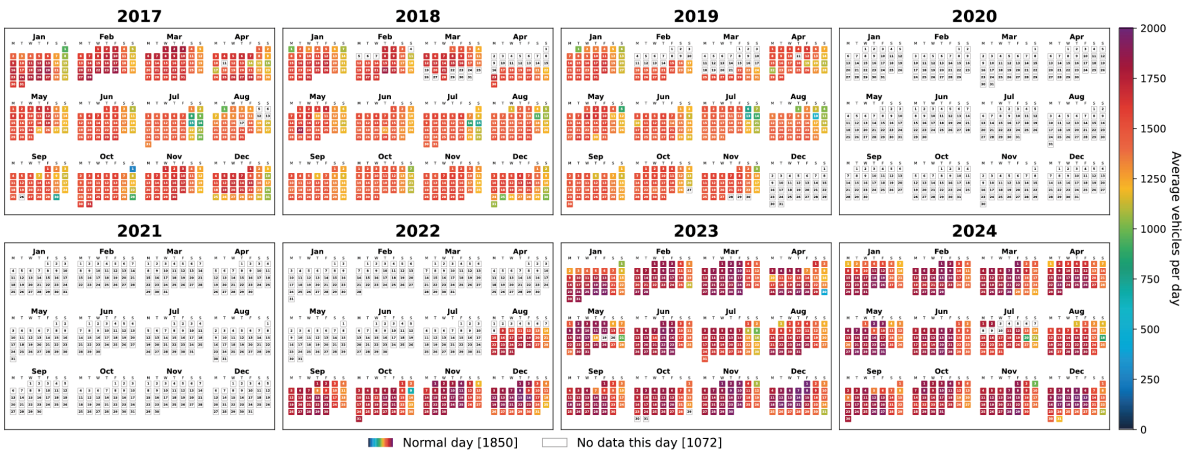
(b) Segment 2, counting station 6041.2.



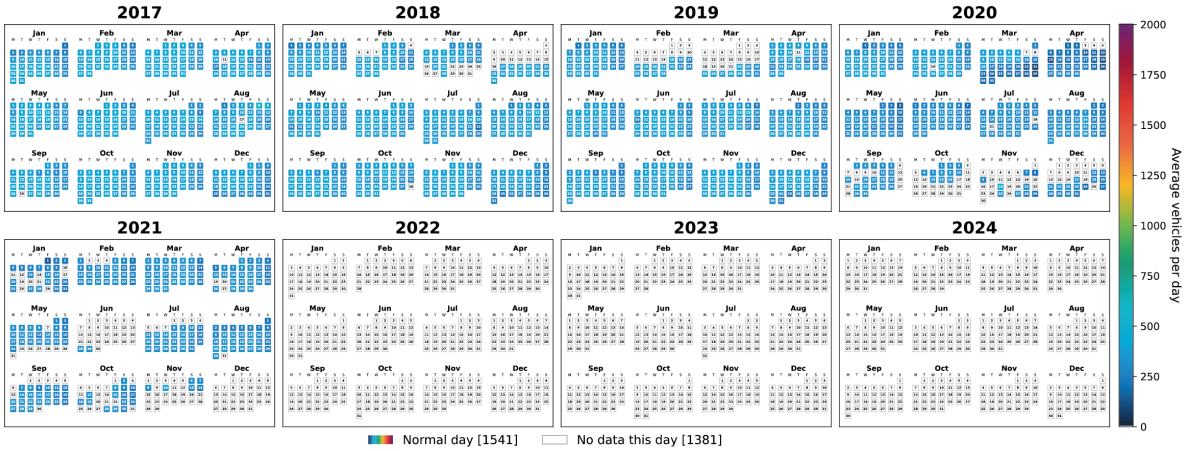
(c) Segment 3, counting station 6079.2.



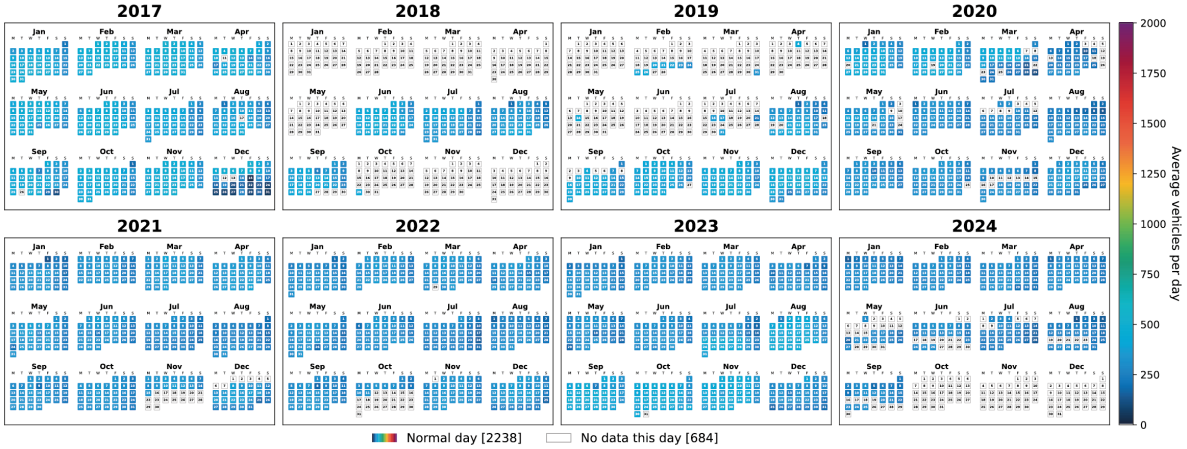
(d) Segment 4, counting station 82.1.



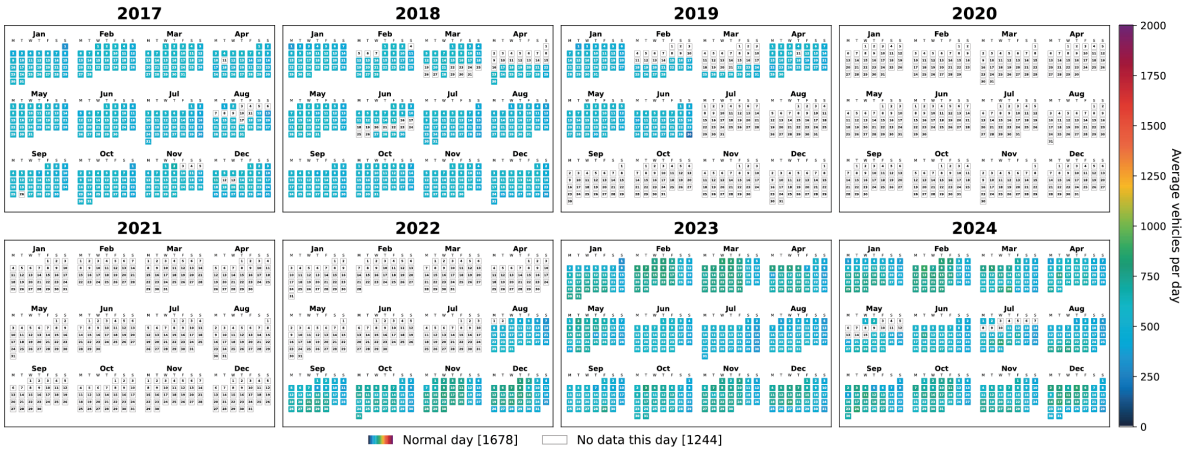
(e) Segment 5, counting station 2005.1.



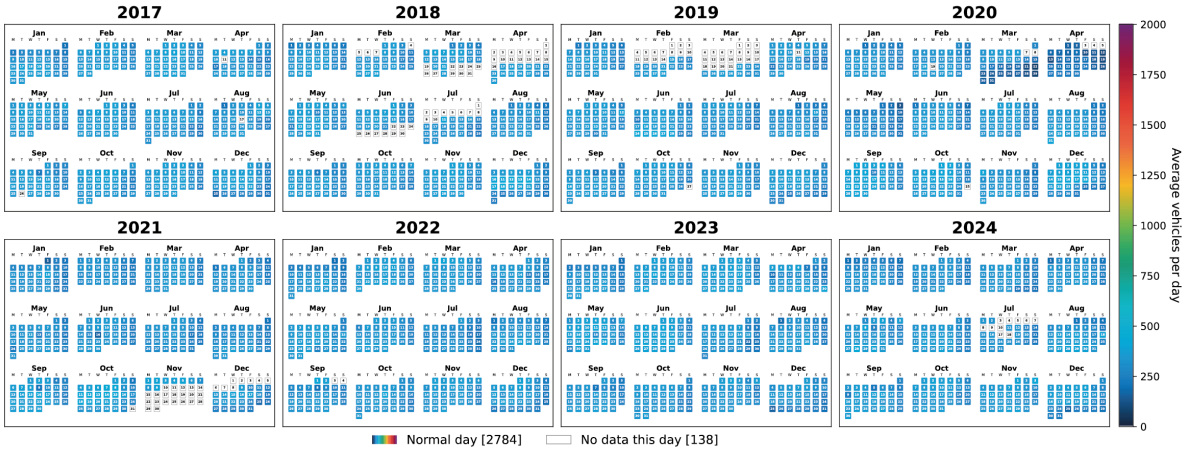
(f) Segment 6, counting station 6021.1.



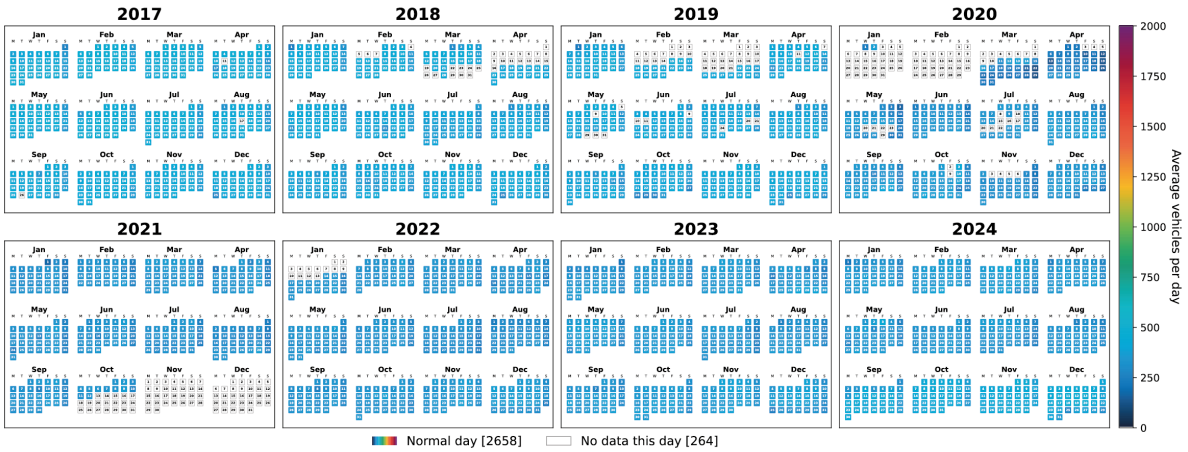
(g) Segment 7, counting station 6007.2.



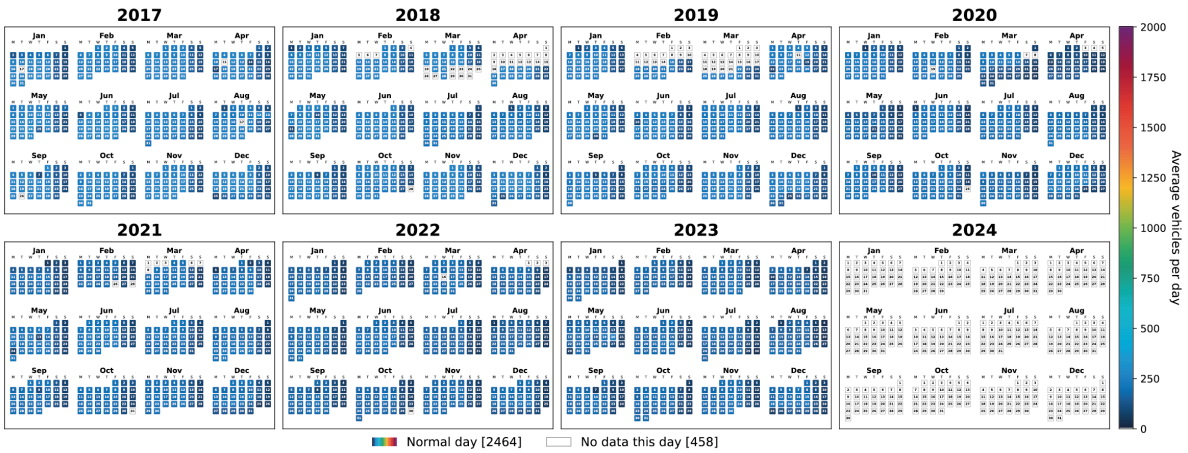
(h) Segment 8, counting station 80.2.



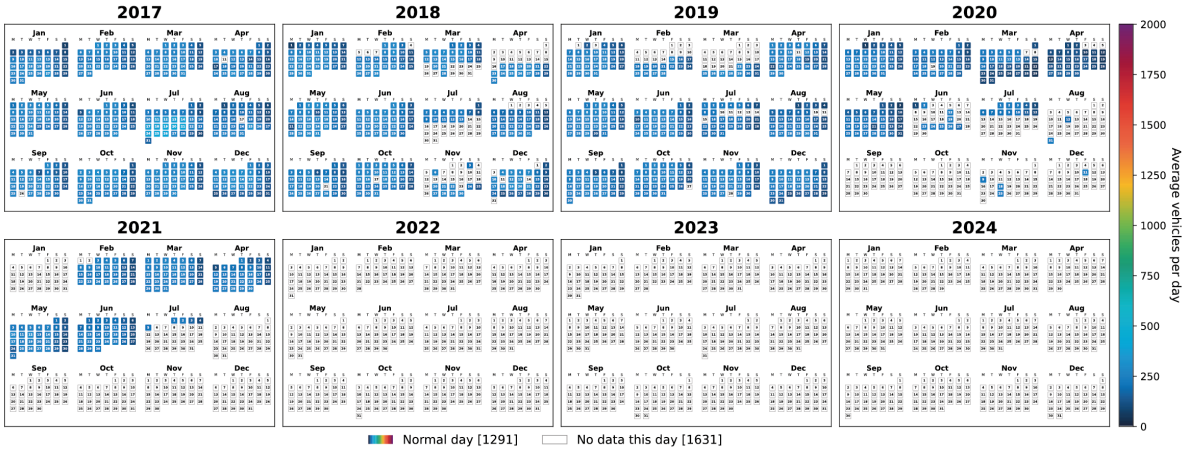
(i) Segment 9, counting station 6021.2.



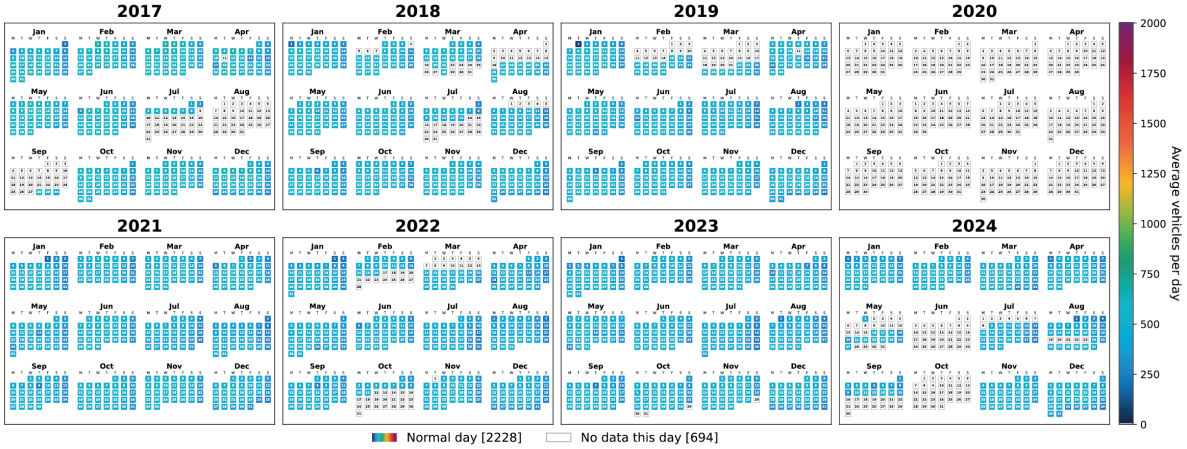
(j) Segment 10, counting station 6076.2.



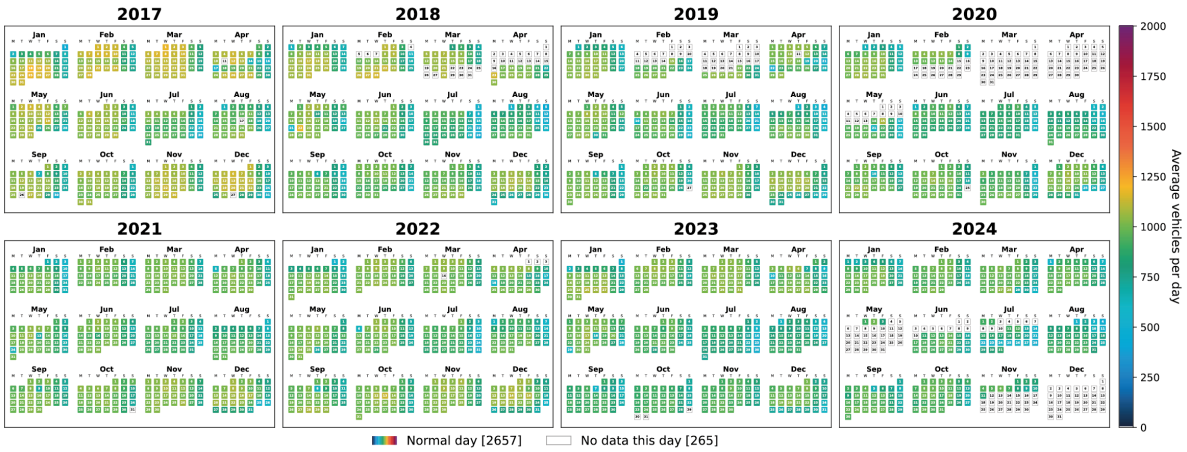
(k) Segment 11, counting station 39.1.



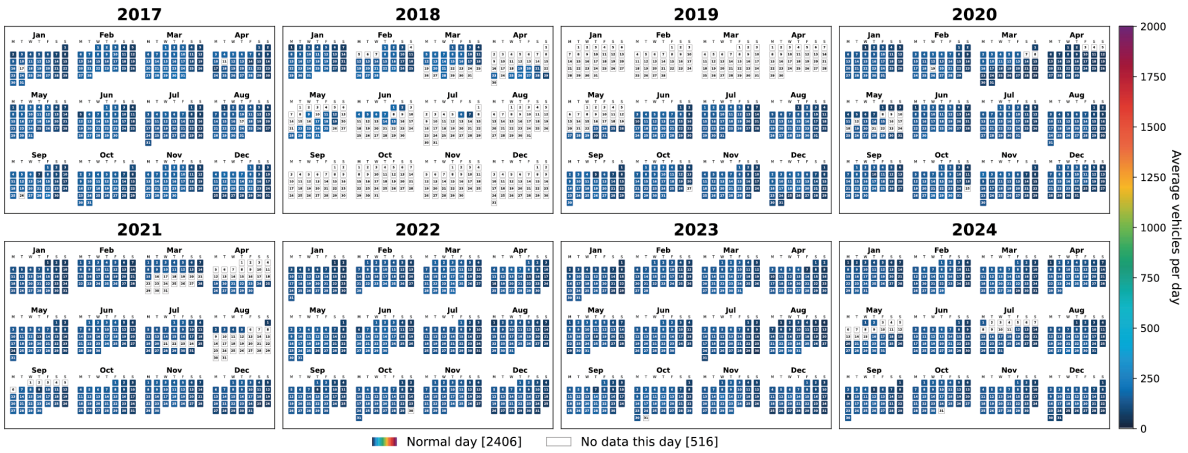
(l) Segment 12, counting station 258.1.



(m) Segment 13, counting station 6062.1.



(n) Segment 14, counting station 2017.2.



(o) Segment 15, counting station 6124.2.

Figure C.1: Daily average vehicle counts from counting stations at all segment locations, 2017–2024.

Descriptive Distributions of Reliability Outcomes

This appendix provides the full segment-level distributions of the three reliability outcomes used in the empirical analysis: on-time performance (OTP), the coefficient of variation of travel time (CV_{TT}), and the coefficient of variation of headway (CV_h). These figures complement the summary provided in Subsection 5.1.2 by showing how the distribution of each reliability indicator differs across the selected directed stop-to-stop segments.

D.1. Distribution of segment-hour OTP by segment

Figure D.1 shows the distribution of segment-hour OTP values for each selected segment. Each observation represents the share of valid public transport movements within a segment-hour that arrived within the predefined punctuality window at the downstream stop.

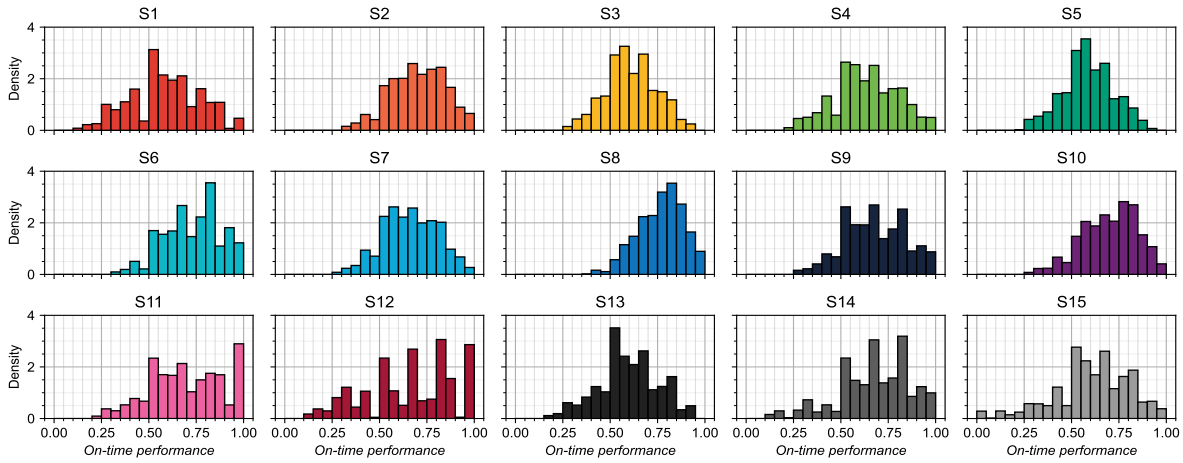


Figure D.1: Distribution of segment-hour OTP values by selected segment.

The distributions indicate that OTP varies substantially across the selected segments. Most observations are concentrated between approximately 0.5 and 0.9, but the location and spread of the distributions differ by segment. Some segments show relatively concentrated distributions, suggesting more stable punctuality outcomes, while others show broader or less regular distributions. For example, several segments contain observations across a wide range of OTP values, indicating that punctuality can vary strongly between segment-hours even within the same selected segment.

This pattern shows that segment-level mean OTP values alone do not fully describe schedule adherence. Two segments may have similar average punctuality while still differing in the consistency of their performance over time.

D.2. Distribution of segment-hour CV_{TT} by segment

Figure D.2 shows the distribution of segment-hour values of the coefficient of variation of travel time for each selected segment. Higher values indicate greater relative variability in realised stop-to-stop travel times.

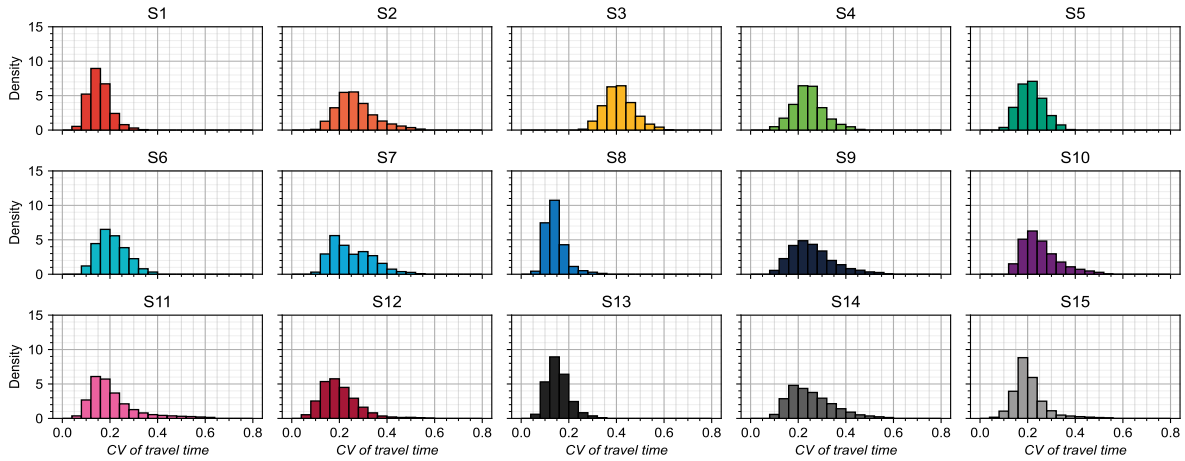


Figure D.2: Distribution of segment-hour CV_{TT} values by selected segment.

The distributions show that travel time variability differs clearly across the selected segments. Most observations are concentrated between approximately 0.1 and 0.35, indicating that realised stop-to-stop travel times are relatively stable in many segment-hours. However, the location and spread of the distributions vary by segment. Several segments show relatively concentrated distributions at lower values, while others show broader distributions or longer right tails, indicating that higher travel time variability occurs more frequently or more strongly on these segments.

The contrast between segments is especially relevant because CV_{TT} is the reliability indicator most directly linked to the analysed stop-to-stop movement. Unlike OTP, which may partly reflect delay accumulated upstream, CV_{TT} captures the relative stability of realised travel times within the selected segment-hour itself.

D.3. Distribution of segment-hour CV_h by segment

Figure D.3 shows the distribution of segment-hour values of the coefficient of variation of headway for each selected segment. Higher values indicate greater variability in realised headways and therefore lower headway regularity.

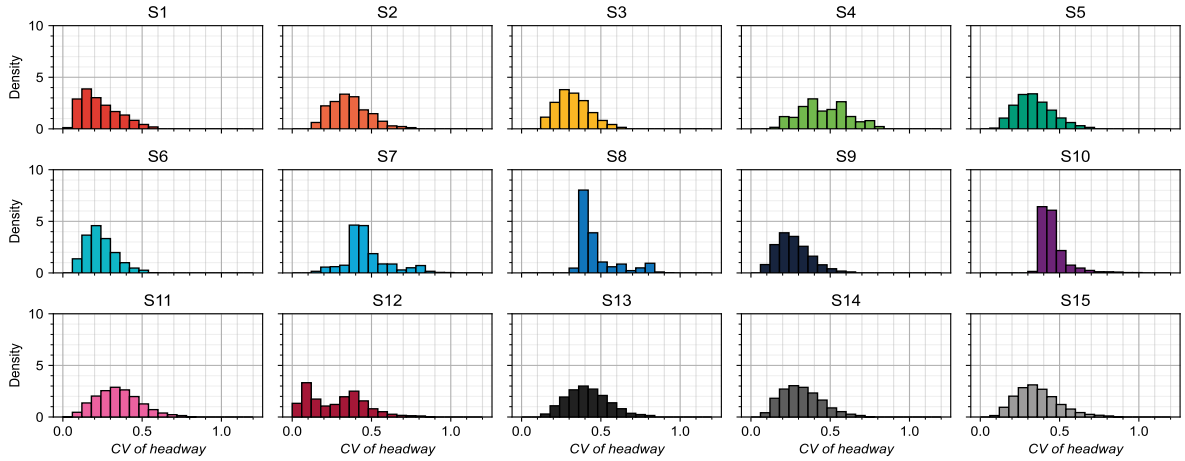


Figure D.3: Distribution of segment-hour CV_h values by selected segment.

The distributions show that headway variability differs substantially across the selected segments. Most observations are concentrated between approximately 0.1 and 0.6, but the location and spread of the distributions vary clearly by segment. Some segments show relatively concentrated distributions, suggesting more consistent headway regularity within those segments. Other segments show broader or less regular distributions, indicating that headway variability differs more strongly between segment-hour observations.

Compared with the distribution of travel time variability, the coefficient of variation of headway shows a wider spread for several segments. This suggests that headway regularity captures a different dimension of reliability than travel time variability. A segment can have relatively stable stop-to-stop travel times while still experiencing irregular vehicle spacing, for example due to upstream delay propagation, bunching, uneven departures, or interactions between multiple lines using the same segment.

[*This page intentionally left blank*]

Regression Model Development

This appendix presents the complete model-development sequence underlying the final regression specifications reported in Section 5.3. For each reliability outcome, the hierarchical blockwise models are presented first, followed by the backward-elimination procedure used to obtain the final parsimonious specification. The blockwise models show how the estimated associations and explanatory power change as infrastructure, service, temporal, traffic, mitigating-measure, and interaction variables are added sequentially. The subsequent elimination models document how variables that did not contribute meaningfully to the final specification were removed, subject to the model-hierarchy constraints described in Section 3.6. The substantive interpretation and comparison of the final specifications are provided in Section 5.3, while complete statistical estimation output is reported in Appendix G.

E.1. On-time performance model development

This section presents the complete model-development sequence for on-time performance (OTP). Six nested models were first estimated using hierarchical blockwise addition, with each subsequent model incorporating an additional thematic block of explanatory variables. The resulting specifications are presented in Table E.1.

Table E.1: Blockwise OLS regression results for On-Time Performance (OTP).

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Infrastructure variables						
`len`	-0.079 **	-0.107 *	-0.102 *	-0.106 *	-0.127 **	-0.149 **
Service variables						
`scheduled_tt_min`		0.028	0.028	0.026	0.034	0.036
`scheduled_hw_min`		0.006 ***	0.005 ***	0.006 ***	0.007 ***	0.007 ***
`mode_bus`		-0.079 **	-0.073 **	-0.079 **	-0.098 ***	-0.099 ***
`DT_with_delay`		-0.211 ***	-0.211 ***	-0.209 ***	-0.208 ***	-0.206 ***
`AT_with_delay`		-0.064 ***	-0.063 ***	-0.064 ***	-0.067 ***	-0.067 ***
Time variables						
`peak_period`			0.007 *	0.009 ***	0.010 ***	0.009 ***
`weekday`			-0.017 *	-0.017 **	-0.017 **	-0.018 ***
Traffic variables						
`IC_ratio`				0.030	0.059 **	0.121 ***
Mitigating measure variables						
`measure_package`					-0.039 **	-0.046 **
Interaction terms						
`traffic_x_measure`						-0.087
Model information						
No. Observations	228,094	228,094	228,094	228,094	228,094	228,094
R-squared	0.010	0.231	0.232	0.235	0.244	0.247

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The infrastructure-only specification provides limited explanatory power, with an R^2 of 0.010. `len` is negatively associated with OTP throughout the blockwise sequence, although the magnitude of the coefficient changes as additional explanatory variables are introduced.

The largest increase in explanatory power occurs after the inclusion of the service variables in Model 2, with R^2 increasing from 0.010 to 0.231. Several service-related variables show stable associations across the subsequent specifications. `scheduled_hw_min` is positively associated with OTP, while bus observations are associated with lower OTP than tram observations in the blockwise models. `DT_with_delay` and `AT_with_delay` show particularly strong negative associations, with departure delay at the upstream stop consistently showing the largest coefficient magnitude.

Adding the temporal variables in Model 3 results in only a small further improvement in model fit, with R^2 increasing from 0.231 to 0.232. Peak-period observations show a small positive association with OTP, whereas weekday observations show a negative association.

Model 4 introduces local traffic exposure through the centred IC-ratio variable. When first included, `IC_ratio` is positive but not statistically significant. After `measure_package` is introduced in Model 5, the coefficient becomes larger and statistically significant. The `measure_package` variable itself has a negative coefficient in this specification.

The full blockwise specification in Model 6 adds the interaction between local traffic exposure and the mitigating-measure package. The interaction coefficient is negative but not statistically significant in the full blockwise model. The inclusion of the final model blocks increases R^2 to 0.247. These coefficient changes illustrate that the estimated association between local traffic exposure and OTP depends on the other characteristics included in the model and should therefore be interpreted within the complete specification rather than in isolation.

Following the blockwise model construction, backward elimination was applied to obtain a more parsimonious specification. Variables that did not meet the retention criteria were removed sequentially, while lower-order terms required by retained interactions were preserved in accordance with the hierarchy principle described in Section 3.6. The resulting sequence is shown in Table E.2.

Table E.2: Backward Elimination OLS regression results for On-Time Performance (OTP).

Variable	Model 7	Model 8	Model 9	Model 10
Infrastructure variables				
`len`	-0.087 ***	-0.079 *	-0.079 *	-0.085 ***
Service variables				
`scheduled_tt_min`	✘			
`scheduled_hw_min`	0.005	-0.001	-0.001	✘
`mode_bus`	-0.066	✘		
`DT_with_delay`	-0.217 ***	-0.229 ***	-0.229 ***	-0.230 ***
`AT_with_delay`	-0.057 ***	-0.053 ***	-0.053 ***	-0.052 ***
Time variables				
`peak_period`	0.009 ***	0.000	✘	
`weekday`	-0.020 **	-0.041 ***	-0.041 ***	-0.039 ***
Traffic variables				
`IC_ratio`	0.098 ***	0.076 **	0.076 **	0.078 **
Mitigating measure variables				
`measure_package`	-0.030 **	-0.019	-0.019	-0.019
Interaction terms				
`traffic_x_measure`	-0.059	-0.068 *	-0.068 *	-0.068 *
Model information				
No. Observations	228,094	228,094	228,094	228,094
R-squared	0.232	0.218	0.218	0.217

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The elimination procedure removes `scheduled_tt_min` in Model 7, `mode_bus` in Model 8, `peak_period` in Model 9, and `scheduled_hw_min` in Model 10. The resulting final specification retains the `len`, `DT_with_delay`, `AT_with_delay`, `weekday`, `IC_ratio`, `measure_package`, and `traffic_x_measure`.

The main associations remain broadly stable during the elimination sequence. `DT_with_delay` continues to show the largest negative association with OTP, while `len` and `weekday` remain negatively associated with the outcome. `IC_ratio` remains positive and statistically significant in the final specification. `traffic_x_measure` becomes weakly significant and remains negative.

The adjusted R^2 of the final parsimonious model is 0.217, compared with 0.247 for the complete blockwise specification. The reduction in explanatory power reflects the removal of several variables while producing a simpler specification for substantive interpretation. The final OTP model corresponds to the specification presented in Section 5.3.

E.2. Travel time variability model development

This section presents the model-development sequence for the coefficient of variation of travel time (CV_{TT}). As for OTP, six hierarchical blockwise models were estimated before backward elimination was applied. Table E.3 presents the complete blockwise sequence.

Table E.3: Blockwise OLS regression results for Coefficient of Variation of Travel Time (CV_{TT}).

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Infrastructure variables						
`len`	-0.131 **	-0.163 ***	-0.166 ***	-0.179 ***	-0.174 ***	-0.195 ***
Service variables						
`scheduled_hw_min`		-0.001	0.000	0.002	0.002	0.002
`mode_bus`		0.068 **	0.063 *	0.047	0.053	0.055
`DT_with_delay`		-0.003	-0.003	0.003	0.003	0.004
`AT_with_delay`		-0.006	-0.006	-0.007	-0.007	-0.007
Time variables						
`peak_period`			0.001	0.004	0.004	0.003
`weekday`			0.012	0.012	0.012	0.011
Traffic variables						
`IC_ratio`				0.068 **	0.055 *	0.125 ***
Mitigating measure variables						
`measure_package`					0.017	0.010
Interaction terms						
`traffic_x_measure`						-0.099 *
Model information						
No. Observations	228,094	228,094	228,094	228,094	228,094	228,094
R-squared	0.090	0.188	0.189	0.228	0.234	0.249

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The infrastructure-only model explains 9.0% of the observed variation in CV_{TT}. `len` shows a negative and statistically significant association throughout the blockwise sequence. Since CV_{TT} is a relative variability measure, the negative coefficient indicates that longer segments are associated with lower relative travel time variability.

The addition of service variables in Model 2 increases R^2 from 0.090 to 0.188. Bus observations initially show higher travel time variability than tram observations, while the variables `scheduled_hw_min` and `DT_with_delay` do not show statistically significant associations across most blockwise specifications. The `mode_bus` coefficient becomes statistically insignificant after the traffic block is introduced, indicating that part of the initial mode association overlaps with differences in the operating context represented by later model blocks.

Temporal variables add very little explanatory power in Model 3, with R^2 increasing from 0.188 to 0.189. Neither `peak_period` nor `weekday` shows a statistically significant association in the blockwise CV_{TT} specifications.

The addition of local traffic exposure in Model 4 produces a larger increase in model fit, with R^2 rising from 0.189 to 0.228. The `IC_ratio` coefficient is positive and statistically significant when introduced and remains positive in all subsequent blockwise specifications. The magnitude of the coefficient increases after the interaction term is added in Model 6.

The `measure_package` variable is positive but not statistically significant when first introduced in Model 5. The full blockwise specification in Model 6 additionally includes `traffic_x_measure`, which is negative and weakly significant. The final blockwise model reaches an adjusted R^2 of 0.248. Overall, the blockwise sequence shows that the traffic-related terms account for a meaningful additional share of the variation in CV_{TT}, whereas the temporal variables contribute comparatively little additional explanatory power.

Backward elimination was subsequently applied to obtain a reduced CV_{TT} specification. Table E.4 presents the successive models.

Table E.4: Backward Elimination OLS regression results for Coefficient of Variation of Travel Time (CV_{TT}).

Variable	Model 7	Model 8	Model 9	Model 10	Model 11
Infrastructure variables					
`len`	-0.195 ***	-0.196 ***	-0.195 ***	-0.188 ***	-0.188 ***
Service variables					
`scheduled_hw_min`	0.002	0.002	0.002	✘	
`mode_bus`	0.056	0.056	0.057	0.066 **	0.066 **
`DT_with_delay`	✘				
`AT_with_delay`	-0.006	✘			
Time variables					
`peak_period`	0.003	0.003	✘		
`weekday`	0.011	0.011	0.011	0.005	✘
Traffic variables					
`IC_ratio`	0.124 ***	0.124 ***	0.124 ***	0.120 ***	0.121 ***
Mitigating measure variables					
`measure_package`	0.010	0.011	0.011	0.012	0.012
Interaction terms					
`traffic_x_measure`	-0.099 *	-0.099 *	-0.099 *	-0.100 *	-0.101 *
Model information					
No. Observations	228,094	228,094	228,094	228,094	228,094
R-squared	0.248	0.248	0.248	0.246	0.245

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The elimination procedure removes the `DT_with_delay` variable in Model 7, `AT_with_delay` in Model 8, `peak_period` in Model 9, `scheduled_hw_min` in Model 10, and `weekday` in Model 11. The final parsimonious specification therefore retains `len`, `mode_bus`, `IC_ratio`, `measure_package`, and `traffic_x_measure`.

The central coefficient patterns remain stable throughout the elimination sequence. `len` remains negatively associated with CV_{TT}, while `IC_ratio` remains positive and statistically significant. `mode_bus` becomes statistically significant in the reduced specification. The `measure_package` coefficient remains small and statistically insignificant in the final model ($\beta = 0.012$), while `traffic_x_measure` remains negative and weakly significant.

The adjusted R^2 of the final parsimonious model is 0.245, compared with 0.248 for the full blockwise specification. The eliminated variables therefore contribute only limited additional explanatory value.

The final CV_{TT} model corresponds to the specification presented in Section 5.3.

E.3. Headway regularity model development

This section presents the complete model-development sequence for the coefficient of variation of headway (CV_h). Six hierarchical blockwise models were estimated before applying backward elimination. Table E.5 presents the blockwise results.

Table E.5: Blockwise OLS regression results for Coefficient of Variation of Headway (CV_h).

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
Infrastructure variables						
`len`	-0.069	-0.045	-0.037	-0.038	-0.015	-0.010
Service variables						
`scheduled_tt_min`		-0.006	-0.010	-0.008	-0.018	-0.019 *
`mode_bus`		-0.039	-0.043	-0.042	-0.024	-0.024
`DT_with_delay`		0.106 ***	0.102 ***	0.101 ***	0.100 ***	0.100 ***
`AT_with_delay`		-0.063 ***	-0.063 ***	-0.063 ***	-0.057 ***	-0.057 ***
Time variables						
`peak_period`			0.025 ***	0.025 ***	0.026 ***	0.027 ***
`weekday`			0.042 **	0.045 ***	0.054 ***	0.054 ***
Traffic variables						
`IC_ratio`				-0.026	-0.071	-0.085
Mitigating measure variables						
`measure_package`					0.069 **	0.071 **
Interaction terms						
`traffic_x_measure`						0.020
Model information						
No. Observations	228,094	228,094	228,094	228,094	228,094	228,094
R-squared	0.012	0.076	0.098	0.101	0.144	0.144

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The infrastructure-only model provides little explanatory power, with an R^2 of 0.012. `len` is not statistically significant in any of the blockwise specifications, indicating that this fixed segment characteristic contributes relatively little to explaining headway variability in the selected sample.

The inclusion of service variables in Model 2 increases R^2 from 0.012 to 0.076. `DT_with_delay` shows a consistently positive and statistically significant association with CV_h, whereas `AT_with_delay` shows a consistently negative and statistically significant coefficient. `scheduled_tt_min` and `mode_bus` show weaker associations across the blockwise specifications.

The temporal block produces a further increase in explanatory power, with R^2 rising from 0.076 to 0.098. Both `peak_period` and `weekday` are positively and statistically significantly associated with CV_h, and these associations remain stable in the subsequent models.

Adding local traffic exposure in Model 4 results in only a small increase in R^2, from 0.098 to 0.101. The `IC_ratio` coefficient is negative and not statistically significant in the blockwise specifications. In contrast, the mitigating-measure block added in Model 5 produces a larger increase in explanatory power, with R^2 rising to 0.144. The `measure_package` coefficient is positive and statistically significant when introduced.

`traffic_x_measure` is added in Model 6. Its coefficient is positive but not statistically significant, while overall model fit remains essentially unchanged. The blockwise sequence therefore indicates that service and temporal variables contribute more strongly to explaining CV_h than local traffic exposure, while the interaction term provides little additional explanatory value.

Backward elimination was subsequently applied to obtain the final reduced specification. The successive models are presented in Table E.6.

Table E.6: Backward Elimination OLS regression results for Coefficient of Variation of Headway (CV_h).

Variable	Model 7	Model 8	Model 9
Infrastructure variables			
`len`	✘		
Service variables			
`scheduled_tt_min`	-0.020 *	-0.020 *	-0.026 **
`mode_bus`	-0.024	-0.024	✘
`DT_with_delay`	0.099 ***	0.100 ***	0.092 ***
`AT_with_delay`	-0.056 ***	-0.056 ***	-0.051 ***
Time variables			
`peak_period`	0.027 ***	0.027 ***	0.026 ***
`weekday`	0.054 ***	0.054 ***	0.054 ***
Traffic variables			
`IC_ratio`	-0.090	-0.072	-0.077 *
Mitigating measure variables			
`measure_package`	0.072 **	0.070 **	0.076 **
Interaction terms			
`traffic_x_measure`	0.026	✘	
Model information			
No. Observations	228,094	228,094	228,094
R-squared	0.144	0.143	0.138

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

The elimination procedure removes `len` in Model 7, `traffic_x_measure` in Model 8, and `mode_bus` in Model 9. The resulting final specification retains `scheduled_tt_min`, `DT_with_delay`, `AT_with_delay`, `peak_period`, `weekday`, `IC_ratio`, and `measure_package`.

The principal coefficient patterns remain stable during the reduction procedure. `DT_with_delay` remains positively associated with CV_h, while arrival delay remains negatively associated. `peak_period` and `weekday` continue to show positive and statistically significant associations. `scheduled_tt_min` becomes more statistically significant in the final specification.

The `IC_ratio` coefficient becomes weakly significant and remains negative in the final model, while `measure_package` remains positive and statistically significant. `traffic_x_measure` is not retained, as it does not contribute meaningfully to the reduced specification.

The adjusted R^2 of the final parsimonious model is 0.138, compared with 0.144 for the complete blockwise model. The reduction therefore results in only a limited loss of explanatory power while providing a simpler final specification. The resulting CV_h model is the specification presented in Section 5.3.

E.4. Multicollinearity during model development

Multicollinearity was assessed during model development using variance inflation factors (VIFs). VIFs were calculated for both the complete blockwise specifications and the final parsimonious models. The purpose of this assessment was to identify overlap between explanatory variables that could affect the stability and interpretation of individual coefficient estimates.

The complete blockwise models in Table E.7 show elevated VIF values for several closely related infrastructure and service variables. The highest values are primarily associated with segment length and scheduled service characteristics, indicating that these variables capture partly overlapping aspects of the segment and service context. Elevated VIF values do not invalidate the models, but they imply

that individual coefficients may be sensitive to the simultaneous inclusion of correlated predictors and should therefore be interpreted conditionally.

Table E.7: Variance inflation factors for the final blockwise model specifications.

(a) Variance inflation factors for the final blockwise OTP model.

Variable	VIF
scheduled_tt_min	12.254
len	10.144
scheduled_hw_min	9.563
IC_ratio	4.452
traffic_x_measure	3.998
weekday	3.913
mode_bus	3.748
measure_package	3.289
AT_with_delay	1.636
peak_period	1.514
DT_with_delay	1.300

(b) Variance inflation factors for the final blockwise CV_{TT} model.

Variable	VIF
scheduled_hw_min	9.517
len	6.899
IC_ratio	4.113
traffic_x_measure	3.650
weekday	3.645
mode_bus	3.348
measure_package	2.692
AT_with_delay	1.538
peak_period	1.497
DT_with_delay	1.281

(c) Variance inflation factors for the final blockwise CV_h model.

Variable	VIF
scheduled_tt_min	12.194
len	6.709
weekday	3.830
traffic_x_measure	3.768
IC_ratio	3.579
measure_package	2.659
mode_bus	2.197
AT_with_delay	1.630
peak_period	1.502
DT_with_delay	1.296

The VIF values are substantially lower in the final parsimonious specifications, as can be seen in Table E.8. In the OTP and CV_{TT} models, the retained predictors show relatively low VIF values. For CV_h, scheduled travel time retains a moderately elevated VIF, while the remaining values are lower. Overall, the final specifications show less overlap between explanatory variables than the complete blockwise models.

Table E.8: Variance inflation factors for the final parsimonious model specifications.

(a) Variance inflation factors for the final parsimonious OTP model.

Variable	VIF
traffic_x_measure	3.406
weekday	3.351
IC_ratio	3.281
len	3.255
measure_package	2.083
AT_with_delay	1.502
DT_with_delay	1.239

(b) Variance inflation factors for the final parsimonious CV_{TT} model.

Variable	VIF
traffic_x_measure	3.166
IC_ratio	3.085
len	2.704
measure_package	1.971
mode_bus	1.776

(c) Variance inflation factors for the final parsimonious CV_h model.

Variable	VIF
scheduled_tt_min	5.072
weekday	3.691
measure_package	2.388
AT_with_delay	1.546
peak_period	1.500
DT_with_delay	1.235
IC_ratio	1.111

Reducing multicollinearity was not the primary purpose of the backward-elimination procedure. The VIF assessment is instead used as an additional diagnostic during model development to evaluate whether severe linear dependence among retained predictors could hinder coefficient interpretation. The remaining correlations between explanatory variables should nevertheless be considered when interpreting individual associations.

[*This page intentionally left blank*]

Additional Outcome-Specific Model Diagnostics

This appendix provides supplementary diagnostics supporting the outcome-specific model specifications discussed in Section 5.4. The residual diagnostics motivating these specifications are presented in the main results chapter. The additional material below documents the distributional effects of the log transformation for the variability indicators and the fitted-value behaviour of the OTP models.

F.1. Distributional characteristics of the variability indicators

Both CV_{TT} and CV_h are non-negative variables and show right-skewed distributions. Table F.1 summarises their distributional characteristics in the final model sample.

Table F.1: Distributional summary of CV_{TT} and CV_h in the final model sample.

variable	n	mean	std	min	25%	median	75%	max	skewness
CV_{TT}	228,094	0.237	0.099	0.058	0.165	0.214	0.290	0.636	0.942
CV_h	228,094	0.357	0.144	0.031	0.250	0.358	0.440	1.000	0.524

The right-skew is more pronounced for CV_{TT} than for CV_h. Log transformations were therefore applied to both outcomes as outcome-specific alternative specifications. Figure F.1 and F.2 show how the transformations affect the respective outcome distributions.

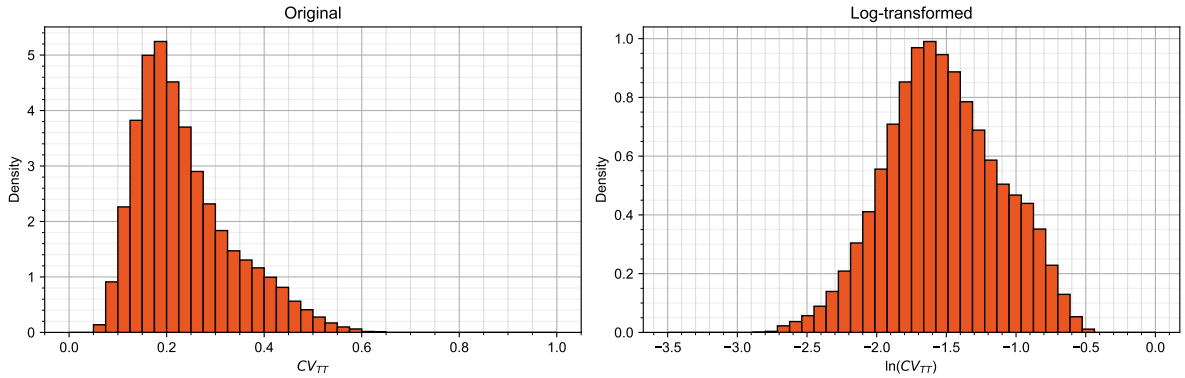


Figure F.1: Original and log-transformed distributions of CV_{TT}.

The log transformation compresses high CV_{TT} values and reduces the pronounced right-skew of the original distribution. This results in a more symmetric outcome distribution and reduces the relative influence of observations with particularly high travel time variability.

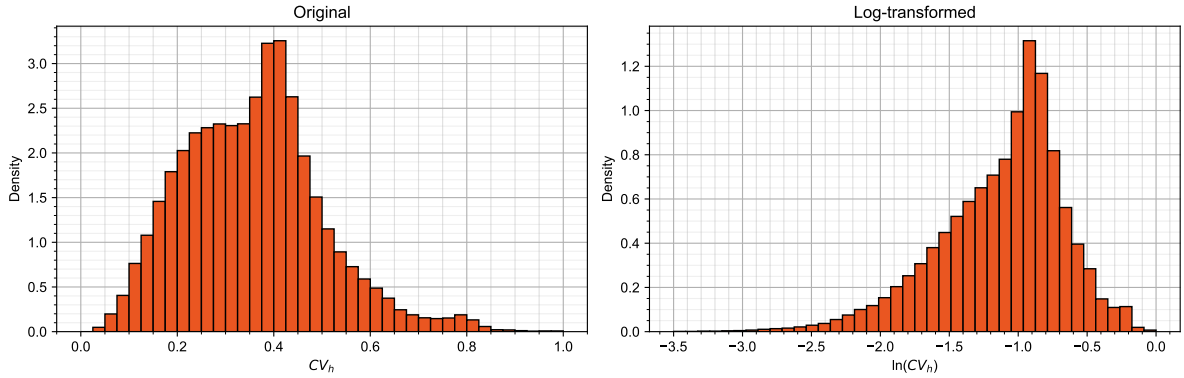


Figure F.2: Original and log-transformed distributions of CV_h.

The original CV_h distribution is less strongly skewed than CV_{TT}, although the log transformation also reduces its asymmetry and compresses high-value observations. The corresponding comparison between the OLS and log-transformed regression specifications is presented in Subsection 5.4.2.

F.2. Fitted-value diagnostics for on-time performance

Because OTP is bounded between 0 and 1, the fitted values of the final parsimonious OLS model were inspected as an additional diagnostic. This check assesses whether the linear specification produces predictions outside the feasible range of the dependent variable.

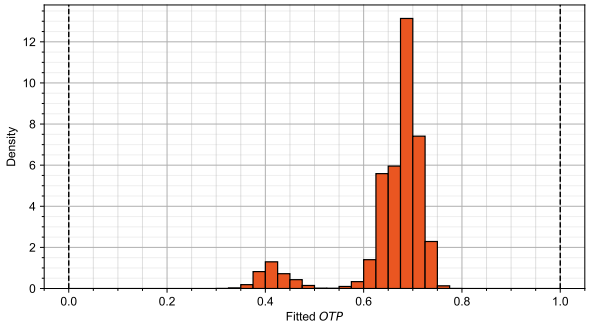


Figure F.3: Distribution of fitted values from the final parsimonious OLS model for OTP.

The fitted values of the OLS specification remain within the feasible $[0, 1]$ range for the observed sample. Table F.2 compares the fitted-value distributions of the OLS and fractional logit specifications.

Table F.2: Summary statistics of fitted values from the final OLS and fractional logit specifications for OTP.

model	n	mean	std	min	25%	median	75%	max
OLS fitted	228,094	0.655	0.082	0.304	0.642	0.679	0.699	0.783
Fractional logit fitted	228,094	0.655	0.082	0.300	0.643	0.679	0.699	0.774

The OLS and fractional logit models produce similar fitted-value distributions. The absence of out-of-range predictions in the OLS model does not remove the statistical limitation associated with modelling a bounded proportion using a linear specification, but it indicates that infeasible predictions are not a practical issue within the observed sample. The substantive comparison between the OLS and fractional logit specifications is presented in Subsection 5.4.3.

Full Regression Output

This appendix provides the complete statistical estimation output underlying the OLS models summarised in Appendix E and the outcome-specific specifications discussed in Section 5.4. The output includes coefficient estimates, cluster-robust standard errors, confidence intervals, model-fit statistics, and additional estimation information.

These results are included for full transparency and reproducibility. The development of the baseline OLS models is summarised in Appendix E, while the substantive interpretation and comparison of the final specifications are presented in Section 5.3 and 5.4.

G.1. Full on-time performance regression output

Model G.1: OTP OLS model 1, addition infrastructure variables.

OLS Regression Results

Dep. Variable:	otp	R-squared:	0.010
Model:	OLS	Adj. R-squared:	0.010
Method:	Least Squares	F-statistic:	5.336
No. Observations:	228094	Prob (F-statistic):	0.0367
Df Residuals:	228092	Log-Likelihood:	74883.
Df Model:	1	AIC:	-1.498e+05
Covariance Type:	cluster	BIC:	-1.497e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.6938	0.023	29.792	0.000	0.648	0.739
len	-0.0792	0.034	-2.310	0.021	-0.146	-0.012

Omnibus:	3450.697	Durbin-Watson:	1.672
Prob(Omnibus):	0.000	Jarque-Bera (JB):	3609.437
Skew:	-0.308	Prob(JB):	0.00
Kurtosis:	3.027	Cond. No.	5.51

*Source: Table E.1.

Model G.2: OTP OLS model 2, addition service variables.

OLS Regression Results

Dep. Variable:	otp	R-squared:	0.231
Model:	OLS	Adj. R-squared:	0.231
Method:	Least Squares	F-statistic:	52.84
No. Observations:	228094	Prob (F-statistic):	8.08e-09
Df Residuals:	228087	Log-Likelihood:	1.0360e+05
Df Model:	6	AIC:	-2.072e+05
Covariance Type:	cluster	BIC:	-2.071e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.6657	0.042	15.711	0.000	0.583	0.749
len	-0.1071	0.060	-1.776	0.076	-0.225	0.011
scheduled_tt_min	0.0280	0.030	0.923	0.356	-0.031	0.087
scheduled_hw_min	0.0056	0.001	3.960	0.000	0.003	0.008
mode_bus	-0.0789	0.031	-2.508	0.012	-0.141	-0.017

DT_with_delay	-0.2107	0.017	-12.755	0.000	-0.243	-0.178
AT_with_delay	-0.0636	0.015	-4.188	0.000	-0.093	-0.034

Omnibus:	1034.244	Durbin-Watson:	1.735
Prob(Omnibus):	0.000	Jarque-Bera (JB):	975.871
Skew:	-0.133	Prob(JB):	1.24e-212
Kurtosis:	2.821	Cond. No.	55.0

*Source: Table E.1.

Model G.3: OTP OLS model 3, addition time variables.

OLS Regression Results

Dep. Variable:	otp	R-squared:	0.232
Model:	OLS	Adj. R-squared:	0.232
Method:	Least Squares	F-statistic:	65.22
No. Observations:	228094	Prob (F-statistic):	9.31e-10
Df Residuals:	228085	Log-Likelihood:	1.0385e+05
Df Model:	8	AIC:	-2.077e+05
Covariance Type:	cluster	BIC:	-2.076e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.6801	0.043	15.902	0.000	0.596	0.764
len	-0.1025	0.060	-1.706	0.088	-0.220	0.015
scheduled_tt_min	0.0276	0.030	0.910	0.363	-0.032	0.087
scheduled_hw_min	0.0047	0.002	2.833	0.005	0.001	0.008
mode_bus	-0.0733	0.031	-2.347	0.019	-0.135	-0.012
DT_with_delay	-0.2110	0.017	-12.702	0.000	-0.244	-0.178
AT_with_delay	-0.0632	0.015	-4.166	0.000	-0.093	-0.033
peak_period	0.0074	0.004	1.824	0.068	-0.001	0.015
weekday	-0.0170	0.009	-1.905	0.057	-0.034	0.000

Omnibus:	1067.699	Durbin-Watson:	1.736
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1014.783
Skew:	-0.138	Prob(JB):	4.39e-221
Kurtosis:	2.825	Cond. No.	61.5

*Source: Table E.1.

Model G.4: OTP OLS model 4, addition traffic variables.

OLS Regression Results

Dep. Variable:	otp	R-squared:	0.235
Model:	OLS	Adj. R-squared:	0.235
Method:	Least Squares	F-statistic:	108.1
No. Observations:	228094	Prob (F-statistic):	2.25e-11
Df Residuals:	228084	Log-Likelihood:	1.0420e+05
Df Model:	9	AIC:	-2.084e+05
Covariance Type:	cluster	BIC:	-2.083e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.6797	0.039	17.237	0.000	0.602	0.757
len	-0.1060	0.060	-1.760	0.078	-0.224	0.012
scheduled_tt_min	0.0261	0.031	0.837	0.403	-0.035	0.087
scheduled_hw_min	0.0055	0.002	3.162	0.002	0.002	0.009
mode_bus	-0.0791	0.031	-2.542	0.011	-0.140	-0.018
DT_with_delay	-0.2090	0.016	-12.765	0.000	-0.241	-0.177
AT_with_delay	-0.0635	0.015	-4.346	0.000	-0.092	-0.035
peak_period	0.0088	0.003	2.654	0.008	0.002	0.015
weekday	-0.0171	0.008	-2.012	0.044	-0.034	-0.000
IC_ratio	0.0299	0.026	1.138	0.255	-0.022	0.081

Omnibus:	895.017	Durbin-Watson:	1.750
Prob(Omnibus):	0.000	Jarque-Bera (JB):	852.907
Skew:	-0.126	Prob(JB):	6.22e-186
Kurtosis:	2.838	Cond. No.	61.9

*Source: Table E.1.

Model G.5: OTP OLS model 5, addition mitigating measure variables.

OLS Regression Results

Dep. Variable:	otp	R-squared:	0.244
Model:	OLS	Adj. R-squared:	0.244
Method:	Least Squares	F-statistic:	147.5
No. Observations:	228094	Prob (F-statistic):	2.07e-12
Df Residuals:	228083	Log-Likelihood:	1.0552e+05
Df Model:	10	AIC:	-2.110e+05
Covariance Type:	cluster	BIC:	-2.109e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.6914	0.039	17.608	0.000	0.614	0.768
len	-0.1270	0.053	-2.376	0.018	-0.232	-0.022
scheduled_tt_min	0.0336	0.028	1.182	0.237	-0.022	0.089
scheduled_hw_min	0.0069	0.002	4.190	0.000	0.004	0.010
mode_bus	-0.0982	0.024	-4.072	0.000	-0.146	-0.051
DT_with_delay	-0.2075	0.016	-12.640	0.000	-0.240	-0.175
AT_with_delay	-0.0669	0.014	-4.792	0.000	-0.094	-0.040
peak_period	0.0096	0.003	3.128	0.002	0.004	0.016
weekday	-0.0175	0.007	-2.457	0.014	-0.031	-0.004
IC_ratio	0.0586	0.024	2.458	0.014	0.012	0.105
measure_package	-0.0392	0.017	-2.294	0.022	-0.073	-0.006

Omnibus:	1144.533	Durbin-Watson:	1.766
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1107.176
Skew:	-0.150	Prob(JB):	3.80e-241
Kurtosis:	2.838	Cond. No.	63.2

*Source: Table E.1.

Model G.6: OTP OLS model 6, addition interaction terms.

OLS Regression Results

Dep. Variable:	otp	R-squared:	0.247
Model:	OLS	Adj. R-squared:	0.247
Method:	Least Squares	F-statistic:	183.9
No. Observations:	228094	Prob (F-statistic):	3.65e-13
Df Residuals:	228082	Log-Likelihood:	1.0605e+05
Df Model:	11	AIC:	-2.121e+05
Covariance Type:	cluster	BIC:	-2.120e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.7060	0.033	21.491	0.000	0.642	0.770
len	-0.1495	0.060	-2.477	0.013	-0.268	-0.031
scheduled_tt_min	0.0364	0.028	1.292	0.196	-0.019	0.092
scheduled_hw_min	0.0070	0.001	4.741	0.000	0.004	0.010
mode_bus	-0.0992	0.023	-4.268	0.000	-0.145	-0.054
DT_with_delay	-0.2060	0.016	-12.523	0.000	-0.238	-0.174
AT_with_delay	-0.0675	0.015	-4.638	0.000	-0.096	-0.039
peak_period	0.0092	0.003	2.707	0.007	0.003	0.016
weekday	-0.0182	0.007	-2.597	0.009	-0.032	-0.004
IC_ratio	0.1206	0.043	2.775	0.006	0.035	0.206
measure_package	-0.0464	0.018	-2.536	0.011	-0.082	-0.011
traffic_x_measure	-0.0873	0.055	-1.597	0.110	-0.194	0.020

Omnibus:	1134.549	Durbin-Watson:	1.769
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1082.175
Skew:	-0.144	Prob(JB):	1.02e-235
Kurtosis:	2.824	Cond. No.	112.

*Source: Table E.1.

Model G.7: OTP OLS model 7, elimination scheduled_tt_min.

OLS Regression Results

Dep. Variable:	otp	R-squared:	0.232
Model:	OLS	Adj. R-squared:	0.232
Method:	Least Squares	F-statistic:	220.6
No. Observations:	228094	Prob (F-statistic):	1.28e-13

Df Residuals:	228083	Log-Likelihood:				1.0381e+05
Df Model:	10	AIC:				-2.076e+05
Covariance Type:	cluster	BIC:				-2.075e+05
	coef	std err	z	P>|z|	[0.025	0.975]
const	0.7504	0.022	34.341	0.000	0.708	0.793
len	-0.0874	0.033	-2.617	0.009	-0.153	-0.022
scheduled_hw_min	0.0046	0.003	1.463	0.143	-0.002	0.011
mode_bus	-0.0658	0.048	-1.364	0.173	-0.160	0.029
DT_with_delay	-0.2168	0.019	-11.155	0.000	-0.255	-0.179
AT_with_delay	-0.0569	0.013	-4.527	0.000	-0.082	-0.032
peak_period	0.0091	0.003	2.615	0.009	0.002	0.016
weekday	-0.0204	0.009	-2.312	0.021	-0.038	-0.003
IC_ratio	0.0982	0.038	2.588	0.010	0.024	0.173
measure_package	-0.0303	0.015	-2.036	0.042	-0.059	-0.001
traffic_x_measure	-0.0592	0.041	-1.461	0.144	-0.139	0.020
Omnibus:	561.856	Durbin-Watson:				1.818
Prob(Omnibus):	0.000	Jarque-Bera (JB):				546.844
Skew:	-0.104	Prob(JB):				1.80e-119
Kurtosis:	2.882	Cond. No.				108.

*Source: Table E.2.

Model G.8: OTP OLS model 8, elimination mode_bus.

Dep. Variable:	otp	R-squared:				0.218
Model:	OLS	Adj. R-squared:				0.218
Method:	Least Squares	F-statistic:				339.8
No. Observations:	228094	Prob (F-statistic):				8.16e-15
Df Residuals:	228084	Log-Likelihood:				1.0170e+05
Df Model:	9	AIC:				-2.034e+05
Covariance Type:	cluster	BIC:				-2.033e+05
	coef	std err	z	P>|z|	[0.025	0.975]
const	0.7837	0.017	45.748	0.000	0.750	0.817
len	-0.0786	0.042	-1.878	0.060	-0.161	0.003
scheduled_hw_min	-0.0009	0.002	-0.392	0.695	-0.006	0.004
DT_with_delay	-0.2289	0.019	-12.093	0.000	-0.266	-0.192
AT_with_delay	-0.0529	0.013	-4.082	0.000	-0.078	-0.028
peak_period	7.128e-05	0.008	0.009	0.993	-0.016	0.016
weekday	-0.0414	0.013	-3.220	0.001	-0.067	-0.016
IC_ratio	0.0759	0.038	1.976	0.048	0.001	0.151
measure_package	-0.0191	0.017	-1.149	0.251	-0.052	0.013
traffic_x_measure	-0.0676	0.040	-1.671	0.095	-0.147	0.012
Omnibus:	1123.602	Durbin-Watson:				1.763
Prob(Omnibus):	0.000	Jarque-Bera (JB):				1073.843
Skew:	-0.144	Prob(JB):				6.58e-234
Kurtosis:	2.826	Cond. No.				108.

*Source: Table E.2.

Model G.9: OTP OLS model 9, elimination peak_period.

Dep. Variable:	otp	R-squared:				0.218
Model:	OLS	Adj. R-squared:				0.218
Method:	Least Squares	F-statistic:				283.8
No. Observations:	228094	Prob (F-statistic):				3.83e-14
Df Residuals:	228085	Log-Likelihood:				1.0170e+05
Df Model:	8	AIC:				-2.034e+05
Covariance Type:	cluster	BIC:				-2.033e+05
	coef	std err	z	P>|z|	[0.025	0.975]
const	0.7837	0.016	47.847	0.000	0.752	0.816
len	-0.0786	0.042	-1.884	0.060	-0.160	0.003
scheduled_hw_min	-0.0009	0.002	-0.402	0.687	-0.005	0.004
DT_with_delay	-0.2289	0.019	-12.022	0.000	-0.266	-0.192

AT_with_delay	-0.0529	0.013	-4.070	0.000	-0.078	-0.027
weekday	-0.0414	0.013	-3.176	0.001	-0.067	-0.016
IC_ratio	0.0759	0.038	1.976	0.048	0.001	0.151
measure_package	-0.0191	0.017	-1.144	0.253	-0.052	0.014
traffic_x_measure	-0.0676	0.040	-1.671	0.095	-0.147	0.012

Omnibus:	1123.397	Durbin-Watson:	1.763
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1073.634
Skew:	-0.144	Prob(JB):	7.30e-234
Kurtosis:	2.826	Cond. No.	108.

*Source: Table E.2.

Model G.10: OTP OLS model 10, elimination scheduled_hw_min.

OLS Regression Results

Dep. Variable:	otp	R-squared:	0.217
Model:	OLS	Adj. R-squared:	0.217
Method:	Least Squares	F-statistic:	265.2
No. Observations:	228094	Prob (F-statistic):	8.75e-14
Df Residuals:	228086	Log-Likelihood:	1.0164e+05
Df Model:	7	AIC:	-2.033e+05
Covariance Type:	cluster	BIC:	-2.032e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.7763	0.017	44.803	0.000	0.742	0.810
len	-0.0854	0.029	-2.952	0.003	-0.142	-0.029
DT_with_delay	-0.2300	0.019	-11.822	0.000	-0.268	-0.192
AT_with_delay	-0.0517	0.014	-3.693	0.000	-0.079	-0.024
weekday	-0.0386	0.009	-4.489	0.000	-0.055	-0.022
IC_ratio	0.0776	0.034	2.278	0.023	0.011	0.144
measure_package	-0.0190	0.017	-1.132	0.258	-0.052	0.014
traffic_x_measure	-0.0681	0.037	-1.823	0.068	-0.141	0.005

Omnibus:	1174.541	Durbin-Watson:	1.763
Prob(Omnibus):	0.000	Jarque-Bera (JB):	1124.611
Skew:	-0.148	Prob(JB):	6.22e-245
Kurtosis:	2.825	Cond. No.	16.5

*Source: Table E.2.

G.2. Full travel time variability regression output

Model G.11: CV_{TT} OLS model 1, addition infrastructure variables.

OLS Regression Results

Dep. Variable:	cvtt	R-squared:	0.090
Model:	OLS	Adj. R-squared:	0.090
Method:	Least Squares	F-statistic:	6.377
No. Observations:	228094	Prob (F-statistic):	0.0243
Df Residuals:	228092	Log-Likelihood:	2.1386e+05
Df Model:	1	AIC:	-4.277e+05
Covariance Type:	cluster	BIC:	-4.277e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3014	0.038	7.968	0.000	0.227	0.375
len	-0.1314	0.052	-2.525	0.012	-0.233	-0.029

Omnibus:	21468.847	Durbin-Watson:	1.039
Prob(Omnibus):	0.000	Jarque-Bera (JB):	28016.898
Skew:	0.835	Prob(JB):	0.00
Kurtosis:	3.400	Cond. No.	5.51

*Source: Table E.3.

Model G.12: CV_{TT} OLS model 2, addition service variables.

OLS Regression Results

Dep. Variable:	cvtt	R-squared:	0.188

Model:	OLS	Adj. R-squared:	0.187
Method:	Least Squares	F-statistic:	4.239
No. Observations:	228094	Prob (F-statistic):	0.0148
Df Residuals:	228088	Log-Likelihood:	2.2681e+05
Df Model:	5	AIC:	-4.536e+05
Covariance Type:	cluster	BIC:	-4.536e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2978	0.035	8.488	0.000	0.229	0.367
len	-0.1628	0.048	-3.378	0.001	-0.257	-0.068
scheduled_hw_min	-0.0006	0.002	-0.366	0.714	-0.004	0.003
mode_bus	0.0684	0.033	2.045	0.041	0.003	0.134
DT_with_delay	-0.0029	0.010	-0.297	0.766	-0.022	0.016
AT_with_delay	-0.0055	0.009	-0.615	0.539	-0.023	0.012

Omnibus:	18502.333	Durbin-Watson:	1.214
Prob(Omnibus):	0.000	Jarque-Bera (JB):	23672.221
Skew:	0.734	Prob(JB):	0.00
Kurtosis:	3.581	Cond. No.	50.3

*Source: Table E.3.

Model G.13: CV_{TT} OLS model 3, addition time variables.

OLS Regression Results

Dep. Variable:	cvtt	R-squared:	0.189
Model:	OLS	Adj. R-squared:	0.189
Method:	Least Squares	F-statistic:	14.98
No. Observations:	228094	Prob (F-statistic):	1.62e-05
Df Residuals:	228086	Log-Likelihood:	2.2707e+05
Df Model:	7	AIC:	-4.541e+05
Covariance Type:	cluster	BIC:	-4.540e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2853	0.043	6.659	0.000	0.201	0.369
len	-0.1659	0.047	-3.538	0.000	-0.258	-0.074
scheduled_hw_min	0.0001	0.002	0.063	0.950	-0.004	0.004
mode_bus	0.0635	0.038	1.672	0.095	-0.011	0.138
DT_with_delay	-0.0030	0.010	-0.309	0.757	-0.022	0.016
AT_with_delay	-0.0056	0.009	-0.617	0.538	-0.023	0.012
peak_period	0.0007	0.004	0.149	0.882	-0.008	0.009
weekday	0.0117	0.010	1.217	0.224	-0.007	0.031

Omnibus:	18743.858	Durbin-Watson:	1.211
Prob(Omnibus):	0.000	Jarque-Bera (JB):	24058.412
Skew:	0.739	Prob(JB):	0.00
Kurtosis:	3.589	Cond. No.	52.4

*Source: Table E.3.

Model G.14: CV_{TT} OLS model 4, addition traffic variables.

OLS Regression Results

Dep. Variable:	cvtt	R-squared:	0.228
Model:	OLS	Adj. R-squared:	0.228
Method:	Least Squares	F-statistic:	24.32
No. Observations:	228094	Prob (F-statistic):	6.14e-07
Df Residuals:	228085	Log-Likelihood:	2.3267e+05
Df Model:	8	AIC:	-4.653e+05
Covariance Type:	cluster	BIC:	-4.652e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2787	0.034	8.137	0.000	0.212	0.346
len	-0.1790	0.044	-4.089	0.000	-0.265	-0.093
scheduled_hw_min	0.0022	0.002	1.110	0.267	-0.002	0.006
mode_bus	0.0474	0.033	1.446	0.148	-0.017	0.112
DT_with_delay	0.0027	0.007	0.374	0.709	-0.011	0.017
AT_with_delay	-0.0074	0.009	-0.822	0.411	-0.025	0.010
peak_period	0.0039	0.004	1.101	0.271	-0.003	0.011
weekday	0.0117	0.009	1.324	0.185	-0.006	0.029

IC_ratio	0.0681	0.032	2.107	0.035	0.005	0.131

Omnibus:	20109.861	Durbin-Watson:	1.256
Prob(Omnibus):	0.000	Jarque-Bera (JB):	26516.850
Skew:	0.760	Prob(JB):	0.00
Kurtosis:	3.691	Cond. No.	52.4

*Source: Table E.3.

Model G.15: CV_{TT} OLS model 5, addition mitigating measure variables.

OLS Regression Results

Dep. Variable:	cvtt	R-squared:	0.234
Model:	OLS	Adj. R-squared:	0.234
Method:	Least Squares	F-statistic:	26.09
No. Observations:	228094	Prob (F-statistic):	3.04e-07
Df Residuals:	228084	Log-Likelihood:	2.3352e+05
Df Model:	9	AIC:	-4.670e+05
Covariance Type:	cluster	BIC:	-4.669e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2682	0.033	8.180	0.000	0.204	0.332
len	-0.1745	0.042	-4.183	0.000	-0.256	-0.093
scheduled_hw_min	0.0017	0.002	0.933	0.351	-0.002	0.005
mode_bus	0.0532	0.037	1.433	0.152	-0.020	0.126
DT_with_delay	0.0030	0.007	0.404	0.686	-0.012	0.018
AT_with_delay	-0.0068	0.009	-0.769	0.442	-0.024	0.011
peak_period	0.0035	0.004	0.935	0.350	-0.004	0.011
weekday	0.0121	0.008	1.444	0.149	-0.004	0.029
IC_ratio	0.0547	0.033	1.666	0.096	-0.010	0.119
measure_package	0.0173	0.028	0.621	0.535	-0.037	0.072

Omnibus:	20780.446	Durbin-Watson:	1.273
Prob(Omnibus):	0.000	Jarque-Bera (JB):	27809.325
Skew:	0.770	Prob(JB):	0.00
Kurtosis:	3.745	Cond. No.	54.8

*Source: Table E.3.

Model G.16: CV_{TT} OLS model 6, addition interaction terms.

OLS Regression Results

Dep. Variable:	cvtt	R-squared:	0.249
Model:	OLS	Adj. R-squared:	0.248
Method:	Least Squares	F-statistic:	11.01
No. Observations:	228094	Prob (F-statistic):	5.07e-05
Df Residuals:	228083	Log-Likelihood:	2.3571e+05
Df Model:	10	AIC:	-4.714e+05
Covariance Type:	cluster	BIC:	-4.713e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2890	0.034	8.412	0.000	0.222	0.356
len	-0.1951	0.045	-4.305	0.000	-0.284	-0.106
scheduled_hw_min	0.0016	0.002	0.853	0.394	-0.002	0.005
mode_bus	0.0550	0.036	1.539	0.124	-0.015	0.125
DT_with_delay	0.0038	0.007	0.516	0.606	-0.011	0.018
AT_with_delay	-0.0066	0.008	-0.793	0.428	-0.023	0.010
peak_period	0.0030	0.004	0.759	0.448	-0.005	0.011
weekday	0.0111	0.009	1.274	0.203	-0.006	0.028
IC_ratio	0.1248	0.045	2.792	0.005	0.037	0.212
measure_package	0.0104	0.026	0.394	0.693	-0.041	0.062
traffic_x_measure	-0.0990	0.051	-1.929	0.054	-0.200	0.002

Omnibus:	20409.548	Durbin-Watson:	1.303
Prob(Omnibus):	0.000	Jarque-Bera (JB):	27222.078
Skew:	0.761	Prob(JB):	0.00
Kurtosis:	3.742	Cond. No.	108.

*Source: Table E.3.

Model G.17: CV_{TT} OLS model 7, elimination DT_with_delay.

OLS Regression Results			
Dep. Variable:	cvtt	R-squared:	0.248
Model:	OLS	Adj. R-squared:	0.248
Method:	Least Squares	F-statistic:	10.42
No. Observations:	228094	Prob (F-statistic):	8.33e-05
Df Residuals:	228084	Log-Likelihood:	2.3570e+05
Df Model:	9	AIC:	-4.714e+05
Covariance Type:	cluster	BIC:	-4.713e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2890	0.034	8.433	0.000	0.222	0.356
len	-0.1949	0.045	-4.302	0.000	-0.284	-0.106
scheduled_hw_min	0.0016	0.002	0.849	0.396	-0.002	0.005
mode_bus	0.0555	0.035	1.577	0.115	-0.013	0.124
AT_with_delay	-0.0057	0.009	-0.666	0.505	-0.023	0.011
peak_period	0.0030	0.004	0.760	0.447	-0.005	0.011
weekday	0.0111	0.009	1.275	0.202	-0.006	0.028
IC_ratio	0.1244	0.045	2.782	0.005	0.037	0.212
measure_package	0.0104	0.026	0.394	0.693	-0.041	0.062
traffic_x_measure	-0.0989	0.051	-1.930	0.054	-0.199	0.002

Omnibus:	20501.860	Durbin-Watson:	1.304
Prob(Omnibus):	0.000	Jarque-Bera (JB):	27393.913
Skew:	0.762	Prob(JB):	0.00
Kurtosis:	3.748	Cond. No.	108.

*Source: Table E.4.

Model G.18: CV_{TT} OLS model 8, elimination AT_with_delay.

OLS Regression Results			
Dep. Variable:	cvtt	R-squared:	0.248
Model:	OLS	Adj. R-squared:	0.248
Method:	Least Squares	F-statistic:	10.61
No. Observations:	228094	Prob (F-statistic):	9.22e-05
Df Residuals:	228085	Log-Likelihood:	2.3560e+05
Df Model:	8	AIC:	-4.712e+05
Covariance Type:	cluster	BIC:	-4.711e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2870	0.032	8.853	0.000	0.223	0.351
len	-0.1955	0.045	-4.305	0.000	-0.285	-0.107
scheduled_hw_min	0.0017	0.002	0.892	0.372	-0.002	0.005
mode_bus	0.0557	0.036	1.569	0.117	-0.014	0.125
peak_period	0.0031	0.004	0.794	0.427	-0.005	0.011
weekday	0.0111	0.009	1.263	0.207	-0.006	0.028
IC_ratio	0.1243	0.044	2.825	0.005	0.038	0.211
measure_package	0.0106	0.026	0.405	0.686	-0.041	0.062
traffic_x_measure	-0.0992	0.051	-1.955	0.051	-0.199	0.000

Omnibus:	20531.105	Durbin-Watson:	1.305
Prob(Omnibus):	0.000	Jarque-Bera (JB):	27442.373
Skew:	0.763	Prob(JB):	0.00
Kurtosis:	3.748	Cond. No.	108.

*Source: Table E.4.

Model G.19: CV_{TT} OLS model 9, elimination peak_period.

OLS Regression Results			
Dep. Variable:	cvtt	R-squared:	0.248
Model:	OLS	Adj. R-squared:	0.248
Method:	Least Squares	F-statistic:	12.15
No. Observations:	228094	Prob (F-statistic):	5.48e-05
Df Residuals:	228086	Log-Likelihood:	2.3557e+05
Df Model:	7	AIC:	-4.711e+05
Covariance Type:	cluster	BIC:	-4.710e+05

	coef	std err	z	P>|z|	[0.025	0.975]

const	0.2888	0.031	9.189	0.000	0.227	0.350
len	-0.1953	0.046	-4.289	0.000	-0.285	-0.106
scheduled_hw_min	0.0016	0.002	0.866	0.386	-0.002	0.005
mode_bus	0.0566	0.035	1.638	0.101	-0.011	0.124
weekday	0.0109	0.009	1.267	0.205	-0.006	0.028
IC_ratio	0.1241	0.044	2.812	0.005	0.038	0.211
measure_package	0.0106	0.026	0.408	0.683	-0.040	0.062
traffic_x_measure	-0.0995	0.051	-1.958	0.050	-0.199	7.62e-05

Omnibus:	20559.740	Durbin-Watson:	1.306
Prob(Omnibus):	0.000	Jarque-Bera (JB):	27491.648
Skew:	0.763	Prob(JB):	0.00
Kurtosis:	3.749	Cond. No.	108.

*Source: Table E.4.

Model G.20: CV_{TT} OLS model 10, elimination scheduled_hw_min.

OLS Regression Results

Dep. Variable:	cvtt	R-squared:	0.246
Model:	OLS	Adj. R-squared:	0.246
Method:	Least Squares	F-statistic:	7.969
No. Observations:	228094	Prob (F-statistic):	0.000708
Df Residuals:	228087	Log-Likelihood:	2.3529e+05
Df Model:	6	AIC:	-4.706e+05
Covariance Type:	cluster	BIC:	-4.705e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.2991	0.030	10.024	0.000	0.241	0.358
len	-0.1881	0.047	-4.045	0.000	-0.279	-0.097
mode_bus	0.0655	0.031	2.107	0.035	0.005	0.127
weekday	0.0054	0.007	0.821	0.412	-0.008	0.018
IC_ratio	0.1195	0.046	2.627	0.009	0.030	0.209
measure_package	0.0121	0.026	0.473	0.636	-0.038	0.062
traffic_x_measure	-0.1002	0.052	-1.927	0.054	-0.202	0.002

Omnibus:	20357.622	Durbin-Watson:	1.310
Prob(Omnibus):	0.000	Jarque-Bera (JB):	27087.783
Skew:	0.761	Prob(JB):	0.00
Kurtosis:	3.731	Cond. No.	16.8

*Source: Table E.4.

Model G.21: CV_{TT} OLS model 11, elimination weekday.

OLS Regression Results

Dep. Variable:	cvtt	R-squared:	0.245
Model:	OLS	Adj. R-squared:	0.245
Method:	Least Squares	F-statistic:	3.948
No. Observations:	228094	Prob (F-statistic):	0.0192
Df Residuals:	228088	Log-Likelihood:	2.3522e+05
Df Model:	5	AIC:	-4.704e+05
Covariance Type:	cluster	BIC:	-4.704e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3037	0.029	10.360	0.000	0.246	0.361
len	-0.1884	0.046	-4.088	0.000	-0.279	-0.098
mode_bus	0.0659	0.031	2.143	0.032	0.006	0.126
IC_ratio	0.1213	0.045	2.694	0.007	0.033	0.210
measure_package	0.0115	0.026	0.449	0.653	-0.039	0.062
traffic_x_measure	-0.1006	0.052	-1.931	0.053	-0.203	0.002

Omnibus:	20283.582	Durbin-Watson:	1.310
Prob(Omnibus):	0.000	Jarque-Bera (JB):	26949.756
Skew:	0.760	Prob(JB):	0.00
Kurtosis:	3.726	Cond. No.	14.3

*Source: Table E.4.

G.3. Full headway regularity regression output

Model G.22: CV_h OLS model 1, addition infrastructure variables.

OLS Regression Results

Dep. Variable:	cvh	R-squared:	0.012
Model:	OLS	Adj. R-squared:	0.012
Method:	Least Squares	F-statistic:	0.8942
No. Observations:	228094	Prob (F-statistic):	0.360
Df Residuals:	228092	Log-Likelihood:	1.1961e+05
Df Model:	1	AIC:	-2.392e+05
Covariance Type:	cluster	BIC:	-2.392e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3906	0.038	10.318	0.000	0.316	0.465
len	-0.0686	0.073	-0.946	0.344	-0.211	0.074

Omnibus:	12359.219	Durbin-Watson:	1.061
Prob(Omnibus):	0.000	Jarque-Bera (JB):	14955.395
Skew:	0.566	Prob(JB):	0.00
Kurtosis:	3.542	Cond. No.	5.51

*Source: Table E.5.

Model G.23: CV_h OLS model 2, addition service variables.

OLS Regression Results

Dep. Variable:	cvh	R-squared:	0.076
Model:	OLS	Adj. R-squared:	0.076
Method:	Least Squares	F-statistic:	37.15
No. Observations:	228094	Prob (F-statistic):	1.35e-07
Df Residuals:	228088	Log-Likelihood:	1.2731e+05
Df Model:	5	AIC:	-2.546e+05
Covariance Type:	cluster	BIC:	-2.545e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.4134	0.037	11.283	0.000	0.342	0.485
len	-0.0451	0.069	-0.653	0.514	-0.180	0.090
scheduled_tt_min	-0.0057	0.011	-0.519	0.604	-0.027	0.016
mode_bus	-0.0392	0.030	-1.328	0.184	-0.097	0.019
DT_with_delay	0.1058	0.010	10.987	0.000	0.087	0.125
AT_with_delay	-0.0628	0.015	-4.139	0.000	-0.093	-0.033

Omnibus:	10920.560	Durbin-Watson:	1.082
Prob(Omnibus):	0.000	Jarque-Bera (JB):	13679.017
Skew:	0.500	Prob(JB):	0.00
Kurtosis:	3.662	Cond. No.	13.0

*Source: Table E.5.

Model G.24: CV_h OLS model 3, addition time variables.

OLS Regression Results

Dep. Variable:	cvh	R-squared:	0.098
Model:	OLS	Adj. R-squared:	0.098
Method:	Least Squares	F-statistic:	51.07
No. Observations:	228094	Prob (F-statistic):	6.72e-09
Df Residuals:	228086	Log-Likelihood:	1.3010e+05
Df Model:	7	AIC:	-2.602e+05
Covariance Type:	cluster	BIC:	-2.601e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3786	0.044	8.616	0.000	0.292	0.465
len	-0.0371	0.071	-0.524	0.600	-0.176	0.101
scheduled_tt_min	-0.0097	0.012	-0.795	0.426	-0.034	0.014
mode_bus	-0.0428	0.028	-1.522	0.128	-0.098	0.012
DT_with_delay	0.1022	0.009	11.926	0.000	0.085	0.119
AT_with_delay	-0.0631	0.015	-4.239	0.000	-0.092	-0.034
peak_period	0.0251	0.008	3.260	0.001	0.010	0.040

weekday	0.0421	0.017	2.420	0.016	0.008	0.076

Omnibus:	16532.316	Durbin-Watson:	1.112
Prob(Omnibus):	0.000	Jarque-Bera (JB):	23884.408
Skew:	0.608	Prob(JB):	0.00
Kurtosis:	4.018	Cond. No.	14.0

*Source: Table E.5.

Model G.25: CV_h OLS model 4, addition traffic variables.

OLS Regression Results

Dep. Variable:	cvh	R-squared:	0.101
Model:	OLS	Adj. R-squared:	0.101
Method:	Least Squares	F-statistic:	58.64
No. Observations:	228094	Prob (F-statistic):	1.91e-09
Df Residuals:	228085	Log-Likelihood:	1.3046e+05
Df Model:	8	AIC:	-2.609e+05
Covariance Type:	cluster	BIC:	-2.608e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3724	0.044	8.511	0.000	0.287	0.458
len	-0.0380	0.071	-0.533	0.594	-0.178	0.102
scheduled_tt_min	-0.0076	0.014	-0.559	0.576	-0.034	0.019
mode_bus	-0.0420	0.027	-1.546	0.122	-0.095	0.011
DT_with_delay	0.1009	0.010	10.336	0.000	0.082	0.120
AT_with_delay	-0.0627	0.015	-4.226	0.000	-0.092	-0.034
peak_period	0.0247	0.008	3.143	0.002	0.009	0.040
weekday	0.0446	0.016	2.825	0.005	0.014	0.076
IC_ratio	-0.0260	0.042	-0.625	0.532	-0.107	0.055

Omnibus:	15006.105	Durbin-Watson:	1.120
Prob(Omnibus):	0.000	Jarque-Bera (JB):	21411.195
Skew:	0.570	Prob(JB):	0.00
Kurtosis:	3.976	Cond. No.	14.0

*Source: Table E.5.

Model G.26: CV_h OLS model 5, addition mitigating measure variables.

OLS Regression Results

Dep. Variable:	cvh	R-squared:	0.144
Model:	OLS	Adj. R-squared:	0.144
Method:	Least Squares	F-statistic:	130.4
No. Observations:	228094	Prob (F-statistic):	6.19e-12
Df Residuals:	228084	Log-Likelihood:	1.3596e+05
Df Model:	9	AIC:	-2.719e+05
Covariance Type:	cluster	BIC:	-2.718e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3292	0.031	10.470	0.000	0.268	0.391
len	-0.0151	0.046	-0.328	0.743	-0.106	0.075
scheduled_tt_min	-0.0183	0.012	-1.589	0.112	-0.041	0.004
mode_bus	-0.0236	0.025	-0.931	0.352	-0.073	0.026
DT_with_delay	0.1002	0.011	9.541	0.000	0.080	0.121
AT_with_delay	-0.0567	0.010	-5.457	0.000	-0.077	-0.036
peak_period	0.0264	0.008	3.173	0.002	0.010	0.043
weekday	0.0537	0.016	3.335	0.001	0.022	0.085
IC_ratio	-0.0710	0.045	-1.585	0.113	-0.159	0.017
measure_package	0.0691	0.031	2.261	0.024	0.009	0.129

Omnibus:	13844.005	Durbin-Watson:	1.176
Prob(Omnibus):	0.000	Jarque-Bera (JB):	20941.084
Skew:	0.517	Prob(JB):	0.00
Kurtosis:	4.065	Cond. No.	15.1

*Source: Table E.5.

Model G.27: CV_h OLS model 6, addition interaction terms.

OLS Regression Results

Dep. Variable:	cvh	R-squared:	0.144
Model:	OLS	Adj. R-squared:	0.144
Method:	Least Squares	F-statistic:	108.0
No. Observations:	228094	Prob (F-statistic):	1.77e-11
Df Residuals:	228083	Log-Likelihood:	1.3600e+05
Df Model:	10	AIC:	-2.720e+05
Covariance Type:	cluster	BIC:	-2.719e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3256	0.041	7.965	0.000	0.246	0.406
len	-0.0101	0.057	-0.178	0.859	-0.121	0.101
scheduled_tt_min	-0.0189	0.011	-1.717	0.086	-0.041	0.003
mode_bus	-0.0235	0.025	-0.933	0.351	-0.073	0.026
DT_with_delay	0.0999	0.011	9.049	0.000	0.078	0.121
AT_with_delay	-0.0565	0.010	-5.660	0.000	-0.076	-0.037
peak_period	0.0266	0.008	3.349	0.001	0.011	0.042
weekday	0.0540	0.016	3.310	0.001	0.022	0.086
IC_ratio	-0.0854	0.082	-1.046	0.295	-0.245	0.075
measure_package	0.0707	0.028	2.489	0.013	0.015	0.126
traffic_x_measure	0.0203	0.104	0.195	0.845	-0.184	0.224

Omnibus:	13867.715	Durbin-Watson:	1.176
Prob(Omnibus):	0.000	Jarque-Bera (JB):	20867.336
Skew:	0.520	Prob(JB):	0.00
Kurtosis:	4.056	Cond. No.	30.0

*Source: Table E.5.

Model G.28: CV_h OLS model 7, elimination len.

OLS Regression Results

Dep. Variable:	cvh	R-squared:	0.144
Model:	OLS	Adj. R-squared:	0.144
Method:	Least Squares	F-statistic:	59.13
No. Observations:	228094	Prob (F-statistic):	1.36e-09
Df Residuals:	228084	Log-Likelihood:	1.3598e+05
Df Model:	9	AIC:	-2.719e+05
Covariance Type:	cluster	BIC:	-2.718e+05

	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3221	0.034	9.499	0.000	0.256	0.389
scheduled_tt_min	-0.0202	0.012	-1.703	0.089	-0.043	0.003
mode_bus	-0.0237	0.025	-0.952	0.341	-0.073	0.025
DT_with_delay	0.0995	0.011	9.238	0.000	0.078	0.121
AT_with_delay	-0.0562	0.011	-5.210	0.000	-0.077	-0.035
peak_period	0.0268	0.009	3.068	0.002	0.010	0.044
weekday	0.0543	0.016	3.336	0.001	0.022	0.086
IC_ratio	-0.0899	0.069	-1.298	0.194	-0.226	0.046
measure_package	0.0718	0.029	2.501	0.012	0.016	0.128
traffic_x_measure	0.0263	0.088	0.298	0.766	-0.147	0.199

Omnibus:	13767.546	Durbin-Watson:	1.176
Prob(Omnibus):	0.000	Jarque-Bera (JB):	20663.664
Skew:	0.518	Prob(JB):	0.00
Kurtosis:	4.050	Cond. No.	26.2

*Source: Table E.6.

Model G.29: CV_h OLS model 8, elimination traffic_x_measure.

OLS Regression Results

Dep. Variable:	cvh	R-squared:	0.143
Model:	OLS	Adj. R-squared:	0.143
Method:	Least Squares	F-statistic:	54.05
No. Observations:	228094	Prob (F-statistic):	3.29e-09
Df Residuals:	228085	Log-Likelihood:	1.3590e+05
Df Model:	8	AIC:	-2.718e+05

Covariance Type:	cluster	BIC:			-2.717e+05
	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3248	0.030	10.845	0.000	0.266	0.384
scheduled_tt_min	-0.0202	0.012	-1.717	0.086	-0.043	0.003
mode_bus	-0.0240	0.025	-0.965	0.335	-0.073	0.025
DT_with_delay	0.0997	0.011	9.463	0.000	0.079	0.120
AT_with_delay	-0.0562	0.011	-5.207	0.000	-0.077	-0.035
peak_period	0.0268	0.009	3.058	0.002	0.010	0.044
weekday	0.0542	0.016	3.324	0.001	0.022	0.086
IC_ratio	-0.0715	0.044	-1.611	0.107	-0.159	0.016
measure_package	0.0702	0.031	2.297	0.022	0.010	0.130
Omnibus:	13644.469		Durbin-Watson:			1.174
Prob(Omnibus):	0.000		Jarque-Bera (JB):			20598.816
Skew:	0.512		Prob(JB):			0.00
Kurtosis:	4.057		Cond. No.			12.4

*Source: Table E.6.

Model G.30: CV_h OLS model 9, elimination mode_bus.

OLS Regression Results					
Dep. Variable:	cvh	R-squared:			0.138
Model:	OLS	Adj. R-squared:			0.138
Method:	Least Squares	F-statistic:			34.18
No. Observations:	228094	Prob (F-statistic):			9.42e-08
Df Residuals:	228086	Log-Likelihood:			1.3527e+05
Df Model:	7	AIC:			-2.705e+05
Covariance Type:	cluster	BIC:			-2.704e+05
	coef	std err	z	P>|z|	[0.025	0.975]
const	0.3247	0.027	12.124	0.000	0.272	0.377
scheduled_tt_min	-0.0263	0.011	-2.340	0.019	-0.048	-0.004
DT_with_delay	0.0920	0.009	10.130	0.000	0.074	0.110
AT_with_delay	-0.0507	0.013	-3.990	0.000	-0.076	-0.026
peak_period	0.0263	0.009	3.052	0.002	0.009	0.043
weekday	0.0537	0.016	3.364	0.001	0.022	0.085
IC_ratio	-0.0771	0.044	-1.768	0.077	-0.163	0.008
measure_package	0.0763	0.031	2.455	0.014	0.015	0.137
Omnibus:	12830.433		Durbin-Watson:			1.174
Prob(Omnibus):	0.000		Jarque-Bera (JB):			18860.140
Skew:	0.498		Prob(JB):			0.00
Kurtosis:	3.996		Cond. No.			12.2

*Source: Table E.6.

G.4. Full log-transformed variability model output

Model G.31: Log-transformed CV_{TT} specification.

OLS Regression Results					
Dep. Variable:	ln_cvtt	R-squared:			0.250
Model:	OLS	Adj. R-squared:			0.250
Method:	Least Squares	F-statistic:			5.176
No. Observations:	228094	Prob (F-statistic):			0.00675
Df Residuals:	228088	Log-Likelihood:			-86810.
Df Model:	5	AIC:			1.736e+05
Covariance Type:	cluster	BIC:			1.737e+05
	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.2096	0.131	-9.210	0.000	-1.467	-0.952
len	-0.8220	0.181	-4.547	0.000	-1.176	-0.468
mode_bus	0.2594	0.106	2.445	0.014	0.051	0.467
IC_ratio	0.5171	0.201	2.567	0.010	0.122	0.912
measure_package	0.0283	0.101	0.280	0.780	-0.170	0.227
traffic_x_measure	-0.4337	0.229	-1.896	0.058	-0.882	0.015

Omnibus:	324.435	Durbin-Watson:	1.357
Prob(Omnibus):	0.000	Jarque-Bera (JB):	314.764
Skew:	-0.076	Prob(JB):	4.47e-69
Kurtosis:	2.900	Cond. No.	14.3

*Source: Table 5.2.

Model G.32: Log-transformed CV_h specification.

OLS Regression Results

Dep. Variable:	ln_cvtt	R-squared:	0.250
Dep. Variable:	ln_cvh	R-squared:	0.137
Model:	OLS	Adj. R-squared:	0.137
Method:	Least Squares	F-statistic:	27.86
No. Observations:	228094	Prob (F-statistic):	3.52e-07
Df Residuals:	228086	Log-Likelihood:	-1.2698e+05
Df Model:	7	AIC:	2.540e+05
Covariance Type:	cluster	BIC:	2.541e+05

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	-1.2124	0.094	-12.908	0.000	-1.396	-1.028
scheduled_tt_min	-0.0944	0.046	-2.056	0.040	-0.184	-0.004
DT_with_delay	0.2491	0.027	9.137	0.000	0.196	0.303
AT_with_delay	-0.1667	0.042	-3.955	0.000	-0.249	-0.084
peak_period	0.0834	0.029	2.875	0.004	0.027	0.140
weekday	0.2033	0.048	4.242	0.000	0.109	0.297
IC_ratio	-0.1742	0.156	-1.119	0.263	-0.479	0.131
measure_package	0.2264	0.111	2.047	0.041	0.010	0.443

Omnibus:	34988.170	Durbin-Watson:	1.218
Prob(Omnibus):	0.000	Jarque-Bera (JB):	78240.729
Skew:	-0.904	Prob(JB):	0.00
Kurtosis:	5.228	Cond. No.	12.2

*Source: Table 5.3.

G.5. Full fractional logit output for on-time performance

Model G.33: Fractional logit specification for OTP.

Generalized Linear Model Regression Results

Dep. Variable:	otp	No. Observations:	228094
Model:	GLM	Df Residuals:	228086
Model Family:	Binomial	Df Model:	7
Link Function:	Logit	Scale:	1.0000
Method:	IRLS	Log-Likelihood:	-1.0048e+05
No. Iterations:	4	Deviance:	28156.
Covariance Type:	cluster	Pearson chi2:	2.51e+04
		Pseudo R-squ. (CS):	0.02786

	coef	std err	z	P>|z|	[0.025	0.975]
Intercept	1.1963	0.080	14.917	0.000	1.039	1.353
len	-0.3842	0.128	-3.012	0.003	-0.634	-0.134
DT_with_delay	-0.9468	0.080	-11.795	0.000	-1.104	-0.789
AT_with_delay	-0.2310	0.063	-3.684	0.000	-0.354	-0.108
weekday	-0.1785	0.041	-4.405	0.000	-0.258	-0.099
IC_ratio	0.3560	0.157	2.266	0.023	0.048	0.664
measure_package	-0.0882	0.076	-1.160	0.246	-0.237	0.061
traffic_x_measure	-0.3127	0.171	-1.830	0.067	-0.648	0.022

*Source: Table 5.4.

Statement on the Use of Generative AI

Generative AI tools, including ChatGPT, Grammarly and NotebookLM, were used during the preparation of this thesis as writing and learning support tools. Their use included improving the structure and clarity of sections, refining academic language, and providing assistance with LaTeX formatting and Python code. AI-generated suggestions were not used without critical review and were adapted where necessary to fit the scope, methodology, and findings of this research.

The analysis, selection of methods, processing of data, interpretation of results, use of literature, and final conclusions were carried out and verified by the author. Generative AI was not used as a primary source of scientific information. The author remains fully responsible for the accuracy, integrity, and originality of the final thesis.