Transit Extraboard Operator Scheduling

by

Qing Yi Wang

B.A.Sc. in Engineering Science, University of Toronto (2018)

Submitted to the Department of Civil and Environmental Engineering
in partial fulfillment of the requirements for the degree of

Master of Science in Transportation

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2020

Author...
Department of Civil and Environmental Engineering
August 17, 2020

Certified by..
Nigel H. M. Wilson
Professor Emeritus of Civil and Environmental Engineering
Thesis Supervisor

Certified by..
Haris N. Koutsopoulos
Professor of Civil and Environmental Engineering, Northeastern University
Thesis Supervisor

Accepted by...
Colette L. Heald
Professor of Civil and Environmental Engineering
Chair, Program Committee

Transit Extraboard Operator Scheduling

by

Qing Yi Wang

Submitted to the Department of Civil and Environmental Engineering on August 17, 2020
in partial fulfillment of the requirements for the degree of Master of Science in Transportation

Abstract

In addition to operators who pick scheduled service duties, transit agencies have a separate group of operators to cover work that becomes open due to absence or other unexpected situations. This group of operators are referred to collectively as the extraboard. Another way to cover the open work is through operator overtime. Therefore, a central challenge of the extraboard planning problem lies in the uncertainty of the amount of the work that will need to be covered, as well as the extent of operators' willingness to work overtime. Due to the critical importance of service reliability, transit agencies seek a systematic approach to schedule extraboard operators to minimize a weighted cost of lost service, overtime, and extraboard operators. This thesis proposes a methodology that systematically addresses the extraboard scheduling problem, focusing on a case study using data from the Massachusetts Bay Transportation Authority (MBTA). The methodology has two components: demand (absence) estimation and schedule optimization.

Absence can be classified as known-in-advance or unexpected, based on both when they are known and the way they are covered. Two negative binomial regression models were formulated based on their different characteristics. Among the variables tested, no significant predictive relationships were found with respect to absences, overtime, or lost service. The resulting models mainly reflect the average behavior on each day of the week.

Multi-stage integer optimization programs were constructed to schedule the extraboard operators. Given the current extraboard size, assignments given by different modelling strategies were similar. When the staffing level constraint was relaxed, compared to deterministic models, the robust solutions achieve more stable level of lost service and overtime, while being less sensitive to model parameters. However, the robust solutions are of higher financial costs to the MBTA, since they included more fixed financial costs from the extraboard operators and less variable costs from overtime and lost service. Therefore, without improvements in the input estimations, decision of extraboard size depends on the tradeoff between financial costs and service reliability.

This thesis contributes to the literature by quantitatively studying operator absence, introducing robust optimization for the extraboard planning problem, and demonstrating the use and the advantages of a systematic assignment procedure.

Thesis Supervisor: Nigel H. M. Wilson
Title: Professor Emeritus of Civil and Environmental Engineering

Thesis Supervisor: Haris N. Koutsopoulos
Title: Professor of Civil and Environmental Engineering, Northeastern University

Acknowledgements

Reflecting on the journey of the past two years, I am immensely grateful for having the opportunity to be guided by Prof. Nigel Wilson and Prof. Haris Koutsopoulos through the ups and downs of this project. Their dedication to research, breadth and depth of knowledge, as well as attention to detail have inspired me far beyond this research project itself. I feel honored to have worked alongside of and learned from Nigel and Haris in my early career as a researcher.

I would like to thank the Massachusetts Bay Transportation Authority for providing this research opportunity, as well as the many people in the agency who answered my many questions in detail and with patience. Without the help of Jen Elise, Angel, Vinny, Charlie, Calvin, Sean, and Anna at the MBTA, this research would not have been possible.

I also wish to express my gratitude towards the JTL and Transit Lab members. They have also been an integral part in the experience. Special thanks to Prof. Jinhua Zhao for his support in my academic endeavors outside of this research, Shenhao for being a resourceful mentor, Baichuan for being a comrade in this journey, Ruben for the always enlightening, spontaneous chats in the office, and the many others in the lab that I will not name for their friendship and academic support. I feel proud and honored to be part of this community.

Finally, my gratitude goes to my family and friends, who have shaped who I am as a person. This marks the tenth year that I have left home for the other side of the world. Thank you to my mother and my grandparents for empowering me with wings when I want to fly, and offering me the warmth of a nest whenever and wherever I need, in all these years.

Table of contents

Chapter 1 Introduction

In the transit industry, the cost of operators makes up 60% of operating expenses and 43% of all transit expenditures (APTA, 2019). Workforce planning is the process used to align the needs and priorities of the organization with those of its workforce to ensure it can meet its legislative, regulatory, and service requirements and organizational objectives. Workforce planning enables evidence-based workforce development strategies that could improve the organization's cost efficiency and service delivery. In this chapter, we motivate and describe the workforce planning problem, and define the analysis scope for this thesis.

1.1 Terminology

Before the workforce and the extraboard planning problem is introduced, the use of some terminology in this thesis is defined below.

Piece of work: The smallest unit of work that can be assigned to an operator.

Shift/run: Shifts and runs are used interchangeably. They refer to a set of pieces of work that are performed by a single operator.

Regular operators: Operators who picked scheduled service runs at the start of the rating.

Extraboard operators: Operators who do not have a run assignment and stand by to cover work when the scheduled regular operator cannot.

Cover list operators: Part of the extraboard operators who are not assigned shifts in advance, but rather are assigned report times to the garage to stand by and cover unexpected open work.

Open work: Work that becomes open after the run-cut and operator assignment, due to the assigned regular operator's inability to deliver, or the need for additional work.

Known-in-advance open work / Absence: Open work or operator absence that becomes known before the report times of the cover list operators are assigned. The garage dispatcher assigns the entire run to an extraboard operator.

Unexpected open work / Absence: Open work or operator absence that is known after the report times of the cover list are assigned. These shifts are usually broken up into pieces which are covered by more than one operator.

1.2 The Workforce Planning Problem for Bus Operators

Transit agencies provide service according to a published schedule. Depending on the size and geographical coverage, an agency usually operates trips out of multiple garages. At each garage, operators are scheduled to meet service requirements. If no operator is available to perform trips at the scheduled times, trips will be lost. Dropped trips significantly impact service reliability and customer satisfaction. For high-frequency routes, dropping a single run may not be a serious issue if subsequent trips can be covered; if not, then the headway will be noticeably longer for the duration of the absent operator's shift. For routes with longer headways, people usually arrive at a stop according to the schedule and the waiting times will be very large (at least one headway) when a trip is missed. Having an adequately sized, efficiently scheduled workforce is a necessity if the agency is to meet their service commitment. Therefore, workforce planning is critical to ensuring service reliability.

Workforce planning, in the transit context, is the process of operator scheduling that deals with aligning the availability of operators with the demand for operators such that service requirements are met at the least cost and with the most reliability. Figure 1-1 shows the basic elements involved in workforce planning and their interactions. At the highest level, the elements can be divided into three categories: operator demand, operator availability, and planning outcomes.

Operator Demand: The demand for operators comes principally from the service requirements, and includes scheduled trips, non-service duties, and extra trips. Non-service duties include training, inspector duties, flagging duties, shuttles, transferring buses between garages (because most agencies operate multiple garages), etc. Extra (unplanned) service requirements include shuttles to replace subways during a shutdown, shuttles for special events, etc. Together, all the work mentioned above constitutes the operator demand.

Operator Availability: The baseline operator availability is the staffing level. In addition, each operator is subject to vacation, making them unavailable for certain weeks in the year. Other factors that could reduce operator availability are absence, training, attrition, and suspension. To deal with the need for extra operators, agencies usually hire more operators than required by the timetable, resulting in an extraboard, consisting of operators who are available to cover runs as needed. Overtime is another way to increase the supply of operators in times of demand surge or availability shortage.

Planning Outcomes: Planning outcomes reflect the match between operator demand and availability. There are two major outcomes for workforce planning. The first is the operator schedule, which matches planned services to individual operators. Second, realized service outcomes are the outcomes after accounting for demand and supply adjustments between the time at which the plan is made and the day of operations. Either service is performed as planned (neutral), covered by other operators (increased cost), or dropped (reduced service reliability).

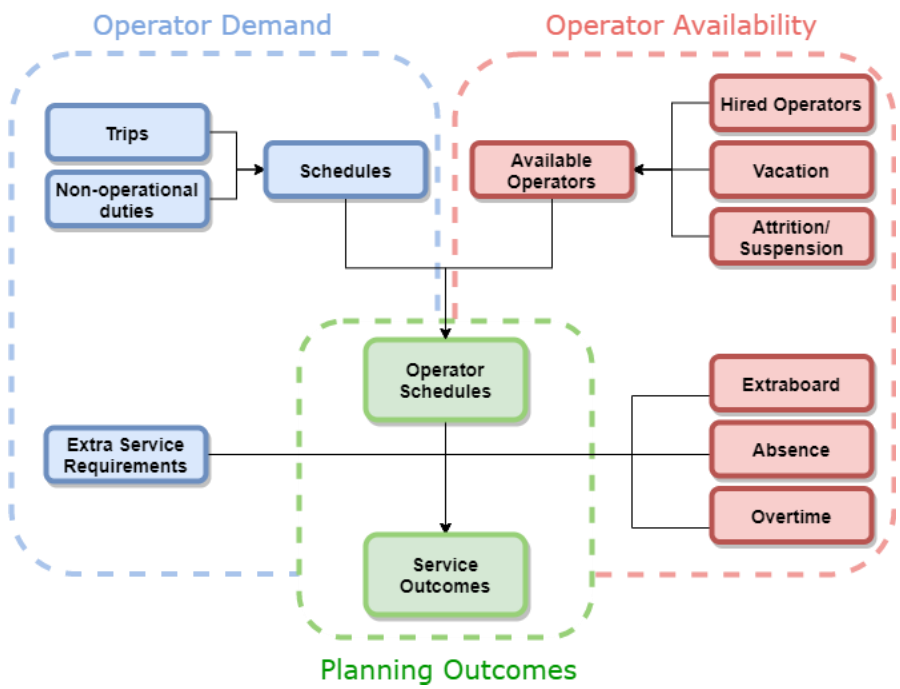


Figure 1-1 Basic Quantities and Their Interactions in Workforce Planning

Decision-making for bus operator planning (and service sector workforce planning in general) can be thought of at three levels: strategic, tactical, and operational (Abernathy et al., 1973; Koutsopoulos and Wilson, 1987). Figure 1-2 shows the inputs to the process and the decisions being made at each level. The right-pointing arrows denote inputs at each stage, and the downward arrows indicate sequential decision-making.

At the strategic level, the goal is to make hiring decisions that minimize overall operator costs while meeting the service requirements, work rules, and budget constraints. An overall goal is set for the planning period. Typically, there are several ratings in a year where the service requirements may differ (for example, the MBTA adjusts its service plan quarterly) and therefore the hiring targets will differ accordingly. The outcome of the strategic level is the number of hired operators, vacation allocation over the year, and services which are scheduled and allocated to operators.

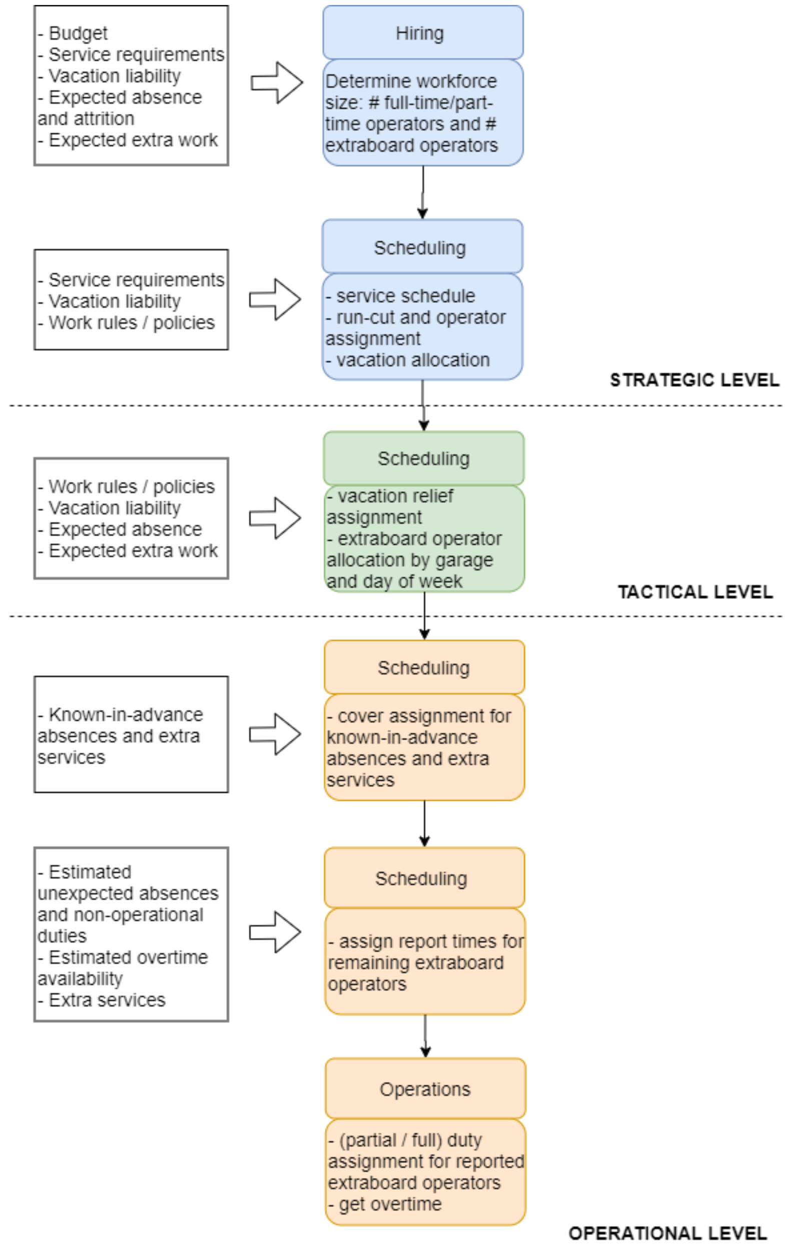


Figure 1-2 Three Levels of Workforce Planning

At the start of each rating, tactical level planning finalizes the assignment of operators who cover vacations, as well as the number of extraboard operators to schedule each day, based on the expected level of absence and overtime availability. Vacation relief operators, who are a separate group from the extraboard operators, cover for operators on vacation.

At the operational level, sequential decisions are made regarding assignments to cover known-in-advance absences, report times for the rest of the extraboard, and overtime acquisition. Some absences and extra work are known-in-advance, meaning that they are known before report times are assigned to the cover list operators. These runs will be assigned directly to extraboard operators. For the remaining cover list, based on the expected time-of-day distribution of unexpected absences and extra work, each operator will be given a time at which they will report to the garage and cover work as needed. If a piece of work becomes open and no cover list operator is available, that piece of work will be offered to available operators as overtime. If an operator accepts the offer, (s)he will work extra time to cover the service; otherwise, the service will be lost.

The workforce planning process involves different stages which can be studied separately. For example, the scheduling problems for regular services and extraboard operators are quite different and can be studied separately. The run-cutting and operator assignment problem is a classical, deterministic problem in operations research. The challenge is to effectively accommodate agency-specific constraints and solve the optimization problem efficiently. The focus of the extraboard scheduling problem is dealing with uncertainty. While making strategic, tactical, and operational level decisions regarding the extraboard, absence and overtime availability are uncertain. This research deals with the extraboard scheduling problem at the tactical and operational levels. The next section describes this problem in more detail.

1.3 The Extraboard Planning Problem

After the run-cut and operator assignment, work may arise due to operator absence, shift swaps, diversions, service disruptions, and extra service. We refer to all work that comes up after the run-cut and operator assignment as open work. Some non-service duties such as transferring buses and serving as substitute inspectors may come up during daily operations. The extraboard operators provide a built-in buffer that makes the schedule more robust; that is, service requirements can still be met when unexpected situations arise. Without extraboard operators, we have no resources to deal with unexpected situations. In this case, whenever open work arises, overtime must be requested. Although overtime has a 50% pay premium, it is generally less expensive than hiring extraboard operators after accounting for benefits and vacations provided to each operator. However, there are two issues associated with relying heavily on overtime. First, not everyone is willing to work overtime and there are work rules around the maximum number of hours for those who want to work overtime. For example, the cap is 20 hours of overtime per operator per week at the MBTA. So, relying too much on overtime risks dropping a lot of service. Second, while evidence in a lot of industries suggest that working a moderate amount of compensated, voluntary overtime is healthy for both the organization and individuals (Beckers et al., 2008, 2004; Holly and Mohnen, 2012), working too much overtime causes fatigue and may compromise safety (Bae and Fabry, 2014; Rogers et al., 2004). Due to the importance of service reliability, an efficient and robust scheduling model that accounts for the uncertainty of the amount in open work and overtime availability is needed.

Scheduling extraboard operators is part of the three-level workforce planning framework (Figure 1-2) and is the main task at the tactical and operational levels. At the strategic level, hiring decisions are made regarding both regular operators and extraboard operators. At the tactical level, the extraboard operators' workdays are determined. At the operational level, first, all known-in-advance open work is assigned to the extraboard. The assigned extraboard operator will inherit all work from the open run, including the splits and built-in overtime. Next, to address unexpected absences, non-service duties, and delays caused by incidents and traffic, the remaining extraboard operators (i.e. the cover list operators) will be assigned report times. They will report to the garage and get paid for an 8-hour straight shift (6 hours for part-timers) regardless of the availability of open work. Having cover list operators present when there is no work results in unproductive but paid operator time. Different levels of tolerance for lost service and unproductive extraboard time will lead to very different scheduling decisions. The importance of effectively sizing and scheduling the extraboard has been established in the literature (Ingels and Maenhout, 2015; Sohoni et al., 2006). In general, having a large extraboard leads to lower utilization rates since there will be more unproductive time. On the other hand, having a small extraboard leads to more overtime requested which jeopardizes service reliability and operator well-being.

Optimizing extraboard planning decisions is a challenging problem because the amount of open work at the time of decision-making is uncertain. While some seasonal and weekly trends exist, the amount of open work that needs to be covered can be highly variable. To further complicate the problem, since overtime is (generally) less expensive and more flexible, a cost-effective extraboard assignment plan should take overtime into account. However, uncertainty exists around how much overtime will be available at a given time, how much available overtime the agency wants to engage (due to the issues mentioned earlier in this section), and how much lost service the agency is willing to accept.

1.4 Scope and Objectives

The principal reason for open work is operator absence (60%-70%) (DeAnnuntis and Morris, 2008; Gupta et al., 2011; Hickman et al., 1988; Shiftan and Wilson, 1993). The passage of the Family and Medical Leave Act (FMLA) of 1993 made operator absence a less predictable and more influential element in workforce planning (DeAnnuntis and Morris, 2008; Strathman and Callas, 2012). The impact is particularly pronounced in transit, because of operators' rigid and oftentimes undesirable work schedules (Bolotnyy and Emanuel, 2019). This thesis specifically addresses the extraboard planning problem with respect to operator absence.

Transit agencies always face tightly constrained resources. It is in their primary interest to fully utilize their existing resources and make informed decisions on a cost-effective workforce size. This thesis approaches the problem at the tactical and operational levels, first taking the strategic level decision (hiring level) as given, and then progressing to make strategic-level recommendations.

The goal of this thesis is to develop a systematic procedure to improve the extraboard planning process for bus operators to reduce lost service, overtime and cost. Breaking down the goal into smaller tasks/objectives, this thesis aims to address:

1) Mining data from HASTUS™ Daily[1] to extract relevant information on absence levels, overtime availability, and service reliability.
2) Characterizing and modelling absence, overtime, lost service and their interrelations.
3) Investigating how the resulting models could be used in extraboard scheduling and evaluate the effectiveness of such scheduling practice.

1.5 Thesis Contents

The thesis is organized as follows: Chapter 2 proposes an overall analysis framework for the extraboard planning problem. Chapter 3 reviews the relevant literature for the building blocks in the framework. Chapter 4 introduces the context for the analysis. Chapters 5 addresses the prediction of absence. Chapter 6 formulates the scheduling model and assesses the performance of different types of models. Lastly, Chapter 7 offers a summary of the major contributions and proposes future research directions.

[1] HASTUS™ Daily is a product of GIRO.

Chapter 2 Overall Analysis Framework

This chapter describes the overall framework employed in this thesis to approach the extraboard planning problem. First, the decision-making process in extraboard planning is described. Then an interaction framework that can be applied at various levels for the demand and supply of operators is presented. Finally, the information feedback loop that cycles performance outcome data back to the earlier stages of planning is illustrated.

2.1 The Decision-Making Process

The full workforce planning timeline can be summarized in Figure 2-1. Specifically, there are three important decisions to be optimized. At the strategic level, the number of extraboard operators to hire and to allocate to each garage is determined. At the tactical level, the number of extraboard operators for each day is determined. At the operational level, the report times of the cover list operators for unexpected absences are decided. In this thesis, only the tactical and the operational level decisions are considered, given a fixed number of extraboard operators available at each garage. At the time of deciding the number of extraboard operators for each day (the tactical level), absence levels (both known-in-advance and unexpected) are uncertain. At the time of deciding the report times of cover list operators (the operational level), the quantity and time-of-day distribution of unexpected absences are uncertain. In Figure 2-1, these two decisions within the analysis scope are highlighted in red. The assignment on the day of operations is assumed to be greedy: extraboard operators will be used whenever available and overtime will be sought when no extraboard operators are available. Whether the greedy approach is optimal is another research topic on its own (Gupta and Li, 2016) and beyond the scope of this thesis.

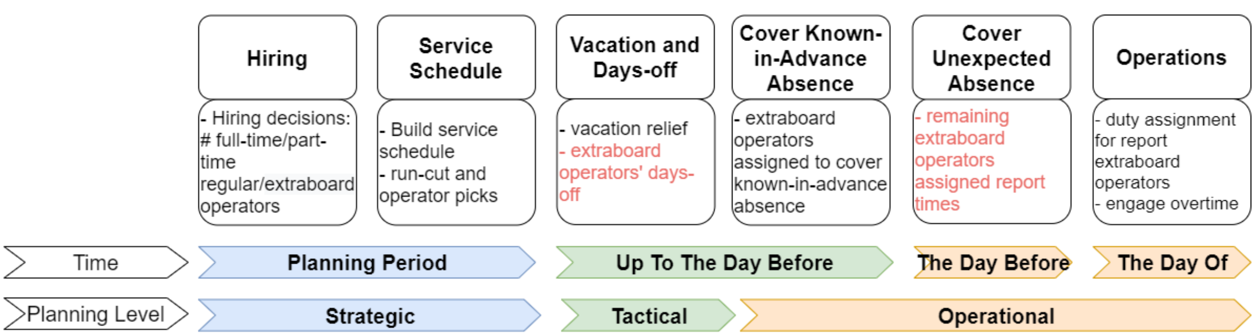


Figure 2-1 The Workforce Planning Process for Bus Operators

2.2 Interaction Framework of Quantities Involved

The problem, at each planning stage, can be decomposed into three building blocks: a demand model, a supply model, and an interaction mechanism. Figure 2-2 shows the proposed analysis framework. All open work is treated as demand for operator-hours, with available time to cover as the supply, and temporally aligning demand and supply produces trip outcomes (either completed or dropped).

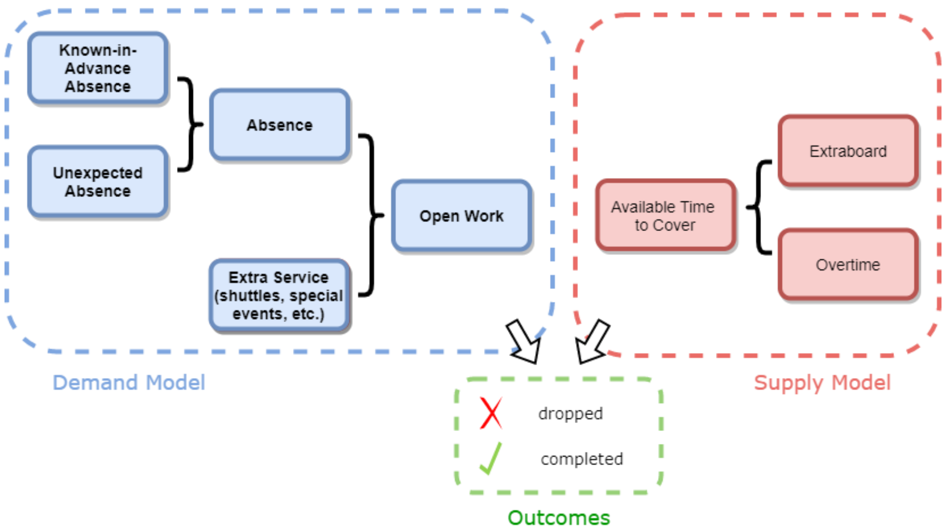


Figure 2-2 Interaction Framework for Interested Quantities

There are several reasons for open work, the biggest one being absence among scheduled operators. Depending on the time of notice, absences are divided into known-in-advance absence and unexpected absence. Different agencies have different criteria to categorize known-in-advance and unexpected absences, and the two types of absences are usually covered with different scheduling procedures. Aside from absence, another source of open work is extra service, which includes shuttles during special events and/or subway shutdowns.

Two options exist to cover open work: the extraboard and overtime. The extraboard is a proactive measure to cover open work, which is scheduled up till the day before the service day; and overtime is a reactive measure, which is acquired as work becomes open on the service day.

Since misalignment of demand and supply produces unproductive cover time and lost service, the interaction mechanism depends on the unit of analysis. For example, if open work and available time to cover are both at the time-of-day level, then directly aligning them temporally produces the times and quantity of personnel shortage. If open work and available time to cover are on the daily level, then assumptions need to be made about the (mis)alignment of the time-of-day profiles.

This analysis framework applies to both tactical and operational level scheduling. In general, the number and report times of the extraboard are the decisions to be made and the open work and overtime availability need to be estimated. Trip outcome is the objective to be optimized. At the tactical level, the number of extraboard operators to schedule for each day needs to be determined. Both the amount of open work and overtime available are unknown and need to be predicted (or aggregated) from the operational level model. At the operational level, the scheduling decisions are the report times for the cover list operators. The number of known-in-advance absences and the majority of extra service, and the number of cover list operators to cover unexpected open work are known; the amount and time-of-day distribution of unexpected open work and available overtime are uncertain. The next section describes how observed trip outcomes can be used to form a feedback loop to inform earlier planning stages.

2.3 Feedback Loop of Decision Variables and Observations

In Figure 2-3 we present an extended version of Figure 1-2 including the information feedback loop. From trip outcomes, information about the known-in-advance absences, unexpected absences, non-service work performed, lost service, and overtime performed can be extracted. First, overtime availability can be estimated historially from overtime performed and lost service. Since overtime is usually the last resort for covering open work, overtime requested is the total of overtime performed and lost service. When overtime performed is less than overtime requested, overtime performed is equal to overtime availability. When overtime performed is equal to overtime requested, available overtime is greater than or equal to overtime performed. Models to forecast unexpected absence and non-service work could be developed directly from observations, thereby providing a better understanding of how many operators are required to cover unexpected open work. Similarly, forecasts for the number of known-in-advance absences can be developed from historical observations. Combining the number of extraboard operators needed for unexpected absences, and the number needed for known-in-advanced absences gives recommendations for the tactical-level scheduling decision. Further, by aggregating the number of extraboard operators needed for each planning period, we could obtain recommendations for the strategic-level decisions on the number of extraboard operators to hire.

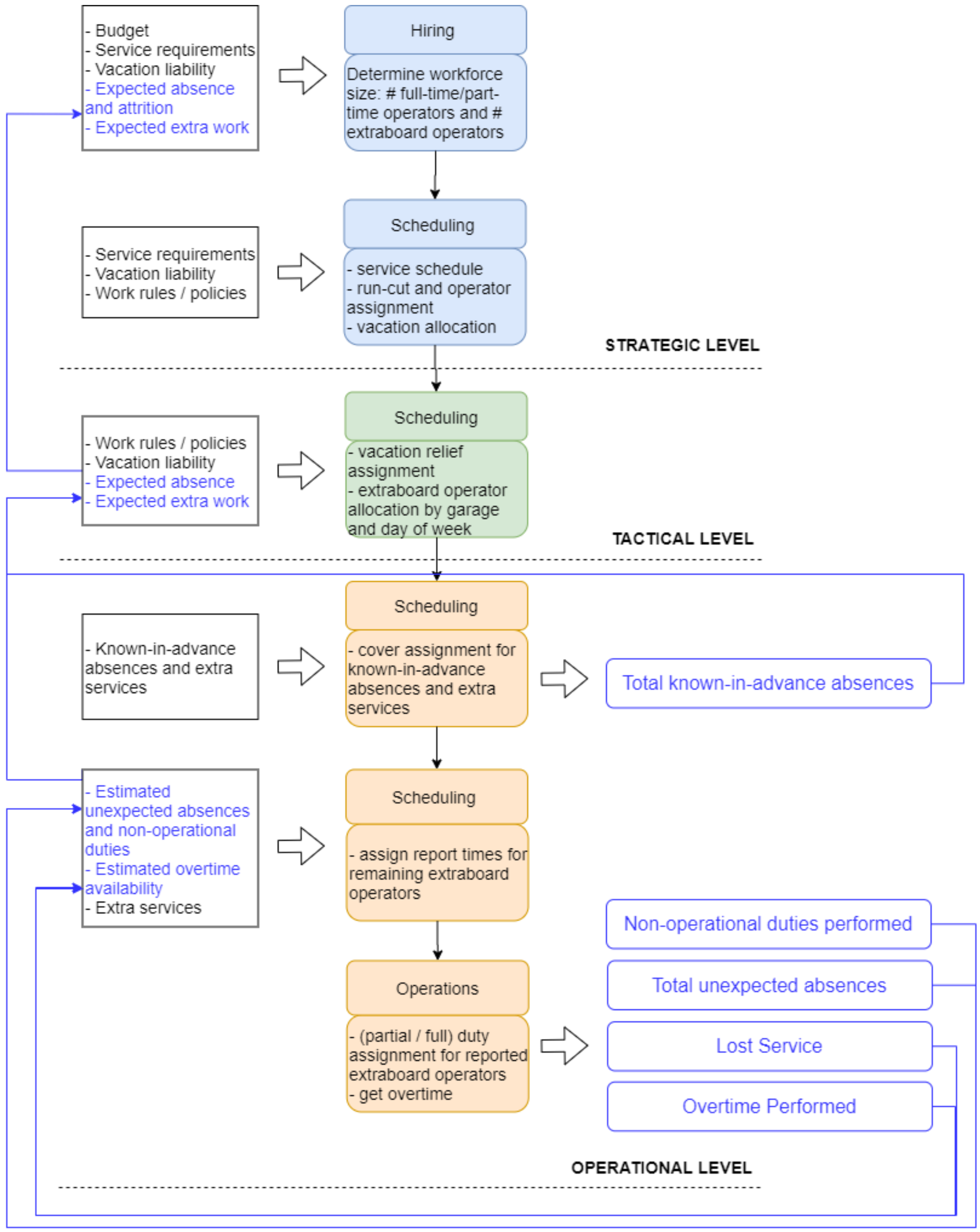


Figure 2-3 Information Feedback Loop

Chapter 3 Literature Review

The proposed analysis framework has two major components: modelling absence and overtime availability, and optimization of extraboard schedules. Sections 3.1 and 3.2 offer an overview of literature in these two areas, respectively. Section 3.3 reviews the prior work on workforce planning at the MBTA. Section 3.4 describes the state of the practice in the transit industry.

3.1 Absence / Overtime Modelling

An ongoing area of research is the study of employee absenteeism and absence management in the social sciences. Behavioral models of absenteeism provide insights into the characteristics of employees who are more likely to be absent, why they are absent, and what incentives can reduce absence. This information can serve as domain knowledge in model and variable selection. There are three perspectives: individual, social, and economic (Kaiser, 1998). Some believe that absenteeism is dependent on the individual's willingness and ability to attend work (Suárez and Muñiz, 2018), others believe that social adaptation (Ahn et al., 2013) and peer pressure (Gaudine and Saks, 2001) influence absence behavior. In economics, an income-leisure tradeoff model formulates absence as a balance between income and leisure, which is directly related to the employee's marginal earnings, perceived safety, and schedule flexibility (Allen, 1981). While there are studies that support this theory (Barmby et al., 2001), there is also empirical evidence showing that pay and absenteeism are unrelated (Winkelmann, 1996). Specific variables that have been used to study absence include personal and family factors (income, health, geographical location, gender, age, etc.), characteristics of work, attitudes towards work, income, social adaptation, firm size, etc. All studies on absenteeism point out that although generalizations can be made at the behavioral level, absenteeism is highly complex and specific to the industry and organization (Ahn et al., 2013; Allen, 1981; Kaiser, 1998; Suárez and Muñiz, 2018). In addition, some types of data, including demographics and attitudes, are difficult to obtain due to privacy and cost concerns. Therefore, while past literature serves as a guide for what to look for, each situation needs to be assessed individually.

Overtime decisions are consistent with utility maximization subject to budget constraints, positively related to the pay premium and the number of hours of work after which overtime must be paid. But this positive correlation is quite weak since overtime decisions are fairly insensitive (Idson and Robins, 1991). An analysis using logistic regression on the 2002 General Social Survey (GSS) Quality of Work Life Module found that, similar to absence, voluntary overtime decisions are also related to certain demographic, job and work characteristics such as being single, satisfaction with one's job, being a union member, employed in the public sector and standard (rather than contingent) jobs and having a say in one's job (Golden and Wiens-Tuers, 2005).

To model absence quantitatively, due to the difficulty of obtaining demographic data, most research is based on aggregate data and uses regression. MacDorman et al. presented the earliest analytical work on this topic and approximated the cumulative distribution of absence by a logit function (MacDorman and MacDorman, 1987). Diab et al. used a multi-level regression to characterize and predict absence using both demographic and historical information (Diab et al., 2014). Poisson regression with underreporting estimated by Monte Carlo Markov Chain (MCMC) (Winkelmann, 1996) and negative binomial regression (Barmby et al., 2001) have also been used to analyze annual absence days. Sturman compared eight different regression models to model the number of excused/unexcused absences in a given year using both simulation and actual data. The models include ordinary least squares (OLS), OLS with transformed variables, Poisson, over-dispersed Poisson, negative binomial, Tobit, ordinal Logit, ordinal Probit. The focus of the study was whether the models yield false positives, that is, identified insignificant variables as being significant. The study concluded that despite methodological expectations, OLS does not produce significantly more false positives; Poisson and Tobit models are more prone to false positives; and negative binomial models are the most conservative when it comes to statistical significance of model coefficients (Sturman, 1996).

3.2 Robust Schedule Optimization

It is important to acknowledge the difference between regular operator scheduling and extraboard scheduling. For regular run-cutting and operator assignment, the problem is often how to effectively formulate and solve a classical set covering problem followed by a set partitioning problem (Constantino et al., 2017). Extraboard operators are assigned to cover regular operator absences and other non-service open work; therefore, while the problem is on a much smaller scale, the major challenge is explicitly accounting for uncertainty. Literature on the specific topic of extraboard scheduling in the transit industry is limited. The question of covering absence also arises in other service industries and has been studied quite extensively in the airline and nursing industries in particular. The rest of the section reviews the methods applied in other industries to account for uncertainty in scheduling.

To build more robustness into the schedules, we could either proactively build robustness into the baseline plan, or reactively seek overtime to cover operational disruptions (Ingels and Maenhout, 2018). In the extraboard scheduling problem we make proactive decisions on the number of extraboard operators to schedule for each day at the tactical scheduling level, and make reactive decisions on the report times for each extraboard operator before the service day, as well as overtime acquisition of the service day. In most of the literature, the strategic, tactical, and operational level models are optimized separately, but the structure of the problem in the scope of the thesis gives rise to a two-stage program that can be solved jointly. Ingels and Maenhout explored this structure in their theoretical analyses of the benefits of scheduling extraboard operators (Ingels and Maenhout, 2015) and the ability to acquire overtime (Ingels and Maenhout, 2018). A systematic evaluation of the effect of having a reserve workforce was performed with simulated uncertain demands using scheduled service as the mean. The notions of 0% robustness (having no extraboard) and 100% robustness (having unlimited extraboard operators) were proposed. Both studies showed that it is important to schedule the extraboard to keep the overall cost under control and the assignment of the extraboard involves a tradeoff between the cost of additional extraboard operators and the cost of requesting overtime and/or dropping service (Ingels and Maenhout, 2018, 2015).

In most analyses, a myopic approach was assumed considering the match between jobs and workers, that is, the extraboard operators will be exhausted before overtime is engaged. Gupta and Li proposed a randomized algorithm that investigated the tradeoff between the myopic approach (of accepting all jobs that can be scheduled) and the strategic approach (of accepting jobs that are longer than some threshold), thus balancing the proactively scheduled extraboard personnel and the potential reactive overtime (Gupta and Li, 2016). However, aside from efficiency, the negative impacts of excessive overtime on the personal lives of the employees and on service reliability must also be considered.

Depending on the availability of data and the amount and overlap of absences known in advance, there are a number of ways that uncertainty can enter the formulation. When a good model to predict uncovered open work is available or if the most absences are known before the scheduling process, an optimal schedule can be developed using a set-covering and partitioning approach without considering uncertainty (Dillon and Kontogiorgis, 1999). When reliable information on the work that needs to be covered is not available, simulation can be used. A simulation of the airline's operations with stochastic journey time and crew absence inputs (without reserve crews) is used to generate input disruption scenarios for the mixed integer programming simulation scenario model (MIPSSM) formulation. Each disruption scenario corresponds to a record of all of the disruptions that may occur on the day of operation (Bayliss et al., 2017). To account for absences, the most common approach is to fit a regression model for the probability of absence for each shift or person (Bayliss et al., 2012; Maass et al., 2017; Wang and Gupta, 2014). Compared to stochastic optimization, another alternative, robust optimization, has the advantage of replacing the probabilistic representation with deterministic, set-based constraints which makes it tractable in high dimensions (Bandi and Bertsimas, 2012). In transportation, robust optimization is widely applied in supply chain management and has demonstrated superior worst-case results, smaller variance in performance with a small loss of optimality when the actual distributions of the model inputs are similar to those used in the nominal model (Bruns et al., 2014; Peters et al., 2016; Van Landeghem and Vanmaele, 2002).

3.3 Prior Work on Workforce Planning at the MBTA

There is a line of research on this topic conducted at the MBTA in the 1980s and 1990s, which set the stage for future transit workforce planning efforts. Looking at absence retrospectively, Koutsopoulos formulated a tactical and an operational level integer programming model to show that significant productivity and reliability improvements could be achieved wkith optimization compared to using rules of thumb or simple deterministic models (Koutsopoulos and Wilson, 1987). Hickman et al. improved the formulation of the estimation of lost service by introducing the concept of “slop time”, which was defined as the nonzero element of uncovered open work and unproductive cover time (since one of them has to be zero for a particular time period), and fitting a “slop time” curve with respect to the difference between available hours to cover and amount of open work (Hickman et al., 1988). Kaysi later represented absence as a binomial process with the probability of a run being open having a constant value. The split between overtime and lost service, when no extraboard operator is available, was assumed to be constant (Kaysi and Wilson, 1990). Building on Kaysi’s work, Shiftan further studied the relationship between absence and overtime, as well as the relationship between overtime and service reliability. This analysis showed that absence was more a habit than a decision made with respect to overtime already performed. At the system level, there was a strong linear relationship between uncovered open work and missed trips at the daily level. In the study, instead of characterizing the “slop time” as one component, “slop time” was decomposed into uncovered open work and unused extraboard hours; the quantities are estimated separately and summed. Then the lesser of the two was taken to be the “slop time” for the day (Shiftan and Wilson, 1993). In the 1990s, data entry and analysis were not fully automated it was very challenging to expand the scale of the study or to do analysis with a finer resolution. This research was not followed up. When more granular data became available from the recently employed scheduling software HASTUS™ Daily, the MBTA became interested in revisiting this topic.

3.4 State of the Practice in the Transit Industry

Some agencies (e.g. Dallas Area Rapid Transit) use historical data and rolling averages of open work to aid in extraboard scheduling but most agencies schedule extraboard report times manually, based on experience (DeAnnuntis and Morris, 2008; Gupta et al., 2011). At the MBTA, all scheduling decisions have to be made before noon on the previous day. After known-in-advance open work is addressed, the garage superintendent usually splits remaining extraboard operators between serving the morning and afternoon peaks (about a 60-40 split, with more emphasis on the mornings), gives assignment times from the start of day in half-hour intervals, and saves one operator for the last shift of the day. In this thesis, we seek to improve this practice by developing systematic, data-driven methods to improve the performance and robustness of extraboard scheduling.

Chapter 4 The MBTA Context

Although the overall decision-making framework is applicable at all transit agencies, specific practices with respect to both data availability and operator scheduling vary. This chapter describes the context in which the application discussed in the following chapters of this thesis takes place. Section 4.1 outlines the scheduling practices at the MBTA and how they influence modelling choices; Section 4.2 describes the data sources and pre-processing procedures; Section 4.3 presents the descriptive analysis of the data to better understand the current state of the practice and the resulting performance.

4.1 Scheduling Practice at the MBTA

This section provides an overview of specific MBTA scheduling practices and how they influence data processing and model formulations. These scheduling practices may not apply to other agencies and need to be revised when applying the models to other agencies. The scheduling practices mentioned here are summarized again in later chapters when the models are presented.

1) Part-time vs. full time extraboard operators: In theory, the MBTA only has full-time extraboard operators. However, part-time extraboard operators may be assigned when there is a shortage of operators (for example, in rating 4 of 2017).
2) Split between known-in-advance and unexpected absence: The MBTA assigns full-day known-in-advance absences to extraboard operators (including splits and built-in overtime). Unexpected absences are covered by the cover list operators. An operator might cover multiple pieces of different runs, and a run might be covered by multiple operators. Since the report times for the cover list operators are assigned at 10am on the prior day, absences reported before 10am on the prior day are classified as known-in-advance, and absences reported later are unexpected. Partial-day known-in-advance absences are not assigned in advance, rather, they are left for the cover list, although anecdotal evidence suggests that partial-day known-in-advance absences are uncommon.
3) Report time assignment: After known-in-advance-absence run assignment, the remaining cover list operators will each be given a report time. At the MBTA, the report times of the cover list operators are based on the garage superintendent's judgement, with the general rule of thumb being that operators will be assigned to cover the first run and the last run in the day, with the rest assigned at half-hour intervals for the morning and afternoon peaks, with more emphasis in the mornings (about a 60-40 split).
4) Vacation: Vacations are mostly handled at the strategic and tactical level with a separate pool of vacation relief operators. Each year in November, operators with three (or more) weeks of vacation, pick their vacation weeks based on seniority and availability, and vacation relief

operators will be assigned to cover their work. Some vacation time in the unit of days can also be picked at the start of each rating and during the rating with valid reasons. These vacation days are covered by the extraboard. Therefore, in data processing, vacations in general were not counted towards operator absence but vacation days were included in operator absence since they need to be covered by the extraboard. Additionally, at times when there are more vacation relief operators than operators on vacation, the extra vacation relief operators are added to the extraboard.

5) Overtime acquisition: Overtime can be sought from both operators on their days-off and operators who are working on the day but not at the hour. In theory, overtime is the last resort to cover services. When there are no extraboard operators available, overtime is offered to regular operators in decreasing orders of seniority. The more senior operators get to decide first if they want to take the overtime. However, at times when the extraboard is too small, operators on their dayoff might be called in for overtime in anticipation of operator shortage. Overtime is paid at 1.5 times of the regular rate. The MBTA used to pay the operators at this 50% premium for all work hours outside of their picked schedule. To avoid overtime-induced absence (where the operators earn enough income at the start of the week through overtime and are more likely to be absent later in the week), the MBTA now pays the premium for additional hours after the operator has accumulated 40 hours of work in any given week.
6) Trippers: Trippers are pieces of work with a run number that are 2-3 hours long. They do not fit into any 8h or 6h shifts and are set aside to be covered by the cover list. In theory they could be covered by different operators each day, but sometimes a specific cover list operator will consistently take a tripper run.
7) Suspension and Attrition: During the rating, open work caused by suspension and attrition will be covered by the extraboard. During data processing, these absences, will be counted towards the open work caused by operator absence.

4.2 Data source and pre-processing

HASTUS™ is a workforce scheduling and management software suite widely adopted in the transit industry for planning, scheduling, operations, passenger information, and analysis. HASTUS™ Daily is a module in the HASTUS™ software suite that develops and monitors daily operator schedules and attendance. The software does not perform extraboard scheduling, but it provides rich data on the state of operations, aiding in the development of effective operator management strategies.

The data in this case study comes from rating 4 (September to December) of 2017, 2018, and 2019 at the Southampton Garage, as well as rating 4 of 2019 at Charlestown Garage at the

MBTA. The MBTA started deploying the software as a pilot at Southampton in December 2016 and rolled it out to all garages in February 2019. The data includes anonymized information by piece of work and scheduled operator, actual operator, start time and end time for each piece of work, and whether the piece is cancelled or covered. It is important to note that the data exports come from the pilot phase when the software was used as a backup system. As a result, consistency and reliability of the data is not guaranteed.

In order to derive useful information from the data, the starting point is to match the original owner (who picked the run) with the actual operator (who performed the run) and identify the status of both operators for each piece of work. When the two are different, the original owner status explains why the work is open (the demand side in the overall framework); and the actual operator status indicates how the work was covered (supply side of the overall framework). A sample classification (specific to the MBTA context) with the quantities in the analysis framework labelled, is shown in Figure 4-1.

One complication with the dataset used was the definition of 'original owner'. The intended definition of this field is the operator who picked the work at the start of the rating. That is, the original owner of any tripper should be empty, and the original owner of each run on a particular day of the week (except for special weeks that are picked separately, for example Thanksgiving week and Christmas week) should be the same person. However, the original owner field of the dataset does not always adhere to these definitions. Trippers, most of the time, have an original owner assigned and an operator could have different assignments for the same day of the week in different weeks. If we stick to using the original rosters as 'original owner', inconsistencies will result while processing other quantities. For example, the absence records will not match with inferred absences from the rosters and actual operator. Therefore, in order to be consistent, the column 'original owner' in HASTUS™ Daily was taken to be the definition of 'original owner'. Besides internal consistency, another argument for using this column in Daily is that updates made so much in advance that the operator is already updated as the original owner in the system are not in the scope of extraboard planning. The drawback is that there is not a consistent timeline on when the original owner field is updated. More consistent and reliable record-keeping is needed to eliminate the need for this assumption.

For open work caused by reasons other than absence, there is no information on whether or not it was essential. It is important to distinguish between essential and non-essential non-service duties because the essential ones are as important as scheduled service duties and should be estimated in conjunction with operator absence, but non-essential ones should have a lower priority. In the dataset, all non-service duties are coded as 'run as directed', 'work as directed', 'cover', or 'availability', and the use of these four codes was not consistent.

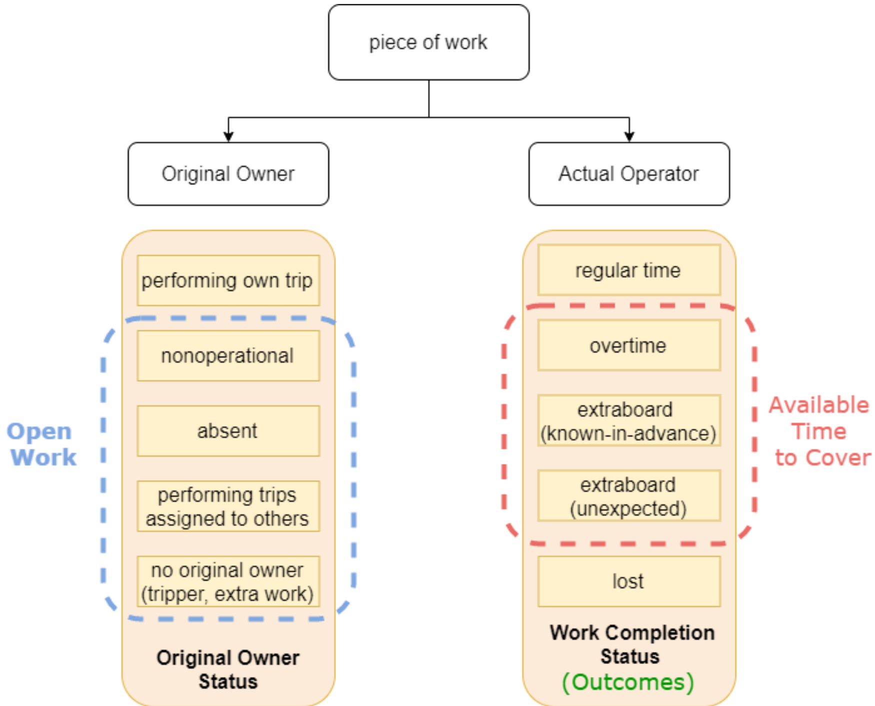


Figure 4-1 Data Pre-Processing

Additionally, sometimes operators were observed to perform trips assigned to others at the same time when (s)he has an assignment. This could happen when the trip they were originally assigned to was less important than the other trip. For example, dropping a trip on a high-frequency route might not be serious problem, but dropping a trip on a 30min headway route will leave passengers waiting for a very long time. Alternatively, during peak periods, when demand is high, missing a trip on a popular route likely leads to overcrowding and cause significant delays. Another route that serves an area with low ridership would be less important. However, these are based on judgments and are ad-hoc, and therefore difficult to anticipate.

In summary, operator absence is the biggest reason for open work (around 80%-85%, as shown in Table 4-2), and the other reasons (non-service duties, extra work, etc.) are difficult to isolate from the data and/or difficult to anticipate from a modeling perspective. Therefore, the rest of the thesis will focus on scheduling extraboard operators to deal with operator absence.

4.3 Descriptive Analysis

This section presents the garage performance statistics, and the analysis of empirical distributions and relationships among absence, overtime, lost service, and the extraboard. This section starts with an overview of the scheduled service in the analysis periods (4.3.1), then proceeds to analyze the day-of-week and time-of-day distributions of absence (4.3.2), the extraboard size (4.3.3), overtime (4.3.4), cover list utilization (4.3.5), and lost service (4.3.6). For all time-of-day analysis, data was aggregated into 10-min intervals and each service day runs from 4am to 1am the next day. On the plots, 1am the next day is plotted following 12am (hour 24) and labelled as hour 25.

4.3.1 Overall Garage/Rating Information

Table 4-1 shows the start and end dates and the number of days in the ratings analyzed, and the amount of scheduled service in each garage/rating. In subsequent sections, absence, overtime, and lost service will be shown as a percentage of scheduled service. Over the three-year analysis period, the amount of service provided has been increasing at Southampton. Charlestown had 50% more scheduled service than Southampton, concentrated on weekdays.

Table 4-1 Rating-Garage Overview

	2017 R4 Southampton	2018 R4 Southampton	2019 R4 Southampton	2019 R4 Charlestown
Start Date	09/03	09/02	09/01	09/01
End Date	12/30	12/29	12/21	12/21
# Days	119	119	112	112
Total Scheduled Service (hours)	86,040	98,030	93,650	140,615
Weekday Scheduled Service (hours)	804	903	914	1,523
Saturday Scheduled Service (hours)	581	689	702	827
Sunday Scheduled Service (hours)	511	611	617	479

Table 4-2 presents rating-level performance metrics derived from HASTUS™ Daily data. Absence is defined as the absences of all operators who are scheduled for service and need to be covered by extraboard operators; that is, absences for both non-service duties and the extraboard were not included in the rate, and long-term absences, where the work had been assigned to another operator, were not included. Additionally, since vacation weeks are covered by dedicated vacation relief operators, they are not included in absence, but single-day vacations are covered by the extraboard, and so are included. Note that the classification for absence (known-in-advance and unexpected) was derived from how absence was covered according to MBTA's scheduling practices outlined in Section 4.1, not by the actual notification time. Absences covered in full by one operator were classified as known-in-advance and absences covered by the cover list in multiple pieces, by overtime, or lost were classified as unexpected. Half-day absences were assigned to the cover list, regardless of the notification time, and therefore classified as unexpected. Since open work not only includes absence, the sum of hours covered by the extraboard, by overtime, and lost does not equal the number of absence hours.

Table 4-2 Rating Level Performance Metrics

	2017 R4 Southampton	2018 R4 Southampton	2019 R4 Southampton	2019 R4 Charlestown
Absence Rate	10.9%	12.6%	14.3 %	13.8%
Known-in-Advance Absence Hours	1,436 (1.7%)	4,787 (4.9%)	5,314 (5.2%)	9,314 (6.6%)
Unexpected Absence Hours	7,909 (9.2%)	7,598 (7.8%)	8,652 (9.1%)	10,039 (7.1%)
Absence as % all open work	82.0%	79.9%	84.3%	83.4%
Overtime Hours (% scheduled service hours)	4,454 (5.2%)	4,283 (4.4%)	4,788 (5.1%)	4,263 (3.0%)
Service Hours Covered by Known-in-Advance Covers (% scheduled service hours)	1,771 (2.1%)	6,817 (6.3%)	5,850 (6.3%)	10,548 (7.5%)
Service Hours Covered by the Cover List (% scheduled service hours)	2,414 (2.8%)	3,213 (3.3%)	3,023 (3.2%)	5,337 (3.8%)
Non-service Duty Hours by the Cover List	1,427	2,299	780	3,092
Cover List Service Utilization Rate[2]	63.4%	58.9%	58.2%	63.9%
Cover List Overall Utilization Rate[3]	74.4%	67.6%	67.9%	70.0%
Lost Service Hours (% scheduled service hours)	1,929 (2.2%)	1,161 (1.2%)	1,945 (2.1%)	2,124 (1.5%)

At Southampton, an increasing absence trend was observed. The absence level for Charlestown was similar to that for Southampton. In Southampton 2017, an unusually small amount of known-in-advance absences occurred (only 15% of all absence was classified as known-in-advance) possibly because of a combination of inaccuracies in record-keeping and the very small extraboard. In 2017, HASTUS™ was in its early pilot phase at the MBTA, so there was no guarantee of accurate record-keeping. Additionally, garages reserved some extraboard operators for the cover list. For example, a cover list operator was scheduled to cover the first run in the

[2] Cover List Service Utilization Rate = Service Hours Covered by the Cover List / (Service Hours + Non-service Hours Covered by the Cover List + Idle Time)

[3] Cover List Overall Utilization Rate = (Service Hours + Non-service Hours Covered by the Cover List) / (Service Hours + Covered by the Cover List + Idle time)

morning to ensure that it would be operated. If there were not enough operators, even when absences were notified beforehand, they would end up covered by the cover list, overtime, or lost and be classified as unexpected absence. In 2018 and 2019 at Southampton, the known-in-advance vs. unexpected split is 4:6 and in 2019 at Charlestown the split is around 5:5. Unexpected absence levels stayed relatively stable over the years (8-9% of all scheduled service). Absence accounted for around 80% - 85% of all open work in all analysis periods, making it an important quantity to estimate for extraboard planning.

Overtime at Southampton was similar across the three years (~5% of all scheduled service), but less was observed at Charlestown (3%) due to the higher percentage of all work done by the extraboard (~11% compared to ~10%).

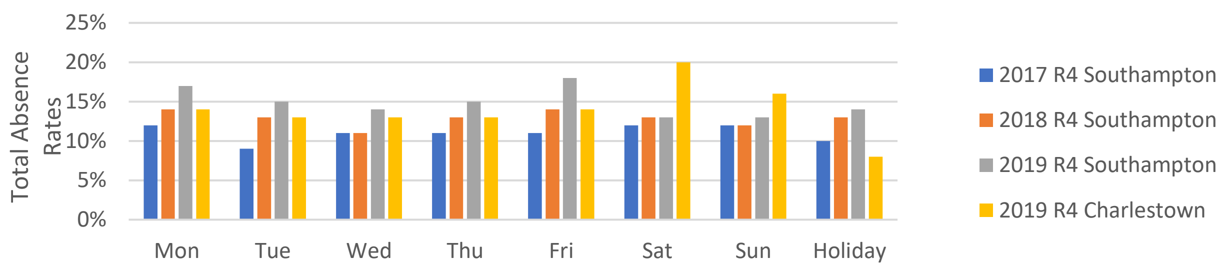
While the amount of work done by the extraboard was low in 2017 (4.9%) due to extraboard shortage, the amount of the work done by the extraboard was similar in 2018 and 2019 at Southampton (9.6% and 9.5%, respectively); and Charlestown had more work done by the extraboard (11.3%).

It is meaningless to calculate utilization rates for covers assigned for known-in-advance open work because the work is always covered with full efficiency. We can categorize the pieces of work that the cover list operators do as service, non-service duties, and idling. The service utilization rate is defined as the amount of time a cover list operator performing service divided by the duration that (s)he is scheduled for. The service utilization rates observed were stable, around 50% to 60%. The cover list operators also do non-service work, such as covering for the inspectors and transferring buses. These pieces of work were not included in service time but were included in the overall utilization rate. The overall utilizations rates were between 65% to 75%. The times that the operators have work coded as 'cover', 'availability', 'run as directed', or 'work as directed' were classified as idle. However, anecdotal evidence suggests that this is a weaker part of the data and sometimes work was done during those times but was not recorded in the system. Therefore, overall utilizations are likely to be higher than the numbers given. At Southampton in 2017, there were fewer cover list operators, and the utilization rate was (slightly) higher compared with other ratings/garages.

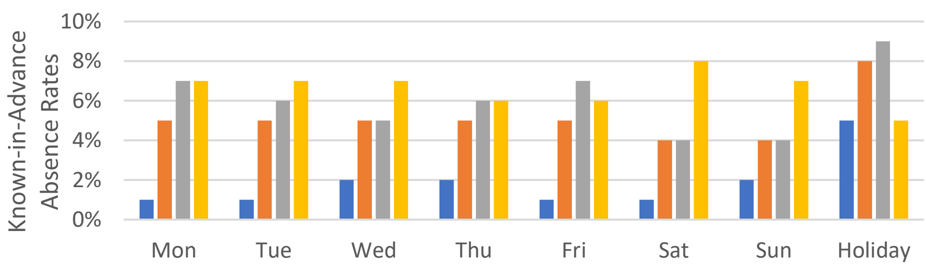
Lost service rate was around 1-2% and was dependent on the absence level, overtime performed, and cover hours. For example, at Southampton in 2017, the absence level was low, but the extraboard staffing level was also low, leading to a high rate of lost service. At Southampton in 2019, although the extraboard staffing level was raised and overtime performed was level, more absences were observed so the lost service rate was as high as 2017. Southampton in 2018 and Charlestown in 2019 had modest amounts of absence, and reasonable amounts of both overtime performed and extraboard hours, therefore less service was lost.

4.3.2 Absence

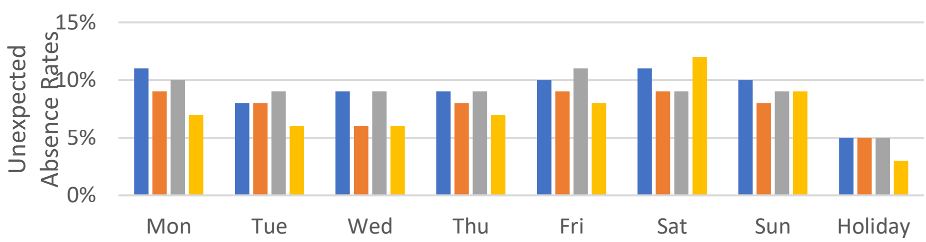
Figure 4-2 shows the absence rates for each garage/rating for different days of the week. Average absence rates (% of scheduled service) are plotted on the y-axis and different days of the week on the x-axis. Here rates are plotted instead of the number of hours, because on weekends and holidays, there were fewer hours scheduled, therefore the number of absence hours was naturally lower, making the comparison using absence hours potentially misleading.



a) Total Absence Rates



b) Known-in-Advance Absence Rates

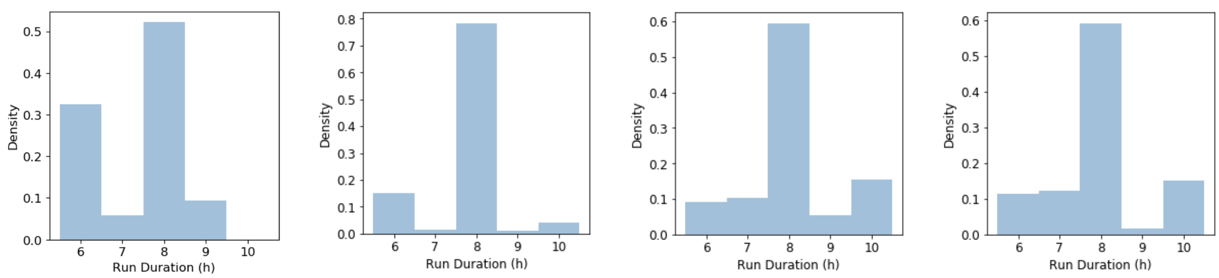


c) Unexpected Absence Rates

Figure 4-2 Absence Rates

Overall, the trends for different ratings/garages were similar: there was less absence on Tuesdays, Wednesdays, and Thursdays; higher absence levels on Mondays and Fridays; and the lowest absence rates on holidays. The difference between the highest and the lowest weekday was around 3%-4% of scheduled service hours. Additionally, at Charlestown, both known-in-advance and unexpected absences rates on weekends were high. At Southampton there were less known-in-advance absences on weekends. On most days, there were more unexpected absences than known-in-advance absences, except for holidays. This is probably due to more single-day vacations requested and granted on holidays and may require separate modelling considerations.

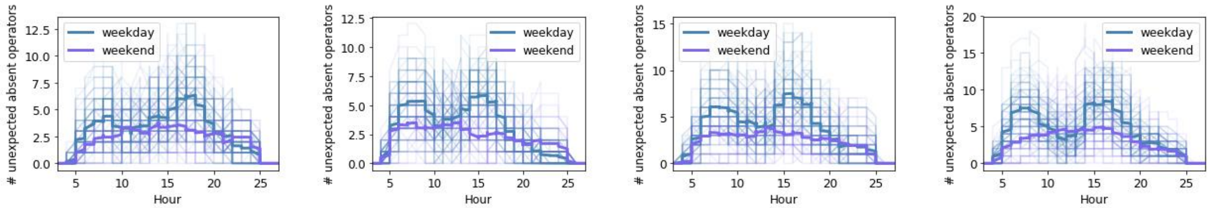
Figure 4-3 shows the distributions of shift durations of the operators who are classified as covering for known-in-advance absences. Usually, only full-time operators (8h shifts) are on the extraboard. For all ratings/garages, 8h shifts were the most common, along with 6h (part-time) and 10h (4-day week) shifts, confirming the validity of the data processing procedure. In 2017 at Southampton, there were a number of part-time operators on the extraboard, due to the significant shortage in extraboard workforce at the time. In the other ratings/ garages, part-timers covered for known-in-advance absences very occasionally, possibly due to special arrangements. Additionally, in ratings after 2017, 10h shifts were scheduled. Known-in-advance extraboard operators also cover those shifts in full, resulting in some 10h shifts in the figure.



a) 2017 R4 Southampton b) 2018 R4 Southampton c) 2019 R4 Southampton d) 2019 R4 Charlestown

Figure 4-3 Known-in-Advance Absence Duration

Figure 4-4 shows the time-of-day distribution of the number of operators covering unexpected absences. Each light blue line represents a day and the thicker line is the mean profile. Known-in-advance absences are excluded, because each is covered by a single operator, so only the number known-in-advance absences per day is of interest, not the time-of-day distribution. On average, the unexpected absence profile has a bi-modal distribution with morning and evening peaks on weekdays and a unimodal distribution on the weekends, and were similar across ratings/garages. Although the time-of-day distribution was highly variable, the average hourly absence was proportional to the amount of scheduled service, meaning that there were no particular times when operators tended to be absent more and the absence rates were similar across the day. At Charlestown, where there were significantly more weekday trips, the peaks in absence profiles were more pronounced than those at Southampton.



a) 2017 R4 Southampton b) 2018 R4 Southampton c) 2019 R4 Southampton d) 2019 R4 Charlestown

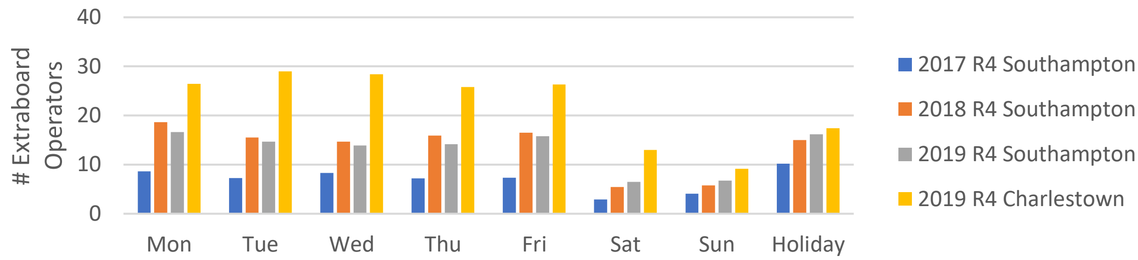
Figure 4-4 Unexpected Absence by Time of Day

4.3.3 The Extraboard

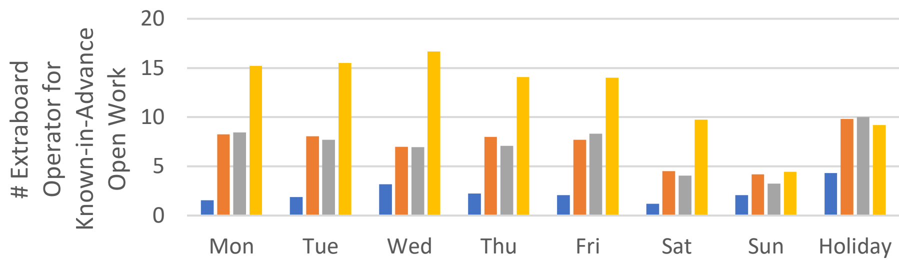
In a similar format to Section 4.3.2, Figure 4-5 shows the number of extraboard operators. Because of the larger amount of scheduled service, Charlestown had a larger extraboard than Southampton. At Southampton, there was a shortage of extraboard operators in 2017, but the numbers caught up in 2018 and 2019, consistent with the rating-level statistics. Absence rates for weekdays and weekends were similar but there was less service scheduled for the weekends.

Across all ratings/garages there were significantly fewer (close to 1/3) extraboard operators scheduled on the weekends, while the scheduled service on the weekends was around 70% and 40% of weekdays for Southampton and Charlestown, respectively. This mismatch might be due to the lack of knowledge on open work distribution, or the lack of extraboard operators, or that weekend trips were strategically given a lower priority. During the week, the number of extraboard operators scheduled generally matched the absence patterns at Southampton (lower on Tuesdays, Wednesdays, and Thursdays), but at Charlestown there were more extraboard operators scheduled on Tuesdays even though there was less absence.

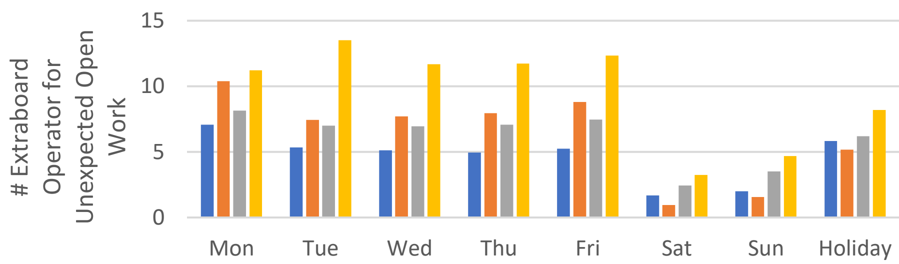
Figure 4-6 shows the time-of-day distribution of the scheduled cover list operators. Due to all cover list operators working straight runs while most regular runs were split, the time-of-day distribution of cover list operators did not closely match the absence profiles and exhibited a more muted bi-modal distribution. In order to match the absence distributions and improve utilization rates, there might be value in scheduling split runs for the cover list and this scenario is tested later in Section 6.6.2.



a) Total Extraboard Operators

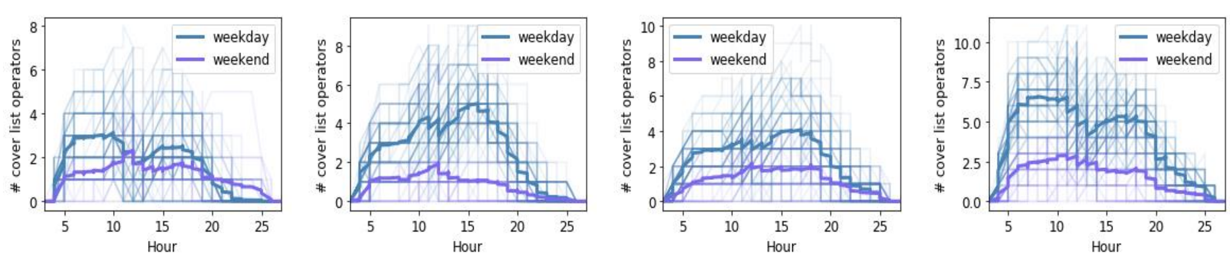


b) Extraboard Operators for Known-in-Advance Open Work



c) Extraboard Operators for Unexpected Open Work

Figure 4-5 Extraboard Operators by Day of Week



a) 2017 R4 Southampton b) 2018 R4 Southampton c) 2019 R4 Southampton d) 2019 R4 Charlestown

Figure 4-6 Time-of-day Distribution of Cover for Unexpected Open Work

4.3.4 Overtime Availability

Figure 4-7 shows the day-of-week patterns of overtime performed. Overtime performed is dependent on absence level and number of operators on the cover list, as well as overtime availability. Performed overtime was inversely correlated with the number of cover list operators. At Charlestown, although the absence levels were higher, overtime performed was lower than that at Southampton, because the size extraboard was also larger. The day-of-week distribution was similar across ratings/garages, with more overtime performed on Thursdays, Fridays and Saturdays, followed by Mondays, Tuesdays, and Wednesdays. Significantly less overtime was performed on Sundays and holidays.

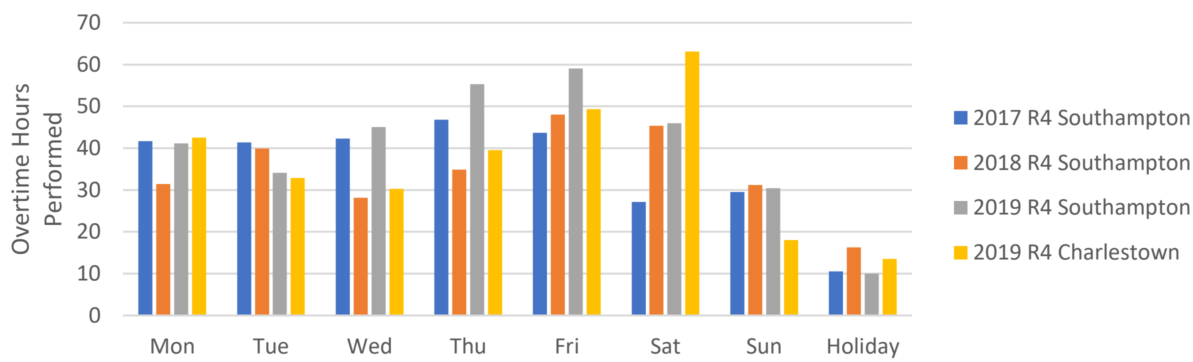
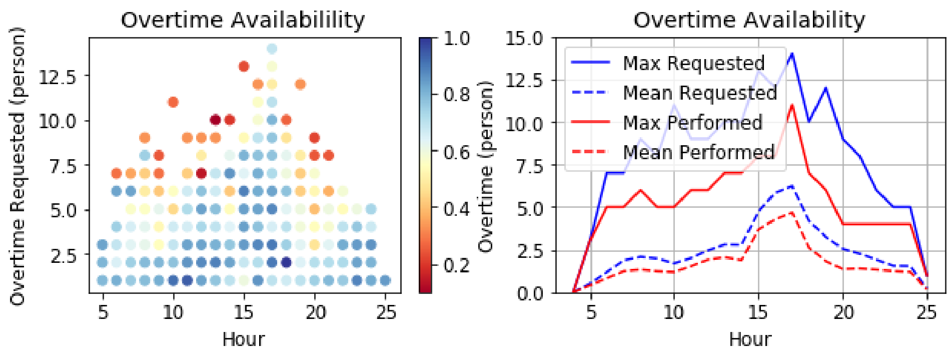


Figure 4-7 Overtime Hours by Day of Week

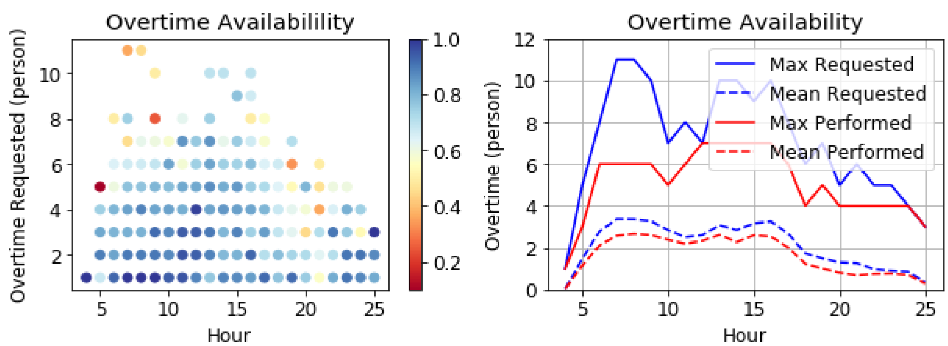
Figure 4-8 plots overtime requests and availability by hour for each rating/garage. In the plots, only overtime to provide service was included; overtime for non-service duties (such as flagging, and inspector duties) was excluded. Since overtime is the last resort to cover service, overtime requested is defined as overtime performed plus lost service. Fulfillment rate is defined as overtime performed divided by overtime requested. The scatter plot on the left shows overtime fulfillment with respect to the hour and number of operators requested. Each point on the plot represents the number of requests made in a particular hour and the color of the point represents the fulfillment rate (from 0% to 100%). A 0% fulfillment rate indicates no overtime is performed with all requests turning into lost service, and a 100% fulfillment rate indicates overtime requested at this hour equals overtime performed and no service is lost. When there are multiple day/hours with the same number of requests, the fulfillment rate shown is the average. The plot on the right shows the maximum and average amount of overtime requested and fulfilled by time of the day.

Similar to overtime performed, the amount of overtime requested is also not an independent quantity. It was derived from the amount of open work and the available cover list at different times. Therefore, the patterns for different ratings/garages differ. For example, Southampton in 2017 and Charlestown in 2019 showed a larger evening peak with respect to overtime requested. There was a higher evening peak in the absence pattern compare to the other two analysis periods

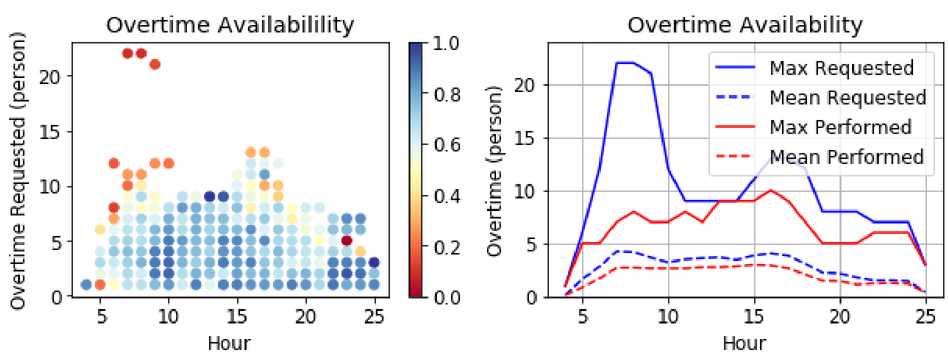
(Figure 4-4) that was not matched by the amount of cover scheduled in the evening peak (Figure 4-6). In general, the number of overtime requests was fewer at the start/end of the day and had a large range from morning peak to evening peak. For example, in mid-afternoon (around 3pm), both 0 requests and 15 requests had been observed on different days. In all cases, a decreasing fulfillment rate was observed with increasing requests at each hour, but the rate and threshold of the decrease were different for each rating/garage/hour.



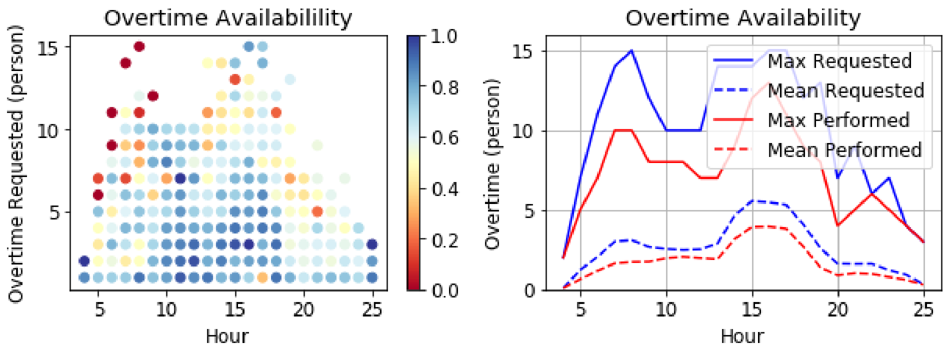
a) 2017 R4 @ Southampton



b) 2018 R4 @ Southampton



c) 2019 R4 @ Southampton

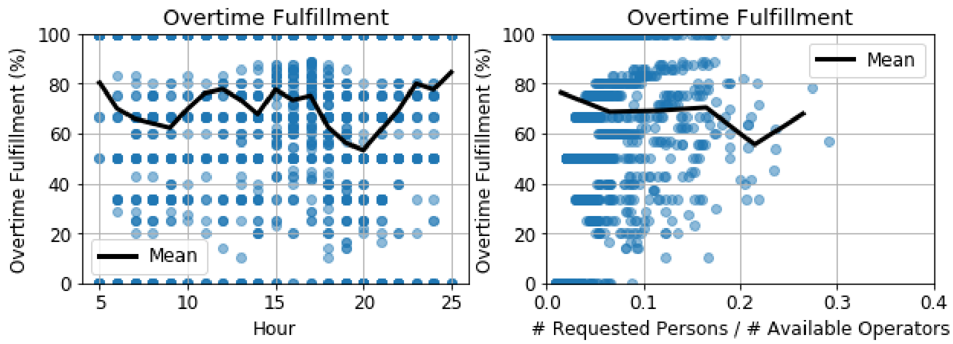


d) 2019 R4 @ Charlestown

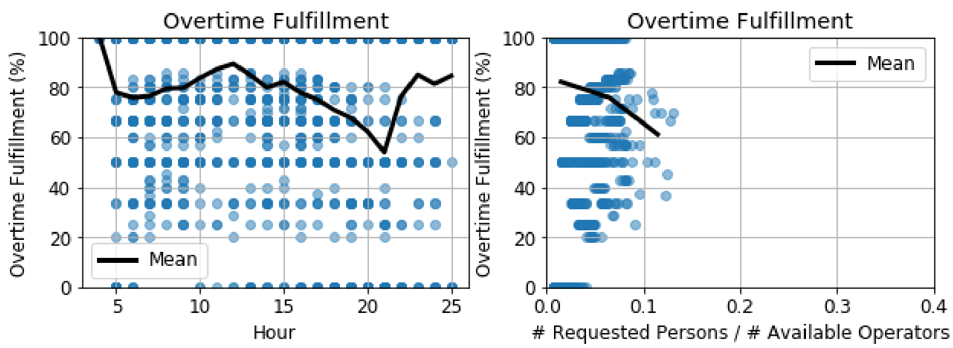
Figure 4-8 Overtime Requests and Availability by Hour

Figure 4-9 shows the relationship between overtime fulfillment rate and hour / number of potential overtime operators. Each point in the figure represents a day/10min period when overtime was requested. On the left, the fulfillment rate is plotted against the hour; on the right the fulfillment rate is plotted against the rate of request and number of potentially available operators. Potential overtime operators are defined as either operators who work on that day but not at that hour, or operators who have the day off but are present for work the previous day or the following day so potentially they could be called in for overtime.

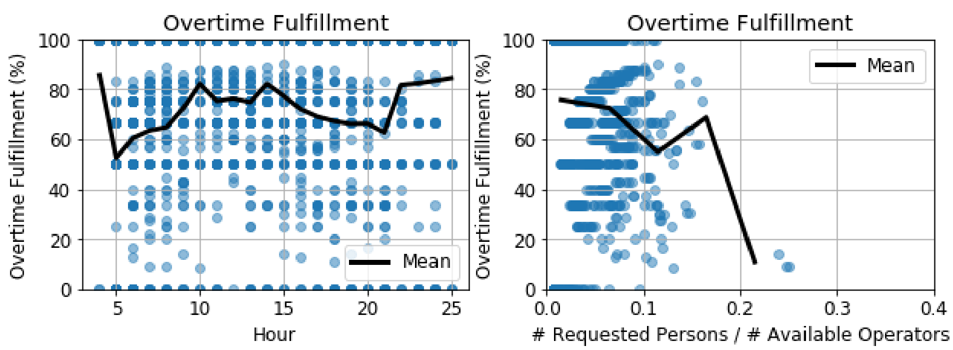
Both relationships for all ratings/garages were highly variable as the points are all over the figures, but we can draw some insights from the average value (plotted in black). During the early mornings and late evenings, the sample size was small since there were often no overtime requests (Figure 4-8); so the numbers were less reliable. On average, the overtime fulfillment rate was lower during the morning and evening peaks. One possible reason is that the number of potential operators being smaller (more operators are already working during peak hours). In the morning it is harder to call people in and in the evening peak there was usually a larger request (Figure 4-8), the combination also resulted in lower fulfillment rates during the peaks. The relationship between overtime fulfillment and request rate was also variable. In general, a higher request rate led to lower fulfillment, but this was more pronounced in some years than others. When the request rate went from 0% to 40%, at Southampton in 2017 the fulfillment rate barely decreased by a couple percentage points, whereas at Southampton in 2019 the fulfillment rate decreased from ~70% to ~30%.



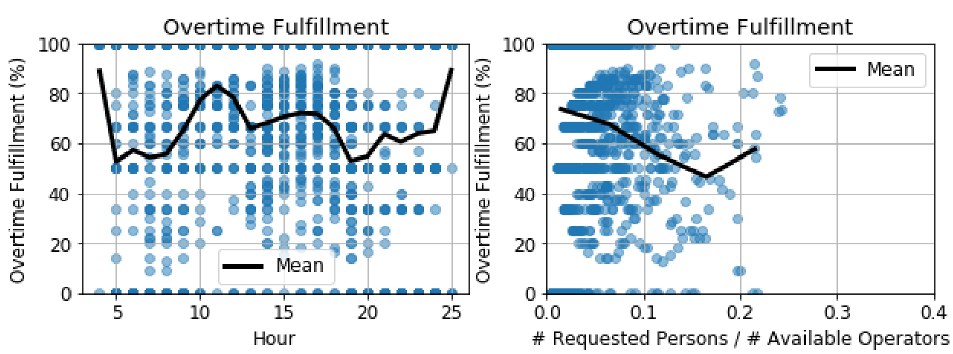
a) 2017 R4 @ Southampton



b) 2018 R4 @ Southampton



c) 2019 R4 @ Southampton



d) 2019 R4 @ Charlestown

Figure 4-9 Overtime Fulfillment w.r.t Hour and Rate of Requests

4.3.5 Cover List Utilization

Figure 4-10 shows the day-of-week cover list service utilization rates. Service utilization rates on Sundays and holidays were lower, possibly a result of less absence to cover. Other than that, there was no common pattern across the ratings/garages.

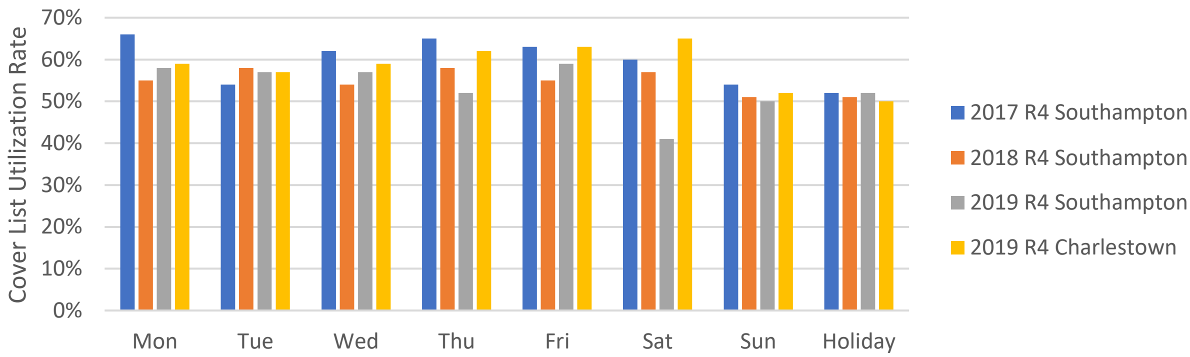
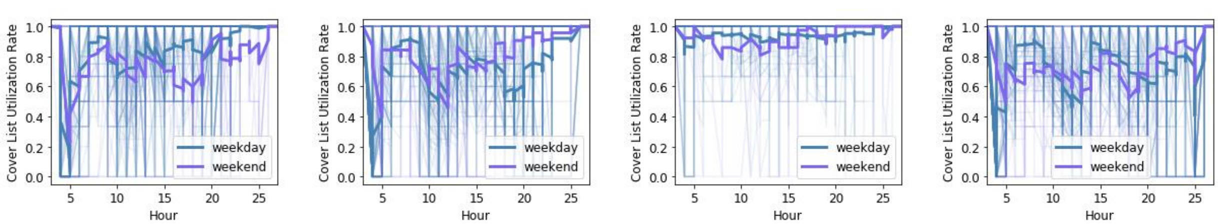


Figure 4-10 Cover List Service Utilization Rate by Day of Week

Figure 4-11 shows the time-of-day distribution of cover list service utilization rates. Similar to overtime fulfillment rates, in the early mornings and late evenings, there were fewer observations (usually one cover list operator was scheduled in both the early morning and late evenings), so the average rates were not consistent. The service utilization rate of the cover list was highest during the morning and evening peaks on weekdays due to there being more open work. A similar, but less pronounced pattern was observed on the weekend. The average service utilization rate was low in the early morning because the garage always scheduled one cover list operator for the first shift in the morning to ensure that it will always run. If the assigned regular operator shows up, there will not another trip for the cover list operator to do, resulting in very low utilization. Options could be explored around this scheduling practice to improve the utilization rate of the cover list operators.



a) 2017 R4 Southampton b) 2018 R4 Southampton c) 2019 R4 Southampton d) 2019 R4 Charlestown

Figure 4-11 Time-of-day Distribution of Cover List Service Utilization Rate

To investigate the assignment of service and non-service duties for the cover list operators, the non-service duty hours for the cover list in Table 4-2 was further broken down with the results shown in Table 4-3. The record-keeping of the work performed for these duties was known to be inconsistent at the MBTA. For a piece of work labelled as ‘run as directed’, which should mean standing by and doing nothing, the operator could be called to do something else, while the records may not get updated. Therefore, all pieces of work done by the cover list that were not service trips were included in this table.

Table 4-3 Breakdown of Non-Service Duty hours

	2017 R4 Southampton	2018 R4 Southampton	2019 R4 Southampton	2019 R4 Charlestown
Non-Service Duty Hours by the Cover List (% cover list hours[4])	1,427 (39.6%)	2,299 (44.2%)	2,248 (45.0%)	3,092 (40.9%)
- **Deadhead**	363	416	368	390
- **Other**	1,064	1,882	1,880	2702
○ **With Lost Service (% all Non-Service Duty Hours by the Cover List)**	256 (18.0%)	352(15.3%)	845 (37.6%)	652 (21.1%)
○ **With Lost Service and/or Overtime (% all Non-Service Duty Hours by the Cover List)**	717 (50.2%)	1,458 (63.4%)	1,561 (69.4%)	1,636 (52.9%)

Non-service duties make up a significant portion of all cover list hours (~40%). While the operators are performing non-service duties, oftentimes services were lost (15%-40%) or overtime was requested to cover service trips (50%-70%). This suggests that some of the non-service duties were essential and/or unexpected absences occurred when the cover list operators had been dispatched for non-service duties and could not be called back for service. However, due to the lack of detailed, reliable records on what was done during these non-service hours and what was essential and what was optional, the non-service hours could not be further classified into the two reasons stated above.

[4] This rate is 1-Cover List Overall Utilization Rate

4.3.6 Lost Service

Lost service is derived from the interaction between open work, the extraboard, and overtime performed. When there was open work and nobody was available to cover, service will be lost. Figure 4-12 shows the day-of-week distribution of the lost service rate. For all ratings/garages, similar trends were observed with the least proportion of service being lost during the mid-week. Mondays, Fridays, and weekends had significantly higher lost service rates than Tuesdays, Wednesdays and Thursdays. It is worth noting that Saturdays and Sundays have a reduced schedule than weekdays, therefore a similar (or higher) number of lost service hours meant the rate of lost service was much higher. On holidays, significantly less service was lost due to lower absence rates and a similar amount of resources (extraboard and overtime) to cover should open work occur.

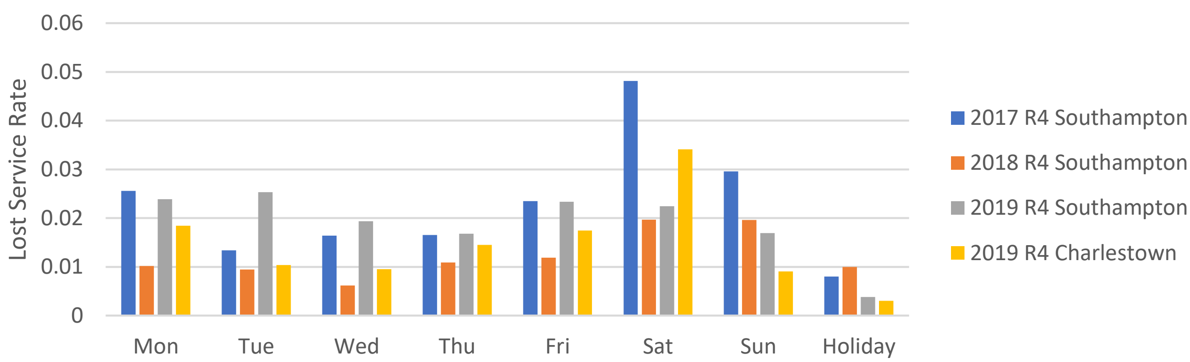
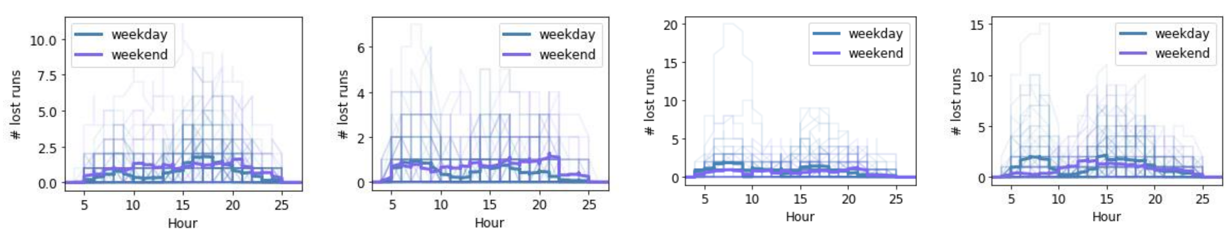


Figure 4-12 Lost Service Rate by Days of Week

Figure 4-13 shows the time-of-day distribution of lost service hours. Across all ratings/garages, distributions with large spreads were observed. The worst case was very bad compared to the average case. On weekdays there were higher lost service hours in the morning and evening peaks, with the distribution on weekends being more uniform.



a) 2017 R4 Southampton b) 2018 R4 Southampton c) 2019 R4 Southampton d) 2019 R4 Charlestown

Figure 4-13 Time-of-day Distribution of Lost Service

In summary, most of the rating-level, day-of-week, and time-of-day patterns of absence and the extraboard are generalizable across ratings and garages. Overtime and lost service were derived from absence and extraboard scheduling, and therefore were more variable. Among weekdays, there was less absence, fewer extraboard operators scheduled, and less lost service on Tuesdays and Wednesdays; rates of absences were similar on the weekends but due to less scheduled service the hours needing to be covered was smaller, the number of extraboard operator scheduled was also smaller on the weekends, resulting in higher lost service rates. Regarding the time-of-day distributions, although highly variable, both average absence and lost service profiles resembled the distribution of scheduled service; due to the straight-run requirement, the modes in the bi-modal distribution of the number of extraboard operators available were less pronounced. Observations at Southampton in 2017 were quite different due both to it being the early pilot of the software, as well as the shortage of extraboard operators.

Chapter 5 Absence Models

This chapter explores one of the most important inputs into extraboard scheduling: operator absence. At the tactical level, we need to determine the number of extraboard operators assigned to each day of the week with only probabilistic information about the amount of open work which will need to be covered. At the operational level, we need to determine the report times for unassigned extraboard operators with only probabilistic information about the time-of-day distribution of unexpected open work. Since most of the need for the extraboard comes from covering absences, developing an accurate model to forecast absence levels on different days of the week, and at different times of day provides valuable information on the demand for extraboard operators, and thus is important in extraboard scheduling.

Section 5.1 discusses the absence modelling problem. Section 5.2 gives an overview of models that could be used and their respective advantages and disadvantages. Sections 5.3 - 5.5 present the model formulations. Section 5.6 presents a case study using data from the MBTA to explore the models' use and effectiveness.

5.1 Issues in Modelling Absence

1) Categorization of Absence

In light of the way the notification time for absence affects how absences are covered, we categorize absences into two types: known-in-advance and unexpected. At the tactical level, both types are unknown. At the operational level, known-in-advance absence is known but unexpected absence remains uncertain.

Known-in-advance absence gives the system (the scheduler as well as operators) enough time that an entire run can be assigned to one operator. Therefore, these full-run absences can be covered without loss of efficiency, as long as there are enough extraboard operators to cover them. If, there are not enough extraboard operators, or if some extraboard operators need to be reserved for other reasons (for example, the MBTA typically reserves two extraboard operators to cover the first and last trips of the day at each garage), known-in-advance absences could end up being covered the same way as unexpected absences. Whether this relationship needs to be modelled depends on the agency scheduling practices and on data availability. Additionally, partial-day known-in-advance absences are not assigned in advance, rather, they are left for the cover list. But anecdotal evidence suggests that partial-day known-in-advance absences are uncommon. In this case study, single-day vacations are categorized as known-in-advance absence since they are allocated to the extraboard. However, the underlying factors explaining single-day vacations will likely be different from those for other absences. If the data permits it would be helpful to model single-day vacation separate from known-in-advance absence.

After assignment of operators to cover known-in-advance absences, the remaining extraboard operators are referred to as the cover list and assigned report times at which they become available to cover unexpected absences for the following 8 hours. Unexpected absences are the ones that managers have no knowledge of at the time of scheduling because they arise after the cover list operators' report times are set. Cover list report times are then determined based on anticipated unexpected absences and overtime availability. When the reality differs significantly from the expectation, losses in the forms of unutilized cover list time, high levels of overtime performed, and lost services can occur. At any time of day, if we schedule for less absence than actually occurs, overtime will be requested, and when overtime is not available, service will be lost. If we schedule for more absence than actually occurs, some extraboard operators will have no productive work to do while getting paid, resulting in higher costs than necessary.

Besides this difference in coverage efficiency, the differences in the duration and time-of-day distributions for known-in-advance and unexpected absences can also be significant. Known-in-advance absences are usually full-day absences and an extraboard operator will inherit all the characteristics of the run including the splits, breaks, and built-in-overtime. The time-of-day distribution follows that of the scheduled runs and because of the one-to-one relationship between scheduled operator duties and extraboard assignments, we are not concerned about the time-of-day distribution of known-in-advance absences for the purpose of extraboard scheduling. However, unexpected absences can occur for only a piece of a run and different pieces could be covered by different operators. Therefore, different pieces of a run could have different outcomes: covered by a cover list operator, covered on overtime, or lost. The time-of-day distribution of unexpected absences is of interest, as the operational level is concerned with matching the time-of-day distribution of unexpected absences with the availability of extraboard operators, which is governed by their report times.

2) Reasons for Absence

There are different reasons for absence. Common reasons include medical leave (self or family), sickness, disability, work rules, vacation, etc. Individual operator characteristics (whether (s)he has a habit of being absent, and job commitment), work characteristics (whether the piece of work is attractive or convenient), and contextual information (for example, when the overall economy is weak, operators may work harder to keep their jobs) can all influence operators' motivation and willlingness to attend. Because the psychology underpinning absence is complex, even when the data includes absence classification by type, the real behavioral intent behind the decision-making process often remains unknown. Psychological factors are a function of the individual, the work, and the environment and are mostly subjective, and difficult to measure.

Therefore, although psychological factors may play a large role in determining absence, they are not easily incorporated into a model to forecast absence.

3) Unit of Analysis

Being absent from work is a time-sensitive, individual-specific decision. While models are an abstraction of reality, the level of abstraction depends on the intended application as well as on data availability. Below we discuss two aspects of abstraction: aggregation and time resolution.

To account for heterogeneity across individuals, we would prefer a disaggregate model. For example, estimating the probability of absence for each individual is a common approach in disaggregate analyses of absence. Disaggregate models require detailed data at the individual level to model the population heterogeneity. In contrast, modelling absence at an aggregate level ignores differences among individuals and models absence hours for a group of individuals. Aggregate models sacrifice individual details but require far less data and are easier to estimate with limited data. In this thesis, the underlying behavioral determinants of absence are beyond our scope and having aggregate absence information is sufficient to make scheduling decisions. Unless data on individual operator characteristics is available and can significantly improve model fit, an aggregate model is appropriate for our purposes.

Absence could be modelled at the time-of-day, daily, weekly, monthly, or annual levels. While more detailed models require more information, depending on the relationships being explored, more detailed models may not necessarily be better. Although time-of-day observations could be summed to daily, weekly, and annual data, it is not the same with modelled results since modelling with different temporal units captures different levels of information. For example, a time-of-day model is useful to explore the correlations between absence at different times of day but is not helpful in characterizing long-term trends. For the purpose of extraboard scheduling, particularly at the operational level, absence by time-of-day is needed, and the long-term trends are less relevant since the analysis should be repeated every season (season, rating, and timetable are used interchangeably in this thesis).

To summarize, in modelling absence for extraboard scheduling at the tactical and operational levels, there are a few things to note: 1) Known-in-advance absence and unexpected absence should be modelled separately. 2) Behavioral factors may be important in modelling absence, but they are difficult, if not impossible, to incorporate into a forecasting model, given the available data. 3) Depending on the application and on data availability, disaggregate and aggregate models each have their advantages and disadvantages. For the purpose of operational-level scheduling, the time resolution should be as high as possible.

5.2 Approaches to Modelling Absence

In this section, we briefly describe alternative absence modelling approaches and discuss their advantages and disadvantages. In general, we can split the approaches into disaggregate and aggregate. In our context, disaggregate models consider differences among individuals and/or runs, whereas aggregate models include only temporal information and exclude individual characteristics. Disaggregate models usually estimate a probability of absence for each individual, run or time period, and overall absence levels can be estimated by aggregating the probabilities. Aggregate models directly estimate the number of absences at a particular time. Figure 5-1 shows the categorization of these alternative modelling approaches.

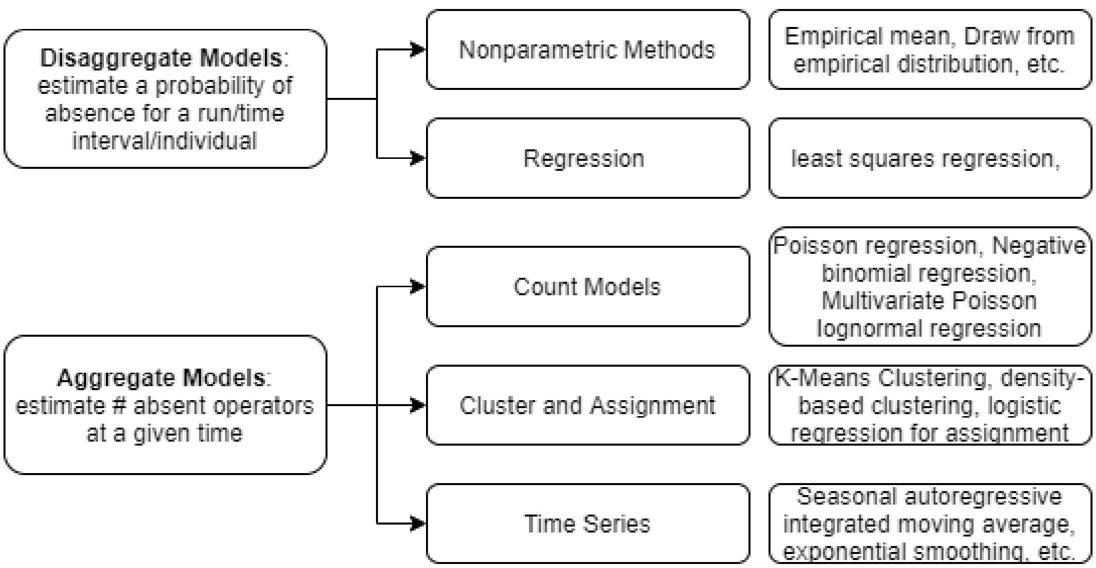


Figure 5-1 Types of Absence Models

For disaggregate models, we could use either nonparametric methods, such as a decision tree, or parametric methods, such as regression to estimate the probability of absence for each operator. For example, if absence is driven by individual circumstances and motivation, we could calculate a probability of absence for each individual based on past performance and aggregate to group performance based on schedule assignment. If the run characteristics outweigh individual differences, we could estimate an absence rate for each run, regardless of the operator. It is hard to verify the assumptions made and these methods are vulnerable to outliers – we need an adequate number of observations in each subdivision to ensure that the rate is representative. When more information, such as demographic characteristics for individual operators is available, regression is usually used for a more comprehensive analysis (Shiftan and Wilson, 1994; Wang and Gupta, 2014).

In aggregate models, depending on the temporal unit of analysis, the data will exhibit different characteristics, and more model choices exist. First, the number of absent operators is integer-valued and non-negative, and the occurrence of absences can be assumed to be random given a rate. This problem description suggests use of count models. Poisson regression, which assumes equi-dispersion (expected value and variance are equal), is the most basic form of count model, but many variations exist to account for different data distributions and structures. In Poisson regression, it is assumed that the rate parameter λ for the Poisson process is deterministic and all the differences in rates are explained by the explanatory variables. If this is not realistic, then we could model the rate parameter with an unobserved random factor η with the resulting rate being $\eta\lambda$. When η has a gamma distribution, $\eta\lambda$ will have a negative binomial distribution (a mixture of Poisson and Gamma). Negative binomial regression can be viewed as a form of Poisson regression that includes a random component reflecting the uncertainty around the true rate (unobserved heterogeneity). Additionally, if we want to model data with a panel structure (for example, correlations among counts at different times of day), we could employ multivariate count models. They are harder to estimate (no closed form solution and more parameters), but they are capable of modelling correlations across days and times of day in the same model.

If the time-of-day distribution of absence is regular and follows a distinctive pattern such as a uniform or bimodal distribution, or follows the pattern of the scheduled work, we could employ a cluster and assign approach. Empirical distributions can be fed into a clustering algorithm to identify representative profiles and then a classification method, such as logistic regression, can be used to assign any given day to the appropriate profile. The advantage of this approach is that it is highly interpretable, since we can directly observe the representative time-of-day distributions. The disadvantage is that two models are needed and it does not work well if the distributions are evenly spread out rather than falling into distinct clusters.

Lastly, the structure of the problem suggests a time-series approach if the most influential factor is time and other factors influencing absence are unobserved. Time series models can accommodate seasonal effects, moving averages, autoregressive error terms, etc. The advantage of time series models is that they can be intuitive and easy to implement, but they do not make use of information other than time and thus have limited explanatory power.

Table 5-1 summarizes and compares the model types. Different model types can be chosen, based on the application and data availability and characteristics. In most cases, disaggregate models will not be feasible due to data limitations, and time-series models have limited modelling capability. Therefore, count or cluster-and-assign models are generally recommended. In the next two sections, we describe in detail two specific models that are used to estimate absence at the tactical and operational levels for extraboard planning in the MBTA case study.

Table 5-1 Comparison of Absence Model Types

Type	Modeling Approach	Advantage	Disadvantage
Disaggregate	Non-parametric Methods	Easy to implement, can model heterogeneity among individuals	Sensitive to outliers, some categories may have very few observations
	Regression	Easy to implement, can model heterogeneity among individuals	Often difficult to obtain required explanatory variables
Aggregate	Count models	Match the problem (nonnegative and integer), can accommodate different data patterns and temporal units of analysis	More versatile models are difficult to implement and require larger sample sizes
	Cluster and Assignment	Intuitive, interpretable, easy to implement	Requires separate models; can only be used when there are a few distinctive patterns
	Time Series	Intuitive, interpretable	Limited modelling capabilities

5.3 Known-in-Advance Absence Prediction with Negative Binomial Regression

Formulation

For known-in-advance absence, we only need to estimate the number of runs for each day since all such absences are full-day runs, each of which will be assigned to a single operator. The operator will cover the full run; therefore, we do not need to know at what times of the day these absences occur.

Known-in-advance absences are modelled as a count-based process using negative binomial regression. The assumption is that the observed value follows a negative binomial distribution given the fitted value. Similar to Poisson regression, absences are assumed to be generated by a memoryless Poisson process, but we account for unobserved heterogeneity and treat the true rate of the process $\tilde{\lambda} = \lambda$ as a random variable. Because of this uncertainty around the true parameter, the observed value assumes a larger variance than in a Poisson regression. The error term ϵ could have many different distributions, but it is mathematically convenient to assume that it has a gamma distribution, in which case the observations follow a negative binomial distribution. For a through discussion of Poisson and negative binomial regression, please refer to Gardner and Mulvey (1995).

We adapt the standard negative binomial formulation to explain differences between rates of absence. The number of known-in-advance absences is directly observed. The count not only depends on the characteristics of the day and the scheduled operators, but also the number of runs scheduled. With more runs scheduled, there is more exposure and the count of absences is likely to be higher. In the formulation, d_i is the number of scheduled runs on day i, introduced as exposure. Let λ_i be the absence rate, then the expected number of absences is $d_i\lambda_i$. Therefore, the regression applies to the rate, making the model more generalizable in the face of service level changes. In mathematical terms, the problem can be formulated as follows.

Define y_i = Number of known-in-advance absences on day i

X_i = Independent variables for day i

β = Regression coefficients

N = Sample size

K = Number of independent variables

λ_i = Absence rate on day i (# absences / # scheduled runs)

d_i = # scheduled runs on day i

where

$$y_i \sim N\text{egBin}(\mu_i, \alpha) \tag{5-1}$$

$$E(y_i) = \mu_i = d_i\lambda_i = d_i\exp(X_i^T\beta) = \exp\left(X_i^T\beta + ln(d_i)\right)$$

$$\text{var}(y_i) = \mu_i + \frac{\mu_i^2}{\alpha}$$

The logarithmic link function transforms μ into a linear combination of the predictors and the log of d_i:

$$\ln(\mu_i) = \ln(d_i) + X_i^T\beta$$

Estimation and Inference

1. Estimate β: The python package statsmodels is used to estimate the model (Seabold and Perktold, 2010). Package statsmodels.glm estimates generalized linear models with an offset, but does not estimate the dispersion α. Since maximum likelihood estimators of β for the linear exponential family (which includes Negative Binomial II) are consistent regardless of the dispersion, we can estimate β and α separately (Gardner and Mulvey, 1995).

2. Estimate α: Given observations y_i with estimated mean μ_i, variance of observations = $\widehat{var(y_i}|\widehat{\mu}_i) = (y_i - \widehat{\mu}_i)^2$. Variance of a negative binomial random vaiable is of the form: $\mu_i + \alpha^{-1}\mu_i^2$. We can find α^{-1} using least squares (with no intercept):

$$(y_i - \mu_i)^2 - \mu_i = \alpha^{-1}\mu_i^2$$

3. Re-estimate β: If $\alpha \leq 0$, then there is no overdispersion in the data, terminate and use the estimated Poisson model. Otherwise, input α as a parameter in the estimation of β. The values of β should not change much as the estimators are consistent regardless of the dispersion.

4. Check dispersion/scale: Calculate observed variance and compare it to the variance assumed by the distribution. $\hat{\phi}$ should be 1 if the distributions match perfectly. A value greater than 1 indicates that there is over-dispersion and there remains a component of y variance that the negative binomial regression model did not capture.

$$\hat{\phi} = \frac{1}{N-J}\sum_{i=1}^{N}\frac{(y_i - \hat{\mu_i})^2}{\mu_i + \alpha^{-1}\widehat{\mu_i}^2}$$

5. Correct for dispersion to make inference: Variance of $\hat{\beta}$ is given by

$$\widehat{Var}(\hat{\beta}) = \left(\sum_{i=1}^{N}\frac{\hat{\mu_i}}{1+\alpha^{-1}\hat{\mu_i}}X_i X_i^{'}\right)^{-1}$$

Assuming that the conditional variance is a linear function of the conditional mean, the corrected $\widehat{Var}(\hat{\beta})$ is given by:

$$\widehat{Var}(\hat{\beta})_{corrected} = \hat{\phi}\left(\sum_{i=1}^{N}\frac{\hat{\mu_i}}{1+\alpha^{-1}\hat{\mu_i}}X_i X_i^{'}\right)^{-1}$$

Evaluation

To evaluate model performance, the standard mean absolute error (MAE) and mean absolute percent error (MAPE) are used, as well as Theil's coefficient which is commonly used to evaluate the fit of time series model (Bliemel, 1973). For an observed time series Y and a fitted time series X, each with n elements, Theil's U is defined as

$$\mathrm{U} = \frac{\sqrt{\frac{1}{n}\sum_i (X_i - Y_i)^2}}{\sqrt{\frac{1}{n}\sum_i X_i^2} + \sqrt{\frac{1}{n}\sum_i Y_i^2}}$$

where U ranges between [0,1]. The smaller the U, the better the fit.

5.4 Unexpected Absence Prediction with Negative Binomial Regression

In this section, we introduce a combination of methods that could be used to estimate unexpected absences. We separate the estimation of the total number of absence hours and their time-of-day distribution. First, the total number of unexpected absence hours is estimated using negative binomial regression. Then representative time-of-day profiles are found with the average time-of-day profile for each day of the week. Although it involves multiple steps, this procedure is intuitive, has relatively few parameters to estimate, and gives interpretable results.

5.4.1 Modelling Daily Unexpected Absence Hours

The model has the same formulation and estimation procedure as that discussed in Section 5.3, except that we substitute number of unexpected absence hours for the number of known-in-advance absences, and the rate of unexpected absence is defined as the unexpected absence hours divided by the total scheduled hours, since unexpected absences can have different lengths.

5.4.2 Modelling Time-of-Day Distribution of Absence

The time-of-day distribution is defined as the number of absent operators for each hour divided by the total number of unexpected absence operator-hours for each day. Based on previous research, the time-of-day distribution of unexpected absence hours has only a small impact on the cover list scheduling results (Kaysi and Wilson, 1990). Descriptive results in Section 4.3.2 suggest that the time-of-day distribution of unexpected absence hours is similar to that of scheduled services. Therefore, a simple approach is used: the time-of-day distributions for each day of the week were averaged and used for prediction. Each day of the week will have a different predicted profile, but will be the same across different weeks in the rating. Holidays are classified based on whether they follow a weekday, Saturday, or Sunday schedule (for example, Columbus Day has a weekday schedule, but Labor Day has a Sunday schedule). For the case study data, this simple approach achieves similar performance to the theoretically more powerful cluster-and-assign approach.

5.5 Unexpected Absence Prediction with Multivariate Analysis

In this section, a model that deals with the day-of-week and time-of-day correlations simultaneously is proposed. First, we recognize that absences at different times on the same day are likely to be correlated since each absence is likely to extend over consecutive time periods. Correlations may also exist between absence rates on different days of the week. The temporal characteristic variables explicitly account for this correlation as well as the absence rates from previous time periods.

Following the standard notation for multivariate data, the counts are organized into a matrix where the indices i and j denote the day (i) and time (j). The parameters β are time-period-specific and a set of day-and-hour-specific latent effects b is included to model the correlations between counts at different times of the day. Conditioned on latent effects b_i and parameters β_j, we assume that the counts y_{ij} follow independent Poisson distributions and the latent effects follow multivariate normal distributions. Under these conditions, the resulting model is a Multivariate Poisson Log-Normal Model (Chib and Winkelmann, 2001). This model has been most commonly applied in car accident count modelling (Bai et al., 2011; Ma et al., 2008; Wang et al., 2018; Zhan et al., 2015; Zhao et al., 2018). We now discuss the details of this model, which primarily uses the methods developed by Chib and Winkelmann (Chib and Winkelmann, 2001), with added regularization on the estimation of D using Bayesian Graphical LASSO (Wang, 2012).

Specific formulation details for our problem can be found below. Note that when we specify one subscript of a multidimensional matrix, the rest is kept in place. For example, $\beta_j \in R^K$ and $X_i \in R^{J\times K}$.

Define N = Sample size (days)

J = Number of time periods

K = Number of independent variables

y = Total number of absence hours (day i, time j) $\in R^{N\times J}$

X = Independent variables (day i, time j, attribute k) $\in R^{N\times J\times K}$

β = Regression coefficients (time j, attribute k) $\in R^{J\times K}$

μ_{ij} = Absence rate on day i at time j

b_i = Latent effects to model correlation among the time periods $\in R^J$

where $y_{ij}|b\_i,\ \beta\_j\ \sim Poisson(\mu_{ij})$

$$\mu_{ij} = \exp\left(X_{ij}^T\beta_j + b_{ij}\right)$$

We assume $b_i \sim \mathcal{N}_J(0, D)$, where D is an unrestricted covariance matrix. To understand features of this model, let $\nu_{ij} = \exp\left(b_{ij}\right)$ and $\nu_i = \left(\nu_{i1}, \nu_{i2}, \dots, \nu_{iJ}\right)$, then $\nu_i \sim LogNormal(\mu, \Sigma)$ with mean $\mu = \exp\left(0.5diag(D)\right)$and covariance $\Sigma = \left(diag(\mu)\right)[exp(D - 11^T)]\left(diag(\mu)\right)$.

Given this, the expectation and covariance of the observations can be derived using the law of iterated expectation as follows:

$$\mathrm{E}\left[y_i\middle|\beta_j, D\right] = \mu_i = \exp\left(X_{ij}^{T\beta_j}\right)\exp\left(0.5diag(D)\right)$$

$$\mathrm{var}\left[y_i\middle|\beta_j, D\right] = \mathrm{diag}(\mu_i) + \mathrm{diag}(\mu_i)[exp(D - 11^T)]\mathrm{diag}(\mu_i)$$

Hence the covariances between the counts are

$$\mathrm{cov}(y_{ij}, y_{ik}) = \mu_{ij}(exp(D_{jk}) - 1)\mu_{ik}$$

which can be positive or negative depending on the signs of D_{jk}.

Estimation and Inference

The likelihood function of the model is

$$\mathrm{P}(y_i|\beta, D) = \int_{b_i} \prod_{j=1}^{J} f(y_{ij}|\beta_j, b_{ij}) \phi_J(b_i|0, D) d_{b_i}$$

where f is the density function of the Poisson distribution and ϕ_J is the density function of a J-variate normal distribution. We cannot use maximum likelihood methods because we cannot solve this multiple integral in closed form. Instead, we resort to simulation-based methods.

The main idea of simulation-based methods is to develop a MCMC chain where the limiting invariant of the chain is the posterior distribution of the parameters of the model β, b, and D. One standard way to construct such a Markov chain is Gibbs sampler, which samples from one full conditional density at a time. Gibbs sampler is used when the joint distribution of β, b, and D is difficult to sample from but the conditional distribution of each variable conditioned on all other variables and the data is easier to sample from. Each full conditional density in the simulation is sampled either directly (if the full conditional density belongs to a known family of distributions) or by utilizing the Metropolis–Hastings (M–H) algorithm. In this case, the posterior distribution of the parameters is

$$\mathrm{P}(\beta, D^{-1}|y, x) \propto f(\beta)\mathrm{f}(D^{-1})\mathrm{P}(y_i|\beta, D)$$

The augmented posterior density of the parameters β, D and the latent effects b can be written as

$$\mathrm{P}(\beta, b, D^{-1}|y, x) \propto f(\beta) f(D^{-1}) \prod_{i=1}^{N} P(y_i|\beta, b_i)\phi_J(b_i|0, D)$$

where $f(\beta), f(D^{-1})$are the prior distributions of β, and D. The full conditionals $\mathrm{P}(b|y, \beta, D)$, $\mathrm{P}(\beta|y, b, D)$, and $\mathrm{P}(D^{-1}|y, b, \beta)$ are

$$\mathrm{P}(b|y, \beta, D) = \prod_{i=1}^{N} P(y_i|\beta, b_i)\phi_J(b_i|0, D) \; \propto \prod_{i=1}^{N} \phi_J(b_i|0, D) \prod_{j=1}^{J} exp(-\mu_{ij})\mu_{ij}^{y_{ij}}$$

$$\mathrm{P}(\beta|y,b,D) \propto f(\beta)\prod_{i=1}^{N}\prod_{j=1}^{J} exp(-\mu_{ij})\mu_{ij}^{y_{ij}}$$

$$\mathrm{P}(D^{-1}|y,b,\beta) = \mathrm{f}(D^{-1})\prod_{i=1}^{N}\phi_J(b_i|0,D)$$

The parameters will be sampled sequentially as follows:

1. Sampling b

To sample from the target density $\prod_{i=1}^{N}\phi_J(b_i|0,D)\prod_{j=1}^{J} exp(-\mu_{ij})\mu_{ij}^{y_{ij}}$, we utilize the M–H algorithm. For more information regarding the M-H algorithm and its use in estimating Poisson models, please refer to the papers by Chib, Greenberg and Winkelmann (Chib et al., 1998; Chiband and Greenberg, 1995). The proposal density is found by approximating the target density around the modal value by a multivariate-t distribution with

Mean $\quad \widehat{b_i} = \mathrm{argmax}\ln\prod_{i=1}^{N}\phi_J(b_i|0,D)\prod_{j=1}^{J} exp(-\mu_{ij})\mu_{ij}^{y_{ij}}$

Variance $\quad V_{b_i} = \left(-H_{b_i}\right)^{-1}$

The modal value $\widehat{b_i}$ and Hessian H_{b_i} can be found by any gradient-based solvers with

Gradient $b_i \quad = -D^{-1}b_i + y_i\exp - ({x_i}^T\beta + b_i)$

Hessian $b_i \quad = -D^{-1}\mathrm{diag} - exp\left(({x_i}^T\beta + b_i)\right)$

2. Sampling β

A normal prior is imposed with mean β_0 and variance B_0. Both β_0 and B_0 are hyperparameters. The target density $\mathrm{f}(\beta)\prod_{i=1}^{N}\prod_{j=1}^{J} exp(-\mu_{ij})\mu_{ij}^{y_{ij}}$ is for $J \times K\ \beta$ parameters at once. However, since the dimension of β is large, we would have a high proportion of rejections. Therefore, we assume that the β_j's are independent of each other and can be sampled one j at a time (B_0 is block-diagonal). The posterior distribution can be sampled by an M-H algorithm similar to that for sampling b's.

For a given j, the proposal density is found by approximating the target density around the modal value by a multivariate-t distribution with

Mean $\quad \widehat{\beta_j} = \mathrm{argmax}\ln\phi_K\left(\beta_j|\beta_0,B_0\right)\prod_{i=1}^{N} exp(-\mu_{ij})\mu_{ij}^{y_{ij}}$

Variance $\quad V_{\beta_j} = \left(-H_{\beta_j}\right)^{-1}$

The modal value $\widehat{\beta_j}$ and Hessian H_{β_j} can be found by any gradient-based solvers with

Gradient $\beta_j \quad = -B_{0_j}\left(\beta_j - \beta_{0_j}\right) + \sum_{i=1}^{N}(y_{ij} - \mu_{ij})(x_{ij})$

Hessian $\beta_j \quad = -B_{0_j} - \sum_{i=1}^{N} \mu_{ij} x_{ij} x_{ij}^{T}$

3. Sampling D^{-1}

D is the covariance matrix and D^{-1} is the precision matrix of b_i's. The target density is $f(D^{-1}) \prod_{i=1}^{N} \phi_J(b_i|0, D)$. Sampling the precision matrix of a Gaussian distribution has long been a standalone topic of discussion in both frequentist and Bayesian statistics. A conjugate Wishart prior is usually used to form a simple and fast procedure for computing the analytic posterior of the precision matrix. Wishart priors are also used in the MVPLN literature (Ma et al., 2008; Wang et al., 2018; Zhan et al., 2015; Zhao et al., 2018). However, in crash injury analyses, the injury level (J) is usually no greater than 5. In our context, we have a large number of time periods (if the temporal unit of analysis is hours, then J = 20 since transit service only operates for around 20 hours a day). The dimension of D is 20 × 20. We have around 120 days ($N \approx 120$) in a rating. The sample size is small compared to the dimensionality of the parameters. Therefore, we seek ways to regularize the precision matrix. Here we used a Gibbs sampler based on a Bayesian Graphical LASSO model (Wang, 2012).

Given observations Y (N × J data matrix), the graphical lasso problem is to maximize the penalized log-likelihood with respect to the precision matrix Ω

$$log(det\boldsymbol{\Omega}) - tr(\frac{S}{n}\boldsymbol{\Omega}) - \rho||\boldsymbol{\Omega}||_1$$

where $S = Y^TY$ and $\rho \geq 0$ is the shrinkage parameter that shrinks the entries of Ω to 0. The graphical lasso problem also has a Bayesian interpretation (Wang, 2012); the graphical lasso estimator is equivalent to the maximum *a posteriori* estimation of the following model

$$\mathrm{P}(y_i|\Omega) = \mathcal{N}(y_i|0, \Omega^{-1})$$

$$\mathrm{P}(\Omega|\lambda) = C^{-1} \prod_{i<j} DE\left(\omega_{ij}|\lambda\right) \prod_{i=1}^{J} \{EXP(\omega_{ii}|\lambda/2)\} 1_{\Omega \in M_+}$$

where the posterior distribution is re-parameterized by λ with the double exponential (DE) and exponential (EXP) density functions. For any fixed values of $\lambda \geq 0$, the posterior mode of Ω is the graphical lasso estimate with $\rho = \lambda/N$.

We observe that if D^{-1} only depends on b, then the posterior distribution $\mathrm{P}(D^{-1}|y, b, \beta)$ is equivalent to $\mathrm{P}(D^{-1}|b)$, where $b_i \sim \mathcal{N}_{\mathcal{J}}(0, D^{-1})$. Therefore, we could parameterize $\mathrm{P}(D^{-1}|b)$ using λ and use Gibbs sampler to estimate D^{-1}.

5.6 The MBTA Case Study

This section presents a case study using MBTA's data on absence applying the models discussed in prior sections of this chapter. The case study aims to explore the effectiveness of the models, interpret the model results, and assess the strengths and limitations of each model. Section 5.6.1 describes the case study and introduces the variables involved in the estimation. Sections 5.6.2, 5.6.3, and 5.6.4 present the model results for known-in-advance absences and unexpected absence using negative binomial regression, and unexpected absence using MVPLN.

5.6.1 Data Description

Using the processed data described in Section 4.2, we built and tested models for the Southampton garage for Rating 4 (September to December) of 2017, 2018, and 2019 and Charlestown garage for Rating 4 of 2019. For each model, detailed results and interpretation will be presented for Southampton garage for 2019 (Rating 4). Having data from previous years for the same rating and the same garage enables us to study the stability of the models over time. Since tactical level scheduling occurs at the beginning of each rating and we have no access to data from the current year, we can only apply the model from prior years. Models trained from prior year's data will be used to forecast 2019 absence levels and the forecast will be compared with actual observations to see how absence behavior changed over time and how well the models generalize. The benchmark error, which is the best error possible using a model class, was taken to be the error from the model retrospectively fitted with observed data (train and test with data for the same period).

Separate models were built for known-in-advance and unexpected absences. On some days, some known-in-advance absences were covered as if they were unexpected absences because of extraboard operator shortage (especially in 2017 R4 @ Southampton) although the data does not facilitate isolation of these instances. For scheduling, we aim to model how absences are covered given the resource constraints, not the notification time of absences. Given a small extraboard, some known-in-advance absences will need to be covered as if they were unexpected absences. Instead of one operator covering a complete run when enough extraboard operators are available, the run would potentially be split into pieces to be covered by multiple operators and some service could be lost. This relationship is accounted for by including both known-in-advance absence and unexpected absence in the model formulation.

In general, the variables included can be divided into four categories: temporal characteristics, reference quantities, interaction terms, and predicted quantities:

Temporal characteristics are dummy variables indicating the time-of-day, day-of-week, and special days such as holidays. At the MBTA, holiday work is paid at a premium: every operator is paid 8 hours of holiday pay whether they are scheduled to work that day or not, and hours worked on holidays are paid separately at straight time rate (1x). However, in order to receive the holiday pay (the 8 hours of pay without working), the contract requires the operator to be present the day preceding, the day of (if scheduled), and the day following the holiday. Vacation and personal days that are scheduled at the start of the rating can be excused from this requirement. For operators who are not scheduled for work on the day preceding or following, this rule does not apply. For example, for a Monday holiday, in addition to being present on Monday if scheduled, an operator also needs to be present for both the Sunday and Tuesday in order to receive holiday pay, but for those who are scheduled to have the weekend off, they only need to be present for the Tuesday, and Monday if scheduled. Therefore, absence rates on and around a holiday are expected to be lower than on the corresponding day in a normal week.

Reference quantities are historical hours of certain types of absence and overtime performed. Overtime is included because more overtime performed might lead to increased levels of absence because of burnout, as well as its effect on any income target. Quantities from two days ago are used, because at the MBTA, the operational level scheduling (setting report times for the cover list) happens at 10am the previous day, when the operational statistics for that day are not yet available.

Intercept terms are added to account for the different effects that different types of reference quantities have on absence on different days of week. These terms add an intercept to specific scenario combinations. For example, although each operator has different days-off based on his/her roster, the weekends have lower service levels, meaning that many operators still have the weekends off. Operators who are sick on Thursday might also consider taking Friday off to get a longer weekend, but this behavior would not happen if the operator is sick on Monday and there are four more workdays ahead.

Predicted Quantities: although not feasible for scheduling, total absence on the previous day was tested and found to be significant. Therefore, we include a predicted total for the previous day in the full model.

In subsequent sections, the base model refers to a model that includes only the temporal characteristics specified in Table 5-2. The base model variables are the same for both known-in-advance absence estimation and unexpected absence estimation since they describe the characteristics of the day. The base model can be used for tactical-level scheduling since no operational information enters the formulation.

Table 5-2 Base Model Variables

Variable	Rationale
Temporal Characteristics	
Dummy for day of week	to account for day-of-week variation
Week number	to account for within-rating trends
Dummy for holiday	to account for the holiday effect
Dummy for holiday extension	1 day before and after a holiday

Besides the base model, additional variables including recent operational-level information are tested in the full model. Table 5-3 describes the rationale for the additional variables tested for the estimation of known-in-advance absence. At the tactical level, the operational-level information is not available. At the operational level, known-in-advance absences are assigned to the extraboard without uncertainty. Therefore, this full model will not be used for scheduling, rather, it can provide insights into the relationships between these variables and known-in-advance absence.

Table 5-3 Additional Variables for Known-in-Advance Absence Estimation

Variable	Rationale
Reference Quantities (1 day prior unless otherwise specified)	
Average overtime hours in the previous week	Based on the hypothesis that too much overtime might lead to operator fatigue. At the same time, working overtime might help the operator reach his/her income target, (s)he is more likely to be absent when income targets are met.
Known-in-advance absence hours	Based on the hypothesis that known-in-advance absences might be autoregressive.
Daily unexpected absence rates	Defined as the daily total unexpected absence hours divided by the total hours of scheduled service. Based on the hypothesis that known-in-advance absences might be related to unexpected absence.

Table 5-4 describes the rationale for the additional variables tested for the full model estimation of unexpected absences. The full model for unexpected absence could be used at the operational level since the report times of the cover list are determined on a daily basis. The reference quantities are for two days prior because at the time of report-time scheduling, the numbers for the previous day are not yet available.

Table 5-4 Additional Variables for Unexpected Absences

Variable	Rationale
Reference Quantities (2 days prior unless otherwise specified)	
Daily unexpected absence rate	Defined as the daily total unexpected absence hours divided by the hours of scheduled service. Based on the hypothesis that unexpected absence might be autoregressive.
Known-in-advance absence rate	Defined as the daily known-in-advance absence runs divided by the number of runs scheduled. Based on the hypothesis that known-in-advance absence may be related to unexpected absence.
Sick absence rate	Based on the hypothesis that sickness might persist over several days
Absence rate related to the Family and Medical Leave Act	Based on the hypothesis that absence for individual and family medical leave might persist over several days
Average overtime hours in the previous week	Based on the hypothesis that overtime might lead to operator fatigue. At the same time, working overtime might help the operator reach his /her income target, (s)he is more likely to be absent when income targets are met.
Overtime hours	same reason as above
Intercept Terms	
Day of week x Reference Quantities	
Predicted Quantities	
Predicted absence hours	for the day before

In subsequent sections, model results are presented in detail for Southampton for 2019; results for the other years and garage are presented in Appendix A and Appendix B, respectively.

5.6.2 Known-in-Advance Absence Prediction with Negative Binomial Regression

This section presents the results from applying the model described in Section 5.3 to estimate known-in-advance absences.

Estimation Results

Table 5-5 shows the model coefficients (with significance level), standard error, and p-value for the models estimated with Southampton's 2019 rating 4 data. Each β_j is the elasticity of x_j with respect to y. For small values of β_j, β_j is approximately the change in y caused by a unit change in x_j. Since the total number of scheduled runs are included as the exposure term, the coefficients represent the elasticity of the explanatory variables with respect to the known-in-advance absence rates.

Table 5-5 Regression Coefficients Known-in-Advance Absence 2019 R4 @ Southampton

	Base Model			Full Model		
	Coefficients (β_j)	Standard Error	p-value	Coefficients (β_j)	Standard Error	p-value
$1/\alpha$	0.12			0.07		
constant	-3.264 ***	0.185	0	-3.1401 ***	0.251	0
holiday	0.0449	0.348	0.897	-0.2394	0.311	0.442
holiday extension	0.2886	0.235	0.22	0.3714	0.221	0.093
week	0.0521 ***	0.012	0	0.0409 **	0.015	0.007
Tue	-0.0736	0.207	0.722	-0.2617	0.188	0.163
Wed	-0.1867	0.209	0.373	-0.4224 *	0.193	0.029
Thu	-0.0247	0.201	0.902	-0.1773	0.18	0.326
Fri	0.0933	0.197	0.636	-0.1079	0.178	0.545
Sat	-0.467 *	0.232	0.044	-0.842 ***	0.229	0
Sun	-0.4701 *	0.241	0.050	-0.5417 *	0.218	0.013
average overtime hours in the previous week				-0.0099	0.005	0.06
daily unexpected absence rate (day of)				0.0086	0.017	0.613
known-in-advance absence (day before)				0.0795 ***	0.016	0
Log Likelihood	-238.43			-227.35		

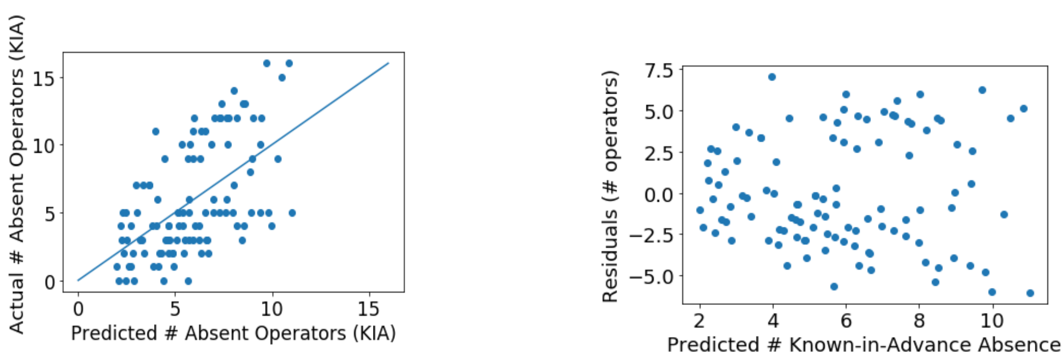
Table notes: *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001.

The log likelihood for the full model is higher than that of the base model, indicating a better overall fit. The log likelihood ratio test statistic is 2*((-227.35) – (-238.43)) = 22.16, the critical χ^2 statistic with degree of freedom 3 at 95% significance level is 7.81. The full model passes the log likelihood test for being a better model. The dispersion factor was 0.12 for the base model and 0.07 for the full model. The full model has less over-dispersion (unobserved heterogeneity) than the base model. The coefficients which are significant at the 95% confidence level are highlighted in the table. The insignificant coefficients are more likely to be unstable and sometimes have opposite effects in the base and full model (e.g. holidays and Fridays in this case). But in general, the coefficients are consistent with the results of the descriptive analysis. For completeness, the insignificant variables were not removed from the final model applied for scheduling in Chapter 6.

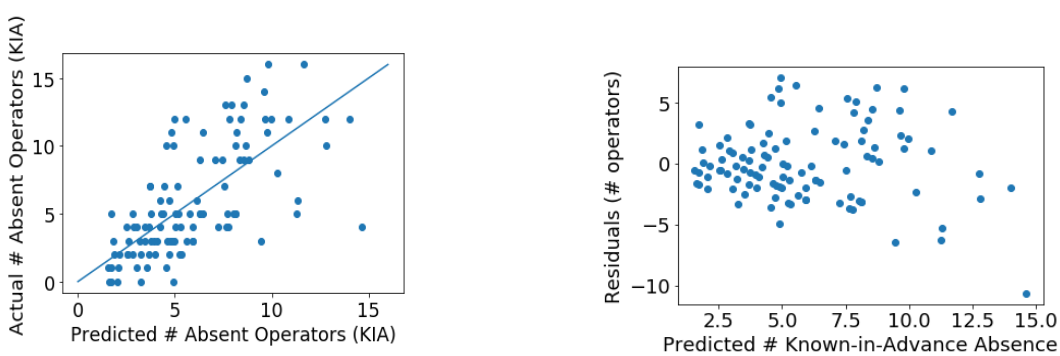
Neither the holidays nor the days around the holidays have a significant impact in this rating on known-in-advance absences. The signs for the coefficients for holidays change between the base model and the full model. For this rating, descriptive analysis suggests that the holidays have higher rates than others, which is inconsistent with the results from the full model. The effect would probably be better captured with quantifying the number of vacation days amongst all known-in-advance absences. The model mainly reflects the average behavior for different days of the week with a weekly trend. There was an increase in absence over the course of the rating, indicated by a positive coefficient for the 'week' variable. The reference day of week is Monday. Compared to Monday, only Friday and holidays had a (insignificant and unstable) higher number of known-in-advance absences and the weekends had the lowest rates, which agrees with the observations in Section 4.3.2.

Two of the three additional variables are insignificant. The effect of overtime in this model was counter-intuitive, the model suggested that more overtime leads to less known-in-advance absence. In the other three ratings, coefficients for rolling overtime were also insignificant, but positive (see Appendix A). Unexpected absences on the day was not a significant indicator for known-in-advance absences across all ratings. On the other hand, known-in-advance absences for the previous day was significant and positive for all ratings, indicating a strong autoregressive effect for known-in-advance absences at an aggregate level.

Validation



a) Base Model



b) Full Model

Figure 5-2 Negative Binomial Regression Results for Known-in-Advance Absences (2019 R4 @ Southampton)

Figure 5-2 shows the number of predicted vs actual known-in-advance absences (left) along with the residual plot (right) for both the base and the full model. The residual plot exhibits spacing patterns of predicted numbers due to most of the explanatory variables being binary. Although the full model has a generally better fit, there was one outlier with an error > 10 operators.

Table 5-6 shows the model fit for each rating/garage. On average, the model prediction is off by 1-3 operators per day. Although the absolute error is lowest for 2017 R4 for Southampton, its percentage error is especially high due to the low reported numbers of known-in-advance absences. Charlestown, having more known-in-advance absences (Figure 4-2: similar absence numbers but much more service), a similar MAE results in much smaller MAPE. The full models, with more variables, yield better results than the base models, but the improvement differs across ratings (ranging from 1% to 9%). Similar observations can be made from Theil's coefficient, where Charlestown had the best fit and the difference in predictive power between the base and full models was minor.

Table 5-6 Negative Binomial Regression Error for Number of Known-in-Advance Absences

		Base			**Full**		
Time Period	Garage	MAE (# operators)	MAPE	Theil's Coefficient	MAE (# operators)	MAPE	Theil's Coefficient
2017 R4	Southampton	1.35	75.8%	0.38	1.23	69.1%	0.34
2018 R4	Southampton	1.80	33.8%	0.20	1.68	31.5%	0.19
2019 R4	Southampton	2.86	47.5%	0.24	2.35	39.0%	0.21
2019 R4	Charlestown	2.17	19.6%	0.12	2.07	18.6%	0.11

Model Transferability

Table 5-7 Model Coefficients Comparison

Variable	**Southampton**			**Charlestown**
	2017 R4	2018 R4	2019 R4	2019R4
average overtime hours in the previous week	0.0183	0.0032	-0.0099	0.0004
daily unexpected absence rate (day of)	-0.0204	-0.0235	0.0086	-0.0035
known-in-advance absence (day before)	0.1215 ***	0.0529 **	0.0795 ***	0.0399 **

Transferability is studied from two perspectives: significance of model coefficients and model performance. Table 5-7 shows the values and significance of the variable coefficients for the additional variables. Only the additional variables from the full model are investigated here as the coefficients for the temporal characteristic variables follow the patterns established in the descriptive analysis (Figure 4-2). In all models, known-in-advance absences have strong autoregressive effects and are not related to overtime or unexpected absences. The degree of dependency varies both spatially and temporally.

Next, transferability was studied by evaluating the performance of the model on data from a different period than the model was trained on. Since 2017 data at Southampton was very different from 2018 and 2019, it did not make sense to try to generalize a model trained on 2017 data to other years. Instead, half of 2019's data at Southampton was used for training and the other half for testing to see the effect of pooling data. Since absence rates are modelled, the model was robust against changes in service levels. Table 5-8 presents the transferability results for both the base and full models. The benchmark was taken as the result from the best fit model for the testing data: in this case, the full model trained on the testing data.

Table 5-8 Transferability Results for Known-in-advance Absences

Training Data	Testing Data	Base Model MAE	Full Model MAE	Benchmark MAE
2018 R4 Southampton	*2019 R4 Southampton*	*3.29*	*2.91*	*2.35*
2018 R4 Southampton	*2019 R4 (second half) Southampton*	*3.39*	*3.22*	*3.07*
2018 R4 + 2019 R4 (first half) Southampton	*2019 R4 (second half) Southampton*	*3.36*	*3.02*	*3.07*

The models yielded a mean error of ±1 operator per day compared to the benchmark. Temporally, the full model generalized better than the base model, since the relationships found for the added variables were similar (Table 5-7). In this case, the improvement from pooling data was negligible in the base model (0.9%), but significant in the full model (6.7%). More data is needed to test the effects of pooling data. Since the model mainly reflected weekly trends and day of week differences, it is important to do the descriptive analysis before generalizing the model.

Limitations

The estimated model was affected by the limited number of explanatory variables and the stochastic nature of absence behavior. Since only temporal variables were available, the model is unable to predict variations beyond day-of-week and holidays. On the other hand, tactical-level scheduling rules do not allow different numbers of extraboard operators for each individual day. Most of the time, for the same day of week, the number of extraboard operators assigned will be the same and the trips not covered by extraboard operators will need to be covered on overtime.

5.6.3 Unexpected Absence Prediction with Negative Binomial Regression

Results from the various models to predict unexpected absences discussed in Section 5.4 are presented here.

Estimation Results

Because of the large number of potential variables listed in Table 5-4, for each dataset, as well as all years and ratings combined, variable selection used the log likelihood test with a 95% confidence level. Variables are sorted based on prior belief in their importance and tested sequentially using the log likelihood test with degree of freedom 1, and the variables that passed the test were included in the final model (Table 5-10). If the confidence level was reduced to 90%, more variables would be significant, but a variable was still significant for at most two ratings.

Table 5-9 Variable Selection Unexpected Absence Regression (90% C.I.)

	Southampton			**Charlestown**	
	2017 R4	2018 R4	2019 R4	2019 R4	All
unexpected absence rate (1 day ago)			√	√	√
unexpected absence rate (2 days ago) x holiday extension	√				√
unexpected absence rate (2 days ago)	√				√
unexpected absence rate (1 week ago)					√
having weekday schedule	√	√			
known-in-advance absence rate (day of)		√			
known-in-advance absence rate (1 day ago)	√		√		
known-in-advance absence rate (1 day ago) x Saturday	√		√		√
known-in-advance absence rate (1 day ago) x Sunday			√	√	
sick absence rate (2 days ago)	√				
sick absence rate (2 days ago) x holiday extension			√		
sick absence rate (2 days ago) x Saturday	√				
FMLA absence rates (2 days ago) x holiday extension	√				√
FMLA absence rates (2 days ago) x Saturday		√			
FMLA absence rates (2 days ago) x Sunday		√			√
rolling overtime hours previous week x holiday extension		√	√		√
rolling overtime hours previous week x Sunday	√				

Table 5-9 shows the selected variables at the 90% confidence level for all ratings/garages. To summarize, the variables significant for two ratings were unexpected absence rates (previous day), having a weekday schedule, known-in-advance absence rates (previous day), known-in-advance absence rates (previous day) interacted with Saturday and Sunday, and rolling overtime interacted extended holidays. Overall, there was not a universal predictive significance pattern regarding the other variables tested.

Table 5-10 Negative Binomial Regression Coefficients for Unexpected Absence 2019 R4 @ Southampton

	Base Model			**Full Model**		
	Coefficient (β_j)	Standard Error	p-value	Coefficient (β_j)	Standard Error	p-value
1/α	0.08			0.07		
constant	-2.5594 ***	0.109	0	-2.7423 ***	0.127	0
holiday	-0.3221	0.222	0.147	-0.3096	0.224	0.166
holiday extension	-0.4661 ***	0.137	0.001	0.186	0.275	0.499
week	0.0301 ***	0.007	0	0.0213 **	0.008	0.006
Tue	-0.0762	0.125	0.543	-0.0882	0.118	0.457
Wed	-0.1066	0.125	0.392	-0.0955	0.119	0.422
Thu	-0.0559	0.123	0.648	-0.0367	0.118	0.755
Fri	0.1545	0.122	0.204	0.1621	0.117	0.165
Sat	-0.0732	0.127	0.563	-0.4277 *	0.183	0.02
Sun	-0.0593	0.129	0.645	-0.3568	0.19	0.061
unexpected absence rate (1 day ago)				0.0265 **	0.01	0.006
rolling overtime hours previous week x holiday extension				-0.0106	0.006	0.068
sick absence rate (2 days ago) x holiday extension				-0.2186	0.116	0.059
known-in-advance absence rate (1 day ago) x Sat				0.0385 *	0.018	0.036
known-in-advance absence rate (1 day ago) x Sun				0.0676	0.035	0.054
Log Likelihood	-428.68			-387.03		

Table notes: *: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001.

Since a general pattern was not identified from the models, a universal full model was not developed. Rather, the full model in subsequent parts included variables significant at the 95% confidence level identified from the dataset. Since unexpected absence from the previous day is not available at the time of operational scheduling, the prediction for the previous day was tested as a proxy, however, the benefits were not significant and therefore was not included in the final model (see Table 5-11).

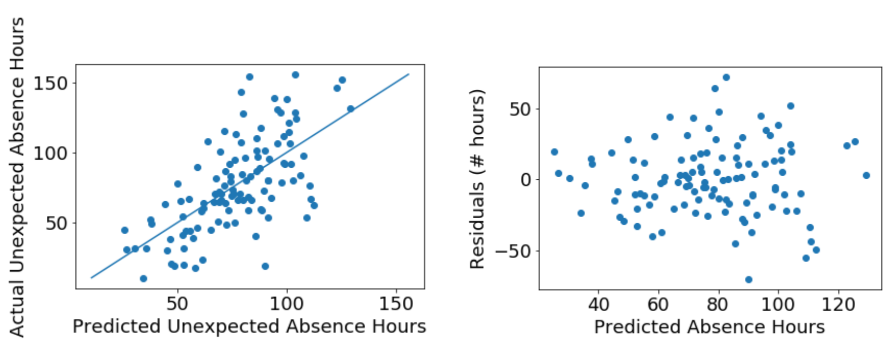
The coefficients for the negative binomial regression model for the number of unexpected absence hours are shown in Table 5-10 for Southampton for 2019 rating 4. While trying to fit the full model, with the inclusion of more explanatory variables, α^{-1} decreases (smaller unobserved heterogeneity, baseline 0.08, full 0.07) and the dispersion ϕ increases (higher over-dispersion with respect to the estimated negative binomial distribution, baseline 1.21, full 1.31). The full model passes the log likelihood ratio test indicates a better fit than the base model.

Since the total number of scheduled hours are included as an exposure term, the coefficients represent the elasticity of the explanatory variables with respect to the unexpected absence rate. Similar to the known-in-advance absence model, the coefficients for significant variables were more stable (week, holiday, and intercept) than those for the non-significant variables. For the 'week' variable. For unexpected absence, the holidays impose a significant negative effect. Unexpected absences were lower both on and around the holidays. Using Monday as a reference, the day-of-week was not significant for this rating in the base model but the signs of the coefficient agree with the descriptive analysis, where in this rating the weekly trend was present but weaker than in the other ratings. With the addition of more variables, day-of-week became more significant.

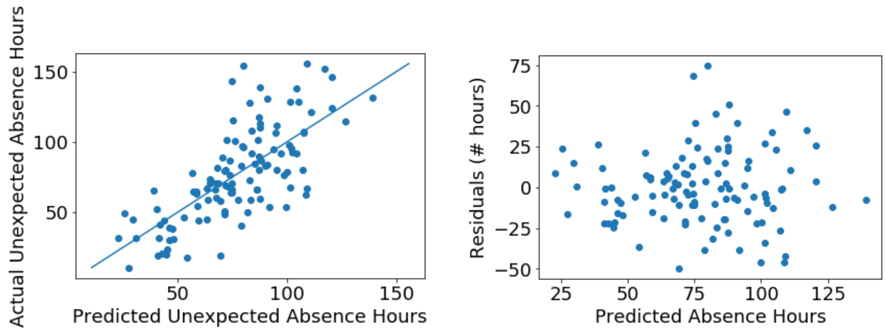
The coefficients for the additional variables in the full model suggest that: 1) Unexpected absences are autoregressive. 2) For weekends, higher previous-day known-in-advance absence rates lead to higher unexpected absence rates. Since there was no universal pattern in the additional variables across other ratings, these relationships are more likely correlational rather than causal.

Validation

Figure 5-3 shows the negative binomial model fit (left) and residuals distribution (right) for the base and full models.



a) Base Model



b) Full Model

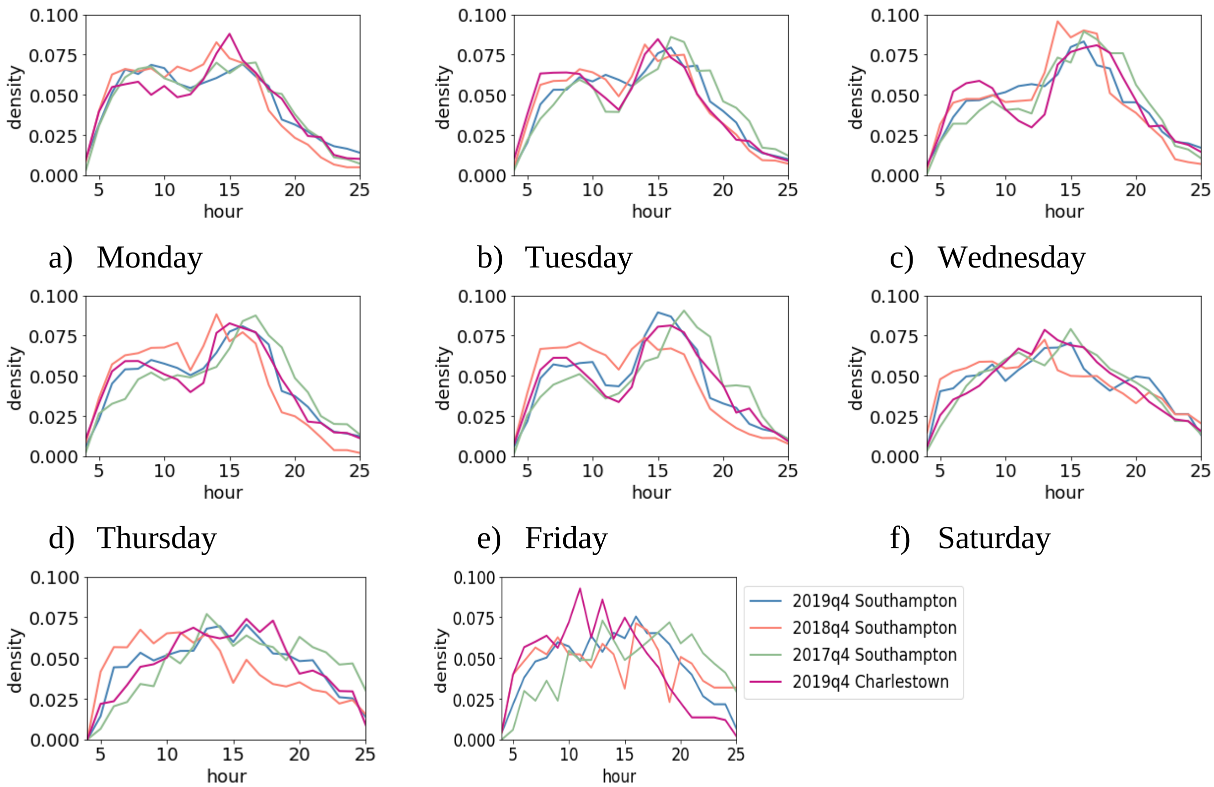
Figure 5-3 Negative Binomial Regression for Unexpected Absence Hours 2019 R4 @ Southampton

Table 5-11 shows the average estimation errors for daily unexpected absence hours for each dataset. The error rate was similar across ratings (20-25%) and Theil's U coefficients were also similar – the amount of variation explained by the model structure was similar. Including more variables drove the error down by about 2%. Since the previous day absence was a significant predictor (Table 5-10) but absence levels for the previous day are not available at the time of cover list scheduling, the prediction for the previous day was included as a surrogate. However, including the previous day prediction did not improve the model performance. The performance gain from including the previous day's absence level was offset by the inaccuracy of the predictions. Therefore, the previous-day prediction was removed from the final formulation.

Table 5-11 Negative Binomial Regression Error Rates for Daily Unexpected Absence Hours

Time Period Garage	Base Model				Full Model			
	w/o Prev Day Prediction		w/ Prev Day Prediction		w/o Prev Day Prediction		w/ Prev Day Prediction	
	MAE (MAPE)	Theil's U	MAE (MAPE)	Theil's U	MAE (MAPE)	Theil's U	MAE (MAPE)	Theil's U
2017 R4 Southampton	15.1 (22.8%)	0.14	15.1 (22.8%)	0.14	14.0 (21.1%)	0.13	13.9 (21.0%)	0.13
2018 R4 Southampton	14.6 (22.9%)	0.14	14.5 (22.6%)	0.14	13.4 (21.0%)	0.13	13.4 (21.0%)	0.13
2019 R4 Southampton	19.2 (24.9%)	0.15	19.1 (24.6%)	0.15	17.3 (22.4%)	0.14	17.1 (22.1%)	0.14
2019 R4 Charlestown	20.5 (22.9%)	0.14	20.5 (22.9%)	0.14	19.2 (21.4%)	0.13	19.4 (21.7%)	0.13

For time-of-day distributions of unexpected absences, average profiles for different days of the week and holidays were used. Figure 5-4 shows the average time-of-day absence profiles for all ratings and garages. Some holidays, in our case Columbus Day, use weekday schedules, and these holidays were categorized by their day of week instead of the holiday profile. On some days of the week, the patterns were similar. For example, the profiles for Mondays and Tuesdays, as well as Thursdays and Fridays, were very similar. But this similarity was not universal therefore the different days of the week were kept separate. The patterns were also similar, across different ratings and garages, especially on weekdays.



a) Monday b) Tuesday c) Wednesday

d) Thursday e) Friday f) Saturday

g) Sunday h) Holiday

Figure 5-4 Time-of-Day Distributions of Unexpected Absences

Cover list scheduling requires estimates for unexpected absences by hour. Integrating daily total unexpected absence hours with their time-of-day distribution gave predictions for unexpected absences by hour. Comparing the modelled results with actual number of unexpected absences, Figure 5-5 shows the model errors by time of day for this rating. The average error for the hourly prediction was 1.51. The hourly prediction, along with the error term, serves as inputs to the scheduling model in Chapter 6. The model produces more negative outliers, i.e. some high absence days were significantly under-predicted. In a pure estimation application, the direction of error does not matter; but for the scheduling application, having a high prediction indicates wasted manpower and a low prediction means heavy demands for overtime and the likelihood of lost service. This tradeoff will be considered further in Chapter 6.

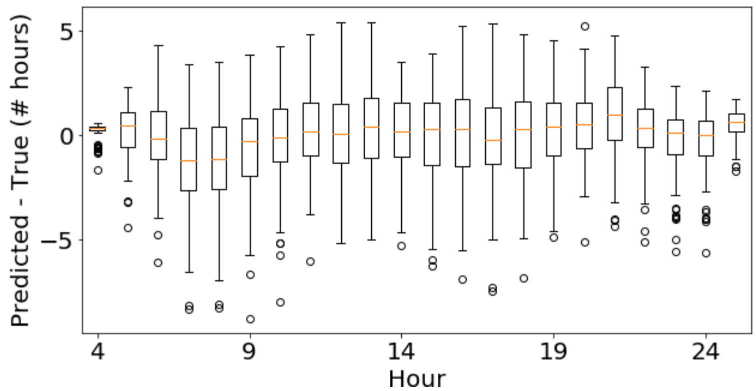


Figure 5-5 Combined Model Error 2019 R4 @ Southampton

Model Transferability

Table 5-12 tabulates the coefficients for the base variables, as well as the additional variables that passed the likelihood test at the 95% confidence level for each year. For the base variables, the day-of-week dummies are generally insignificant. Among the coefficients identified, the signs and magnitudes are similar across years with a few exceptions. The effect of holidays was not significant. There was an increasing trend in the amount of unexpected absences going from September to December as the end of year approaches. As for holiday extensions, in three rating/garages, the variable had a negative value. In 2019R4 at Southampton, although the holiday extension variable was positive in the full model, this variable was negative with the highest level of significance (***), which leads us to believe that the inclusion of other variables in the full model has offset this effect. The holiday pay provisions are effective in reducing absence.

There was no additional variable identified as significant in the prediction of unexpected absence hours for more than two rating/garages. The variables that were significant in two rating/garages are highlighted in the table. The two common factors are unexpected absence the previous day, as well as the intercept term of known-in-advance absence on Friday for Saturday absence prediction. First, unexpected absence was likely to be autoregressive and the coefficients estimated were similar. Besides the autoregressive effect, the other variables passing the test are all intercept terms with holiday extensions and weekends. Having more known-in-advance absences on Friday could lead to more unexpected absences on Saturdays. In 2018, where known-in-advance absence x Saturday was not included, FMLA x Saturday was strongly significant (p-value less than 0.001), which could be accounting for the same effect in which a longer weekend might be necessary (or desired).

For only one rating (2018 Southampton) did same-day known-in-advance absence rates pass the likelihood test, but was not significant at the 95% confidence level in the final model. If known-in-advance absence exceeds the extraboard capacity and overflows into unexpected absences, known-in-advance absence rates on the day of should have a positive and significant coefficient in the models, but there was no evidence to support this hypothesis.

Table 5-12 Variable Selection and Estimation Results Unexpected Absence Regression (95% C.I.)

	Southampton			Charlestown
	2017 R4	2018 R4	2019 R4	2019 R4
Base Variables				
constant	-2.4363***	-2.4829***	-2.7423***	-3.1550***
holiday	0.4384*	-0.1327	-0.3096	-0.2726
holiday extension	-0.6870***	-0.3570***	0.1860	-0.2471
week	0.0244***	0.0148**	0.0213**	0.0210**
Tue	-0.2532*	-0.0227	-0.0882	-0.0567
Wed	-0.1736	-0.2723**	-0.0955	-0.0976
Thu	-0.1869	-0.1227	-0.0367	0.0605
Fri	-0.0411	0.0222	0.1621	0.2184
Sat	0.0794	-0.6506**	-0.4277*	0.6531***
Sun	0.0829	0.3598*	-0.3568	0.1642
Additional Variables				
unexpected absence (1 day ago)			0.0265 **	0.0337 **
unexpected absence (2 days ago)				
unexpected absence (1 week ago)				
known-in-advance (same day)		-0.0183		
known-in-advance (1 day ago) x holiday extension	-0.0207			
known-in-advance absence (1 day ago) x Saturday	0.0647 *		0.0385 *	
known-in-advance absence (1 day ago) x Sunday			0.0676	
sick absence (2 days ago) x holiday extension			-0.2186	
sick absence (2 days ago) x Saturday	-0.3858 **			
rolling overtime previous week x holiday extension			-0.0106	
FMLA absence rates (2 days ago) x holiday extension	-0.1828 *			
FMLA absence rates (2 days ago) x Saturday		0.1682 ***		
FMLA absence rates (2 days ago) x Sunday		-0.1380 ***		

Table notes: ***: p-value < 0.05; **: p-value < 0.01; ***: p-value < 0.001.**

Table 5-13 shows prediction errors for different training and testing datasets. For example, both training and testing for 2019 show the model fit on the same dataset and serves as a best possible model performance benchmark with the current model structure. All base models had the same specification. The specifications used for the full models were the best ones found using the training dataset. The error from the model trained on 2018 data and tested on 2019 data shows how much the relationships between unexpected absence and the explanatory variables changed from 2018 to 2019. To include as much data as possible, as well as showing the value of pooling data and updating the model using data from the current rating, the second half of 2019 data was chosen as the test set.

Table 5-13 Transferability for Unexpected Absence Hours Model

		Base Model		**Full Model**	
Training Data[5]	Testing Data	Negative Binomial–MAE (MAPE)	Theil's U	Negative Binomial–MAE (MAPE)	Theil's U
***2019* (second half)**	2019 (second half)	21.24 (25.2%)	0.14	18.11 (21.5%)	0.12
2017+2018+2019 (first half)		23.64 (28.0%)	0.18	23.76 (28.2%)	0.18
2017+2018		23.95 (28.4%)	0.19	25.28 (30.0%)	0.19
2018		26.90 (31.9%)	0.21	28.96 (34.3%)	0.22
2017		21.59 (25.6%)	0.15	28.29(33.6%)	0.2

Unlike known-in-advance absences, although the full model had a better fit on the training data, it had worse testing errors. 2017 and 2019 had similar absence rates (Table 4-2), therefore using 2017 data to predict 2019 (MAPE is 27%, 4% worse than the benchmark) did better than using 2018 (30%). When the training distribution and the testing distribution were farther from each other (larger errors), the full model did worse because it was overfitted to the training data. Pooling effectively hedged against this effect. When data from the current rating was included in the model formulation, both models had similar performance.

Note that at the start of the rating, at the tactical level, we only have information to apply the base model. However, as we get into operational-level assignment, the full model is still not recommended since the full model is overfitted to the training data and has poor interpretability and transferability. In summary, the unexpected absence patterns differed slightly from year to year and data from a time period closer to the testing time period did not necessarily lead to better results. Combining the available training data from different time periods effectively hedges the effect. When there are policy changes that would affect unexpected absence behavior, caution must be used when selecting an appropriate training dataset.

[5] The garage was Southampton.

5.6.4 Unexpected Absence Prediction with Multivariate Analysis

In this section, results from the multivariate model discussed in section 5.5 are presented and analyzed. Since the explanatory variables were the same, and MVPLN is a variation of count models, the transferability patterns are similar to the model system discussed in 5.5, therefore we omit the discussion on model transferability.

Estimation Results

One of the disadvantages of this model is the number of parameters that need to be estimated. There are 21 hours * 10 explanatory variables = 210 β's, and a 21x21 covariance matrix must be estimated. For the rating shown (2019 R4 @ Southampton), the sample size was 112 days * 21 hours per day = 2352. Therefore, it is challenging to estimate the coefficients accurately. With a larger sample the benefit of the model would be easier to assess.

Figure 5-6 shows the beta values on the left and the covariance matrix on the right from Southampton in 2019 R4. Due to the large numbers of coefficients estimated, the magnitude of the coefficients is shown using a heatmap with a color scale centered at 0 (white). On the x-axis are the explanatory variables, the y-axis the hour of day, and the colors the magnitude of the coefficients. For example, the cell corresponding to 5-intercept shows the value of the intercept at 5am.

Similar to the previous models, due to the limitations in the information included in the explanatory variables, the model only captured the day-of-week and time-of-day effects not in the descriptive analysis. In this rating, there was one very clear trend: late evening absence increased as the rating progressed, as reflected in the coefficient for the week number. Correspondingly, the effect of the intercept, which includes constant effects, decreased. This pattern was not common over all years (for results from other years, please refer to Appendix C). This model used Monday as a reference. Only Fridays had either comparable (white for the hours 5-13 and 21-25) or higher (orange-red colors for the hours 14-20) levels of unexpected absence compared to Mondays. Thursdays had lower morning absences and slightly higher evening absences. All the other days of the week had lower unexpected absences at all times of the day, especially holidays.

The correlation matrix presents the correlation between different hours of the day that is not modelled explicitly. The correlations were generally positive, meaning the unobserved effects generally applied to the whole day with higher correlation between hours which are close together. The correlation matrix patterns in this rating were similar to those for other ratings/garages. For results from other ratings/garages, please refer to Appendix C.

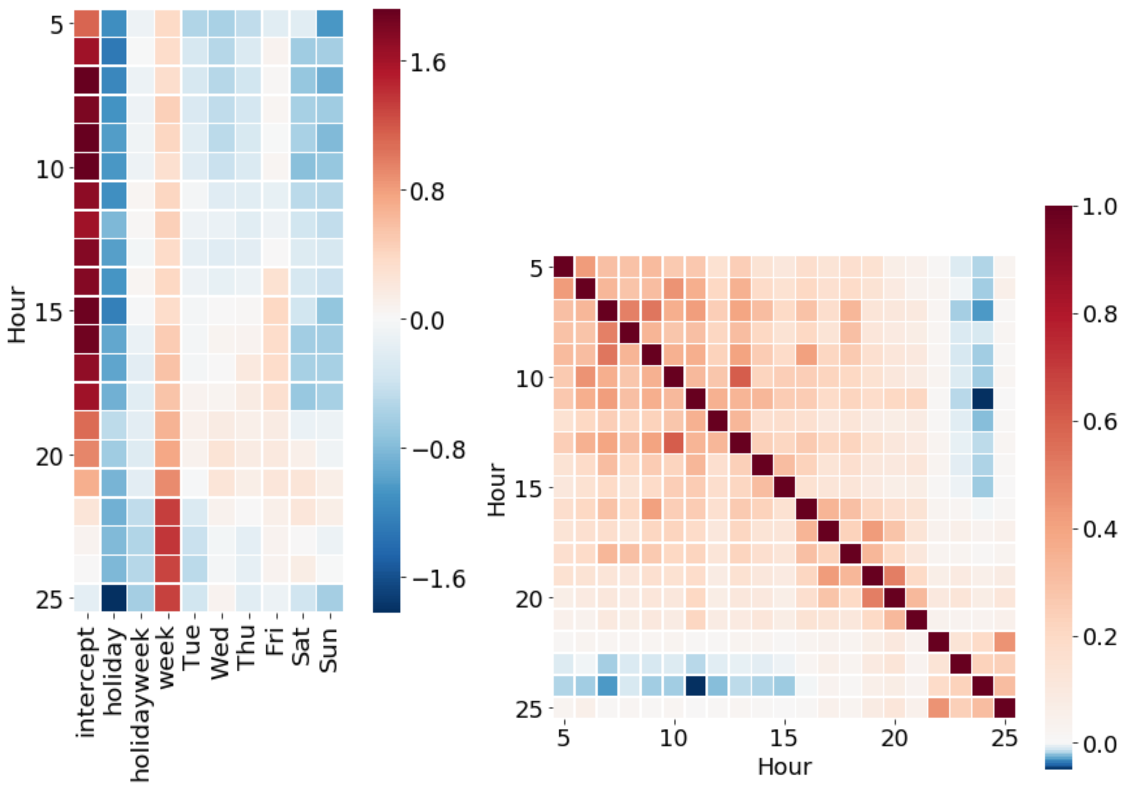


Figure 5-6 MVPLN Beta Values (left) And Correlation Matrix (right) 2019 R4 @ Southampton

To illustrate the effects of increased sample size on this model, Figure 5-7 shows the correlation matrix estimated from the combined 2017, 2018 and 2019 datasets at Southampton (350 days). The colors fade steadily from the diagonal to the lower left and upper right corner, indicating that the unobserved heterogeneity was more strongly correlated between hours which are close together.

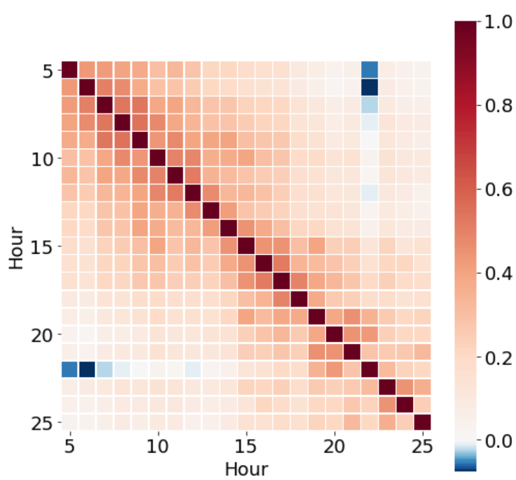


Figure 5-7 MVPLN Correlation Matrix 2017, 2018 and 2019 @ Southampton

Validation

Figure 5-8 shows the model errors by time of day for this rating. The average error for the prediction of hourly unexpected absence was 1.67, higher than the value obtained from the previous method (1.51), but the error range was more balanced and there were less extreme negative errors from this model. Because of the large numbers of parameters estimated, this method would benefit from larger datasets.

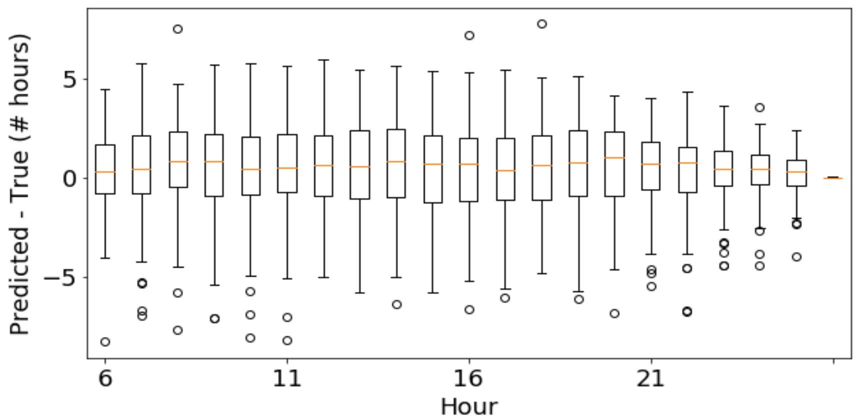


Figure 5-8 MVPLN Error 2019 R4 @ Southampton

Operator absence makes up the majority of the demand for extraboard operator-hours and will serve as an input to the extraboard scheduling model. In order to achieve better schedules, accurately predicting absence is an integral step. This chapter presented different forms of count models to predict aggregate absence. The key takeaways are: 1) The holiday pay incentive, explained in Section 5.6.1 with the model results shown in Table 5-12, was shown to be effective. Therefore, underlying behavioral factors play an important role in absence. Count models can model aggregate absences with mean absolute errors at around 20% to 30%. The lack of other behavioral and demographic data leaves room for improvement. 2) Both known-in-advance and unexpected absences are autoregressive, but past overtime and different types of absence are not significant predictors. 3) Unexpected absences at different hours in the day are correlated, and this correlation can either be captured by modelling a daily total and imposing a time-of-day distribution, or modelled explicitly with latent factors

Chapter 6 The Extraboard Scheduling Optimization Model

Extraboard operators are intended to provide cover for both service trips and non-service duties (such as transferring buses, flagging, yard, inspector duties, etc.). All work that the extraboard is intended to cover is referred to as open work. This chapter addresses the assignment of extraboard operators under uncertainty at the tactical and operational levels. With the strategic level in the three-step workforce planning process (Figure 1-2) beyond the scope of this research, the overall extraboard size is fixed. Figure 6-1 shows a modified version of Figure 1-2 where only the tactical and operational levels are included. At the tactical level, given the estimated levels of absence, overtime availability, essential non-service duties, and the number of available extraboard operator-days, the allocation of extraboard operators across garages and then the allocation of extraboard operators across days of the week are determined. Extraboard operators are trained and work at specific garages and can operate any route in the garage but are not shared between garages. Although it would be interesting to study the costs, and benefits of sharing extraboard operators across garages, due to data not being available from all garages, this is left for future research. At the operational level, on a daily basis, known-in-advance absences are assigned directly to the extraboard, assuming that the extraboard is large enough to cover all known-in-advance absences. The remaining operators (the cover list) are assigned report times for the next day to cover unexpected absences, and other open work as required. Finally, on the day of operations, as unexpected absences and non-service duties appear, outcomes in terms of lost service, overtime performed, and cover utilization follow. Although strategic-level decisions are taken as an input, results in terms of lost service and overtime performed for different staffing levels can inform strategic level decisions such as whether to adjust the extraboard size.

Section 6.1 discusses the treatment of another important input into extraboard scheduling: overtime availability. Section 6.2 discusses the policy constraints on extraboard scheduling, as well as some assumptions made in the model formulation. Section 6.3 presents the nominal model formulation without uncertainty and Section 6.4 presents the robust version of the model. Section 6.5 presents the numerical results of the MBTA case study using the models developed in Chapter 5, and Section 6.5.6 analyzes the benefits of two alternative scheduling policies which could be considered by the MBTA.

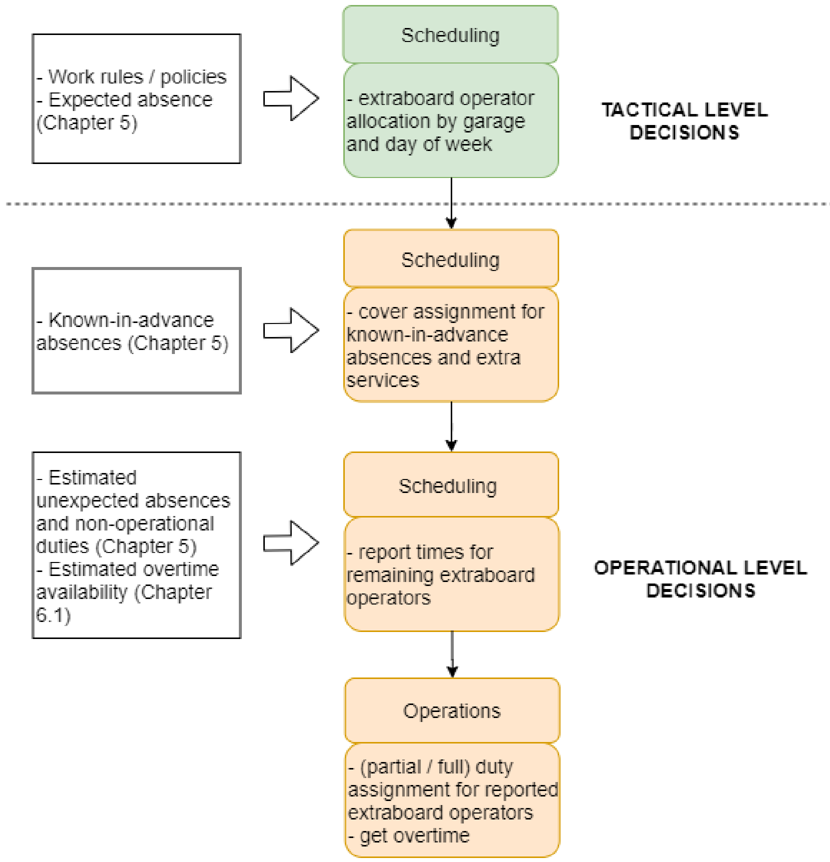


Figure 6-1 Tactical and Operational Level Planning Decisions

6.1 Treatment of Overtime Availability

Operator overtime plays a major role in covering absences to maximize service coverage, especially when there is a shortage in the number of extraboard operators (see Table 4-2). Although paid at a 50% pay premium over the regular rate, overtime is often a less expensive option than increasing the size of the cover list due to the benefits and vacation liability associated with additional operators. Overtime can also be a more flexible and efficient option than the extraboard because calls for overtime are made only when there is open work, thus no wasted time is incurred. However, relying too heavily on overtime can lead to operator fatigue and more operator absence. If no one volunteers to work overtime when it is offered, service will be lost, affecting reliability of service and public perception of the agency. Therefore, a key question is what is the amount of overtime which can be reliably counted on. Correctly anticipating the level of overtime available at different times of the day and on different days of the week can help in scheduling cover list operators such that maximum coverage is obtained at minimum cost.

A few models are tested for estimating overtime availability, but no satisfactory models resulted from this research, mainly due to data limitations. The attempts made are described in Appendix D. We will discuss some modelling considerations that formed the basis of the approximations made in this research, mainly for future reference. There are two potential sources that the garage superintendent could seek overtime from: operators on their day-off and extension of work hours for operators who are performing a shift that day. Most overtime (from both sources) is worked as single pieces with no breaks. The decision-making processes behind the two sources of overtime are different. The factors correlated with the quantity on the aggregate level were also expected to be different. In the regression models considered, separate formulations were constructed for each type.

The biggest issue in modelling overtime availability is that it is not directly observed, so no ground truth is available. From the data, overtime performed is observed, which depends on both overtime availability and overtime requested and is therefore a lower bound on overtime availability. At the MBTA, overtime is only requested if there is no cover list operator available to do the work. Assuming that this practice is followed rigorously and that the garage superintendent always attempts to fill all trips, we could infer how much overtime is requested as the sum of lost service and overtime performed. When there is lost service, we can conclude that overtime performed is equal to overtime available, since if there was more overtime available, more service would have been covered. However, when there is no lost service, observed overtime is just a lower bound of available overtime.

In this research, the empirical distribution of overtime availability was used. First, the hours with lost service are selected, since for these times, overtime performed is also available overtime. Second, the numbers of operators available for overtime are grouped by day of week and hour of day. For deterministic models, the mean μ_{dt}^{o} is taken as the overtime availability for the hour and day of week. For robust models, standard deviations σ_{dt}^{o} are taken from the empirical distributions to form the range of values in the uncertainty set for each level of robustness.

6.2 Scheduling Policy Constraints and Modelling Assumptions

This section discusses the model constraints associated with the labor agreement, common scheduling practices at the MBTA, and assumptions made in model formulation. Other agencies might have different policies and practices and the associated constraints would need to be modified accordingly. Depending on the objectives of the analysis, other assumptions may also need to be modified.

Tactical Level

1) Analysis unit: number of operators per day. Tactical-level analysis is based on expected absences, overtime availability, and lost service on a daily basis, which is aggregated from the operational-level model.

2) Scheduling policy: Most vacation cover is scheduled with dedicated vacation relief operators, which are part of a separate group of operators, in a separate process from extraboard allocation. The remaining vacations are scheduled in the form of single vacation days, and are covered in the same way as regular absences. In any week when there are more vacation relief operators than operators on vacation, vacation relief operators are added to the extraboard.

3) Scheduling policy: Each day of the week has the same scheduled number of extraboard operators in all weeks in the rating. Note that the actual number of operators available on any given day will depend on the absences among extraboard operators, as well as any unassigned vacation relief operators. At the tactical level, the MBTA builds the extraboard duties into rosters which operators pick in order of (decreasing) seniority. Each operator has the same work days in all weeks of the rating. Although it is important to take week-to-week variations in absence and overtime availability into account while scheduling, the tactical-level assignment plan cannot respond to these variations by changing the scheduled number of extraboard operators from week to week within the rating. Data suggests that the actual number of extraboard operators available on any given day varies greatly, because of extraboard operators updated in the data as the original owner of the shifts, extraboard absences, unassigned vacation relief operators, operator suspension and/or attrition.

4) Scheduling policy: Generally, only full-time operators are assigned to the extraboard. Many agencies employ full-time and part-time operators with different duty requirements and constraints (as well as benefits). Based on current MBTA practice, the extraboard consists mainly of full-time operators, taking on scheduled runs (for known-in-advance absence) and eight-hour straight shifts (for the cover list). Although in special circumstances, such as significant shortages of full-time operators (for example rating 4 at Southampton in 2017), part-time operators could be assigned to the extraboard. In the base case, it is assumed that all cover list operators are full-time operators working eight-hour straight shifts.

Operational Level

1) Analysis unit: hourly workload curve. Operational-level analysis is based on an hourly workload curve, since the absence and overtime analysis are at an aggregate level. Since the

daily total and time-of-day profile were estimated separately, non-integer numbers naturally arise for the number of unexpected absences. Hourly overtime availability was integer-valued. The cover list assignment was modelled by the report time for each extraboard operator; therefore, the cover list coverage curve was integer valued.

2) Assumption: Part-day known-in-advance absences, which are uncommon, are treated as if they were unexpected absences. At the MBTA, cover is not arranged in advance for these absences and they end up on the cover list. In theory, while determining report times for the following day, known-in-advance, part-day absences give concrete information on times that cover is required, reducing the amount of uncertainty on unexpected absences. From Figure 4-3, there are few known-in-advance part-day absences. Therefore, they are not accounted for during operational-level scheduling. All absences covered by the cover list are assumed to be unexpected and any knowledge on part-day known-in-advance absences is not used in the assignment of report times.

3) Assumption: Rest rules between shifts do not affect scheduling at the aggregate level. Rest rules exist to protect operators from becoming dangerously fatigued and thereby raising safety concerns. For example, the MBTA has requirements on both rest time between days and cumulative work hours in a given time period. All operators need to have at least 6 hours of rest between consecutive workdays and at least 8 hours of rest if the previous day shift was 14 hours. In a 16-hour spread on any day (including breaks), the operator cannot work for more than 14 hours, as well as 42 hours in a 72-hour period, and 60 hours in a week (from Saturday to Friday). These rules apply to every operator. Since this analysis is at the aggregate level, the rest rules are not explicitly included. It is assumed that the rest rules can be satisfied by swapping shifts between operators after the shifts are made if the rest rules would be violated.

4) Assumption: During operations, overtime will not be requested unless no cover list operator is available. On the day of operations, the MBTA follows a greedy approach when assigning work: when unexpected demands occur, work is assigned to the cover list operators before overtime is offered. The scheduling model adopts this assignment strategy in the calculation of overtime performed and lost service on the day of operations. Whether the greedy approach is optimal is an open question (Gupta and Li, 2016), but this issue is beyond the scope of this thesis.

5) Assumption: It is assumed that all service trips take precedence over non-service duties and that non-service duties can be scheduled in between required service duties or not performed without incurring losses. Besides service trips, cover list operators also perform non-service duties such as transferring buses between garages, doing flagging work, substituting for inspectors, etc. Based on empirical evidence, the amount of non-service duties is not negligible (Table 4-3). In terms of importance, non-service duties can be divided into essential and non-essential. Some essential non-service duties may be required even at times when service trips also need to be covered. Although in practice, once an operator has started a non-essential non-service duty and, if an unexpected absence occurs, the operator may not be able to terminate the non-service duty and be available for the service trip, this dispatching issue is beyond the scope for this formulation and it is assumed that no conflict will be incurred. Similar to absences, essential non-service duties can also be divided into known-in-advance and unexpected. It is assumed that, in extraboard scheduling, the known-in-advance and unexpected non-service duties are treated the same as known-in-advance absences and unexpected absences, respectively.

6.3 Nominal Problem Formulation

This process involves two levels of decision making that can be formulated as a two-stage optimization problem. The first stage (tactical level) involves deciding the number of extraboard operators to assign to each day of the week in the analysis period (rating) based on predicted absences and overtime availability. During the rating, on a daily basis, extraboard operators are assigned to cover known-in-advance absences. At the second stage (the operational level), the remaining extraboard operators are assigned report times for the following day to be available to cover unexpected absences. At both stages, the goal is to minimize lost service and overtime performed and maximize extraboard productivity.

This section introduces the nominal (deterministic) formulation of the problem. In the nominal model, absence levels and overtime availability are treated as deterministic quantities. Although not directly applicable to scheduling (since absence and overtime availability are uncertain at the time of scheduling), this formulation is used to get the optimal assignment strategy given hindsight, and is used as a benchmark for the scheduling strategies developed subsequently. This formulation also serves as a starting point for introducing uncertainty into the formulation, as discussed in Section 6.4.

The variables, objective functions, and constraints for the two levels of scheduling are introduced separately below, and then combined in the complete formulation.

6.3.1 Tactical Level

Model Parameters and Variables

The decision variables for the tactical level are $\mathrm{X} = [x_0, \dots, x_6]$, a vector that represents the number of extraboard operators scheduled for Monday to Sunday $(x_0, \dots, x_6)$. Each subscript represents a day of the week that, in theory, should have the same number of extraboard operators scheduled in all weeks in the rating.

The parameters and inputs for the models are shown below.

c_{ls}, c_{ot}, c_{ex}	cost coefficients for lost service, overtime performed, and extraboard hours
D	the dates in scope
T	the hours in a day
N	maximum number of extraboard operators that can be scheduled in any day
M	number of available extraboard-days in a week
s	shift length in hours
δ_{di}	indicator for the number of extraboard operators scheduled for date d; $\delta_{di} = 1$ if date d corresponds to category i in a week, $\delta_{di} = 0$ otherwise; $\mathrm{i} = 0, \dots, 7$; $\sum_{i=0}^{6} \delta_{di} = 1 \quad \forall d$ since each day can belong to only one category.

For notational convenience, the following intermediate variables are defined:

x_d^e	number of extraboard operators assigned for date d
x_d^c	number of cover list operators for date d
z_{dt}^{ls}	number of lost service hours on date d hour t
z_{dt}^{o}	number of operators performing overtime on date d hour t

Objective Function

The objective is to minimize a weighted sum of the cost of lost service, overtime performed, and extraboard operators for all days in the analysis period:

$$\min_X \sum_{\mathrm{d}=1}^{\mathrm{D}} \left(c_{ex}\, \mathrm{s}\, x_d^e + \min_Y \; c_{ls} \sum_{t=1}^{T} z_{dt}^{ls} + c_{ot} \sum_{t=1}^{T} z_{dt}^{o} \right)$$

where c_{ls}, c_{ot}, c_{ex} are weights assigned to an hour of lost service, an hour of overtime performed, and an hour of work performed by an extraboard operator. However, there is no direct monetary cost to the agency associated with lost service (besides lost passenger revenue, which involves demand prediction to quantify), and the cost of the extraboard includes fringe benefits which makes the cost of each extraboard hour higher than the nominal pay rate. Although overtime appears to be a less expensive option, with the cost of fringe benefits associated with extraboard operators, relying too heavily on overtime risks losing service, and influencing the well-being of operators, both of which are influential and hard to quantify. Therefore, c_{ls}, c_{ot}, c_{ex} are weights based on agency-specific judgment. For example, if minimizing lost service is the ultimate goal, then c_{ls} can take a large value relative to c_{ot} and c_{ex}, thereby ensuring that lost service is minimized, given the resource constraints. Within the resource constraints (the staffing level), overtime performed, and extraboard assignment costs are relevant only when the same level of lost service is attained.

Constraints

Two resource scenarios may be of interest: constrained resources and unconstrained resources. The constrained resource model can be used to efficiently allocate known resources, whereas the unconstrained model can be used to inform strategic-level decisions. In the constrained resource case, the total number of extraboard operators is fixed. In the unconstrained resource case, the total number of extraboard operators is not restricted and the scheduling model is used to recommend the staffing level to minimize overall weighted cost objective. However, constraints on the relationship between the number of operators hired and the number of days worked also need to be satisfied. In this model, it is assumed that every extraboard operator works five days a week. For example, if we have x extraboard-days assigned in one week, then we cannot have more than x/5 operators in a day.

- Apply to both scenarios
    - The same number of extraboard operators is scheduled for each day of the week across all weeks in the rating.

$$x_d^e = \sum_{i \in I} \delta_{di} x_i$$

where $\delta_{di} = 1$ when date d belongs to day-of-week category i.

- Constrained Resources
  - Total number of extraboard operator-days available in a week. Either M or N needs to be specified, or both.

$$\sum_{i=0}^{6} x_i \leq \text{M} \leq 5\text{N}$$

  - Number of extraboard operators available in a day cannot exceed the number of extraboard operators.

$$x_i \leq N \quad \forall i$$

- Unconstrained Resources
  - The number of extraboard operators assigned each day is less than the total number of extraboard days in the week divided by 5.

$$x_d^e \leq \frac{1}{5}\sum_{i=0}^{6} x_i \quad \forall \text{d}$$

6.3.2 Operational Level

Model Parameters and Variables

Y $\in \{0,1\}^{\text{N} \times \text{D} \times T}$ is a matrix of binary values that represents the report times of each cover list operator ($y_{jdt} = 1$ if cover list operator j reports at hour t on date d, $y_{jdt} = 0$ otherwise). Operator j on date d and operator j on date d' do not have to be the same person, as long as the number of operators scheduled is greater than, or equal to, the number of operators assigned.

During operations, when faced with unexpected open work, in anticipation of more open work arising, it is the garage's dispatcher's decision whether to dispatch a cover list operator, request overtime, or drop the work. Since the real-time dispatching problem during operations is beyond the scope of this thesis, a greedy approach (where open work is assigned to the cover list on a first come, first serve basis until no more cover list operators are available, and then the remaining work is offered as overtime) was assumed for the assignment. The costs of lost service and overtime performed were assumed to be the same for service and non-service duties: if the costs are different, one type of duty will take precedence over the other. The variables in the model are defined as follows.

c_{ls}, c_{ot} cost coefficients for lost service and overtime performed

T the hours in a day

s shift length in hours

δ_{di} indicator for the number of extraboard operators scheduled for date d; $\delta_{di} = 1$ if date d corresponds to category i in a week, $\delta_{di} = 0$ otherwise; $i = 0, \dots, 7$; $\sum_{i=0}^{6} \delta_{di} = 1 \quad \forall d$ since each day can belong to only one category.

n_{dt}^{uow} number of unexpected open work on date d hour t

n_{d}^{kow} number of known-in-advance open work on date d

n_{dt}^{ao} number of operators available for overtime on date d hour t

n_{dt}^{u} number of unexpected absences on date d hour t

n_{d}^{k} number of known-in-advance absences on date d

n_{dt}^{uns} number of unexpected essential non-service duties on date d hour t

n_{d}^{kns} number of operators needed from the extraboard for known-in-advance essential non-service duties on date d

Objective Function

At the operational level, each day is optimized independently. The number of extraboard operators is inherited from the tactical level, so the cost of the extraboard is constant no matter the assignment strategy; therefore, the objective function contains only costs related to lost service and overtime performed.

$$Q\left(\mathrm{n_d^{kow}},\ \mathrm{n_d^{uow}}, \mathrm{n_d^{o}}, \mathrm{X}\right) = \min_{Y} \ c_{ls} \sum_{t=1}^{T} z_{dt}^{ls} + c_{ot} \sum_{t=1}^{T} z_{dt}^{o}$$

Constraints

- Work rules
  - Open work consists of absence and essential non-service duties

$$n_{d}^{kow} = n_{d}^{kns} + n_{d}^{k}$$
$$n_{dt}^{uow} = n_{dt}^{uns} + n_{dt}^{u}$$

- Extraboard operators x_d^e cover both types of open work: known-in-advance and unexpected open work.

$$x_d^e \leq x_d^c + n_d^{kow}$$

$$x_d^c = \sum_{t=0}^{T}\sum_{j=1}^{N} y_{jdt}$$

- At the MBTA, one cover list operator is reserved to cover the first shift in the morning and another for last shift at night. This rule does not apply when there are no cover list operators and if there is only one cover list operator, (s)he is scheduled to cover the first shift in the morning.

$$\sum_{j=1}^{N} y_{jd0} \geq 1 \text{ , } \sum_{j=1}^{N} y_{jd(T-s)} \geq 1$$

- The number of cover list operators available in hour t is the number of operators that started their duties before this time and are still active, and those who start their duties in the current hour.

$$x_{dt}^c = \sum_{j=1}^{x_d^c}\sum_{k=max(0,t-s)}^{t} y_{jd\mathrm{k}}$$

- Lost service is calculated from the shortage of operators in each hour. Demand for cover operators is the unexpected absence and essential non-service duties at time t, n_{dt}^{uow}, and cover availability is the sum of available overtime and active cover list operators $n_{dt}^o + x_{dt}^c$. When there are more available operators than absences, $n_{dt}^{uow} - n_{dt}^o - x_{dt}^c$ is negative and there is no lost service.

$$z_{dt}^{ls} = \max(n_{dt}^{uow} - n_{dt}^o - x_{dt}^c, 0)$$

- Overtime performed is calculated as the hourly number of operators working overtime. Since overtime is the last resort to cover, $z_{dt}^{po} > 0$ only when $n_{dt}^{uow} = n_{dt}^u + n_{dt}^{uns} > x_{dt}^c$.

$$z_{dt}^o = \max\left(n_{dt}^{uow} - x_{dt}^c - z_{dt}^{ls},\ 0\right)$$

- Non-negativity and uniqueness constraints
  - A scheduled operator can have only one report time.

$$\sum_{t=1}^{T} y_{djt} \leq 1 \quad \forall j, \mathrm{d}$$

6.3.3 Combined Model

Combining and linearizing the constraints, the nominal problem formulation is:

Tactical Level:

$$\min_{X} \sum_{\mathrm{d}=1}^{\mathrm{D}} \left(c_{ex} s x_d^e + \min_{Y} \; c_{ls} \sum_{t=1}^{T} z_{dt}^{ls} + c_{ot} \sum_{t=1}^{T} z_{dt}^{o} \right)$$

	Constrained resources:	Unconstrained resources:	
s.t.	$\sum_{i=0}^{6} x_i \leq \mathrm{M} \leq 5\mathrm{N}$		
	$x_i \leq N$	$x_d^e \leq x_i$	$\forall$i
	$x_d^e = \sum_{i \in I} \delta_{di} x_i$	$x_d^e \leq \frac{1}{5} \sum_{i=0}^{6} x_i$	$\forall$d
	$x_i \in \mathrm{N}$	$x_i \in \mathrm{N}$	

Operational Level (for a specific date d):

$$\mathrm{Q}\left(\mathrm{n_d^{kow}}, n_d^{uow}, \mathrm{n_d^{ao}}, \mathrm{X}\right) = \min_{Y} \; c_{ls} \sum_{t=1}^{T} z_{dt}^{ls} + c_{ot} \sum_{t=1}^{T} z_{dt}^{po}$$

s.t. $n_d^{kow} = n_d^{kns} + n_d^{k}$

$n_{dt}^{uow} = n_{dt}^{uns} + n_{dt}^{u}$ $\quad \forall$t

$x_d^e = \sum_{i \in I} \delta_{di} x_i$

$x_d^e \leq x_d^c + n_d^{kow}$

$x_d^c = \sum_{t=0}^{T} \sum_{j=1}^{N} y_{jdt}$

$x_{dt}^c = \sum_{j=1}^{x_d^c} \sum_{k=max(0,t-s)}^{t} y_{jdk}$ $\quad \forall$t

$z_{dt}^{ls} \geq n_{dt}^{uow} - n_{dt}^{ao} - x_{dt}^{cl}$ $\quad \forall$t

$z_{dt}^{o} \geq n_{dt}^{uow} - x_{dt}^{c} - z_{dt}^{ls}$ $\quad \forall$t

$z_{dt}^{ls} \geq 0$ $\quad \forall$t

$z_{dt}^{o} \geq 0$ $\quad \forall$t

$\sum_{j=1}^{N} y_{jd0} \geq 1$ $\quad$ when $x_d^c \geq 1$

$\sum_{j=1}^{N} y_{jd(T-s)} \geq 1$ $\quad$ when $x_d^c \geq 2$

$\sum_{t=1}^{T} y_{djt} \leq 1$ $\quad \forall j$

$\mathrm{Y} \in \{0,1\}^{\mathrm{N} \times T}$

6.4 Adaptive Robust Problem Formulation

If we know n_d^k, n_{dt}^u, n_d^{kns}, n_{dt}^{uns} and n_{dt}^o, then the nominal formulation (6.3) should give the optimal assignment strategy. However, at the tactical level, n_d^k, n_{dt}^u, n_d^{kns}, n_{dt}^{uns}, and n_{dt}^o are all uncertain. While n_d^k and n_d^{kns} are realized before the operational level assignment, n_{dt}^u , n_{dt}^{uns}, and n_{dt}^o remain uncertain. To deal with this uncertainty, two approaches are feasible. In the general stochastic optimization literature, the expected value of the objective function, or a specified percentile of the objective function is optimized. Another approach is to use robust optimization, that attempts to develop solutions that are robust to the specified uncertainty. The objective function is optimized while maintaining feasibility for all parameter values in the specified uncertainty sets. In this thesis, robust optimization was selected for two reasons: first, transit agencies would want to have a plan that hedges worst-case risks; second, the expected objective function value of an integer optimization problem with uncertainty is intractable. This section introduces the robust formulation of the scheduling problem and proposes a solution algorithm.

6.4.1 Formulation

The extraboard scheduling problem is formulated as a two-stage adaptive robust optimization problem since the operational-level decisions depend on the tactical level decisions. The tactical-level problem is a robust adaptive problem, and the operational-level problem is a standard integer robust optimization problem.

The decision variables and parameters are defined as in the nominal version. X is the here-and-now decision and Y is the wait-and-see decision. Since the cost of service and essential non-service duties are the same, n^k and n^{kns} are combined to form n^{kow} and n^u and n^{uns} are combined to form n^{uow}outside the optimization program. Let $Z_{n^{kow}}, Z_{n^{uow}}, Z_{n^{ao}}$ represent the uncertainty sets for known-in-advance open work, unexpected open work, and overtime availability for date *d* and time *t* respectively. The adaptive robust model can be formulated as follows.

Tactical Level:

$$\min_X \sum_{\mathrm{d}=1}^{\mathrm{D}} \left(c_{ex} s x_d^e + \min_Y \; c_{ls} \sum_{t=1}^{T} z_{dt}^{ls} + c_{ot} \sum_{t=1}^{T} z_{dt}^{o} \right) \tag{6-1}$$

	Constrained resources:	Unconstrained resources:	
s.t.	$\sum_{i=0}^{6} x_i \leq \mathrm{M} \leq 5\mathrm{N}$		
	$x_i \leq N$	$x_d^e \leq x_i$	$\forall$i
	$x_d^e = \sum_{i \in I} \delta_{di} x_i$	$x_d^e \leq \frac{1}{5} \sum_{i=0}^{6} x_i$	$\forall$d
	$x_i \in \mathrm{N}$	$x_i \in \mathrm{N}$	

Operational Level (for a specific date d):

$$\mathrm{Q}\left(\mathrm{n_d^{kow}}, n_d^{uow}, \mathrm{n_d^o}, \mathrm{X}\right) = \min_Y \; c_{ls} \sum_{t=1}^{T} z_{dt}^{ls} + c_{ot} \sum_{t=1}^{T} z_{dt}^{o} \tag{6-2}$$

s.t.	$n_d^{kow} = n_d^{kns} + n_d^{k}$		
	$n_{dt}^{uow} = n_{dt}^{uns} + n_{dt}^{u}$	$\forall t$	
	$x_d^e = \sum_{i \in I} \delta_{di} x_i$		
	$x_d^e \leq x_d^c + n_d^{kow}$		
	$x_d^c = \sum_{t=0}^{T} \sum_{j=1}^{N} y_{jdt}$		
	$x_{dt}^c = \sum_{j=1}^{x_d^c} \sum_{k=max(0,t-s)}^{t} y_{jd\mathrm{k}}$	$\forall t$	
	$z_{dt}^{ls} \geq n_{dt}^{uow} - n_{dt}^{o} - x_{dt}^{cl}$ $z_{dt}^{o} \geq n_{dt}^{uow} - x_{dt}^{c} - z_{dt}^{ls}$	$\forall t; n_{dt}^{uow} \in Z_{n^{uow}};\; n_{dt}^{o} \in Z_{n^{ao}}$	(6-3)
	$z_{dt}^{ls} \geq 0$	$\forall t$	(6-4)
	$z_{dt}^{o} \geq 0$	$\forall t$	
	$\sum_{j=1}^{N} y_{jd0} \geq 1$	when $x_d^c \geq 1$	(6-5)
	$\sum_{j=1}^{N} y_{jd(T-s)} \geq 1$	when $x_d^c \geq 2$	(6-6)
	$\sum_{t=1}^{T} y_{djt} \leq 1$	$\forall j$	
	$\mathrm{Y} \in \{0,1\}^{\mathrm{N} \times T}$		

6.4.2 Solution Algorithm

The adaptive robust programs (6-1) and (6-2) are a two-stage adaptive integer program. Decision variables for both levels (X, Y) are integer, therefore, affine decision rules cannot be applied, and duality cannot be used to simplify the minimax formulation, either. However, note that the tactical-level decisions depend on the operational-level objective function value $\mathrm{Q}(\mathrm{n_d^{kow}},\ \mathrm{n_d^{uow}}, \mathrm{n_d^o}, \mathrm{X})$, which depends on the known-in-advance absences, unexpected open work, available overtime, and the number of extraboard operators scheduled. Note also that unexpected open work $\mathrm{n_d^{uow}}$ and available overtime $n_{\mathrm{dt}}^{\mathrm{o}}$ are inputs to the model and do not depend on the tactical level decisions, and that both X and n_d^k can only take on integer values in a reasonable range. In this case, the operational-level objective function value's dependence on the tactical-level information only lies in one scalar variable: the number of cover list operators $x_d^e - \mathrm{n_d^{kow}}$. Therefore, the tactical-level integer program can be written as a minimax formulation $\min_{X} \max_{n_d^k \in Z_{n^k}} \mathrm{Q}(n^{kow}, \mathrm{n_d^{uow}},\ \mathrm{n_d^o}, \mathrm{X},\ \mathrm{Y})$, which can be solved by enumerating all values of $x_d^e - \mathrm{n_d^{kow}}$. Now we need to find a solution to the operational-level problem (6-2).

The inner minimization (6-2) is a robust integer optimization since $\mathrm{n_d^{uow}}$ and $\mathrm{n_d^o}$ are still uncertain at this stage. $Z_{n^{uow}}$ and Z_{n^o} are disjoint uncertainty sets. However, the robust counterparts for the constraints are not trivial to derive. The basic assumption in the derivation of robust counterparts is that uncertainty is constraint-wise. However, constraint (6-3) includes two independent uncertainty sets, and the subtraction of two uncertain variables with budget uncertainty sets is not of standard form. At the same time, $Z_{n^{uow}}$ also appears in constraint (6-4). Since the concept of robust optimization is constraint-wise, splitting the uncertainty in n^{uow} across two constraints implies the worst case lost service and overtime are optimized with potentially different n^{uow} values. The program will be not only unnecessarily conservative, but also impractical since different n^{uow} values are used for the same day.

Since robust counterparts for the program are difficult to derive, a cutting-plane approach is used to solve the inner minimization. Let w represent the set of uncertain parameters $\mathrm{w} = [n_{\mathrm{dt}}^{\mathrm{uow}}, \mathrm{n_{dt}^o}]$ and $\mathrm{Z} = [Z_{n^{uow}}, Z_{n^o}]$ the uncertainty set. The proposed algorithm is summarized below:

1. Set iteration number k = 1
2. The master problem (6-2) is solved to get Y^k and the objective function value (obj). The reduced uncertainty set $\tilde{Z}$ consists of only the nominal values in this iteration.
3. Y^k is substituted to solve the adversarial problem (6-7) below to get the worse-case uncertainty parameters w and objective function value (obj_adv).

$$w^k = \underset{n_{dt}^{uow}, n_{dt}^{ao}}{\operatorname{argmax}} \quad c_{ls} \sum_{t=1}^{T} z_{dt}^{ls} + c_{ot} \sum_{t=1}^{T} z_{dt}^{o} \tag{6-7}$$

$$\text{s.t.} \quad x_{dt}^{c} = \sum_{j=1}^{x_d^c} \sum_{k=max(0,t-s)}^{t} y_{jdk} \qquad \forall t$$

$$z_{dt}^{ls} = \max(n_{dt}^{uow} - n_{dt}^{o} - x_{dt}^{c}, 0)$$

$$z_{dt}^{o} = \max\left(n_{dt}^{uow} - x_{dt}^{c} - z_{dt}^{ls},\ 0\right) \qquad \forall t$$

$$n_{dt}^{uow} \in Z_{n^{uow}}, n_{dt}^{o} \in Z_{n^{ao}} \qquad \forall t$$

4. w^{k} is added to the master model uncertainty set $\tilde{Z}$. Solve the master problem (6-2) to obtain the assignment for the k$^{\text{th}}$ iteration Y^k.
5. The master problem (6-2) with uncertainty set $\tilde{Z}$ gives a lower bound for the optimal solution in each iteration, and the adversarial problem gives an upper bound. The algorithm terminates if the bounds are within a predefined error range. Otherwise, k=k+1, go to step 2 and iterate with the updated uncertainty set $\tilde{Z}$.

Note that the adversarial problem (6-7) has a maximization in the objective with another maximization in the constraints (the nonnegativity constraints of lost service and overtime performed). To ensure non-negativity, indicator decision variables i^{ls}, i^{o} were added to indicate the signs of lost service and overtime (1 for positive and 0 for negative). Since we are maximizing the objective function, when there is excess capacity for service and lost service or overtime performed is negative, the indicator variable will take value 0 to maximize the objective. The result is a quadratic program (6-8).

$$\max_{n_{dt}^{uow}, n_{dt}^{o}} \quad c_{ls} \sum_{t=1}^{T} z_{dt}^{ls} \times i_t^{ls} + c_{ot} \sum_{t=1}^{T} z_{dt}^{o} \times i_t^{o} \tag{6-8}$$

$$\text{s.t.} \quad x_{dt}^{c} = \sum_{j=1}^{x_d^c} \sum_{k=max(0,t-s)}^{t} y_{jd\mathrm{k}} \qquad \forall t$$

$$z_{dt}^{ls} = n_{dt}^{uow} - n_{dt}^{o} - x_{dt}^{c} \qquad \forall t$$

$$z_{dt}^{o} = n_{dt}^{uow} - x_{dt}^{c} - z_{dt}^{ls} \qquad \forall t$$

$$n_{dt}^{uow} \in Z_{n^{uow}}, n_{dt}^{o} \in Z_{n^{o}} \qquad \forall t$$

$$i^{ls} \in \{0,1\}^T, i^{o} \in \{0,1\}^T$$

Additionally, the calculation of overtime z_{dt}^{o} involves lost service z_{dt}^{ls}, which is not a problem when the formulation is linear, but the nonlinearity of lost service in the new formulation (6-8) makes it infeasible for standard solvers. Instead, note that the sum of the lost service and overtime performed is the difference between uncovered open work and available cover or 0 when we have more cover than work, that is $z_{dt}^{ls} + z_{dt}^{o} = \max(n_{dt}^{uow} - x_{dt}^{c},\ 0)$. A new variable $z_{dt}^{lspo} = z_{dt}^{ls} + z_{dt}^{o} = \max(n_{dt}^{uow} - x_{dt}^{c},\ 0)$ was defined to represent the sum of lost service and overtime, substituting it for z_{dt}^{o}, we have the formulation (6-9) shown below, which can be solved with a standard solver.

$$\max_{n_{dt}^{uow}, n_{dt}^{o}} \quad (c_{ls} - c_{ot}) \sum_{t=1}^{T} z_{dt}^{ls} \times i_{t}^{ls} + c_{ot} \sum_{t=1}^{T} z_{dt}^{lspo} \times i_{t}^{lspo} \tag{6-9}$$

$$\begin{aligned} \text{s.t.} \quad & x_{dt}^{c} = \sum_{j=1}^{x_d^c} \sum_{k=max(0,t-s)}^{t} y_{jdk} && \forall t \\ & z_{dt}^{ls} = n_{dt}^{uow} - n_{dt}^{o} - x_{dt}^{c} && \forall t \\ & z_{dt}^{lspo} = n_{dt}^{uow} - x_{dt}^{c} && \forall t \\ & n_{dt}^{uow} \in Z_{n^{uow}}, n_{dt}^{o} \in Z_{n^{o}} && \forall t \\ & i^{ls} \in \{0,1\}^{T}, i^{lspo} \in \{0,1\}^{T} \end{aligned}$$

6.5 The MBTA Case Study

This section compares the performance of the different scheduling models presented in the previous section, making use of the absence models estimated in Chapter 5. First the dataset and its limitations are described (Section 6.5.1), followed by discussion of uncertainty set formation (Section 6.5.2). Then the experimental setup (Section 6.5.3) is presented, including the characteristics of the three models: the hindsight model, the nominal model, and the adaptive robust model. The scheduling results and sensitivity analyses of the cost parameters for the constrained and unconstrained resources scenarios are presented in Sections 6.5.4 and 6.5.5. Finally, the comparison between constrained and unconstrained resource scenarios and model selection considerations are presented in Section 6.5.6.

6.5.1 Data Description and Limitations

The dataset used in this section is the same as in previous chapters. In this section, the focus is on Southampton Garage in rating 4 of 2019 and the results are compared with the actual performance in the rating. Since each holiday has different characteristics and the holidays typically have enough operators, holidays are not considered in this case study. Some modelling adjustments were made to the formulations in Sections 6.3 and 6.4 to accommodate data limitations and/or the chosen modelling scope and framework:

1) The time horizon for the models is taken to be a week. Extraboard availability (N and M) was taken to be the maximum number of extraboard operators for one day and the total number of observed extraboard operator-days in the week.

Theoretically, the number of extraboard operators to schedule should be the same for each week, while the actual number of extraboard operators available also reflect the extraboard operators being updated in the system as the original owners of the work, the number of excess vacation relief operators, extraboard operators who are assigned to cover long-term absences, and extraboard operator absence. If the extraboard operator has been updated as the original owner of the shift, (s)he will not be included in the extraboard count for the day. Besides the number of extraboard operators scheduled during the pick, excess vacation relief operators will be added to the extraboard and absent extraboard operators will decrease the count. The result is that, empirically, the week-to-week variation of the number of extraboard operators was large. For example, in R4 of 2019 at Southampton, week 12 had the most extraboard operators scheduled, with 124 operator-days, while in week 5 the total was 60.

To make the model results comparable to the current assignment outcomes, the constraint that each week should have the same number of extraboard operators was relaxed and the optimization model was run for each week. Then the results for the whole rating were aggregated to form the rating-level statistics. Empirical observation of actual availability was taken as the model constraint. Additionally, because of this update, the relationship between the number of extraboard operators and the number of extraboard-days worked (i.e. $x_d^e = \sum_{i \in I} \delta_{di} x_i$ and $x_d^e \leq \frac{1}{5} \sum_{i=0}^{6} x_i$ for each day d) is not enforced in the model. In weeks with holidays, the total number of days was then 5 (both Thursday and Friday of Thanksgiving week were holidays) or 6. In R4 of 2019 at Southampton, there were 16 weeks with the total number of operators for each week shown in Table 6-1.

Table 6-1 Extraboard Days (by week)

Week #	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
# extraboard-days observed	80	74	82	81	75	60	72	108
# days in the week	6	7	7	7	7	7	6	7
Week #	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
# extraboard-days observed	74	78	87	124	63	88	111	77
# days in the week	7	7	6	7	5	7	7	7

2) All extraboard operators are all full-time operators.

This is one of the MBTA's scheduling policies, although it is sometimes relaxed especially when the extraboard size would otherwise be inadequate. Since any relaxation is a matter of judgment, it is assumed in the model that this policy is followed strictly.

3) During the evaluation phase, an hour of time is randomly made unavailable for each cover list operator. That is, a maximum of 7 hours of work is possible for a cover operator while they are paid for 8 hours.

Travel time is incurred when the trip starts at a different location from the previous location of the cover list operator. The smallest unit of coverage is a round trip from (and back to) the garage. Some short durations of cover availability cannot be utilized because they are too short to cover a round trip, or the trip has already started when the cover became available. Since open work is modelled as a workload curve, this work piece-level information cannot be incorporated. Rather, an hour of time for each operator-day was randomly removed to approximate this loss.

4) Known-in-advance absences that exceed the number of extraboard operators are lost.

On rare occasions, the number of known-in-advance absences can exceed the number of extraboard operators. In practice, since no cover list operators remain, any uncovered known-in-advance absences, along with unexpected absences, are offered as overtime. However, the time-of-day distribution of known-in-advance absences was not modelled, so this solution is impractical under the current modelling framework. Instead, any uncovered known-in-advance absences are treated as eight hours of lost service. Although this known-in-advance absence piece can likely be covered on overtime since it is known in advance, the supply of overtime will decrease accordingly. Because of the shortage of extraboard operators and reliance on overtime, unexpected absences are likely going to be lost. Therefore, not reducing the overtime availability and directly treating the known-in-advance absence as lost service is a reasonable approximation.

5) Service trips take precedence over all non-service duties and non-service duties can be scheduled in between service trips without conflict.

This assumption is *not* correct. If this assumption were to hold, when the cover list operators perform non-service duties, no overtime or lost service should occur. However, empirical evidence suggests that there was a significant amount of overtime performed and/or services lost while the cover list operators were on non-service duties (Table 4-3). However, because of data limitations (see Section 4.3.5), models for essential non-service hours are beyond the scope of this thesis and we only consider the scheduling of extraboard operators for absences. Both known-in-advance and unexpected open work are assumed to consist only of absence ($n^{kow} = n^k$, $n^{uow} = n^u$) in this case study. From this point on, we will limit our scope to account only for absence and overtime availability when evaluating the effectiveness of the proposed models.

6.5.2 Uncertainty Sets

The uncertainty sets ($Z_{n^{kow}}, Z_{n^{uow}}, Z_{n^o}$) in the robust formulation are the range of values that the parameters could take. The uncertainty sets were derived from the models estimated in Chapter 5 and the approximations described in Section 6.1. Since probability distributions were estimated, the modeler could choose the probability that the uncertain parameters (absence and overtime availability) will fall into the uncertainty set, i.e. the degree of robustness, denoted as α. α is the significance level parameter, meaning that the actual absence and overtime availability fall into the uncertainty sets with probability $(1 - \alpha)$ according to the estimated models. Therefore, for the scheduling models, the smaller α, the greater the robustness. High robustness usually results in more conservative solutions as more resources are allocated to cover extreme cases. In this case, having high robustness would imply having more extraboard operators such that the worst case can be covered. However, having more extraboard operators to cover extreme cases would result in a higher fixed cost and (probably) higher overall cost since extraboard operators are paid regardless of whether, or not, there is productive work to do. Both Poisson and negative binomial distributions are discrete distributions. Normal approximations are used where appropriate to reduce computational complexity. Suppose that the level of significance we are scheduling for is α, $z_{\alpha/2}$ represents the z-score of the standard normal distribution.

Known-in-advance Absences:

Estimated model:

$$n_d^k \sim N\text{egBin}\left(\mu_d^k,\ \alpha_k\right)$$

The model estimated for known-in-advance absence was the negative binomial regression model given by Equation (5-1). Given the dependent variable values of date d, the number of known-in-advance absences has mean μ_d^k and dispersion α_k. The normal approximation is used for the negative binomial distribution $n_d^k \sim \mathcal{N}\left(\mu_d^k,\ \mu_d^k(1 + \alpha_k)\right)$. Known-in-advance absences are independent for each date d. Therefore, for each date d, the uncertainty set for the scalar n_d^k is

$$Z_{n^k} = \left\{n_d^k \in N: \left|n_d^k - \mu_d^k\right| \le z_{\alpha/2}\sqrt{\mu_d^k(1 + \alpha_k)}\right\}$$

The uncertainty set is visualized below. The x-axis is the days in the rating and the y-axis is the number of known-in-advance absences for each day as well as the range of values for different robustness levels. Because the model only captures the day-of-week and seasonal trends in the rating, even the highest level of robustness cannot always capture the true value.

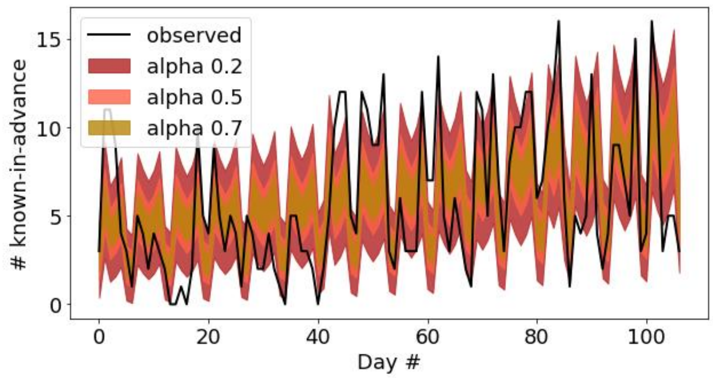


Figure 6-2 Uncertainty Range for Known-in-Advance Absences

<u>*Unexpected Absences*</u>:

Estimated model:

$$n_{d.}^{u} = \sum_{t} n_{dt}^{u} \sim N\text{egBin}(\mu_{d}^{u}, \alpha_{u})$$

$$\text{a}b_{dt}^{u} = \text{a}b_{d.}^{u} \times p_{dt}$$

where p_{dt} is the time-of-day distribution of unexpected absences for date d ($\sum_t p_{dt} = 1$).

Similar to the number of known-in-advance absences, the daily total of unexpected absence hours follows a negative binomial distribution with mean μ_d^u and dispersion α_u. The normal approximation is used for the negative binomial distribution of daily total hours $n_{d.}^{u} \sim \mathcal{N}(\mu_d^u, \mu_d^u(1+\alpha_u))$. Since there is also uncertainty on the time-of-day distribution, uncertainty limits also need to be constructed for unexpected absence by time-of-day. For continuous distributions such as a normal distribution, we can directly scale the mean and variance. It follows that each n_{dt}^{u} also follows a normal distribution $\mathcal{N}(\mu_d^u p_{dt},\ p_{dt}^2 \mu_d^u(1+\alpha_u))$. Constraining both the hourly number of unexpected absences and the daily total hours of unexpected absence, the result was a budget uncertainty set:

$$Z_{n^{ua}} = \left\{ n_{dt}^{u} \in \mathbb{R} \colon |n_{dt}^{u} - \mu_d^u p_{dt}| \le z_{\alpha/2} p_{dt} \sqrt{\mu_d^u (1+\alpha_u)}, |n_{d.}^{u} - \mu_d^u| \le z_{\alpha/2} \sqrt{\mu_d^u(1+\alpha_u)} \right\}$$

The uncertainty set for the daily total is visualized below. The x-axis is the days in the rating and the y-axis is the unexpected absence hours for each day as well as the range of values for different robustness levels considered. Because the model only captures the day-of-week and seasonal trend in the rating, even the highest level of robustness cannot always capture the true value.

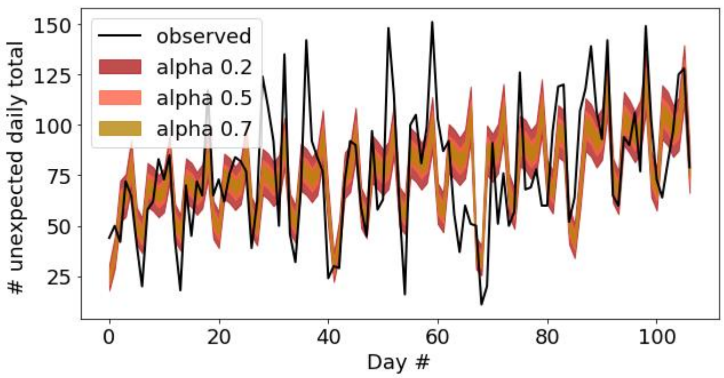


Figure 6-3 Uncertainty Range for Daily Total Unexpected Absence Hours

<u>*Overtime Availability*</u>:

As described in Section 6.1, no satisfactory model was estimated for overtime. Instead, the empirical distribution is used for scheduling. Assuming that the overtime availability for each hour on each day of the week follows a normal distribution, let μ_{it}^{o} and σ_{it}^{o} represent the mean and standard deviation of the empirical distribution for day of week i and hour t, using only the observations where lost service was observed (indicating overtime performed = overtime available).

Since valid observations do not necessarily come from the same day, a daily total of available overtime operator-hours cannot be directly observed. Although ideally, we would like to constrain the daily total of available overtime hours, it cannot be done in this case. Instead, a more conservative box uncertainty set was used:

$$Z_{n^o} = \{n_{it}^{a} \in R: |n_{it}^{o} - \mu_{it}^{o}| \leq z_{\alpha/2}\sigma_{it}^{o}\}$$

The uncertainty set is visualized below. Unlike absences, overtime availability was taken from the day-of-week and hourly average. For each day of week and hour the availability was taken to be the same. Therefore, the timeseries for a week was plotted. On the x-axis is day of week and on the y-axis is the overtime availability for each hour as well as the range of values for different robustness levels considered.

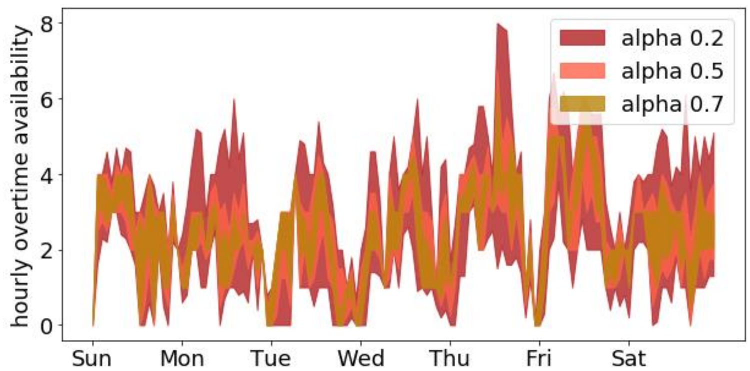


Figure 6-4 Uncertainty Range for Hourly Overtime Availability

6.5.3 Experimental Setup

Models Estimated:

Three models were estimated: the hindsight model, the nominal model, and the adaptive robust model. In **the hindsight model**, the realized absence (both known-in-advance and unexpected) was entered into the nominal model. Because of the requirement that cover list operators work straight eight-hour shifts, there is unavoidable loss and inefficiency in the scheduling process. The hindsight model quantifies this loss and serves as a benchmark for other models when evaluated on the realized outcome. **The nominal model** is the model discussed in Section 6.3. Absences and overtime availabilities (n_d^k, n_d^u, n_{dt}^o) were taken as the nominal values of regression means (μ_d^k, μ_d^u) and the mean of the observed availability μ_{dt}^o. No uncertainty was considered. **The adaptive robust model** was discussed in Section 6.4. Three adaptive robust models with different levels of robustness ($\alpha = 0.2, 0.5, 0.7$) were tested. For example, when α=0.2, the actual absence and overtime availability fall into the uncertainty sets of the robust model with probability 0.8 according to the estimated models. Increasing levels of robustness (decreasing values of α) means that the model took into account larger ranges of absence and overtime availability. Experimenting with the level of robustness offers insights into the tradeoff between robustness and efficiency. Assignment plans from the more robust models should be more hedged under extreme (worst) cases but are not necessarily better in average cases because of the redundancy built into the assignment plan to cover these extremes.

Model Parameters:

The initial cost of lost service hours (c_{ls}), overtime hours (c_{ot}), and extraboard operator hours (c_{ex}) are taken to be 5, 2, and 1, respectively. First, extraboard operators are paid at the regular rate, so $c_{ex} = 1$. Overtime is paid at 1.5x the regular rate, but considering the employee benefits associated with hiring additional extraboard operators, an hour of overtime costs about the same as an hour of extraboard operator time. However, c_{ot} was set to 2 because transit agencies would not want to rely too heavily on operator overtime to deliver service. Relying too heavily on overtime has non-monetary costs related to operator fatigue, low job satisfaction, and vulnerability to missing service. Lastly, the cost of lost service hours was chosen such that it is the least desired outcome; yet, c_{ls} was not so large that costs of overtime and extraboard operators did not matter, resulting in a great deal of unproductive extraboard time. Since the selection of these parameters was somewhat arbitrary, a sensitivity analysis was performed (see Sections 6.5.4 and 6.5.5). In all models, including the hindsight model, overtime availability was taken to be the empirical distribution. Since known-in-advance absences are known at the time of operational level scheduling and they are covered with 100% efficiency, only the cover list (# extraboard operators - # known-in-advance absences) utilization was calculated.

Model Evaluation:

All three of the models were evaluated by the financial cost, the amount of lost service, performed overtime, scheduled extraboard hours, and cover list utilization. Because of the aggregate representation of the hourly workload curve, some work piece level information cannot be accurately reflected, which results in an assumed 100% maximum cover list utilization. However, in reality there are unavoidable losses in cover hours (see #3 in Section 6.5.1). To account for the unavoidable loses, during the evaluation phase, an hour of time is randomly made unavailable for each cover list operator. That is, a maximum of 7 hours of work is allowed for a cover operator while they are paid for 8 hours.

The direct financial cost is the monetary cost paid to the operators and is expressed as a percentage of the current financial cost. It is simply an index to show the expected changes with any proposed solutions relative to the current cost. Because of the benefits associated with the extraboard operators, an extraboard hour and an overtime hour are considered to have the same financial cost. Although lost service greatly hurts the service reliability and has large indirect costs, it is not part of the financial cost consideration. Lost service is considered in the objective function, which is an artificial weighted cost for combining the direct costs of operator wages and the indirect costs because of unreliable service.

The models are evaluated in three different cases: the realized outcome, the average case, and the worst case. **The realized outcome** was based on the observed absences in hindsight. To evaluate the model assignments under different realizations of absence and overtime availability, 100 draws were made independently from the estimated absence and overtime availability. Each draw contains an instance of the entire rating with each day drawn independently (112 days, for example, in rating 4 of 2019). **The average case** was obtained by taking the average performance from the 100 draws of the rating. **The worst case** was the case with the worst objective function value among the 100 draws. For each assigned operator on each day, an hour was randomly blocked as unusable to approximate the time lost due to travel and available duration not always being enough to cover a piece of work.

Since the realized outcome was used to build the hindsight model, it is expected that the best assignment will be achieved with the hindsight model in the realized outcome. Since the nominal model was built based on the expected value of modelled absence, it is expected that it will have better average performance than the others, but there will also be edge cases that might affect the nominal model performance more than for the robust models. The average cost of the robust models might be similar to the nominal model and the tradeoff between efficiency and robustness is what we want to observe. Finally, since neither the hindsight model nor the nominal model account for uncertainty, it is expected that the robust models will achieve the lowest objective function value for the worst case.

Scenarios Considered:

First, the current garage performance is quoted from Table 4-2. The current assignment plan cannot be evaluated for the simulated ratings because the number of known-in-advance absences, and therefore the number of cover list operators, vary across different simulation trials. The current assignment plan only gives the assignment for one of those possibilities, therefore it only applies to the *realized outcome* evaluation case. However, there are a few assumptions in the models that affect the quality of this comparison: the biggest being the inability to model non-service duties, and the assumption that every cover operator works for eight hours.

In addition to comparing with the current performance, the models are evaluated under two scenarios: constrained resources (Section 6.5.4) and unconstrained resources (Section 6.5.5). In the constrained resources case, the weekly number of extraboard-days available and the maximum number of extraboard operators that can be scheduled in any day were set to the observed values during the rating. Model results from the constrained case can be used to assess the current assignment plan. In the scheduling model with unconstrained resources, no upper limits were set for the number of extraboard-days in a week or the number of extraboard operators in a day. Model results from the unconstrained case can be used to get recommended extraboard staffing levels in accordance with the scheduling policies and model assumptions set out in Sections 6.1 and 6.5.1.

Sensitivity Analysis with respect to the Cost Coefficients:

Since the cost coefficients in the model ($c_{ls} = 5,\ c_{ot} = 2,\ c_{ex} = 1$) were reasonable but arbitrary, the sensitivity of the results is analyzed. To make the objective function values for the assignment plans comparable, the sum of the costs was constrained to be 8. Additionally, while constructing different test scenarios, the cost coefficients were constrained by:

cost of cover list hours < cost of overtime hours < cost of lost service hours.

For example, (4, 2.5, 1.5) is a valid combination, but (4, 2, 2) is not.

Presentation of Results:

All models and tests were performed on R4 of 2019 at Southampton. All costs and hours presented in the model results are totals for the entire rating unless otherwise stated. Since we assumed that known-in-advance absences were assigned in advance, the results table only shows the hours and utilization rates for the cover list operators.

6.5.4 Scheduling with Constrained Resources

Rating-Level Metrics:

First, the effective use of the current resources was tested. The rating-level performance metrics for the different models and evaluation scenarios are shown in Table 6-2. The number of extraboard hours were the same across all cases and models, but slight differences exist in cover list hours due to the lost known-in-advance absence hours (see #4 in Section 6.5.1). When there are known-in-advance absence runs lost, the absence runs are counted in lost service, instead of covered using the extraboard operators. Since the total number of extraboard operators are the same, on other days there will be more cover list operators.

For the realized outcome, performance of the current assignment plan was calculated. The worked hours in the dataset are not a clean cut at eight hours so there are some differences in **the scheduled cover list hours**, even though the total number of weekly extraboard-days was constrained to be the same. Since information on essential non-service duties is missing, **the realized lost service and overtime** are higher than the modelled amount, whereas **the productive cover time and utilization** are lower. The model assumes that all existing resources can be used to cover services; however, in reality some of the scheduled cover hours are assigned to essential non-service work and should be considered productive but are currently labeled as unproductive. As a result, the productive cover hour and utilization rate are lower bounds on the actual values. Because some cover time is dedicated to non-service duties, cover hours available for covering services decrease and more services needs to be covered by overtime or are dropped. Therefore, the number of overtime and lost service hours are likely larger than the modelled amounts. Since overtime and cover list are considered the same cost financially, the **financial costs** are not influenced by the hours that should be covered by the cover list but are covered by overtime. When service is dropped, those hours are excluded from the financial cost and included in the lost service hours, making the actual financial cost of covering for absences larger than the current figure. At the same time, some cover list hours are used for non-service duties, therefore, if only absence is considered, the scheduled cover list hours are likely smaller, making the actual financial cost of covering absences smaller than the current figure. Therefore, without the number of hours that the cover list operators were working on essential non-service duties, it is hard to decide whether the true financial cost would be higher or lower.

Table 6-2 Rating-Level Metrics (Constrained Resources)

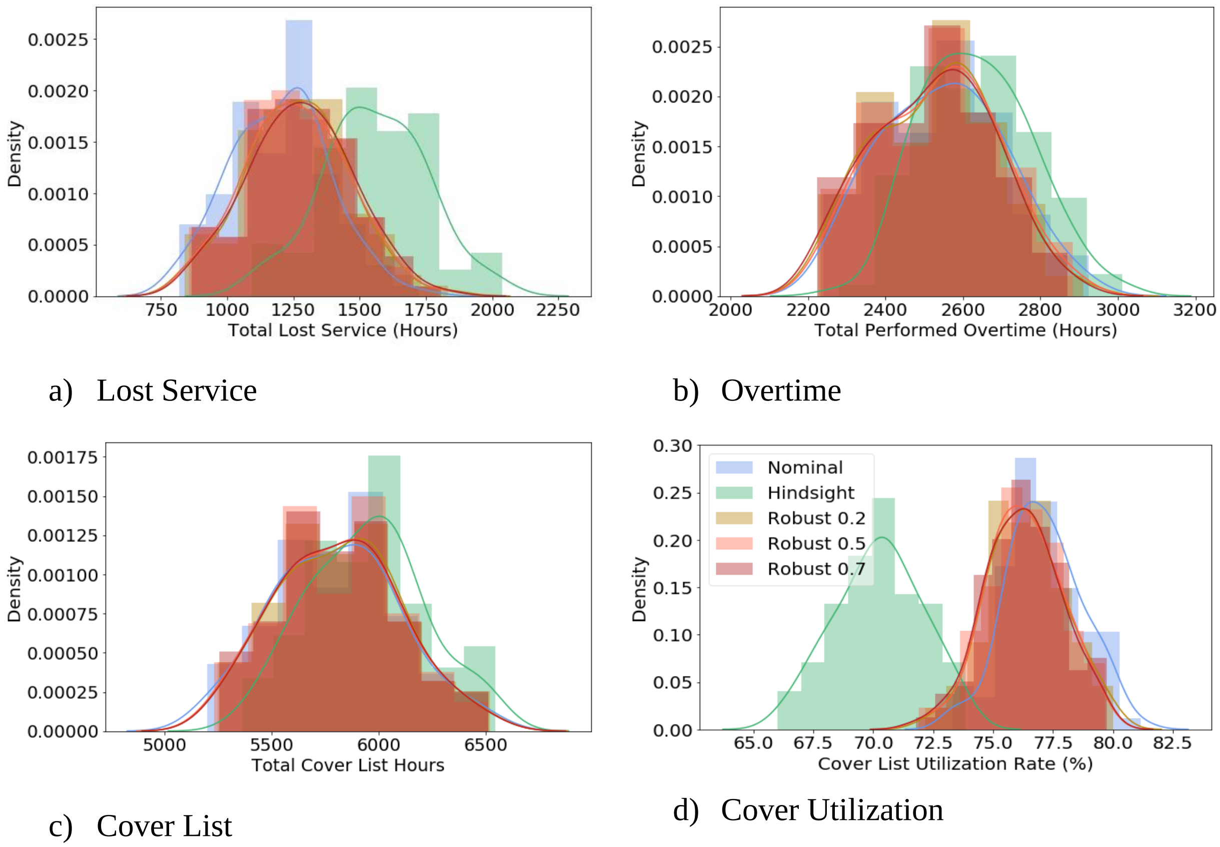
		Current	Hindsight	Nominal Model	Adaptive Robust (α)		
					0.7	0.5	0.2
Realized Outcome	Financial cost	100%	90%	89%	89%	89%	89%
	Lost Service (h)	1945	992	1383	1433	1428	1450
	Overtime (h)	4788	2805	2748	2743	2754	2729
	Scheduled Cover List (h)	5194	5648	5704	5728	5728	5728
	Productive Cover List Time (h)	3023	4433	4155	4134	4128	4130
	Cover List Utilization (%)	58%	78%	72%	72%	72%	72%
Average Behavior	Financial cost	n/a	89%	88%	88%	88%	88%
	Lost Service (h)	n/a	1555	1197	1274	1260	1263
	Overtime (h)	n/a	2633	2548	2524	2535	2528
	Scheduled Cover List (h)	n/a	5946	5800	5811	5807	5814
	Productive Cover List Time (h)	n/a	4173	4472	4429	4429	4439
	Cover List Utilization (%)	n/a	70%	77%	76%	76%	76%
Worst Case	Financial cost	n/a	89%	89%	89%	89%	88%
	Lost Service (h)	n/a	2037	1723	1805	1797	1832
	Overtime (h)	n/a	2750	2616	2638	2635	2596
	Scheduled Cover List (h)	n/a	5632	5480	5504	5488	5496
	Productive Cover List Time (h)	n/a	3905	4201	4120	4116	4127
	Cover List Utilization (%)	n/a	69%	76%	74%	75%	75%

The amounts of **lost service, overtime,** and **cover utilization** were similar for the nominal and robust models. The assignment results from different modelling strategies are also similar. Since the cost of the extraboard is the same in this case, the **financial cost** is lower with less overtime performed. The amount of overtime performed is also similar across all models. Either the current staffing level was more than enough (such that all services are covered no matter what), or the current staffing level was inadequate (such that all strategies will assign the extraboard to cover the most basic needs and the rest is left to overtime). Since the amount of lost service was not negligible, the second reason is more plausible. The financial costs estimated by the models are around 90% of the current cost. Whether the actual savings is above or below 10% depends on the number of hours of essential non-service duties worked by the cover list.

The absence and overtime availability distributions used for evaluation were simulated from the modelled absence and overtime availability, which are the inputs for the nominal model and the robust models. The nominal model and the robust models resulted in similar average performances and cover list utilizations, with the nominal model doing slightly better. The distributions of lost service, overtime, scheduled cover list, and cover list utilization for the 100 simulated cases are shown in Figure 6-5. With the resources constrained, robust models are not better than the nominal. Robustness levels are increased by introducing more extraboard operators such that when unusually high levels of absence occur, some of the surge can be absorbed. In this case, with the number of extraboard-days constrained, resources can only be moved around. Therefore, the differences in lost service, overtime, scheduled cover, and cover utilization between the nominal and the robust models, as well as with different robustness levels, are insignificant.

The hindsight model was overfitted to the realized outcome and achieves the lowest lost service and overtime in the realized outcome. When the evaluation scenario is the modelled absences, the hindsight model yields the lowest utilization rates and the most lost service (the green curves). The nominal model, on the contrary, achieves the lowest lost service. The accuracy of the inputs is critical to the deterministic models, even in the constrained resource case.

The average **cover list utilization rate** was calculated as the total number of worked service hours by the cover list in the 100 draws divided by the total hours of scheduled cover list in the 100 draws. Because 1 hour of time was made unavailable for every extraboard operator to account for travel time and unavoidable losses, the theoretical maximum cover list utilization rate is 87.5%. If the modelled distribution was the same as the evaluation set, utilization was 76-78% (hindsight model evaluated on realized outcome; nominal and robust models evaluated on average simulated behavior). When the distribution differed from the modelled distribution, the utilization rates were lower – around 70-72%. Most of the cover list was matched to open work regardless of the changing distribution of absence, and the covers can only meet the basic needs with the variations leading to a combination of overtime and lost service.



a) Lost Service

b) Overtime

c) Cover List

d) Cover Utilization

Figure 6-5 Model Results with Constrained Resources

<u>*Tactical-Level Assignment:*</u>

Figure 6-6 shows the tactical-level assignment results for different models as well as the current assignment. The overall results produced by the different models are similar. The current assignment assigns more extraboard operators on weekdays and fewer on weekends. The two deterministic models (hindsight and nominal) produced similar results and the three robust models produced similar results. Some differences exist in the treatment of Fridays and Sundays between the deterministic and robust models. More resources were allocated to cover Sundays when the robustness level was increased since Sundays have lower but more variable operational-level costs whereas Fridays have higher operational-level costs but are the least variable.

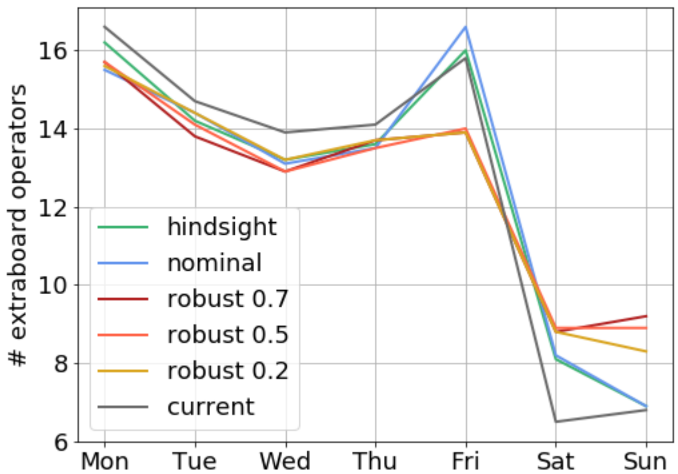


Figure 6-6 Tactical-Level Assignments (with constrained resources)

<u>Operational-Level Assignment:</u>

Figure 6-7 shows the average operational-level assignments for different days of the week. Since the time-of-day distribution was taken as the average profile for each day of the week, the time-of-day distributions of unexpected absence averaged by day of week were the same as those for hindsight, while differences exist between individual days. As a result, the assignment plan was very similar across all models.

The purple line in each figure is the observed time-of-day distribution of cover list assignments and the black line is the distribution of unexpected absence by time-of-day. The model results were similar to the observed distribution except for Mondays. The differences in assignments were due to the differences in estimated overtime availability. Despite the differences, the performances of these assignment plans were almost identical since the resources were constrained.

Since the time-of-day distribution of unexpected absence on weekdays was typically bi-modal, theoretically, the cover list time-of-day distribution should also be bi-modal. However, especially in the constrained case where overtime is heavily used, the assigned time-of-day distribution of cover list was a product of both unexpected absence and overtime availability, and overtime availability was generally unimodal. The resulting pattern was more unimodal with a small dip at midday. Another factor is that the straight eight-hour shift rule might be reducing the efficiency of the cover list: this rule will be tested in Section 6.6.2 by allowing two independent four-hour duties, which should enable the operational-level assignment to adhere more closely to the absence pattens.

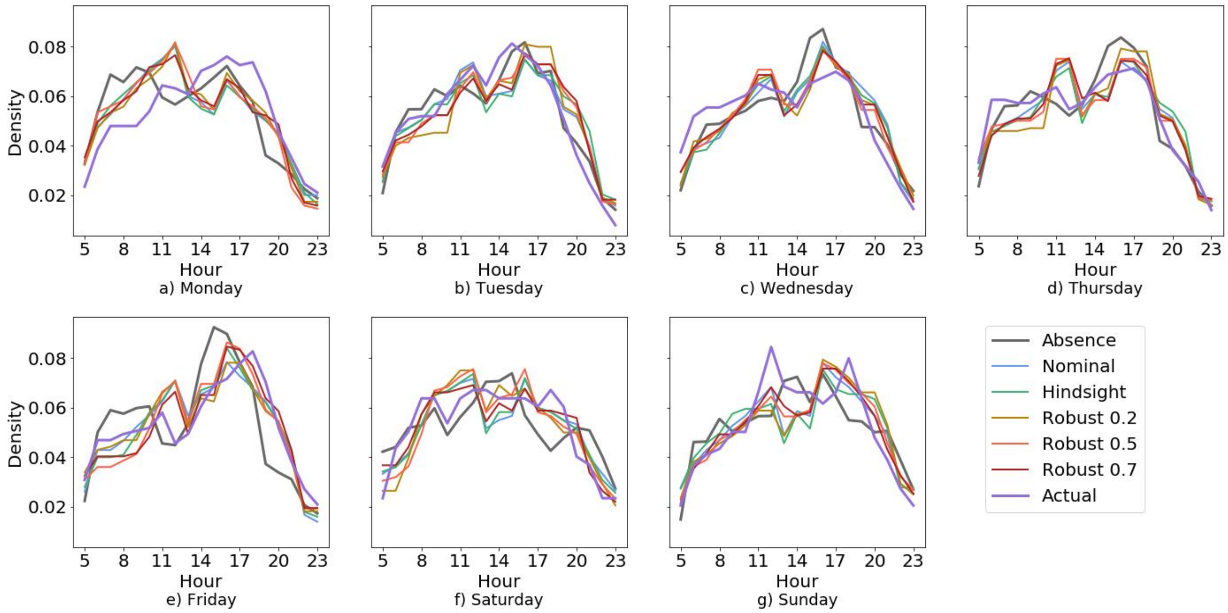


Figure 6-7 Operational-Level Assignments (with constrained resources)

Sensitivity Analysis of Cost Coefficients:

Since there are 3 covariates (cost of cover list hours, overtime hours, and lost service hours), we isolate each to see the effect on the assignment results. Figure 6-8 show the sensitivity of lost service hours, performed overtime hours, and scheduled cover list hours, with respect to the cost of cover list hours.

All models are insensitivities to the cost coefficients. Due to the resource constraint, the service outcomes are unaffected by the cost parameters (Figure 6-8), illustrating again that the current extraboard-days constraint is the factor limiting performance. Due to the inevitable need for overtime and the cost of overtime being lower than the cost of lost service, the amount of overtime performed cannot be reduced by varying overtime costs. Because the service outcomes are constant, sensitivity plots with respect to cost of overtime and lost service are omitted. For the realized outcome and worst case, the assignment results were also insensitive to the cost parameters, similar to the average behavior, and the corresponding plots are shown in Appendix E.

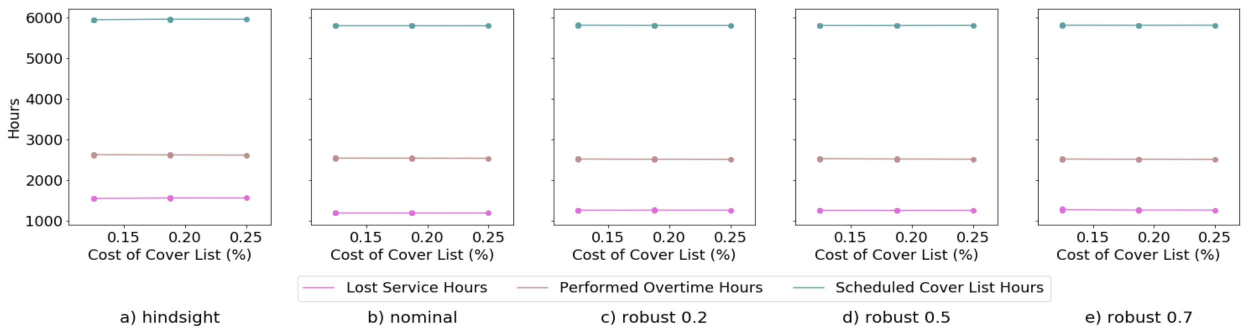


Figure 6-8 Sensitivity w.r.t. Cost of Cover List Hours (with constrained resources)

In summary, with the resources constrained at the current level, all models give similar results, with only the basic needs being covered. The assignment strategies are similar across all models and are insensitive to the cost coefficients. To further study performance as a function of staffing levels, as well as the characteristics of the different models, results with the constraint on weekly extraboard-days relaxed are presented in the next section.

6.5.5 Scheduling with Unconstrained Resources

<u>*Rating-Level Metrics:*</u>

To determine the recommended levels of extraboard staffing, the constraint on the maximum number of operators was relaxed with the model results shown in Table 6-3. When the resources are unconstrained, the different characteristics of assignment strategies are shown more clearly.

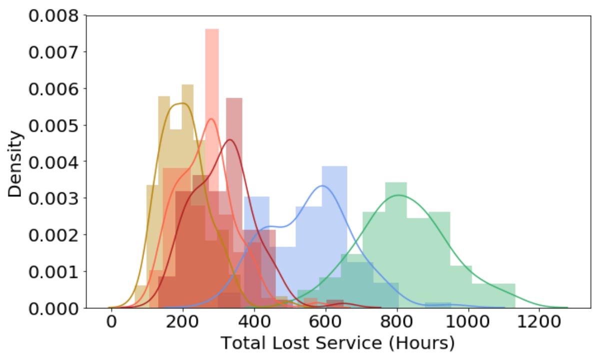
A model with robustness level α meant that the model took into account estimated inputs at $(1-\alpha)$ confidence level. Therefore, a smaller α meant more robustness. Being more robust meant scheduling more cover list operators such that when there are unusually high levels of absence, cover list operators are more likely to be available. In this case, with the inputs having the same expected values, the robust models, on average, scheduled 25-45% more extraboard operators than the nominal model in exchange for 44-63% reduction in lost service and 34-54% reduction in overtime. However, having more cover list operators increases financial costs. When there are fewer absences, the unproductive cover list operators still get paid and the **cover list utilization rates** decrease with increasing robustness levels. The reduction in overtime alone cannot offset the increase in the cover list operator cost. Therefore, the **financial costs** incurred by the MBTA increase as robustness levels increase. The financial costs are expressed as a percentage of the current financial cost. The deterministic models increased their costs by 10% compared to the constrained case, and are now on par with the current costs, but the robust models would recommend 9-16% more financial costs than the deterministic models.

Table 6-3 Rating-Level Metrics (Unconstrained Resources)

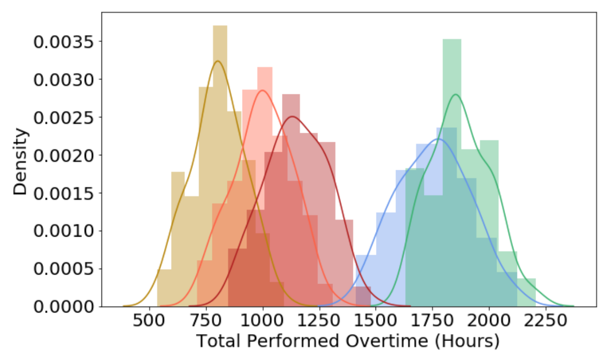
		Hindsight	Nominal Model	Adaptive Robust (α)		
				0.7	0.5	0.2
Realized Outcome	Financial cost	100%	99%	109%	112%	116%
	Lost Service (h)	302	771	444	367	299
	Overtime (h)	1715	2141	1560	1399	1160
	Scheduled Cover List (h)	8328	7768	9736	10336	11296
	Productive Cover List Time (h)	6213	5334	6226	6463	6771
	Cover List Utilization (%)	74%	68%	63%	62%	59%
Average Behavior	Financial cost	101%	97%	106%	109%	114%
	Lost Service (h)	820	555	312	263	204
	Overtime (h)	1872	1754	1149	1001	804
	Scheduled Cover List (h)	8438	7802	9771	10370	11328
	Productive Cover List Time (h)	5482	5805	6639	6834	7088
	Cover List Utilization (%)	64%	74%	67%	65%	62%
Worst Case	Financial cost	103%	97%	107%	110%	116%
	Lost Service (h)	1133	952	651	578	493
	Overtime (h)	2156	1847	1306	1192	1034
	Scheduled Cover List (h)	8440	7328	9312	9912	10872
	Productive Cover List Time (h)	5643	5484	6327	6514	6756
	Cover List Utilization (%)	66%	74%	67%	65%	62%

Since the errors in the modelled absences for R4 at Southampton are around 48% for known-in-advance absences and 25% for unexpected absences, the realized outcome has a different distribution than the simulations using the estimated models. The hindsight model, naively fitted to the realized outcome, observed increases in overtime performed (increases of 157 hours) and lost service (increase of 518 hours) from the realized outcome scenario (model input) to the average-case modelled scenario (evaluation scenario). Similarly, the other deterministic model, the nominal model, which is naively fitted to the modelled absences, overtime and lost service increased by 387 and 216 hours, respectively, from the average-case modelled scenario (model input) to the realized outcome (evaluation scenario). On the other hand, for the robust models tested, the amounts of overtime and lost service are smaller, but the increases are 356 – 421, 95 – 132 hours, for overtime and lost service, respectively, only slightly better than the nominal models. Therefore, it is important to estimate the demand accurately. Otherwise high fixed costs from scheduling more operators have to be incurred to reduce variable costs by the performed overtime and lost service.

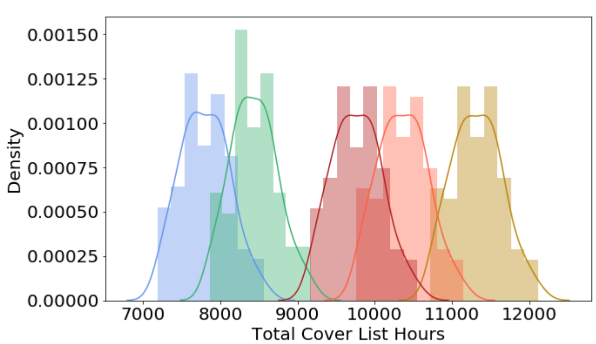
A similar pattern could be observed in Figure 6-9, where the models are evaluated on different realizations of the rating by simulation. Since the simulation was based on the estimated model, the hindsight model achieves the highest lost service and overtime. The expected value of inputs of both the nominal and the robust models match the mean of the simulations. More robust models tend to provide solutions with less lost service and overtime both in terms of the average value and the variance at similar objective function values as the nominal, with larger scheduled cover list.



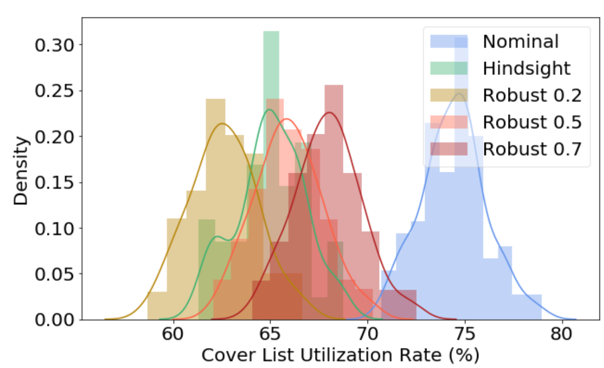
a) Lost Service



b) Overtime



c) Cover List



d) Cover Utilization

Figure 6-9 Model Results with Unconstrained Resources

Tactical-Level Assignment:

Figure 6-10 shows the tactical-level assignment results for the different models in the unconstrained case. The current average number of extraboard days per week is 88, which is far fewer than recommended by any of the models. The weekly totals suggested by the models are

110, 106, 122, 127, 135, for the hindsight, nominal, robust $\alpha = 0.7, 0.5, 0.2$ models respectively. The nominal model suggested the fewest operators, but it would still be 20% more than the current level. Among the robust models, the more robust models scheduled significantly more extraboard-days. The most robust model suggested that the extraboard staffing level should increase by more than 50%. The differences in tactical level assignment between different models were in the weekly number of extraboard-days scheduled: the day-of-week patterns were similar across the different models.

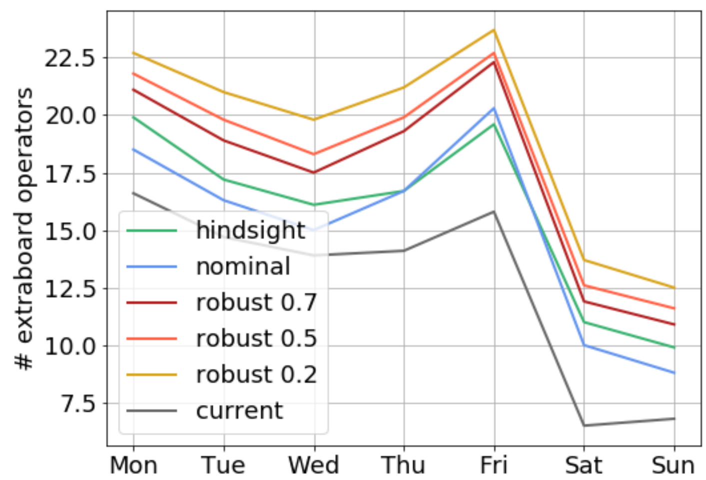


Figure 6-10 Tactical-Level Assignments (with unconstrained resources)

<u>*Operational-Level Assignment:*</u>

Figure 6-11 shows the average operational-level assignments for different days of the week. Since the time-of-day distributions were normalized, they are independent of the resource constraints, as well as of the predicted daily unexpected absence hours. As a result, the assignment plans were similar to those of the constrained case. Therefore, the observations made in Section 6.5.4 hold.

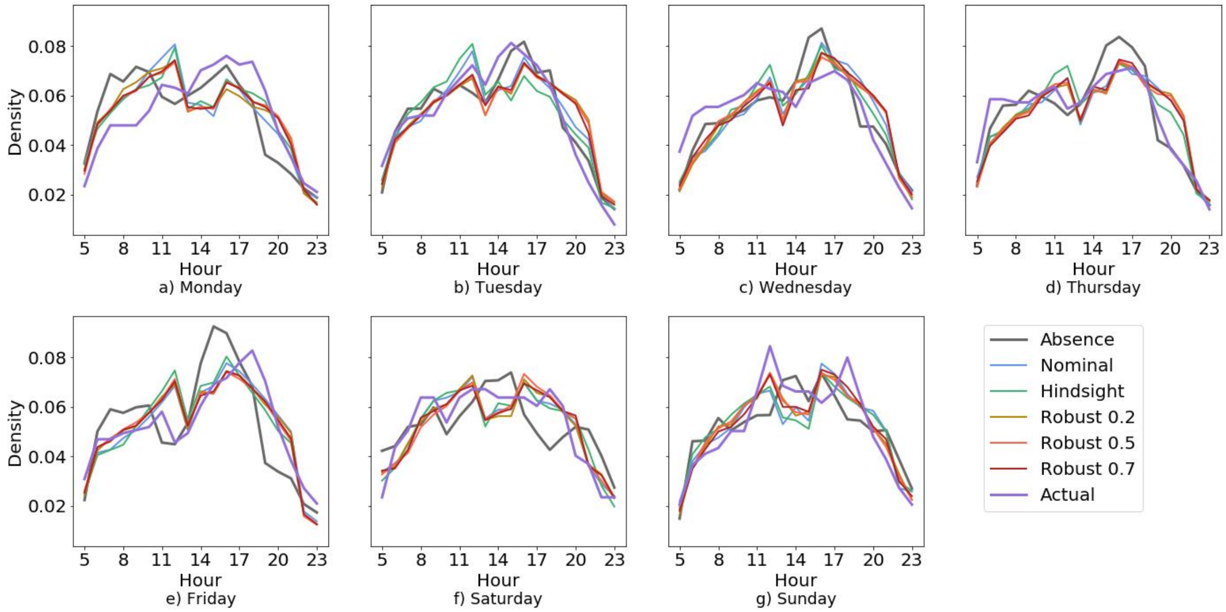


Figure 6-11 Operational-Level Assignments (with unconstrained resources)

<u>*Sensitivity Analysis of Cost Coefficients:*</u>

Figure 6-12, Figure 6-13, and Figure 6-14 show the sensitivities of assignment results (lost service hours, overtime hours, scheduled cover list hours) with respect to the cost of cover list hours, overtime hours, and lost service hours, respectively. The x-axis was taken to be the cost expressed as a percentage of the total. For example, in the original case (5, 2, 1), the percentage values in the plots would be (0.675, 0.25, 0.125). The lines represent the average cover list hours, overtime hours, and lost service hours in the 100 simulated runs. The sensitivity patterns for the realized outcome and worst case was similar to the average behavior (the plots are shown in Appendix E).

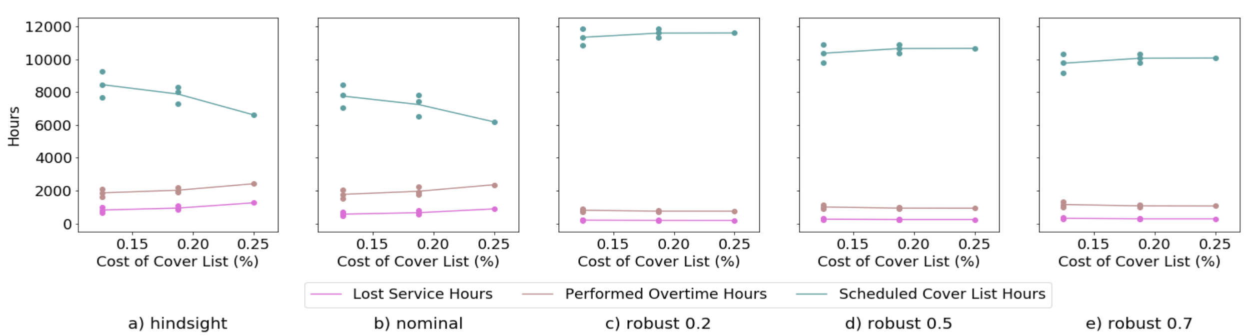


Figure 6-12 Sensitivity w.r.t. Cost of Cover List Hours (with unconstrained resources)

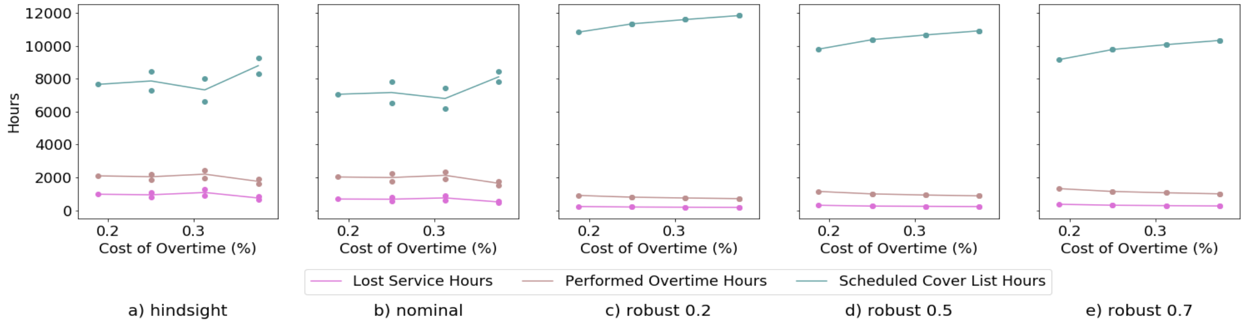


Figure 6-13 Sensitivity w.r.t. Cost of Overtime Hours (with unconstrained resources)

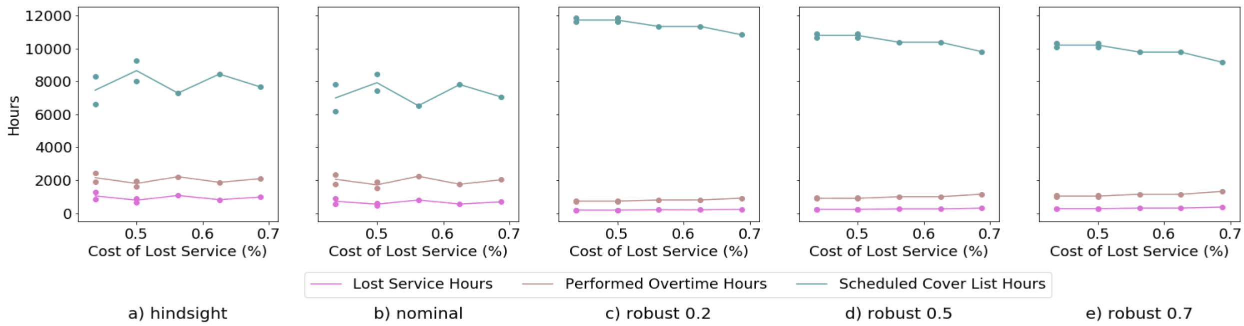


Figure 6-14 Sensitivity w.r.t. Cost of Lost Service Hours (with unconstrained resources)

Unlike the constrained scenario where all models are insensitive to the model parameters, in the unconstrained scenario, the deterministic models (hindsight and nominal model) were more sensitive to changes in the cost coefficient than the robust models (shown by higher slopes in Figure 6-12). The deterministic models not only have larger variance with respect to uncertainty in model inputs, but also to model parameters.

The number of scheduled cover list hours from the hindsight and nominal models were quite sensitive to the cost coefficients. A significant decrease in the number of cover list operators can be observed with increasing costs of the cover list. However, the assignment plans from the robust models were insensitive to the cost coefficients. Rather, they were more sensitive to the level of robustness (α). For each α, the assignment plan curves were mostly flat and the differences in costs were mostly due to the different distributions of costs.

In the unconstrained case, the sensitivity of different parameters can be observed. In this case, the cost of the cover list hours was more sensitive than the cost of overtime and lost service hours since the amount of cover list hours was determined by the model and the amount of overtime and lost service were derived, intermediate variables. In Figure 6-12, the number of scheduled cover list hours monotonically decreases with increasing cost of the cover list. However, in Figure 6-13 and Figure 6-14, the number of overtime and lost service hours did not have a clean, negative correlation, suggesting that there were other factors contributing to these results.

6.5.6 Comparing the Constrained and Unconstrained Scenarios and Model Selection

Comparing the performances across different scenarios for each model, the percent differences between values in Table 6-2 and Table 6-3 are tabulated in Table 6-4. Bigger improvements in lost service and performed overtime are observed for the deterministic models when the input data is more similar to the evaluation data for all models. For example, a 70% reduction in lost service was observed for the hindsight model with a 10% increase in the financial cost. For robust models, the effect of removing resource constraints was also pronounced with the most robust model having the largest increase in the scheduled cover list, along with the greatest reductions in lost service, overtime, and cover utilization rates. For example, while evaluating on the realized outcome, the most robust model suggested a 95% increase in cover list hours, while reductions of 79%, 57%, and 13% occurred in lost service, overtime, and cover utilization rates, respectively. With the same absences to be covered and lower utilization rates, the amount of overtime and lost service in the unconstrained case both decreased compared to the constrained case. Therefore, higher utilization rates are not necessarily better: higher utilization could be a result of either effective scheduling, or an inadequate cover list. Therefore, the cover list utilization rates have to looked at with other performance metrics to form strategic-level recommendations.

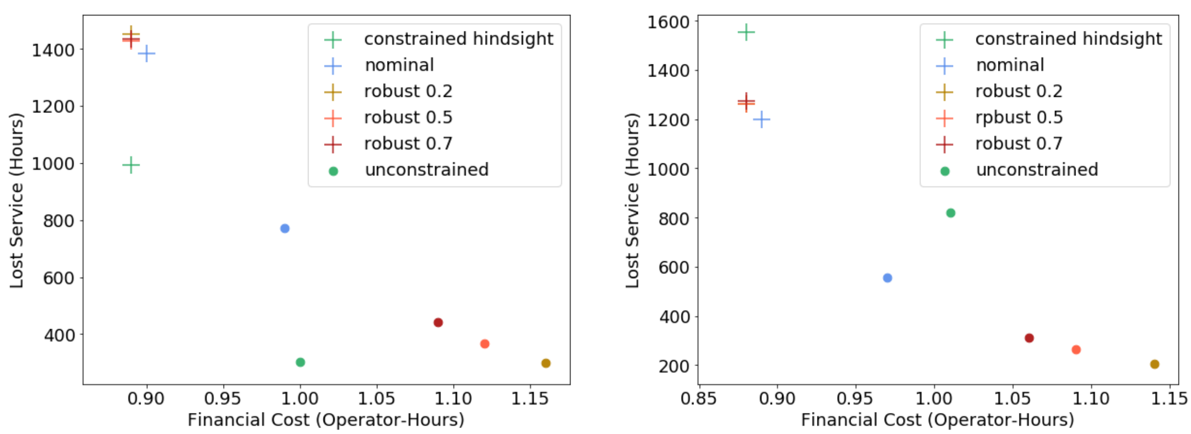
Table 6-4 Constrained vs. Unconstrained Performance Comparison

		Hindsight	Nominal Model	Adaptive Robust (α)		
				0.7	0.5	0.2
Realized Outcome	Financial cost	10%	10%	20%	23%	27%
	Lost Service (h)	-70%	-44%	-69%	-74%	-79%
	Overtime (h)	-39%	-22%	-43%	-49%	-57%
	Scheduled Cover List (h)	47%	36%	70%	80%	97%
	Productive Cover List Time (h)	40%	28%	51%	57%	64%
	Cover List Utilization (%)	-4%	-4%	-9%	-10%	-13%
Average Behavior	Financial cost	12%	9%	18%	21%	26%
	Lost Service (h)	-47%	-54%	-76%	-79%	-84%
	Overtime (h)	-29%	-31%	-54%	-61%	-68%
	Scheduled Cover List (h)	42%	35%	68%	79%	95%
	Productive Cover List Time (h)	-6%	-3%	-9%	-11%	-14%
	Cover List Utilization (%)	-11%	-13%	-17%	-17%	-17%
Worst Case	Financial cost	14%	8%	18%	21%	28%
	Lost Service (h)	-22%	-29%	-50%	-55%	-60%
	Overtime (h)	50%	34%	69%	81%	98%
	Scheduled Cover List (h)	45%	31%	54%	58%	64%
	Productive Cover List Time (h)	-14%	-9%	-14%	-14%	-14%
	Cover List Utilization (%)	-70%	-44%	-69%	-74%	-79%

In order to inform model selection, Figure 6-15 shows the tradeoff between financial costs and the amount of lost service in both the constrained and the unconstrained scenario for all model formulations. The colors indicate the different model formulations, the plus sign represent the constrained scenario and the dot represents the unconstrained scenario. The constrained models are all concentrated at the upper left corner, where the financial costs are low but lost service rates are high. The unconstrained models are scattered towards the higher end of financial costs with lower lost service rates.

The best model should reside in the lower left corner, with the lowest financial cost and lost service levels. But this total dominance of the models only happens when the input distribution matches the evaluation distribution. For example, the hindsight model dominates the other models when evaluated on the realized outcome, but is dominated by other models when evaluated on the modelled absences.

With the nominal and robust models, an efficient frontier was observed. No model strictly dominates the others (less financial cost and less lost service). With the resource constraint relaxed, given the current modelling assumptions, more cover list hours are scheduled, resulting in reduced lost service and overtime at the same time, but direct financial costs to the agency will increase as the decrease in overtime does not offset the increase in costs from additional extraboard operators.



a) Realized Outcome b) Average Simulated Behavior

Figure 6-15 Lost Service vs. Financial Cost

A deterministic model works well when the actual distribution is centered on the input of the deterministic model. The robust models produce results with slightly better objective function values but with very different characteristics: larger extraboards, less variance, less lost service and overtime, and less sensitivity to the cost coefficients in the model. Making robustness levels too high results in too many cover list operators to deal with situations that are unlikely to occur, giving the transit agency a very high financial cost (the lower right corner in Figure 6-15). To achieve decreased financial costs and lost service at the same time, the input models must be improved, otherwise, determining the optimal extraboard size and assignment strategy is a choice of balancing direct financial costs and lost service. In this case, we choose to take the value of the objective function, the overall weighted cost, into consideration and choose the model with a moderate amount of robustness. In the next section, the adaptive robust model with α=0.7 is used for several scenario analyses.

6.6 The MBTA Case Study Scenario Analysis

This section tests possible policy scenarios to see if, and by how much performance can be improved using the robust model with $\alpha = 0.7$. Two policy scenarios are tested: removal of the requirement for early-morning and late-night covers and the introduction of split covers.

6.6.1 Early-morning and Late-night Covers

The MBTA scheduling practice is that one cover list operator is reserved for both the first shift in the morning and last shift at night to ensure that the first run and the last run of the day are operated. However, during those times there are very few runs, and if the scheduled operator shows up, the cover hours will be wasted. Therefore, we want to test whether removing the requirement will help improve the assignment efficiency. In this scenario, constraints (6-5) and (6-6) were relaxed with the results for models both with and without the resource constraint, shown in Table 6-5. The “with requirement” column was taken from Table 6-2 and Table 6-3 for ease of reference.

In the constrained case, the removal of the requirement yielded a slight improvement. With the constraint removed, on 83% of the days, there are operators scheduled for the first hour; on 50% of the days there are operators scheduled for the last hour. In the unconstrained case, the effect was minimal. The small variations come from the random draw of the unavailable hour in the evaluation phase. In fact, with the requirement removed, on 97% of the days, there are operators scheduled for the first hour and on 90% of the days there are operators scheduled for the last hour. Additionally, even if the operator for the first and/or last run is present, part of the morning/evening peak is included in the report time. The saving in unproductive cover time is likely less than 3 hours.

Table 6-5 Evaluation of Dropping Early Morning and Late Night Covers

Adaptive Robust ($\alpha = 0.7$)		**Constrained resources**		**Unconstrained resources**	
		With Requirement	Requirement Removed	**With Requirement**	Requirement Removed
Realized Outcome	Financial cost	**89%**	89%	**109%**	109%
	Lost Service (h)	**1433**	1422	**444**	435
	Overtime (h)	**2743**	2725	**1560**	1559
	Scheduled Cover List (h)	**5728**	5720	**9736**	9736
	Productive Cover List Time (h)	**4134**	4155	**6226**	6236
	Cover List Utilization (%)	**72%**	72%	**63%**	64%
Average Case (100 simulations)	Financial cost	**88%**	88%	**106%**	106%
	Lost Service (h)	**1274**	1239	**312**	312
	Overtime (h)	**2524**	2536	**1149**	1146
	Scheduled Cover List (h)	**5811**	5808	**9771**	9771
	Productive Cover List Time (h)	**4429**	4449	**6639**	6641
	Cover List Utilization (%)	**76%**	76%	**67%**	67%
Worst Case	Financial cost	**89%**	88%	**107%**	107%
	Lost Service (h)	**1805**	1795	**651**	654
	Overtime (h)	**2638**	2608	**1306**	1320
	Scheduled Cover List (h)	**5504**	5512	**9312**	9312
	Productive Cover List Time (h)	**4120**	4168	**6327**	6310
	Cover List Utilization (%)	**74%**	75%	**67%**	67%

6.6.2 Split Covers

Because of the straight eight-hour cover list shift requirement, the time-of-day coverage patterns are mostly unimodal. However, time-of-day unexpected absence has a bi-modal distribution: in order to match the bi-modal distribution of ridership, most the scheduled runs are split runs. Therefore, whether having split covers will improve the scheduling performance as well as cover list utilization is investigated. Table 6-6 shows the performance comparison with split runs allowed and shows the coverage generated by the model. Slight improvements were observed in financial costs and lost service in all scenarios when split runs are allowed, with reduction in the number of cover list operators recommended, reduced lost service, and overtime. A slight improvement in cover list utilization was also observed in all cases. Although the improvements are not large, there are no tradeoffs being made and assignment plan is strictly better than the 8h straight assignment. Similar performance was achieved by the nominal model with the results shown in Appendix F.

The time-of-day distribution of coverage shown in Figure 6-16 better approximates the absence distributions than before (Figure 6-7 and Figure 6-11). The time-of-day distribution of coverage for the unconstrained case was similar to the constrained case and is shown in Appendix F.

Table 6-6 Evaluation of Split Cover List (Robust Adaptive Model)

Adaptive Robust ($\alpha = 0.7$)		**Constrained Resources**		**Unconstrained Resources**	
		8h straight	4h splits	**8h straight**	4h splits
Realized Outcome	Financial cost	**89%**	89%	**109%**	108%
	Lost Service (h)	**1433**	1377	**444**	404
	Overtime (h)	**2743**	2690	**1560**	1520
	Scheduled Cover List (h)	**5728**	5712	**9736**	9672
	Productive Cover List Time (h)	**4134**	4227	**6226**	6306
	Cover List Utilization (%)	**72%**	74%	**63%**	65%
Average Case (100 simulations)	Financial cost	**88%**	88%	**106%**	105%
	Lost Service (h)	**1274**	1251	**312**	290
	Overtime (h)	**2524**	2503	**1149**	1096
	Scheduled Cover List (h)	**5811**	5814	**9771**	9708
	Productive Cover List Time (h)	**4429**	4476	**6639**	6714
	Cover List Utilization (%)	**76%**	76%	**67%**	69%
Worst Case	Financial cost	**89%**	88%	**107%**	106%
	Lost Service (h)	**1805**	1816	**651**	637
	Overtime (h)	**2638**	2603	**1306**	1270
	Scheduled Cover List (h)	**5504**	5512	**9312**	9248
	Productive Cover List Time (h)	**4120**	4153	**6327**	6377
	Cover List Utilization (%)	**74%**	75%	**67%**	68%

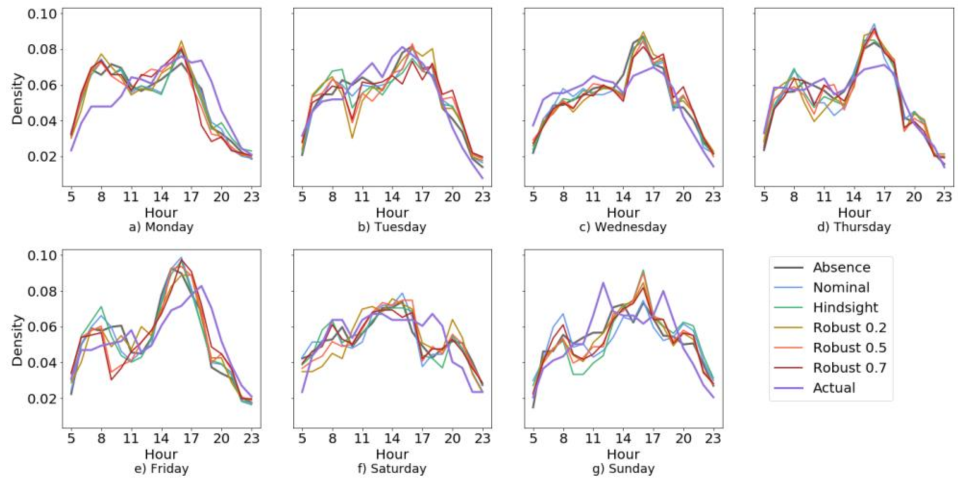


Figure 6-16 Evaluation of Split Cover List – Assignments (with constrained resources)

Although the time-of-day distribution of absences were approximated much better by the coverage curve, improvements in all performance metrics were small. However, the value of split runs lies not only in the anticipated cost savings, but also in the ability to react to situations better. For example, as the unexpected absences become known during the day, adjustments can be made to the report times of the second half of the assignment to match with the known absences.

In this chapter, three models were constructed for extraboard scheduling at the tactical and operational levels: a hindsight model that by definition is the best given the knowledge of the absence levels; a deterministic, nominal model that only takes the estimated absence and overtime availability profiles into account; and an adaptive robust integer program that includes input uncertainty in the formulation. The model was applied to the data from Southampton in 2019R4. The key takeaways from the case study are,

1) Right now, the current assignment plan is similar to the modelled plan, but the performance metrics are much worse than the outcomes anticipated in the models. This is due to the extraboard operators performing essential non-service duties that are not clearly identified in the data. At the same time, there should not be services lost or overtime performed at times extraboard operators are idle. In the dataset, there were services lost at times when there are unassigned extraboard operators, which could be an issue of dispatching, or record-keeping. More consistent and reliable record-keeping is needed to identify the cause and, to make the modelled results directly comparable to the current performance metrics.
2) When there are not enough extraboard operators, different assignment strategies and model parameters result in similar performance.

3) Having an adequately sized extraboard can help reduce overtime and lost service, at the expense of greater fixed costs associated with the additional extraboard operators. More robustness is **not** always better. As robustness increases, costs are added to cover less and less likely cases. Determining the optimal extraboard size and assignment strategy depends on the tradeoff between financial costs and service reliability (represented by the lost service).
4) It is important to estimate accurate input models. Deterministic models are overfitted to their input distribution. When the input distribution matches the evaluation distribution, deterministic models perform well. As the evaluation distribution deviates away from the input distribution, performances of deterministic models deteriorate quickly. Robust solutions can accommodate the specified level of uncertainty at the expense of extra fixed costs. To achieve decreased financial cost and lost service at the same time, the input models (in this case absence and overtime availability) need to be improved.
5) The robust models are less sensitive to model parameters, but they are quite sensitive to the level of robustness parameter.
6) Early-morning and late-night covers do not significantly affect performance.
7) Allowing split assignments better approximates the absence distributions than eight-hour straight runs. However, improvements in financial cost and lost service were small. This scenario shows that the time-of-day assignments of cover list operators only affect model performance slightly. On the other hand, allowing split covers has the potential to improve performance in ways not modelled in this thesis (e.g. by making the operational-level assignment more reactive).

Chapter 7 Conclusion and Future Work

This chapter summarizes the research, outlines the contributions, acknowledges the limitations, and suggests future research directions in extraboard operator planning.

7.1 Summary of Major Findings

In this section, the goal and objectives set out in Section 1.3 are revisited and the research outcomes discussed. The overall goal of this thesis was to develop a systematic procedure to improve the extraboard planning process for bus operators to reduce lost service, overtime and extraboard cost. A demand and supply interaction framework was formulated to address this problem. First, relevant demand and supply quantities other than the schedule of the extraboard operators are estimated. Then the interaction framework is formulated as an optimization model to solve for the optimal assignment plan that achieves the best weighted cost of lost service, overtime and the extraboard.

Breaking down the goal into smaller tasks/objectives, this thesis aimed to mine data from the newly introduced scheduling software HASTUS™ Daily, characterizing and modelling absence, overtime, lost service and their relations, as well as investigating how resulting models could be used in extraboard scheduling and evaluate the effectiveness of such scheduling practice.

First, pre-processing and descriptive analysis showed that the consistency and reliability of record-keeping could be improved. In particular, two areas are identified for further improvement: the consistency of original owner, and the reliability and detail of non-service duty records. Currently, since the original owner definition is unclear, assumptions have to be made during data preprocessing. Additionally, the non-service duties are not reliably documented. Different codes are used interchangeably, and the codes are not informative with respect to the necessity of these duties. Not being able to further classify non-essential duties creates comparison difficulties during scheduling model evaluations.

Second, absence and overtime availability are difficult to predict. The underlying mechanism cannot be captured by existing information. Negative binomial regression is used to estimate both the known-in-advance and unexpected absences. The historical quantities of absence, overtime, or lost service were not significant variables in the prediction of absences and the resulting model mainly accounts for the day-of-week and week in rating variations. The resulting error rates are in the 20%-40% range. Overtime availability is a more difficult quantity to model because the performed overtime is only a lower bound on available overtime. With the current information, no satisfactory models were found and the empirical distribution of records where performed overtime is equal to available overtime for each day of the week and hour of the day was taken as input to the scheduling model.

Lastly, robust optimization was used to schedule the extraboard operators. The models showed that at the current staffing levels, different assignment strategies yield similar results. Adjusting the number of extraboard operators can reduce lost service and overtime, but with added extraboard costs and therefore a higher financial cost. Robust models provide solutions that are more conservative. With a similar weighted objective function value accounting for both service reliability and financial cost, robust solutions tend to have more fixed costs from more extraboard operators. Because of the higher extraboard size, robust solutions result in more stable financial cost and lost service in different realizations of absence and are less sensitive to model parameters. The robust solutions are sensitive to the robustness levels defined. Robustness levels must be chosen carefully: higher robustness implies higher financial costs to cover increasingly rare scenarios. Joint reductions in both lost service and financial cost can only be achieved when the input models are improved. Therefore, tradeoff between financial costs and reliability must be considered while determining the extraboard size. In this case, the model with a moderate amount of robustness α=0.7 was used to evaluate two policy scenarios. Removing the requirement of early morning and late night covers achieved some improvement in the constrained case but when there was no constraint on the number of operators, more than 90% of the time there was a cover reserved for the first and last shift regardless. Therefore, there was no improvement upon relaxing the requirement. Allowing split covers significantly improved the model's ability to match the coverage curve to the time-of-day absence distribution, however, the improvements in performance metrics were limited.

With the major findings summarized above, the contributions made by this thesis are

1) Quantification of absence and overtime availability levels.

The vast majority of absence and overtime availability literature focuses on the motivation and psychological decision-making aspects of operator behavior. While these are central in developing management strategies to reduce absence, equally important is acknowledging that operator absence will not be totally eliminated and preparing the garages to cover operator absences more efficiently. This thesis made an attempt at quantifying absence and overtime availability for extraboard scheduling purposes, and gave insight into the factors that influence absence and overtime availability at an aggregate level.

2) Application of robust optimization in extraboard planning.

Having an extraboard is itself a way to build robustness into the overall scheduling system. Although optimization has been used widely in literature, and some prior work has demonstrated the importance of using extraboard operators to improve the system's robustness, previous literature has not looked at how to optimize the extraboard assignment to maximize the robustness impacts of the extraboard . This thesis places this problem in a robustness optimization framework and demonstrates the benefits of optimizing the extraboard scheduling process given uncertainty in absence and overtime availability.

3) An end-to-end framework for transit extraboard scheduling.

In previous work on scheduling, the demand side estimation and supply side optimization were not integrated. This thesis took a holistic approach and developed a procedure that takes the raw data from the scheduling software to an optimized extraboard assignment.

7.1 Limitations

The approaches developed here, for absence modelling and extraboard scheduling, have several limitations that must be acknowledged.

1) No operator level demographic information was used in the absence and overtime availability models. Although it was demonstrated in the literature that demographic information such as gender, age, marital status, and income can have significant impacts on absence rates and willingness to accept overtime, due to privacy considerations, this data is not released by the MBTA and therefore significantly limited the explanatory powers of the absence and overtime availability models.
2) An hourly workload curve was assumed for the extraboard scheduling model. Although this is a common practice in the workforce scheduling literature, using an hourly workload curve instead of individual pieces of work makes the model less detailed regarding the compatibility of pieces of work. This loss was approximated by setting aside a random hour of unavailable time for each operator. Although this loss of detail will not impact the overall scheduling recommendations, comparisons between the modelled performance metrics and the observed performance metrics will be inaccurate.
3) Essential non-service duties are not modelled. Essential non-service duties are an integral part of the extraboard's work. Although they are performed at (or from) the garage, there are no accurate records of them in the dataset. Assuming that all service duties take precedence will make the modelled performance metrics overly optimistic comparing the observed performance metrics.
4) The garage dispatcher makes many real-time, on-the-fly decisions that are not necessarily reflected in the optimization model. For example, overtime may be requested before unexpected absences are realized in anticipation of operator shortage. Although half-day absences end up being covered by the cover list, regardless of the time the garage is notified, having this information can aid in scheduling report times.

7.2 Future Research

The above limitations are mostly in terms of data quantity and quality. The models developed in this research also allude to other modelling and policy scenarios that are beyond the scope of this research but are of value in the bigger picture of better extraboard performance. Several attractive directions for future research are:

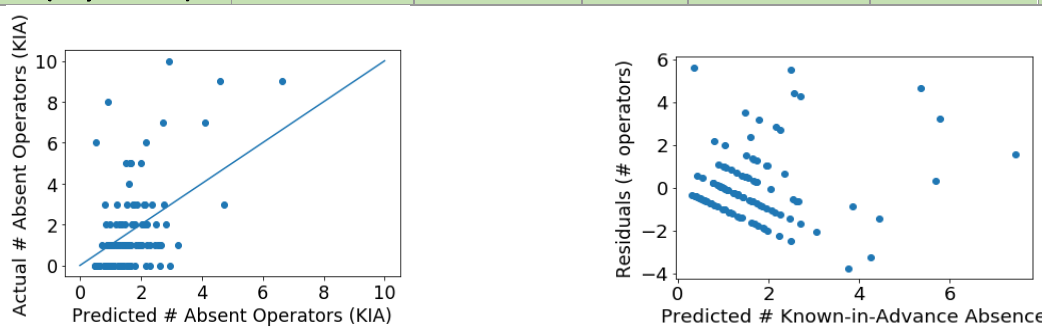
1) Incorporate the dispatch strategy on the day-of: This thesis develops an end-to-end framework from data to planning. A final step of operations (dispatching) is missing. Whether to call for overtime, divert a regular operator, or use the extraboard when open work arises is another area that is worth investigating and might affect overall extraboard effectiveness.
2) With data from different garages, sharing extraboard operators between garages may be investigated. Allowing sharing between the garages will introduce more resources to use by the garages to cover extreme conditions. For example, if an incident happens, idle operators from other garages could be dispatched to help. Additionally, operator sharing among different garages could allow the agency to exploit different travel patterns at different garages during scheduling. For example, pairing garages that experience peak demands at different times of day could even out the demand peaks. In order to achieve sharing between the garages, travel time between garage, compatibility between pieces of work, operator training across routes in different garages, as well as real-time information sharing, will then become important considerations.
3) Besides 4hour shifts, the cost and benefits of more variable shift lengths for the cover list may be investigated. This will give more flexibility to the garage superintendents to adjust the coverage. At the same time, more variable hours will likely create inconvenience and stress for the cover list operators. Therefore, the cost and benefit of different combination of shift lengths and the break time between the shifts need to be quantified.
4) Different costs for lost service at different times of day and on different days can be considered. Currently, the Southampton garage experiences higher rates of lost service on weekends. The models suggest that currently larger shortages of extraboard operators are more likely to occur on weekends. This recommendation is made on the assumption that lost services are of equal disutility at all times of the week. As mentioned in Section 4.1, the garage superintendent does sometimes take operators off their own assignments to cover work that are deemed more important. Therefore, under some resource constraint, the issue of which services are more important to be covered is worth investigating.

Appendix A Full Results of Known-in-Advance Absence Models

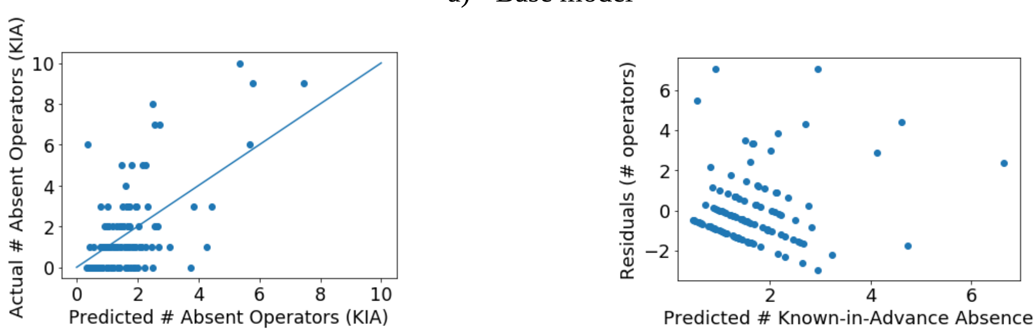
2017 Rating 4 Southampton Garage

Table A-1 Regression Coefficients for Known-in-Advance Absences 2017 R4 @ Southampton

	Base Model			Full Model		
	Coefficient (β_j)	Standard Error	p-value	Coefficient (β_j)	Standard Error	p-value
1/α	0.35			0.21		
constant	-4.8504 ***	0.419	0	-5.2323 ***	0.704	0
holiday	0.1147	0.619	0.853	-0.1544	0.566	0.785
holiday extension	0.9991 *	0.407	0.014	0.5496	0.465	0.237
week	0.0507 *	0.025	0.043	0.0196	0.028	0.487
Tue	-0.0684	0.486	0.888	-0.1639	0.47	0.727
Wed	0.488	0.451	0.279	0.4722	0.422	0.263
Thu	0.5521	0.438	0.207	0.4112	0.407	0.312
Fri	0.1848	0.453	0.684	0.1319	0.418	0.752
Sat	-0.3377	0.535	0.528	-0.6101	0.534	0.253
Sun	0.3495	0.499	0.483	0.4013	0.463	0.386
rolling overtime previous week				0.0183	0.016	0.253
unexpected absence (day of)				-0.0204	0.043	0.636
Known-in-advance Absence (day before)				0.1215 ***	0.036	0.001



a) Base model



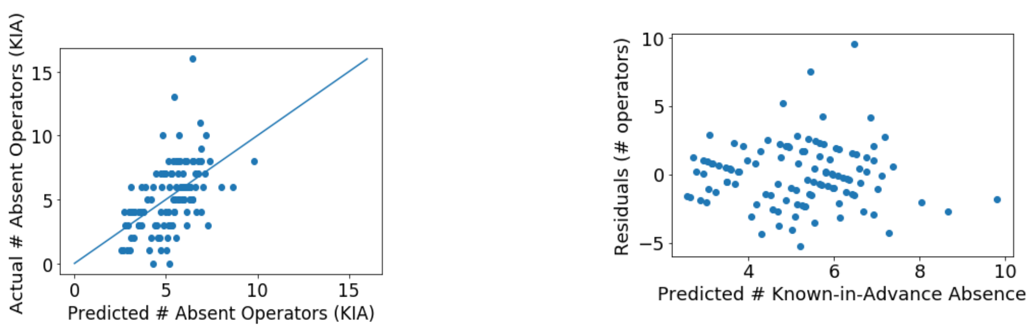
b) Full model

Figure A-1 Negative Binomial Regression Results for Known-in-Advance Absences 2017 R4 @ Southampton

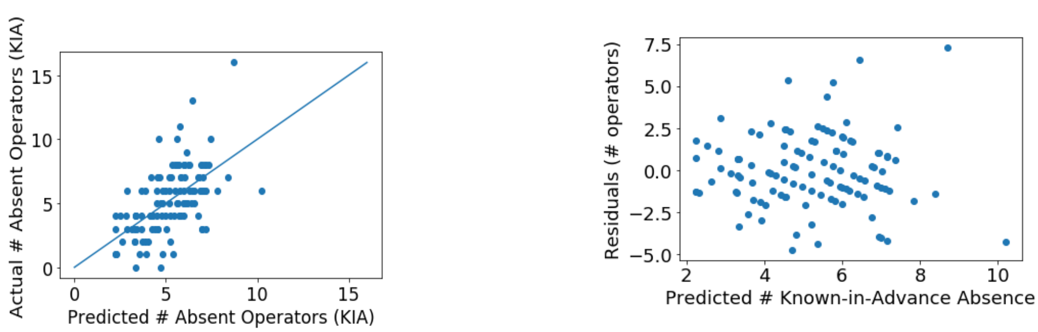
2018 Rating 4 Southampton Garage

Table A-2 Regression Coefficients for Known-in-Advance Absences 2018 R4 @ Southampton

	Base Model			Full Model		
	Coefficient (β_j)	Standard Error	p-value	Coefficient (β_j)	Standard Error	p-value
1/α	0			0		
constant	-2.8162 ***	0.122	0	-2.9416	0.241 ***	0
holiday	0.1813	0.199	0.363	-0.0437	0.214	0.838
holiday extension	0.3283 **	0.135	0.015	0.2185	0.144	0.129
week	-0.0194 **	0.008	0.017	-0.0181 *	0.009	0.05
Tue	0.0509	0.137	0.71	-0.008	0.137	0.953
Wed	-0.1436	0.146	0.324	-0.292	0.152	0.055
Thu	-0.0366	0.143	0.797	-0.0972	0.143	0.498
Fri	0.0631	0.138	0.647	-0.0128	0.14	0.927
Sat	-0.218	0.161	0.175	-0.3198	0.164	0.051
Sun	-0.268	0.165	0.103	-0.2797	0.165	0.089
rolling overtime previous week				0.0032	0.005	0.53
unexpected absence (day of)				-0.0235	0.018	0.186
Known-in-advance Absence (day before)				0.0529 **	0.018	0.004



a) Base model



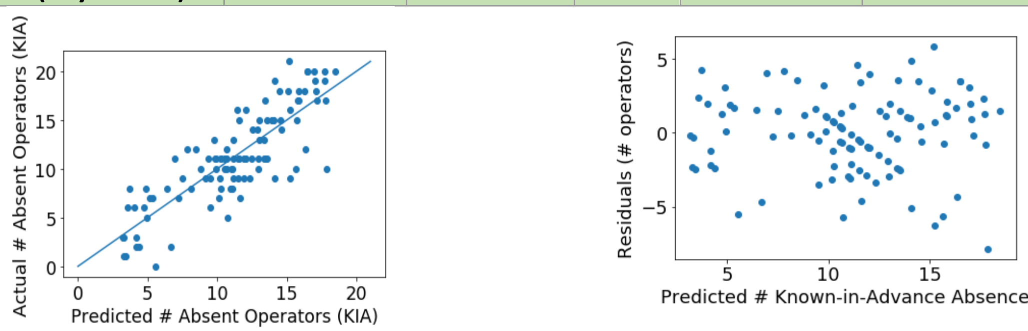
b) Full model

Figure A-2 Negative Binomial Regression Results for Known-in-Advance Absences 2018 R4 @ Southampton

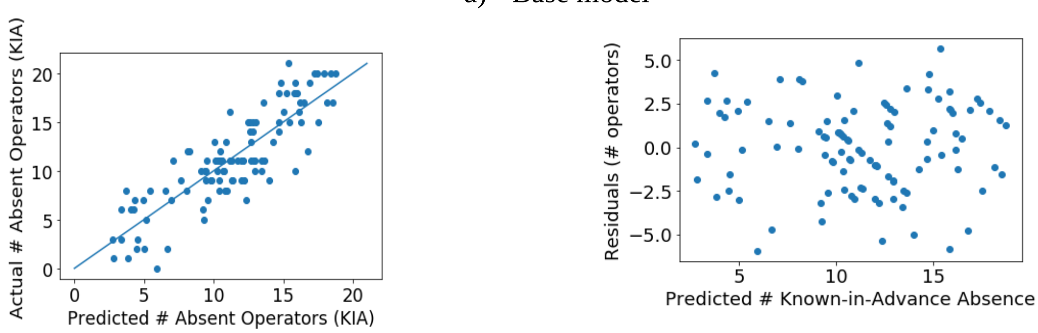
2019 Rating 4 Charlestown Garage

Table A-3 Regression Coefficients for Known-in-Advance Absences 2019 R4 @ Charlestown

	Base Model			Full Model		
	Coefficient (β_j)	Standard Error	p-value	Coefficient (β_j)	Standard Error	p-value
1/α	0			0		
constant	-3.0632 ***	0.094	0	-3.2354 ***	0.134	0
holiday	-0.2952	0.207	0.154	-0.4148 *	0.21	0.049
holiday extension	0.0748	0.124	0.545	0.1284	0.13	0.325
week	0.0406 ***	0.006	0	0.029 **	0.011	0.01
Tue	-0.0275	0.1	0.784	-0.0186	0.101	0.853
Wed	0.0156	0.099	0.875	0.0231	0.101	0.819
Thu	-0.0728	0.101	0.472	-0.066	0.101	0.515
Fri	-0.1157	0.101	0.25	-0.0856	0.102	0.399
Sat	0.2099	0.112	0.061	0.2619 *	0.132	0.048
Sun	0.048	0.143	0.737	-0.0127	0.147	0.931
rolling overtime previous week				0.0004	0.003	0.88
unexpected absence (day of)				-0.0035	0.012	0.776
Known-in-advance Absence (day before)				0.0399 **	0.013	0.002



a) Base model

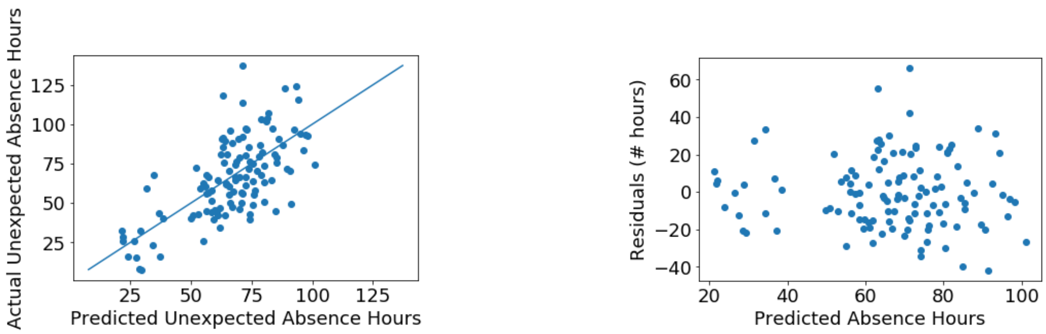


b) Full model

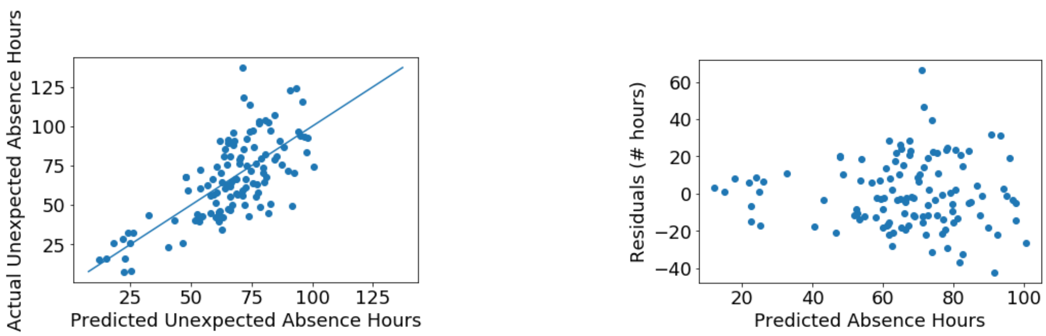
Figure A-3 Negative Binomial Regression Results for Known-in-Advance Absences 2017 R4 @ Charlestown

Appendix B Full Results of Unexpected Absence Models

2017 Rating 4 Southampton Garage



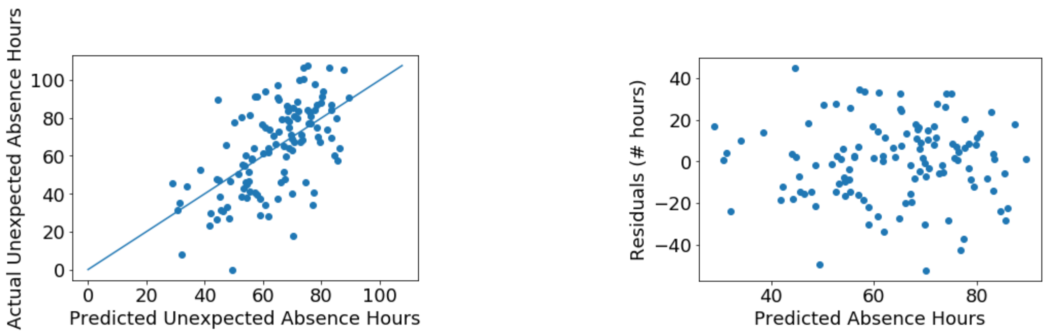
a) Base Model



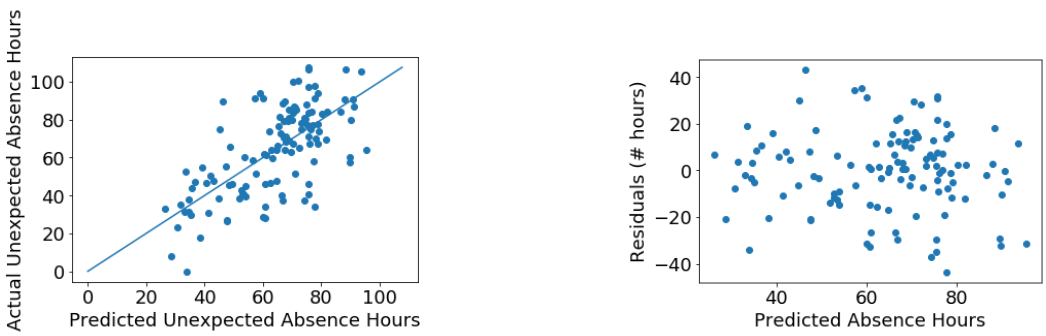
b) Full Model

Figure B-1 Negative Binomial Regression for Unexpected Absence Hours 2017 R4 @ Southampton

2018 Rating 4 Southampton Garage



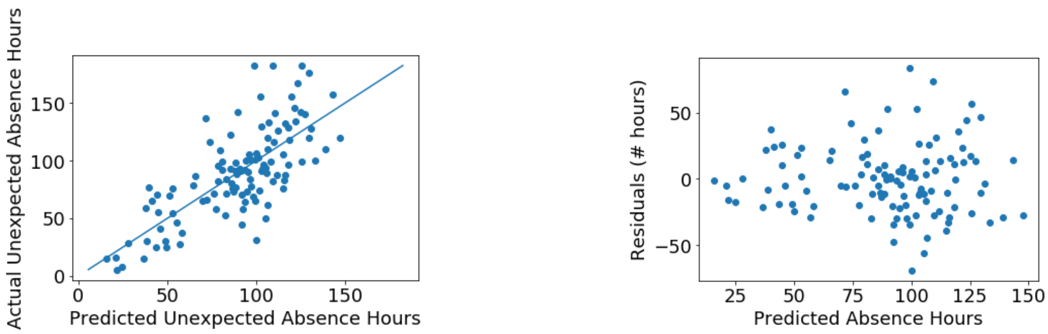
a) Base Model



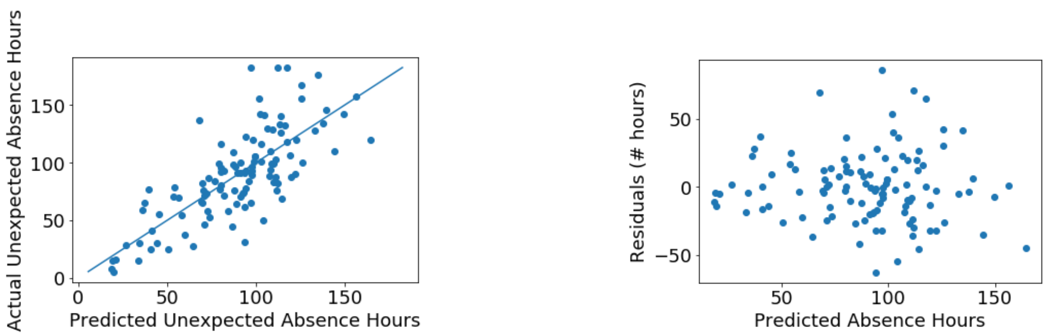
b) Full Model

Figure B-2 Negative Binomial Regression for Unexpected Absence Hours 2018 R4 @ Southampton

2019 Rating 4 Charlestown Garage



a) Base Model



b) Full Model

Figure B-3 Negative Binomial Regression for Unexpected Absence Hours 2019 R4 @ Charlestown

Appendix C Full Results of Multivariate Unexpected Absence Models

2017 Rating 4 Southampton Garage

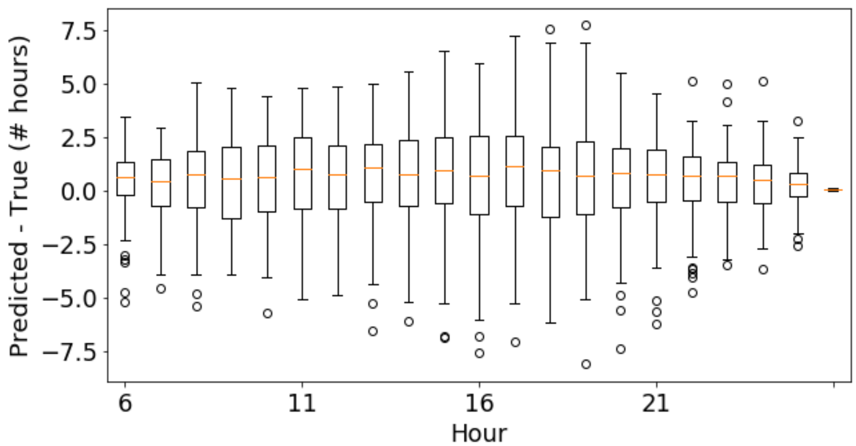


Figure C-1 MVPLN Error 2017 R4 @ Southampton

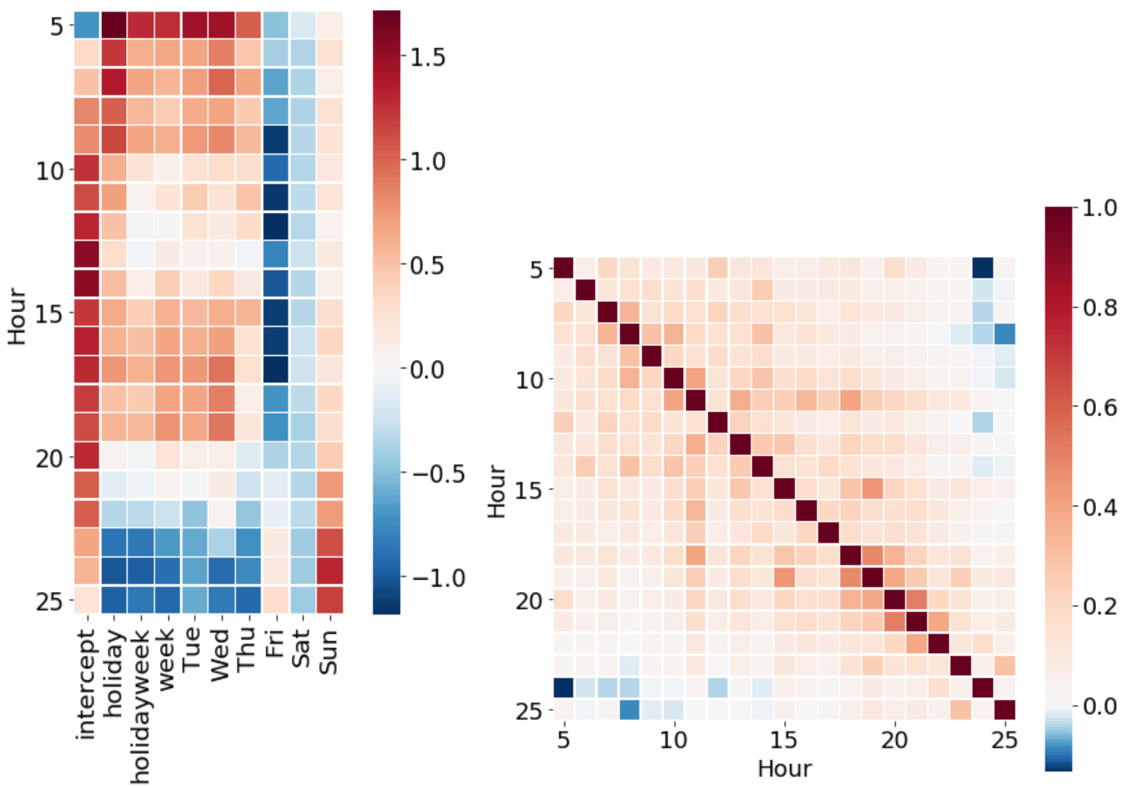


Figure C-2 MVPLN Beta Values (left) and Correlation Matrix (right) 2017 R4 @ Southampton

2018 Rating 4 Southampton Garage

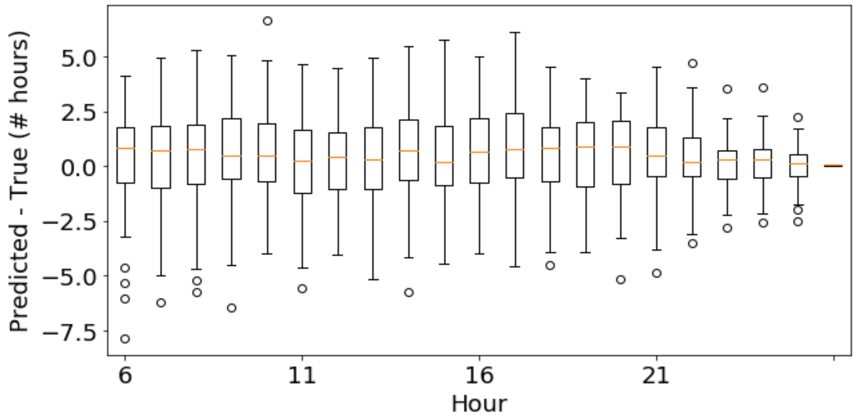


Figure C-3 MVPLN Error 2018 R4 @ Southampton

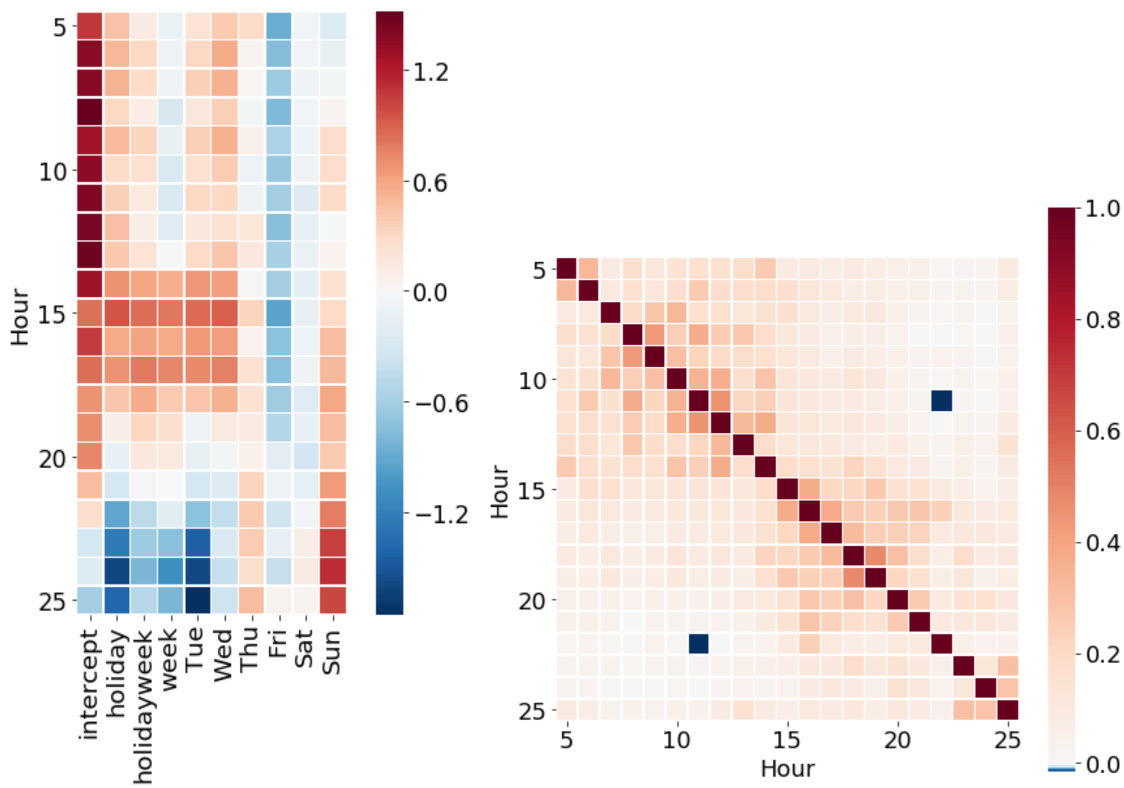


Figure C-4 MVPLN Beta Values (left) and Correlation Matrix (right) 2018 R4 @ Southampton

2019 Rating 4 Charlestown Garage

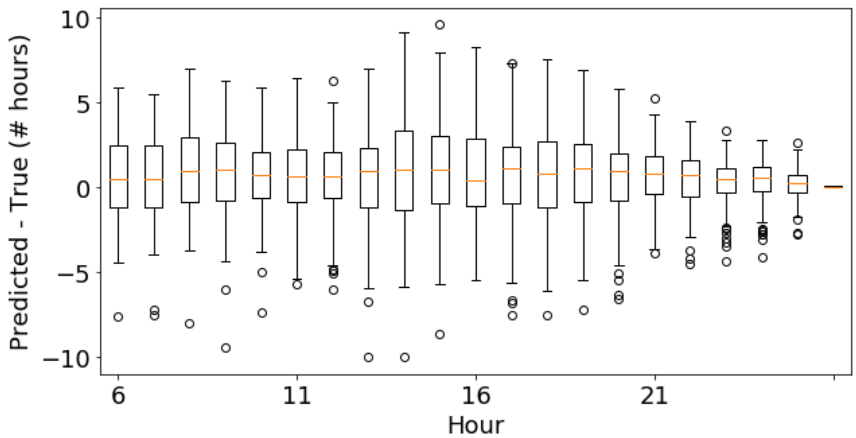


Figure C-5 MVPLN Error 2019 R4 @ Charlestown

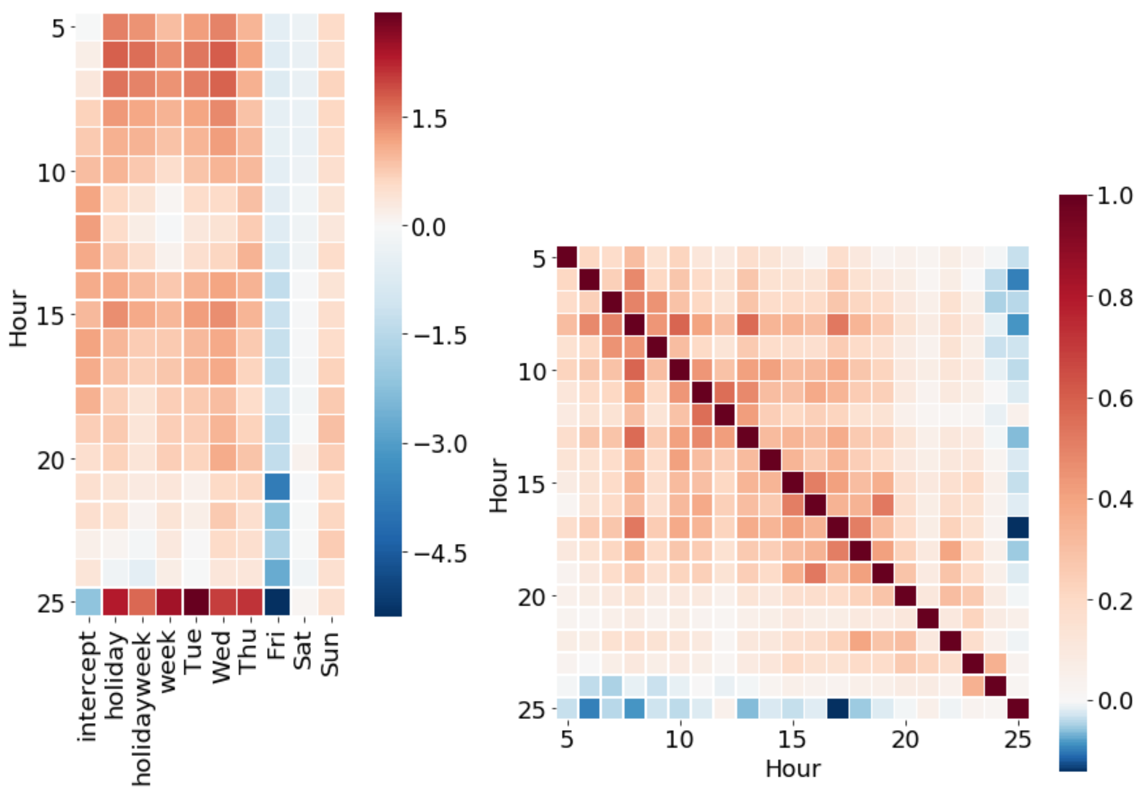


Figure C-6 MVPLN Beta Values (left) and Correlation Matrix (right) 2019 R4 @ Charlestown

Appendix D Modelling Approaches for Estimating Overtime Availability

Two model formulations were tested for overtime availability models. The formulation, variables tested, the results, and why they are not used in the scheduling model are briefly explained in this section.

Censored Poisson Regression

Since we are interested in the amount of overtime available only for the purpose of extraboard scheduling, not characterizing the decision-making process, only aggregate models are considered. Due to the discrete nature of the number of operators performing overtime, count models are used. Since overtime available is different from overtime performed, the time-of-day profile for overtime availability cannot be directly taken from the observed hourly overtime. The two-step approach for absence is not applicable, and overtime availability at different hours is modelled directly. The number of operators available for overtime depends both on the number of potential operators available for overtime, and their willingness to accept overtime. Define y_{it}, λ_{it}, $and\ n_{it}$ to be the number of available operators for overtime, the proportion of potential operators willing to perform overtime, and the number of potential operators for overtime, respectively, for day i and hour t. Then their relationship is $y_{it} = \lambda_{it} \times n_{it}$.

The main difference between absence and overtime availability modelling is that overtime observations are performed overtime which is a function of overtime requested and overtime available. The quantity of interest, available overtime y_{it} , is unobserved. For example, if the number of cover list operators is sufficient to cover all open work, then overtime observed will be 0, but this does not mean that there is no overtime available. Operators willing to work overtime are unobserved if no overtime is requested. To account for the unobserved available overtime, a censored model is used.

The common form of truncation is zero (left)-censored models where the value of 0 cannot be observed. For example, if we are modelling the number of hospitalization days, only hospitalized individuals will exist in the dataset and 0 days cannot occur and we left-truncate the model. In this case of overtime availability, the observations are right-censored. This form of truncation is often observed in surveys when the "x or more" option is included. The same principles for left truncation can be applied to right truncation problems. Let Y represent a discrete random variable (overtime availability) with probability mass function f_Y. Since realizations greater than r (number of requests for overtime) are omitted, r will be observed for every value of y that is larger than r, and the likelihood of observing r is the sum of the probabilities of Y taking values larger than or equal to r. The right-censored distribution representing the probability $P(Y = y)$ is given by

$$P(Y = y) = \begin{cases} \sum_{a=r}^{\infty} f_Y(Y = a) & if\ y = r \\ f_Y(Y = y) & if\ y \leq r \end{cases}$$

There are two types of overtime availability: overtime from operators who extend their workday and operators who come in for overtime on their days-off. Since the availability pool for the two types of overtime are separate, and the rate at which they can be obtained are likely different, they are estimated by two separate models and then summed to get the overall availability profile, to ensure the representation of the amount of overtime from both sources is accurate. Superscripts "on" and "off" are used in the formulation below to denote the two types of overtime availability. Since requests are made to both pools of operators together, the number of requests r_{it} and indicator variable d_{it} are shared between the two models. Before the mathematical formulation is presented, the rest of the regression parameters and inputs are defined below (for the "on" case).

$y *_{it}^{on}$ = number of operators (on their workday) performing overtime on day i at hour t

μ_{it}= average number of operators available for overtime on day i at hour t

X_{it} = independent variables for day i and hour t

β^{on} = regression coefficients

N = number of observations

K = number of independent variables

r_{it}= number of overtime requests at day i hour t (lost service + overtime performed)

d_{it}= indicator whether the observation is the actual value (1) or a lower bound (0)

It is assumed that y_{it}^{on} has a Poisson distribution where the mean value is a function of the explanatory variables X_{it}.

$$P(y_{it}^{on}) \sim Poisson(\mu_{it}^{on})$$

$$E(y_{it}^{on}) = \mu_{it}^{on} = n_{it}^{on} \lambda_{it}^{on} = n_{it}^{on} exp(X_{it}^{T} \beta^{on})$$

$$var(y_{it}^{on}) = \mu_{it}^{on}$$

$$\begin{cases} d_{it} = 0 & if y *_{it}^{on} + y *_{it}^{off} = r_{it} \\ d_{it} = 1 & if\ y *_{it}^{on} + y *_{it}^{off} < r_{it} \end{cases}$$

The likelihood function can be written as,

$$L^{on} = \prod_{i=1}^{N}\prod_{t=1}^{T} f_p(y_{it}^{on} = y*_{it}^{on} \mid n_{it}^{on}, X_{it}, \beta^{on})^{d_{it}} f_p(y_{it}^{on} \geq y*_{it}^{on} \mid n_{it}^{on}, X_{it}, \beta^{on})^{1-d_{it}}$$

where f_p is the Poisson distribution.

The log likelihood is

$$l^{on}(\beta) = \sum_{i=1}^{N}\sum_{t=1}^{T}\Big[d_{it} log\left(f_p(y_{it}^{\mathrm{on}} = y*_{it}^{\mathrm{on}} \mid \mathrm{n}_{\mathrm{it}}^{\mathrm{on}}, X_{it}, \beta^{\mathrm{on}})\right) + (1-d_{it}) log\left(f_p(y_{it}^{\mathrm{on}} \geq y*_{it}^{\mathrm{on}} \mid \mathrm{n}_{\mathrm{it}}^{\mathrm{on}}, X_{it}, \beta^{\mathrm{on}})\right)\Big]$$

Similarly, if we substitute the superscript with "*off*", then we get the same model with parameters β^{off} to be estimated and the log likelihood $l^{off}(\beta)$ for the overtime availability among operators who have the day off. The overall log likelihood of the model is

$$l(\beta) = l^{on}(\beta) + l^{off}(\beta)$$

Since the log likelihood function contains the cumulative distribution function of the Poisson distribution, analytical solutions are not available. The Nelder-Mead algorithm (Gao and Han, 2012) in scipy.minimize (Hill, 2016) was used to minimize the log likelihood. Variables are selected based on a priori expectations aided by the likelihood tests.

Binomial Regression

Filtered observations to the ones where requested overtime > performed overtime, that is, at hours when observed overtime = available overtime, such that we can directly see the fit of the model and the explanatory power of the exogeneous variables.

On day d at hour t:

p_{dt} = probability of potential operators accept overtime (assuming homogeneous among operators)

n_{dt} = number of potential operators doing overtime

y_{dt} = observed number of people performing overtime

$$y_{dt} \sim Binom(n_{dt}, p_{dt})$$

Separate models are estimated for available overtime from operators working and operators on their days-off.

Variables Considered

Besides temporal characteristics, daily and hourly overtime performed/available rates and lost service rates were tested to see if they help with the fitting. The rationale for including overtime rates is that overtime availability may be autoregressive and past overtime rates might be indicative of current overtime rates. Past lost service rates are also indicative of overtime availability since with the extraboard resources approximately the same, more lost service is indicative of less overtime availability.

Hourly vs. Daily: Since the rate is an hourly rate, overtime and lost service rates for the same hour is indicative of the levels we could anticipate for the same hour since overtime availability for different times of the day is different. However, daily rates were tested as well since hourly rates may not be stable and are subject to outliers that may affect model results.

Overtime availability rate vs overtime performed rate: The quantity being modelled is hourly availability rates. Therefore, the referenced hourly rates are hourly. However, the same definition does not work when the availability is aggregated to a daily resolution. An operator cannot be available for all hours of the day, and therefore on the daily level, overtime performed rate was quoted. Although not the same definition, higher overtime availability directly leads to overtime performed. The modelled relationship is not affected.

Same/different day-category: The day-category refers to weekdays and weekends. The availability rates for weekends and weekdays should be different. Therefore, two variables are constructed for the cases where the referenced day (1 or 2 days ago) is the same with or different from the day of interest, such that the coefficients for the two variables represent the effect of past observed values when the dates was of the same or different categories, respectively. There are three separate cases: first, the target day and the referenced day are both weekdays or weekends; second, the target day is on the weekend but the referenced day is a week day; third, the target day is a weekday but the referenced day is on the weekend.

In models of both y_{dt}^{on} and y_{dt}^{off}, the variables tested are shown in Table D-1.

Table D-1 Variables Tested for both Working and Off Operators

Variable	Resolution	Definition
Temporal Characteristics		
Day of week dummy	daily	To account for day-of-week variation
Week number	daily	To account for within-rating trends
(Weekday/Weekend) x Time period dummy	by defined time period	To model the time-of-day distribution of overtime. The distribution should be different for days with weekday schedules and weekend schedules. Two types of time periods were tested: morning/midday/afternoon/evening vs. hourly
Additional Variables:		
average hourly lost service rate in the past week	hourly	$\overline{ls_{dt}}$ = lost / scheduled
average daily lost service rate in the past week	daily	$\overline{ls_{d}}$ = lost / scheduled
lost service rate 2 days ago during the same hour $ls_{(d-2)t}$	hourly	same day category
		target is weekend; reference is weekday
		target is weekday; reference is weekend
daily lost service rate 2 days ago $ls_{(d-2)}$	daily	same day category
		target is weekend; reference is weekday
		target is weekday; reference is weekend
average hourly overtime availability rate in the past week	hourly	$\overline{p_{dt}}$ = overtime performed / potential operators
average daily overtime availability rate in the past week	daily	
overtime availability rate 2 days ago during the same hour ($p_{(d-2)t}$ = overtime performed / potential operators)	hourly	same day category
		target is weekend; reference is weekday
		target is weekday; reference is weekend
overtime performed rate 2 days ago (p_{d-2} = daily overtime performed / scheduled service)	daily	same day category
		target is weekend; reference is weekday
		target is weekday; reference is weekend
Analysis Terms:		
lost service rate 1 day ago during the same hour $ls_{(d-1)t}$	hourly	same day category
		target is weekend; reference is weekday
		target is weekday; reference is weekend
daily lost service rate 1 day ago $ls_{(d-1)}$	daily	same day category
		target is weekend; reference is weekday
		target is weekday; reference is weekend
overtime availability rate 1 days ago during the same hour ($p_{(d-1)t}$ = overtime performed / potential operators)	hourly	same day category
		target is weekend; reference is weekday
		target is weekday; reference is weekend
overtime performed rate 1 days ago (p_{d-1} = daily overtime performed / scheduled service)	daily	same day category
		target is weekend; reference is weekday
		target is weekday; reference is weekend

Estimation Results:

Variables: the daily variables are not significant and often have the wrong signs. They are then excluded. The additional variables do not have a common, interpretable, and significant pattern besides the first variable (only average hourly overtime availability rate in the past week is consistently positive and significant).

Censored Poisson: Due to the survival function in the likelihood formulation, the minimization of negative log likelihood is difficult and likely to run into numerical issues. The resulting coefficients were not stable, and the signs were often opposite to prior beliefs. Additionally, the profiles that were estimated were difficult to comprehend.

Since in this formulation, no ground truth exists, in Figure D-1 the average estimated overtime availability (black) and observed overtime (red) are plotted for rating 4 at Southampton garage. The rates in the plots are defined as the estimated (black) / observed (red) operators / all available operators and weekdays and weekends are estimated separately. In the plots the time-of-day dummy was taken to be 4 periods: morning (before 10am), midday (10am-4pm), afternoon (4pm – 9pm), evening (after 9pm). Theoretically, the estimated availability should be above the observed overtime and it is true in most cases while there are some patterns estimated that are difficult to interpret. However, since there is no ground truth to compare the model results to. It is unknown whether the problem lies in numerical estimation, the noise in the data, or the sample size being too small.

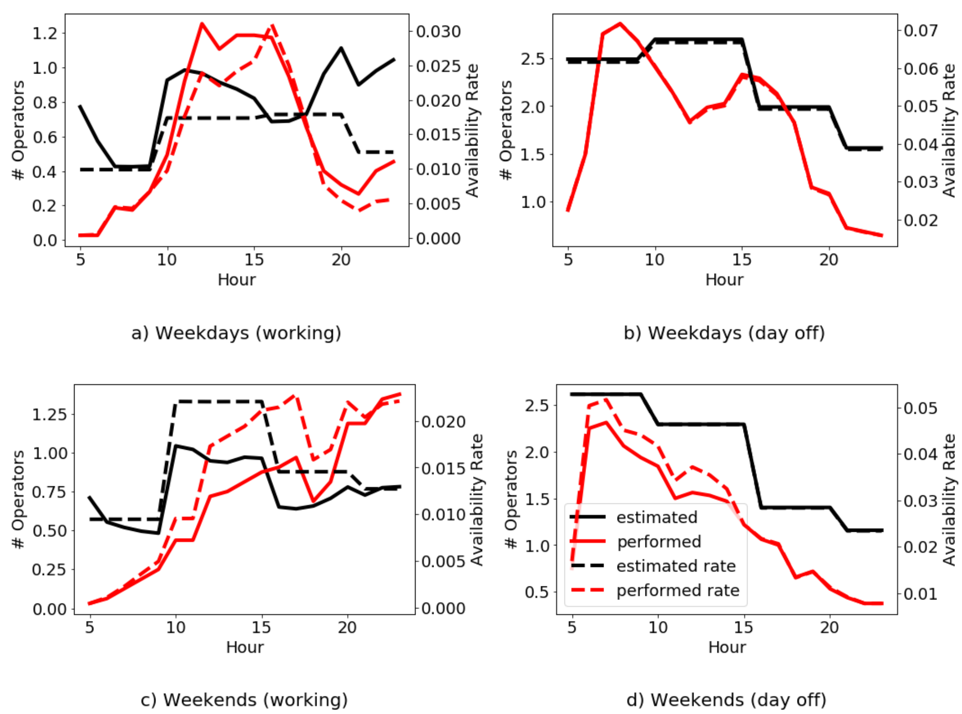


Figure D-1 Overtime Availability Censored Poisson Formulation 2019R4 @ Southampton

Binomial: To verify the data quality, the binomial formulation was tested. The data is filtered to the records where observed is equal to available, therefore the tails of the distribution should not affect the results estimation. However, the model fit only slightly better than estimating the average. Some years/garages are better than others and the plots are shown below. Therefore, knowing that with the existing explanatory variables, we are not capable of explaining the variance in overtime availability. Therefore, in the scheduling model, empirical distributions were used instead of modelled results. Models for overtime availability is left for future research.

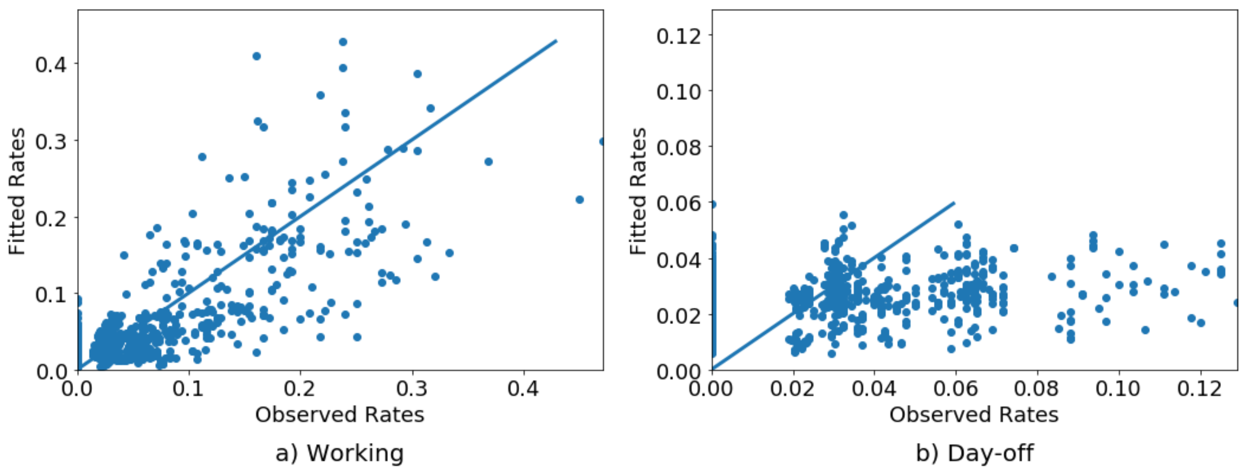


Figure D-2 Overtime Availability Binomial Formulation Fit 2017R4 @ Southampton

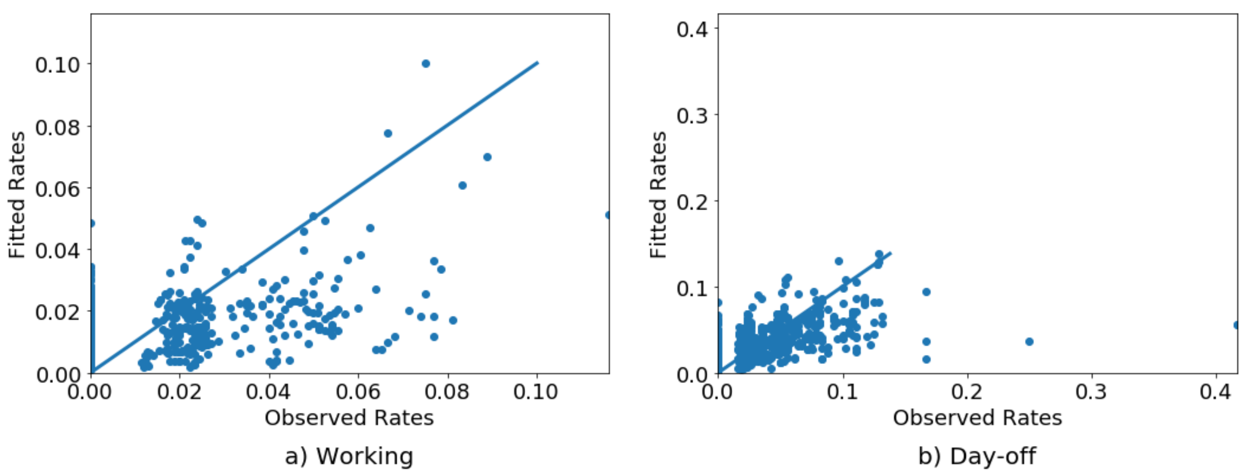


Figure D-3 Overtime Availability Binomial Formulation Fit 2018R4 @ Southampton

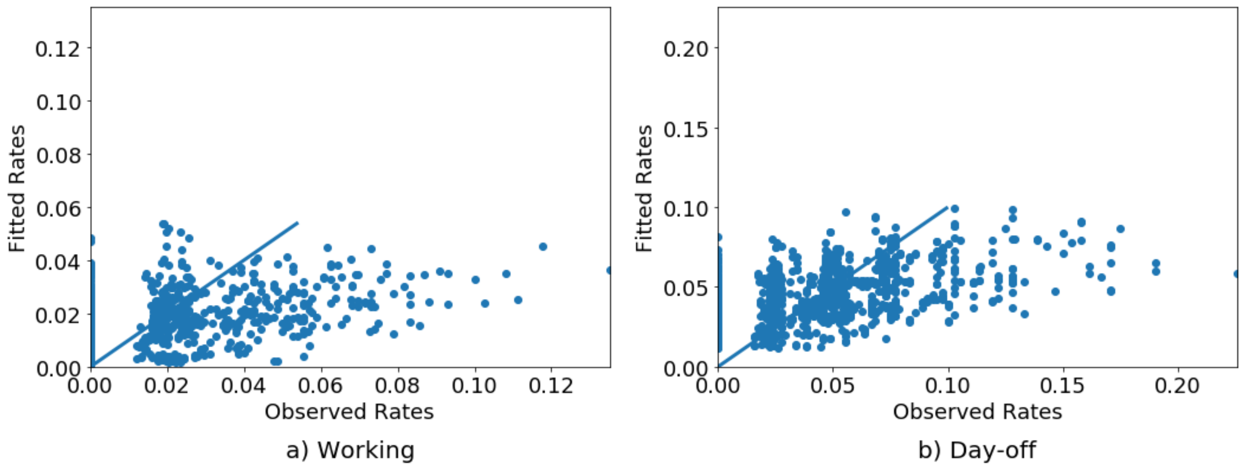


Figure D-4 Overtime Availability Binomial Formulation Fit 2019R4 @ Southampton

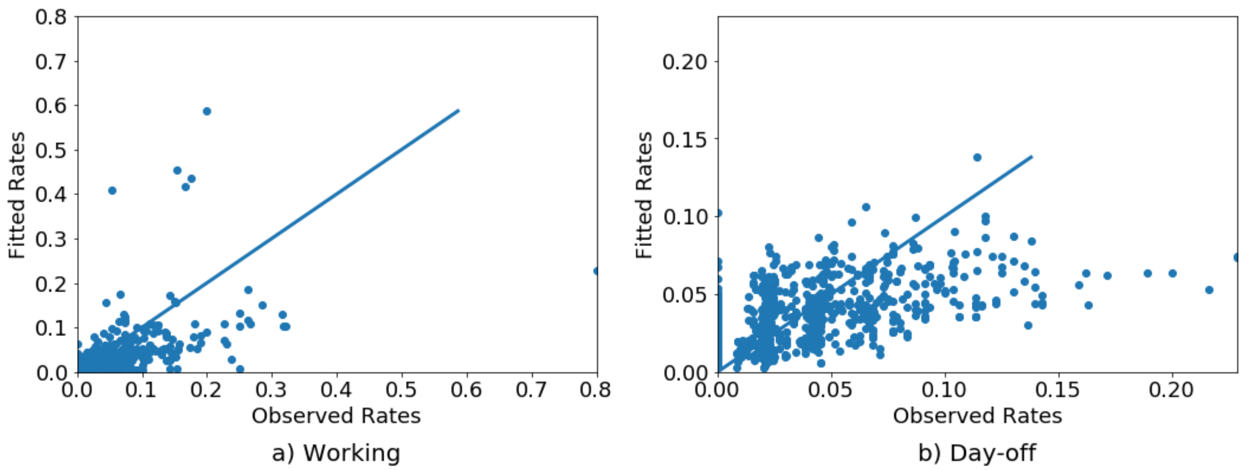
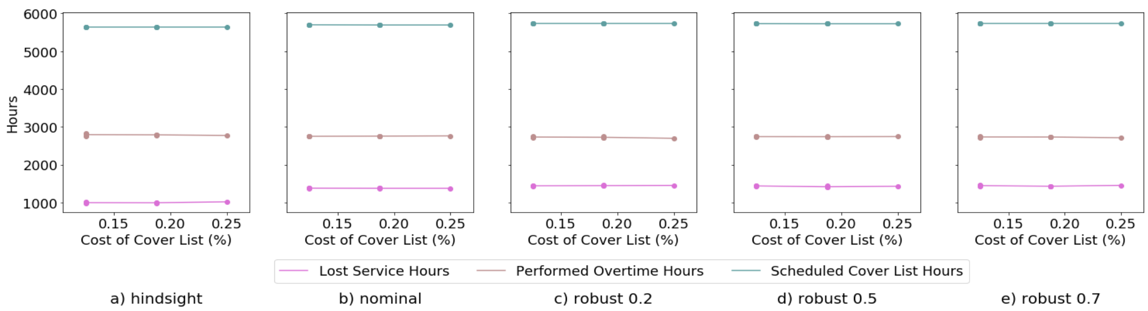


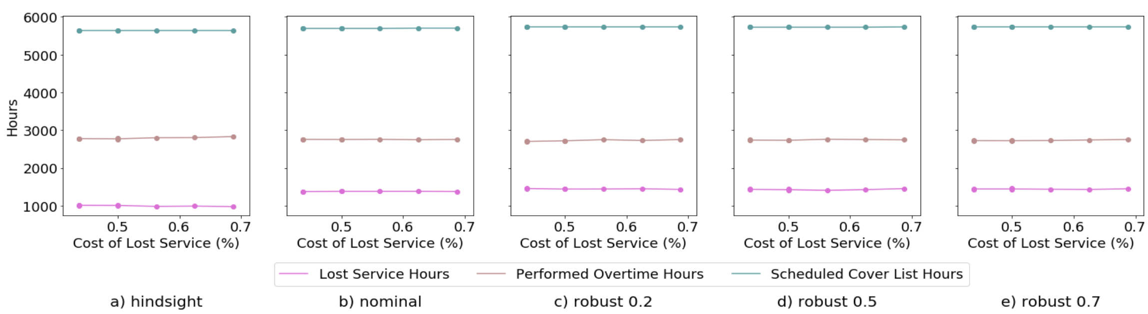
Figure D-5 Overtime Availability Binomial Formulation Fit 2019R4 @ Charlestown

Appendix E Sensitivity Analysis of Cost Coefficients in the Realized Outcome and Worst-case Scenarios

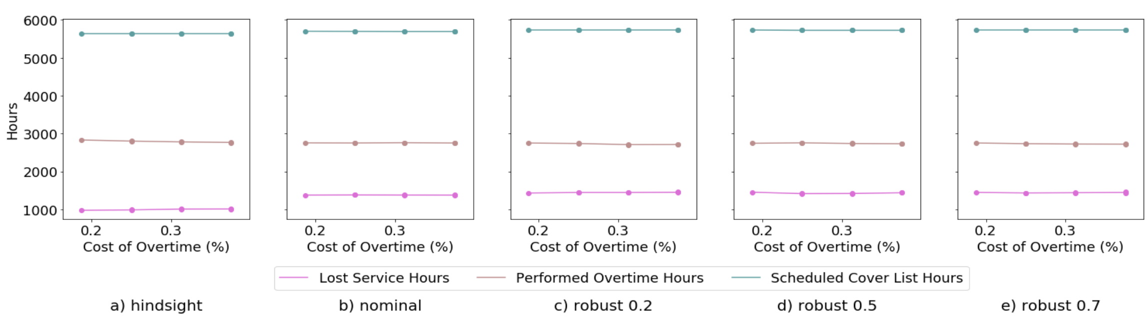
Realized Outcome (Constrained Resources):



i) w.r.t cost of cover list hours



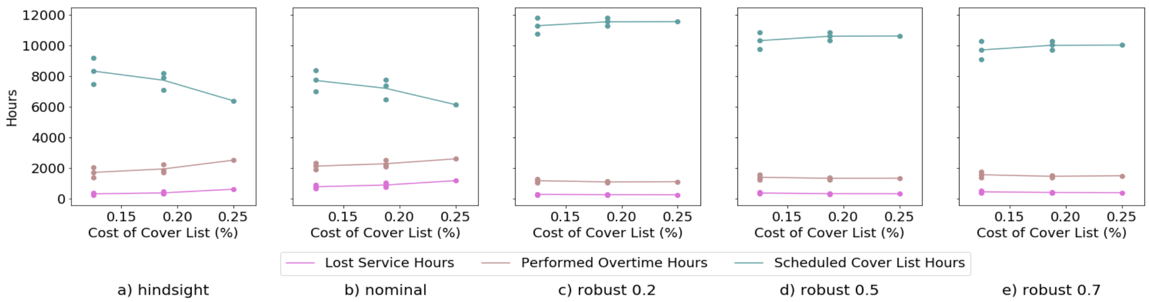
ii) w.r.t cost of overtime hours



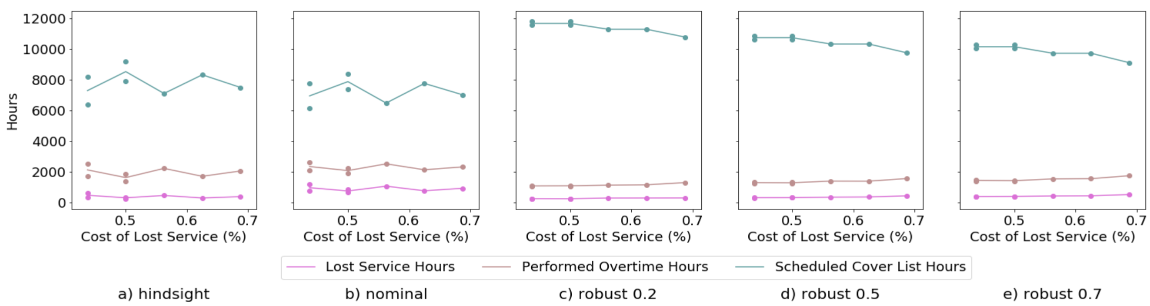
iii) w.r.t. cost of lost service hours

Figure E-1 Sensitivity Analysis of Realized Outcome (constrained)

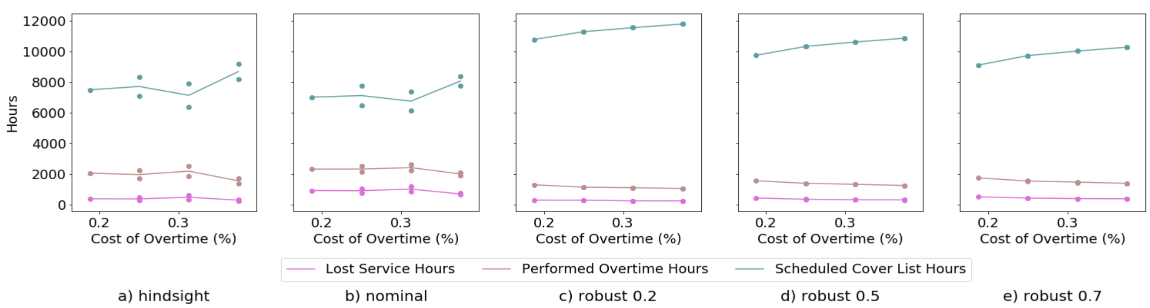
Realized Outcome (Unconstrained Resources):



i) w.r.t cost of cover list hours



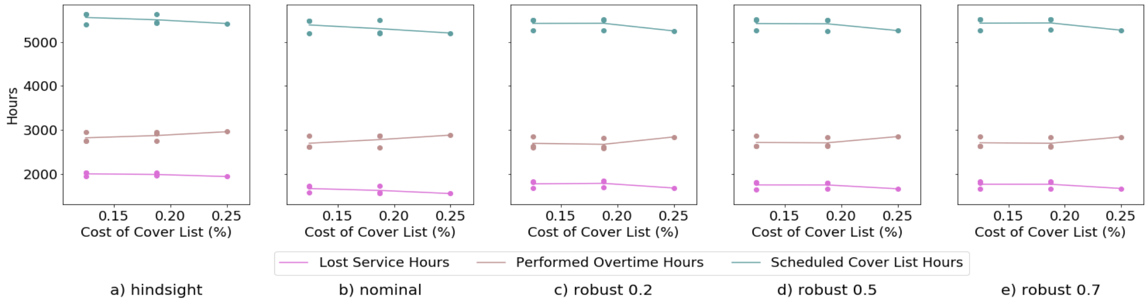
ii) w.r.t cost of overtime hours



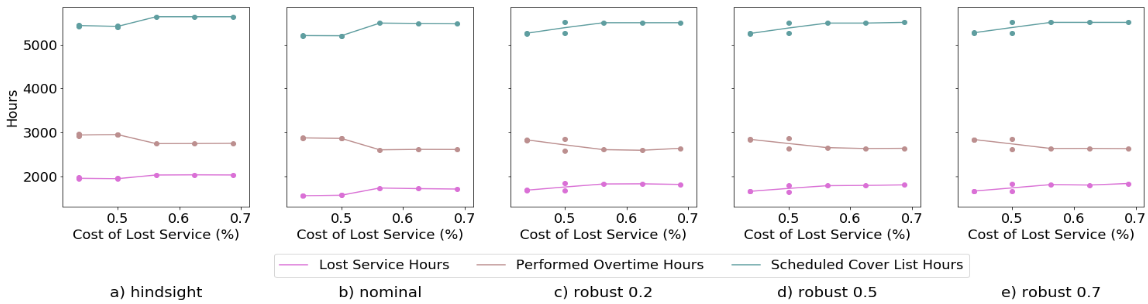
iii) w.r.t. cost of lost service hours

Figure E-2 Sensitivity Analysis of Realized Outcome (unconstrained)

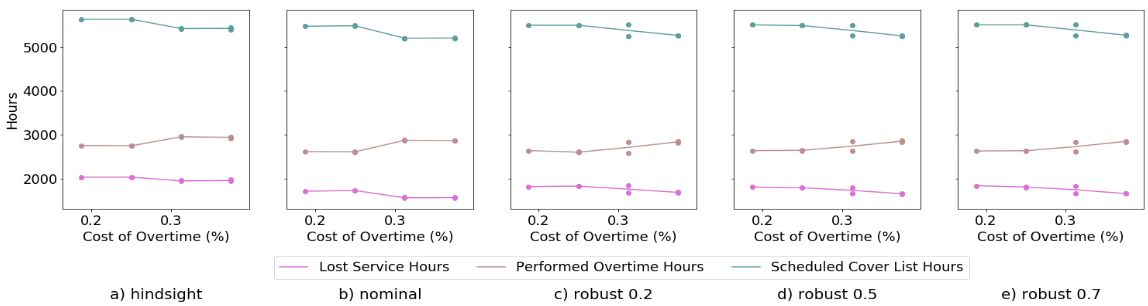
Worst-case Outcome (Constrained Resources):



i) w.r.t cost of cover list hours



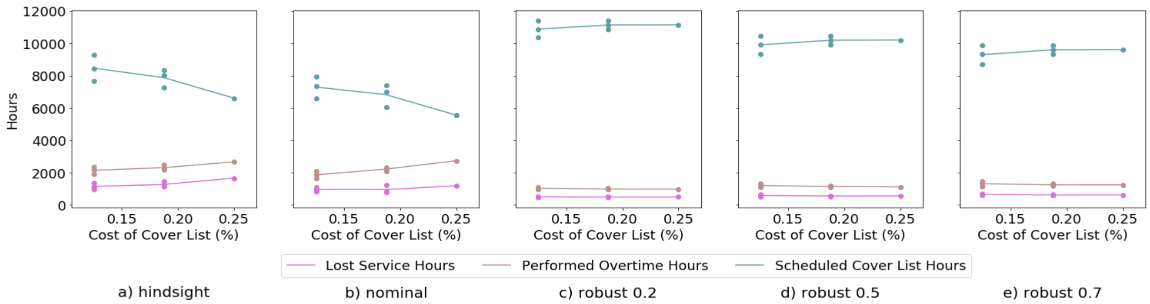
ii) w.r.t cost of overtime hours



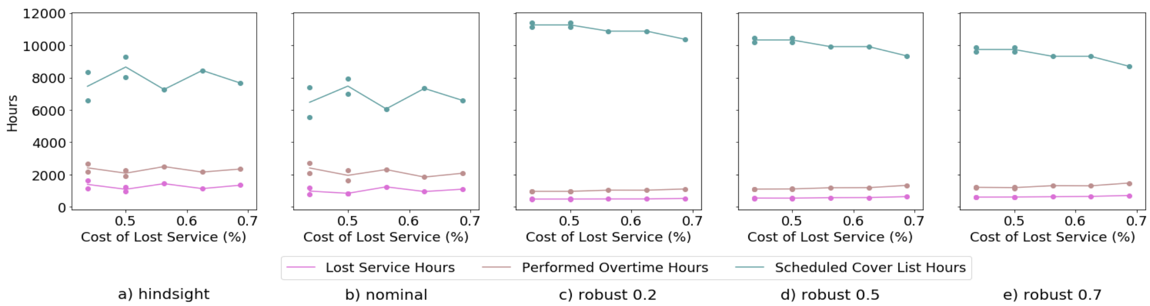
iii) w.r.t. cost of lost service hours

Figure E-3 Sensitivity Analysis of Worst-case Outcome (constrained)

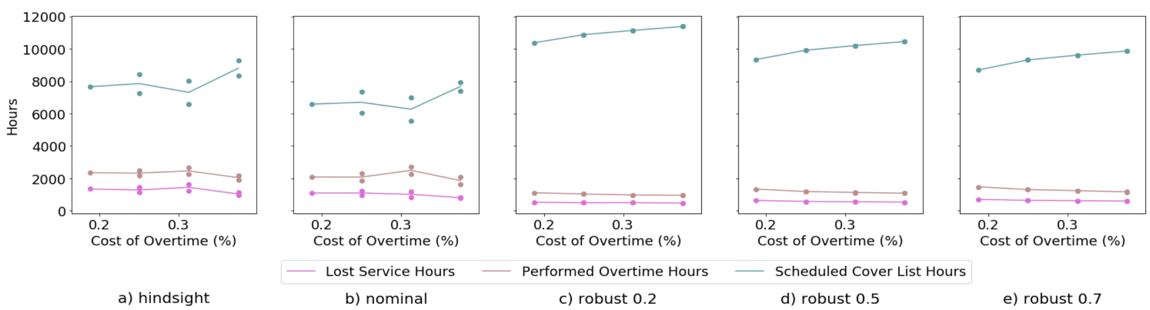
Worst-case Outcome (Unconstrained Resources):



i) w.r.t cost of cover list hours



ii) w.r.t cost of overtime hours



iii) w.r.t. cost of lost service hours

Figure E-4 Sensitivity Analysis of Worst-case Outcome (unconstrained)

Appendix F Scenario Analysis of the Nominal Model

Table F-1 Split Covers (Nominal Model)

Adaptive Robust ($\alpha = 0.7$)		**Constrained Resources**		**Unconstrained Resources**	
		8h straight shifts	4h splits allowed	**8h straight shifts**	4h splits allowed
Realized Outcome	Financial cost	**89%**	89%	**99%**	99%
	Lost Service (h)	**1383**	1304	**771**	712
	Overtime (h)	**2748**	2736	**2141**	2047
	Scheduled Cover List (h)	**5704**	5704	**7768**	7840
	Productive Cover List Time (h)	**4155**	4246	**5334**	5479
	Cover List Utilization (%)	**72%**	74%	**68%**	69%
Average Case (100 simulations)	Financial cost	**88%**	88%	**97%**	97%
	Lost Service (h)	**1197**	1162	**555**	513
	Overtime (h)	**2548**	2557	**1754**	1697
	Scheduled Cover List (h)	**5800**	5802	**7802**	7880
	Productive Cover List Time (h)	**4472**	4500	**5805**	5903
	Cover List Utilization (%)	**77%**	77%	**74%**	74%
Worst Case	Financial cost	**89%**	89%	**97%**	98%
	Lost Service (h)	**1723**	1706	**952**	909
	Overtime (h)	**2616**	2604	**1847**	1793
	Scheduled Cover List (h)	**5480**	5504	**7328**	7416
	Productive Cover List Time (h)	**4201**	4254	**5484**	5589
	Cover List Utilization (%)	**76%**	77%	**74%**	75%

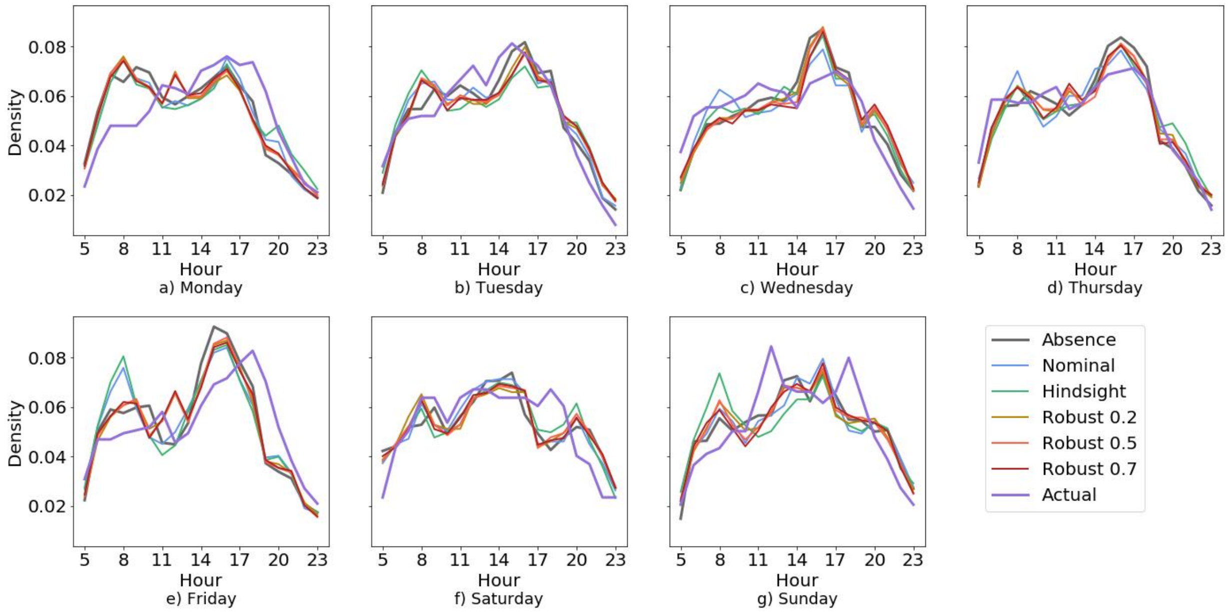


Figure F-1 Split Covers - Operational-Level Assignment (with Unconstrained Resources)

References

Abernathy, W.J., Baloff, N., Hershey, J.C., Wandel, S., 1973. THREE-STAGE MANPOWER PLANNING AND SCHEDULING MODEL - A SERVICE-SECTOR EXAMPLE. Oper. Res. https://doi.org/10.1287/opre.21.3.693

Ahn, S., Lee, S., Steel, R.P., 2013. Effects of Workers' Social Learning: Focusing on Absence Behavior. J. Constr. Eng. Manag. 139, 1015–1025. https://doi.org/10.1061/(asce)co.1943-7862.0000680

Allen, S.G., 1981. An Empirical Model of Work Attendance. Rev. Econ. Stat. 63, 77–87.

APTA, 2019. Public Transportation Fact Book 2019 23.

Bae, S.H., Fabry, D., 2014. Assessing the relationships between nurse work hours/overtime and nurse and patient outcomes: Systematic literature review. Nurs. Outlook. https://doi.org/10.1016/j.outlook.2013.10.009

Bai, L., Liu, P., Li, Z., Xu, C., 2011. Using Multivariate Poisson-Lognormal Regression Method for Modeling Crash Frequency by Severity on Freeway Diverge Areas. Proc. 11th Int. Conf. Chinese Transp. Prof. 346–359.

Bandi, C., Bertsimas, D., 2012. Tractable stochastic analysis in high dimensions via robust optimization, in: Mathematical Programming. https://doi.org/10.1007/s10107-012-0567-2

Barmby, T., Nolan, M., Winkelmann, R., 2001. Contracted workdays and absence. Manchester Sch. 69, 269–275. https://doi.org/10.1111/1467-9957.00247

Baron, O., Milner, J., Naseraldin, H., 2011. Facility location: A robust optimization approach. Prod. Oper. Manag. 20, 772–785. https://doi.org/10.1111/j.1937-5956.2010.01194.x

Bayliss, C., De Maere, G., Atkin, J., Paelinck, M., 2012. Probabilistic Airline Reserve Crew Scheduling Model, in: 12th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems. pp. 132–143. https://doi.org/10.4230/OASIcs.ATMOS.2012.132

Bayliss, C., De Maere, G., Atkin, J.A.D., Paelinck, M., 2017. A simulation scenario based mixed integer programming approach to airline reserve crew scheduling under uncertainty. Ann. Oper. Res. 252, 335–363. https://doi.org/10.1007/s10479-016-2174-8

Beckers, D.G.J., Van Der Linden, D., Smulders, P.G.W., Kompier, M.A.J., Taris, T.W., Geurts, S.A.E., 2008. Voluntary or involuntary? Control over overtime and rewards for overtime in relation to fatigue and work satisfaction. Work Stress. https://doi.org/10.1080/02678370801984927

Beckers, D.G.J., Van Der Linden, D., Smulders, P.G.W., Kompier, M.A.J., Van Veldhoven, M.J.P.M., Van Yperen, N.W., 2004. Working overtime hours: Relations with fatigue, work motivation, and the quality of work. J. Occup. Environ. Med. https://doi.org/10.1097/01.jom.0000147210.95602.50

Bliemel, F., 1973. Theil's Forecast Accuracy Coefficient: A Clarification. J. Mark. Res. https://doi.org/10.2307/3149394

Bolotnyy, V., Emanuel, N., 2019. Why Do Women Earn Less Than Men ? Evidence from Bus and Train Operators, Working paper.

Bruns, F., Goerigk, M., Knust, S., Schöbel, A., 2014. Robust load planning of trains in intermodal transportation. OR Spectr. https://doi.org/10.1007/s00291-013-0341-8

Chib, S., Greenberg, E., Winkelmann, R., 1998. Posterior simulation and Bayes factors in panel count data models. J. Econom. 86, 33–54.

Chib, S., Winkelmann, R., 2001. Markov chain Monte Carlo analysis of correlated count data. J. Bus. Econ. Stat. 19, 428–435. https://doi.org/10.1198/07350010152596673

Chiband, S., Greenberg, E., 1995. Understanding the Metropolis-Hastings Algorithm. Am. Stat. 49, 327–335.

Constantino, A.A., de Mendonça Neto, C.F.X., de Araujo, S.A., Landa-Silva, D., Calvi, R., dos Santos, A.F., 2017. Solving a large real-world bus driver scheduling problem with a multi-assignment based heuristic algorithm. J. Univers. Comput. Sci. 23, 479–504.

DeAnnuntis, C.P., Morris, W.P., 2008. Transit Extraboard Management. Transp. Res. Rec. J. Transp. Res. Board 2072, 110–124. https://doi.org/10.3141/2072-12

Diab, E.I., Wasfi, R.A., El-Geneidy, A.M., 2014. Extraboard team sizing: An analysis of short unscheduled absences among regular transit drivers. Transp. Res. Part A Policy Pract. 66, 27–38. https://doi.org/10.1016/j.tra.2014.04.019

Dillon, J.E., Kontogiorgis, S., 1999. US airways optimizes the scheduling of reserve flight crews. Interfaces (Providence). 29, 123–131. https://doi.org/10.1287/inte.29.5.123

Gardner, W., Mulvey, E.P., 1995. Regression Analyses of Counts and Rates : Poisson , Overdispersed Poisson , and Negative Binomial Models. Quant. Methods Psychol. 118, 392–404. https://doi.org/10.1037/0033-2909.118.3.392

Gaudine, A.P., Saks, A.M., 2001. Effects of an absenteeism feedback intervention on employee absence behavior. J. Organ. Behav. 22, 15–29. https://doi.org/10.1002/job.73

Golden, L., Wiens-Tuers, B., 2005. Mandatory or Not Mandatory - Is that the Difference? The Nature of Overtime Work by Characteristics of Workers, Jobs and Employers. SSRN Electron. J. https://doi.org/10.2139/ssrn.995216

Gupta, D., Li, F., 2016. Reserve driver scheduling. IIE Trans. 48, 193–204. https://doi.org/10.1080/0740817X.2015.1078016

Gupta, D., Li, F., Wilson, N.H.M., 2011. Extraboard-Driver Workforce Planning for Bus Transit Operations. CURA Report. 11–18.

Hickman, M.D., Koutsopoulos, H.N., Wilson, N.H.M., 1988. Strategic Model for Operator Work-Force Planning in the Transit Industry. Transp. Res. Rec. 1165, 60–68.

Holly, S., Mohnen, A., 2012. Impact of Working Hours on Work-Life Balance. SSRN Electron. J. https://doi.org/10.2139/ssrn.2135453

Idson, T.L., Robins, P.K., 1991. Determinants of Voluntary Overtime Decisions. Econ. Inq. 29, 79–

91. https://doi.org/10.1111/j.1465-7295.1991.tb01254.x

Ingels, J., Maenhout, B., 2018. The impact of overtime as a time-based proactive scheduling and reactive allocation strategy on the robustness of a personnel shift roster. J. Sched. 21, 143–165. https://doi.org/10.1007/s10951-017-0512-6

Ingels, J., Maenhout, B., 2015. The impact of reserve duties on the robustness of a personnel shift roster: An empirical investigation. Comput. Oper. Res. 61, 153–169. https://doi.org/10.1016/j.cor.2015.03.010

Kaiser, C.P., 1998. What do we know about employee absence behavior? An interdisciplinary interpretation. J. Socio. Econ. https://doi.org/10.1016/S1053-5357(99)80078-X

Kaysi, I., Wilson, N.H.M., 1990. Scheduling Transit Extraboard Personnel. Transp. Res. Rec. 1266, 31–43.

Koutsopoulos, H.N., Wilson, N.H.M., 1987. Operator Workforce Planning in the Transit Industry. Transp. Res. Part A 21A, 127–138.

Ma, J., Kockelman, K.M., Damien, P., 2008. A multivariate Poisson-lognormal regression model for prediction of crash counts by severity, using Bayesian methods. Accid. Anal. Prev. 40, 964–975. https://doi.org/10.1016/j.aap.2007.11.002

Maass, K.L., Liu, B., Daskin, M.S., Duck, M., Wang, Z., Mwenesi, R., Schapiro, H., 2017. Incorporating nurse absenteeism into staffing with demand uncertainty. Health Care Manag. Sci. 20, 141–155. https://doi.org/10.1007/s10729-015-9345-z

MacDorman, L.C., MacDorman, J.C., 1987. COST-EFFECTIVE SIZING OF THE TRANSIT EXTRABOARD.

Peters, K., Fleuren, H., Kavelj, M., Silva, Ss., Gonnalves, R., Ergun, O., 2016. The Nutritious Supply Chain: Optimizing Humanitarian Food Aid. SSRN Electron. J. https://doi.org/10.2139/ssrn.2880438

Rogers, A.E., Hwang, W.T., Scott, L.D., Aiken, L.H., Dinges, D.F., 2004. The working hours of hospital staff nurses and patient safety. Health Aff. https://doi.org/10.1377/hlthaff.23.4.202

Seabold, S., Perktold, J., 2010. Statsmodels: Econometric and Statistical Modeling with Python. PROC. 9th PYTHON Sci. CONF.

Shiftan, Y., Wilson, N.H.M., 1994. Absence, Overtime, and Reliability Relationships in Transit Workforce Planning. Transp. Res. Part A 28A, 245–258.

Shiftan, Y., Wilson, N.H.M., 1993. Strategic transit work force planning model incorporating overtime, absence, and reliability relationships. Transp. Res. Rec. 98–106.

Sohoni, M.G., Johnson, E.L., Bailey, T.G., 2006. Operational airline reserve crew planning. J. Sched. 9, 203–221. https://doi.org/10.1007/s10951-006-6778-8

Strathman, J.G., Callas, S., 2012. Extraboard Performance: Trimet Case Study. Transp. Res. Board, 91st Annu. Meet.

Sturman, M.C., 1996. Multiple Approaches to Absenteeism Analysis. CAHRS Work. Pap. Ser.

Suárez, M.J., Muñiz, C., 2018. Unobserved heterogeneity in work absence. Eur. J. Heal. Econ. 19, 1137–1148. https://doi.org/10.1007/s10198-018-0962-6

Van Landeghem, H., Vanmaele, H., 2002. Robust planning: A new paradigm for demand chain planning. J. Oper. Manag. 20, 769–783. https://doi.org/10.1016/S0272-6963(02)00039-6

Wang, H., 2012. Bayesian Graphical Lasso Models and Efficient Posterior Computation. Bayesian Anal. 867–886. https://doi.org/10.1214/12-BA729

Wang, K., Zhao, S., Jackson, E., 2018. Multivariate Poisson Lognormal Modeling of Weather-Related Crashes on Freeways. Transp. Res. Rec. 2672, 184–198. https://doi.org/10.1177/0361198118776523

Wang, W.-Y., Gupta, D., 2014. Nurse Absenteeism and Staffing Strategies for Hospital Inpatient Units. Manuf. Serv. Oper. Manag. 16, 439–454. https://doi.org/10.1287/msom.2014.0486

Winkelmann, R., 1996. Markov Chain Monte Carlo analysis of underreported count data with an application to worker absenteeism. Empir. Econ. 21, 575–587. https://doi.org/10.1007/BF01180702

Zhan, X., Aziz, H.M.A., Ukkusuri, S. V., 2015. An efficient parallel sampling technique for Multivariate Poisson-Lognormal model: Analysis with two crash count datasets. Anal. Methods Accid. Res. 8, 45–60.

Zhao, M., Liu, C., Li, W., Sharma, A., 2018. Multivariate Poisson-lognormal model for analysis of crashes on urban signalized intersections approach. J. Transp. Saf. Secur. 10, 251–265. https://doi.org/10.1080/19439962.2017.1323059