LUND UNIVERSITY

An Empirical Study of Timetable Strategies and Their Effects on Punctuality

Palmqvist, Carl-William; Olsson, Nils; Hiselius, Lena

2017

Document Version:
Peer reviewed version (aka post-print)

Link to publication

Citation for published version (APA):
Palmqvist, C.-W., Olsson, N., & Hiselius, L. (2017). *An Empirical Study of Timetable Strategies and Their Effects on Punctuality*. Paper presented at 7th International Conference on Railway Operations Modelling and Analysis, Lille, France.

Total number of authors:
3

General rights
Unless other specific re-use rights are stated the following general rights apply:
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.
• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Read more about Creative commons licenses: https://creativecommons.org/licenses/

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

LUND UNIVERSITY

PO Box 117
221 00 Lund
+46 46-222 00 00

An Empirical Study of Timetable Strategies and Their Effects on Punctuality

Carl-William Palmqvist [a, 1], Nils O.E. Olsson [a, b], Lena Hiselius [a]
[a] Department of Technology and Society, Lund University
P.O. Box 118, 221 00 Lund, Sweden
[1] E-mail: carl-william.palmqvist@tft.lth.se Phone: +46 (0)70 438 57 37
[b] Department of Production and Quality Engineering, Norwegian University of Science and Technology
S. P. Andersens veg 5, 7031 Trondheim, Norway

Abstract
Punctuality in Sweden has been too low for several years, and there is now a goal in the industry that it must improve by more than five percentage points by 2020. Research has shown that timetable properties affect punctuality, and timetables can be changed relatively quickly and easily, compared to infrastructure and rolling stock. In this paper we study some strategies for assigning margins in timetables, and how they affect punctuality. The aim is to identify changes that can improve the punctuality of passenger trains. The study is based on a dataset containing detailed timetables and records of train movements for all trains in Sweden during the year of 2015, over 1.1 million departures. We find that every additional percentage point of margins improves punctuality by about 0.1 percentage points, and about the same for every percentage point increase in the weighted average distance of margins. It is common for margins on some line sections to be negative, which lowers punctuality by on average 4.1 percentage points. Small time supplements directly following station stops are associated with a punctuality drop of 2.5 percentage points. We also find that the number of interactions between trains at stations, and on line sections, lowers punctuality by 1.2 and 3.9 percentage points per instance, respectively. From these results we propose a package of changes to timetable planning which we estimate would improve overall punctuality by up to five percentage points.

Keywords
Railway, timetable, punctuality, time supplements, margins.

1 Introduction

Punctuality is an important factor for the attractiveness and efficiency of the railway sector. In Sweden a target has been set and agreed upon by the industry, that by 2020 the punctuality across all trains should be 95%. This is measured as arriving at the final destination with a delay of at most five minutes. Since 2012 the punctuality has been steady at around 90 % for passenger trains and slightly below 80% for freight trains (Trafikanalys, 2016). Large and rapid improvements are thus required, if the target of 95% is to be reached on time. As Parbo et al. (2016) show, timetable characteristics are important influencing factors for delays and robustness in railway traffic.

Previous studies indicate that properties of the timetable have an impact on delays and punctuality (Palmqvist, et al., 2017). These often stem from a number of strategic decisions a planner must consider when designing a timetable. On a high level these strategies include the balance between precision and slack (Olsson, et al., 2015), a balance between using headways to assign buffers *between* trains or using time supplements to assign margins *within* train journeys (Nelldal, 2009), the degree to which a cyclic timetable is desired, the heterogeneity of traffic and the degree to which homogenization measures are to be employed (Nelldal, et al., 2009), questions of geographical accessibility and the design of the network to be utilized, the balance between stops at end points or intermediate stations, and so on.

On a more detailed level that individual timetable planners operate on, the decisions to be made are often about the size of the margins and how they should be distributed. For both of these decisions, there are a number of different approaches or strategies, which can be identified both in the literature and in practice.

The punctuality effects of these decisions are often difficult to foresee for the planner, and because of the complexity of traffic, changing conditions, and sheer number of decisions each timetable planner makes, it is difficult to assess their impacts. Because of this difficulty, there has not been a convergence around best practice in timetabling, based on our interviews with planners. Furthermore, there have been no steady improvements in punctuality (Trafikanalys, 2016), which one might expect as planners learn what works and what does not.

1.1 Aim and Delimitations

The purpose of this paper is to study the effects on punctuality of some strategies for allocation of margins in timetables for passenger trains, by analyzing empirical data.

The analysis focuses on aspects that relate to the planning of a timetable, i.e. the size and distribution of margins. We also focus on margins “within” train paths rather than those “between” them. As such, this study does not address headway or buffer times.

Freight and service trains are not included, neither are strategic dimensions such as cyclicality, homogenization of traffic, frequency and network structure, or the specific causes for and distributions of delays.

1.2 Previous Research

An overview of the state of the art in timetable research is provided by Hansen (2009). The author concludes that the key issue for high quality timetables is a precise estimation of blocking times, taking into account the signals, platforms, train processing, and using

realistic run and dwell times. This is often not the case in practice, or when using deterministic models. Similarly, queuing and simulation models inadequately reflect speed variations and the behavior of railway staff. According to Parbo et al. (2016), planners have tools to make timetables robust against delays, for example by adding time supplements, lowering heterogeneity in the timetable by having uniform stopping patterns, finding optimal speed and reducing interdependencies between trains.

Carey (1999) discusses a number of different heuristic measures of timetable reliability, with special consideration to knock-on delays. These include probabilities of delays, calculated in a number of ways, and different headway based measures. The latter are found to be easier to use and calculate, because they do not require nearly as much data.

Influencing factors on train punctuality in Norway are presented by Olsson & Haugland (2004). In short, the authors find that in congested areas the management of boarding and alighting passengers is the key factor, while on single track lines the management of train crossings is the key success factor. Gorman (2009) used statistical analysis to study which factors contributed the most to delays for freight trains in the US. He found that the number of meets, passes and overtakes consistently had the highest impact, suggesting that congestion was the primary cause for delays.

Wiggenraad (2001) studied seven Dutch train stations in detail. He found that dwell times are longer than scheduled, that the dwell times at peak and off-peak were the same, and that passengers concentrated around platform access points. This suggests an improvement potential, of shorter real dwell times if travelers could be distributed more evenly along the platform. Along the same lines, Nie & Hansen (2005) studied trains in the station area of The Hague. They found that trains operate at lower than design speeds, and that dwell times at platforms are systematically extended because of other trains blocking their routes, and because of the behavior of train personnel.

Precision and Slack

In the literature on train timetables, delays and punctuality, Olsson et al. (2015) identify two major strategies.

One strategy is precision, which focuses on making sure that the trains run when they should by finding what causes variation in the run and dwell times, such as the vehicles, infrastructure, and traveler behavior at stations, and addressing those issues so that the disturbances are reduced. This leads to greater efficiency and shorter travel times, among other things, but it is a difficult to implement and maintain.

The other strategy is slack, and consists of introducing buffer times and redundancies to ensure that when a disturbance occurs and delays appear, the timetable has enough margins for the train to recover, and for the delays not to spread throughout the system. This strategy also improves reliability, but reduce efficiency.

One way that slack reduces efficiency is discussed in Carey (1998): the commonly observed fact that, as more time is allowed for an activity, there is a behavioral response that makes the activity take a longer time. This response reduces or eliminates the expected benefit from adding the time supplement, in terms of for example increased reliability. This is important to consider when assigning both dwell times and running time supplements. The paper presents a framework which arrives at optimal time allowances based on two ratios: the behavioral response ratio, and the ratio of scheduled time costs to lateness costs. The latter is touched upon by Rietveld et al. (2001), who studied the valuation of certain travel time increases compared to uncertain ones, in the Netherlands. They found that a certain delay of 1 minute was valued at 27 cents, while a 50% chance of a delay of 2 minutes was 64 cents, which implies a strong attitude of risk aversion.

Headway Times and Time Supplements

One way to introduce slack into the timetable is by using headway times to separate trains from each other, trying to ensure that even if one train is delayed, that delay does not spread to the surrounding traffic. The tradeoff with headway times is that the increased reliability comes at the cost of reduced capacity, as the number of trains running is necessarily decreased. Yuan & Hansen (2008) find headway times to be effective in reducing knock-on delays. In an earlier paper, Yuan & Hansen (2007) propose an analytic model for estimating knock-on delays of trains in complex stations, with the intent to optimize capacity utilization in stations. They find that as the buffer time between trains decreases, the knock-on delays increase exponentially. Similarly, Dewilde et al. (2013) introduce a method to increase timetable robustness in complex stations, basically by maximizing the minimum headway time between trains. They find that this approach improves the robustness in the station zone of Brussels by 8% and reduces knock-on delays in the area by half.

The second way of adding slack in a timetable is by adding time supplements to trains, so that if they are delayed for whatever reason, they have some time reserve that they can use to catch up to the schedule. The cost of time supplements is simply that the travel times increase, which in turn consumes capacity, increases costs, and reduces the competitiveness of the trains compared to other modes of travel. This is discussed by Cicerone et al. (2009) and Schöbel & Kratz (2009).

To an extent, time supplements are mandatory. The International Union of Railways (2000) requires recovery margins to be a minimum of 1.0 min/100 km for passenger trains that are motor-coach trainsets, and 1.5 min/100 km for other passenger trains. If the margins are expressed as a percentage of journey-time, values from 3 to 6 % are required, based on maximum speed and tonnage. The most commonly occurring combinations in Sweden require recovery margins of 5 %.

Also in a Swedish context, Nelldal et al. (2009) performed simulation experiments on high speed trains (speed limit 200 km/h) between Gothenburg and Stockholm, finding that by ensuring that the minimum headway between these trains and the surrounding traffic is at least 5 minutes, punctuality would improve by 5 to 10 percentage points, depending on the direction. Alternatively, additional running time supplements of four minutes would yield the same gains. They conclude that increasing headways is preferable, because it does not extend the journey time.

Distribution of Margins

When adding time supplements, we can identify five different strategies regarding their distribution from the literature, see for example Vromans (2005), Andersson (2014), and Palmqvist (2014).

1. A uniform distribution of margins

When timetables are created in Sweden, the runtimes are calculated as if the train had a technical top speed 3% lower than what it does, which practically speaking adds a uniformly distributed margin of 3%. This is often motivated by differences in train driver behavior, and in varying weather conditions. This is often described as the primary source of margins in timetables, but in fact only makes up a small fraction of the total margins. In addition to this, the signaling system permits trains to exceed the speed limit by a percentage point or two, which in effect adds a further margin, that is also uniformly distributed. Using simulation-based methods, Vromans (2005) and Vekas et al. (2012) found that a uniform

distribution was sub-optimal for delay recovery, given some assumptions of the delay distributions. However, Scheepmaker & Goverde (2015) show that from an energy-efficiency perspective it is better to distribute the running time supplements evenly.

2. Shifting margins towards beginning or end

If the supplements are not evenly distributed, there must by definition be more either towards the beginning or the end of the journey. The balance is this: If delays happen towards the end of the journey, margins towards the beginning of the journey are "wasted" and can no longer be used. If, however, a delay happens towards the beginning of the journey and there are not enough supplements there, the delay will be carried for a longer part of the journey affecting passengers and surrounding trains. Vromans (2005) uses the measure of *Weighted Average Distance*, or WAD, to describe how the supplements are distributed along the journey, and attempted to optimize this using both analytical and numerical methods for some hypothetical and real cases, concluding that a slight shift towards the beginning is best. This is further discussed in Andersson (2014).

3. Margins can be placed at or near strategic locations

These could be places where the risk of delays spreading is high, where a lot of travelers would be affected, or where punctuality is measured. In Sweden (Palmqvist, 2014) and Switzerland (Vromans, 2005; Haldeman, 2003), among others, node supplements are intended to increase punctuality into these previously defined strategic locations, called "nodes".

4. Where disturbances happen most frequently

Attempts can also be made to place the margins where disturbances happen most frequently. This is often seen as the preferred alternative, as it minimizes both the risk of supplements being "wasted", and the risk of delays being "carried" for long stretches. Naturally, the difficulty is in identifying these places, because disturbances are random to a large degree, and delays that occur systematically should be managed directly as the dwell- and runtimes are calculated instead of by added margins. It is complicated further by the fact that reported delays contain both primary and secondary delays, while only the primary delays are useful when locating disturbances (Vromans, 2005). However, as shown by Wiggenraad (2001), Dewilde et al. (2013) and Olsson & Haugland (2004) among others, there is ample evidence that delays often occur during scheduled stops at stations, and it may be reasonable to add margins directly after these stops.

5. Placing margins at "critical points"

A more advanced version of distributing margins is using the concept of "critical points", introduced in Andersson et al. (2013), further described and utilized in Andersson et al. (2015). The authors define critical points as points in a timetable when and where two trains are traveling on the same line in the same direction, either when one train enters a line after a preceding train, or when one train overtakes another. These points are especially sensitive to delays, both because existing delays tend to increase in these cases, and because any delays here easily spread to other trains. The first paper introduces the concept and a way to calculate the robustness in critical points, and compares the measure to other reliability measurements. The second paper describes a mixed integer linear programming approach

to increase the robustness in critical points, by reallocating the existing time supplements of relevant trains.

2 Method

2.1 Study Design

This study consists of three components carried out in sequence: a literature review, an interview study, and data analysis.

Firstly, a literature review was carried out covering the topics of timetabling, robustness, punctuality and delays. The emphasis was especially on the size and distribution of margins, and to factors linked to punctuality. The result of this review was then used in designing the interview study and the data analysis.

The interview study aimed to further inform the data analysis, and to help improve the understanding of its results. It contained interviews with all long term timetable planners in Southern Sweden, four in number, and the senior timetable planner for the largest passenger train operating company in Sweden. Four themes were explored: feedback, guidelines, trade-offs, and rules of thumb. Each interview was approximately an hour long, recorded, and transcribed in full, resulting in a written material of around 50 pages. Throughout the paper, we will make reference to statements made by the planners during these interviews, to help inform the results and discussion. We believe that the planners we interviewed, and the issues they mention, are representative for those in the rest of the country. However, the point of the interviews was not to be comprehensive or exhaustive, but to provide some interesting input and understanding into the wider study.

The data analysis was carried out on a dataset containing detailed timetables and train control data spanning one year and the totality of the Swedish railway network. The intent behind this analysis is to provide a means of studying the effects of decisions and strategies discussed in both the literature and the interviews.

We believe that the empirical aspect is our primary contribution to the research field. The rest of the paper is essentially structured as a presentation and discussion of the data analysis, with reflections and input from the literature and interview studies.

2.2 Dataset

The quantitative part of this study is based on analysis of a dataset containing detailed exports of all timetables used for train traffic in Sweden during the timetable year of 2015, provided by the Swedish Transport Administration.

This dataset covers almost 46,000 distinct timetable versions and over 1.1 million departures. Of these, 83 % were passenger trains, 14 % freight trains, and 3 % service trains. 43 % of journeys were longer than 100 km, 31 % shorter than 50, and 25 % between 50 and 100. All of the changes made during the ad hoc-process during the year are included. The detailed timetables are used to identify the size and distribution of margins.

We also have data for all train movements in Sweden, for the corresponding period, derived from the track blocking and signaling systems. We use this data to determine whether or not trains were punctual, allowing us to link together timetable properties with real outcomes.

A number of filters were the applied on the data, in order to focus on only passenger

trains, and to exclude incomplete observations and outliers according to a wide range of parameters. The remaining data covers over 470 000 completed passenger train journeys across Sweden during one year.

2.3 Analyzed Variables

In this paper, we are concerned with the problem of margin allocation in timetables for passenger trains, their size and how they should be distributed. The study seeks to analyze the strategies for answering those questions, and the effectiveness of those strategies in improving punctuality.

The strategies are studied through an analysis of the relationship between punctuality and a selection of influencing factors summarized in Table 1. Based on the literature review and interview study, six influencing factors were selected for use in the data analysis.

Variable	**Definition/Unit**	**Analyzed Range**	**Proposed by**
Punctuality	Fraction	0.00 – 1.00	
Influencing Factors			
Total size of margins	Fraction of run time added	0.00 – 0.50	The International Union of Railways (2000)
Distribution of time supplements	Weighted average distance, as a fraction	0.00 – 1.00	Vromans (2005)
Negative time supplements	Yes, or no	1, 0	Interview study
Average supplement directly following stops	Seconds	0 – 120	Interview study, Wiggenraad (2001)
Interactions at stations	Count	0 – 20	Gorman (2009)
Interactions on line sections	Count	0 – 5	Andersson, et al. (2013)

Table 1 Studied variables.

Punctuality

In this study we use punctuality as the measure against which we compare different strategies and influencing factors.

We define punctuality in the following way: trains arriving at their final destinations with a delay not exceeding five minutes are considered punctual and are given a value of 1, while trains arriving at their final destinations with a delay of more than five minutes are not considered punctual and are given a value of 0. In any aggregate of trains, the punctuality is calculated as the average of these values, either 1 or 0, present in the aggregate, and presented as a fraction. To illustrate: in a group of four trains where three are punctual and one is not, the punctuality is 0.75.

In some cases, the overall punctuality is discussed. This we define as punctuality across all trains, not just that of the trains in any particular aggregate. For instance, in Sweden

during the year of 2015, 83 % of all trains carried passengers. If punctuality for passenger trains were improved by five percentage points, punctuality overall would only increase by 4.15 percentage points, because the punctuality of the other 17 % of the trains would not necessarily increase. In this way we can estimate what impact the punctuality improvements for some passenger trains would have on the overall punctuality.

Margins and Time Supplements

Throughout this paper, we discuss margins and time supplements, regarding both their sizes and their distribution. In both cases, we do so in terms of either percent or fractions of minimum run times. We are able to do this because of the detailed timetables we have access to, which contain the minimum run time in one field and various time supplements in other, separate fields.

The terms "margins" and "time supplements" are similar in meaning, and could perhaps be used interchangeably, we will try to explain how we interpret them. In Sweden a timetable is made up of a series of line sections and stations, for which there are run times and dwell times, respectively. For every line section there is a non-zero run time, which is in turn made up of a technical minimum, calculated by a physics-based computer program, and time supplements. These time supplements are set manually by the timetable planner and can be zero, positive and even negative. When we write of "time supplements" we mean these specific and individual supplements. When we instead write of "margins" we mean the totality of all time supplements in a given timetable, or across a number of trains.

To illustrate both how we use the two terms, and to demonstrate how we calculate them, we will give an example. If a train journey is made up of 10 line sections where the technical minimum runtime is five minutes each, suppose that on two sections there are time supplements of two minutes each. On those two sections, the supplements are 0.4, if expressed as a fraction, while on the other eight sections the supplements expressed as fractions are 0.0. Across the whole journey, the train has margins of 8%.

When aggregating margins across trains, there are two conceivable ways of doing so, yielding slightly different results. One is to first calculate the margins for each train individually, and then average across those values. If one train has 10% margins and another 20%, the average across the two is 15%.

The other way is to take the sum of all time supplements across the trains, and divide it by the sum of the minimum run time for all trains. If the train that has 10% margins has a technical minimum runtime of 100 minutes, it has 10 minutes in time supplements. If the second train, with 20% margins, only had a technical minimum runtime of 50 minutes, it also has time supplements of 10 minutes. The sum of time supplements for both trains is then 20 minutes, and the technical minimum run time is 150 minutes, which makes for 13.33% in margins, on average.

We use the first method in this paper, calculating the margins for trains individually first, and averaging across those individual margins. The second method gives a higher weight to trains traveling longer distances, and at lower speeds, which we do not necessarily consider justified.

In this paper, we focus on margins "within" train paths rather than those "between" them. As such, we do not look at headway times, or buffer times as described in Pachl (2002). We know that these are also important factors, and leave them for future work.

Negative Time Supplements

Throughout our research, both in the interview study and in the data analysis, we have come across instances of negative time supplements in train timetables. These are cases where the

planner has subtracted run time, so that the scheduled time is shorter than the technical minimum. This is described further in section 3.1 below.

Assessing the impact of these negative supplements, we have chosen to use a dummy variable with the value of 1 if there are any instances of this in the corresponding timetable, and 0 if there are not. Thus, we have not looked at the number of negative margins in a timetable, the sum, or the average size, but merely identified the timetables that contain them and those that do not.

Distribution of Margins

As we saw in section 1.2, there are several different ways of distributing margins when planning a timetable. There are also a number of measures that can be used to describe how the margins have been distributed across a timetable, in retrospect. In this paper, we do this in two ways.

The first method we use is the Weighted Average Distance (WAD) measure described in Vromans (2005). This is used to measure how the various time supplements in a timetable are balanced, being more towards the beginning or end of the journey, or in between. It is expressed as a fraction and normally takes values between 0 and 1, with lower values expressing a shift towards the beginning and higher values a shift towards the end of the journey. A value of 0.5 implies a perfect balance. However, there are many different ways one can reach such a balance. Having a single time supplement in the very beginning and an equally large one at the very end will yield a 0.5, as will a single supplement in the middle, or a series of equally large supplements across the whole journey. The measure does not distinguish between these alternatives, although one might expect them to have different effects.

We found that the WAD sometimes behaves unpredictably when used with negative time supplements. They can cause the measure to take values both below 0 and above 1, which we do not believe was the intention. Rather than developing our own measure or tweaking the formula for calculating WAD, we have chosen to exclude observations where the WAD is negative or larger than one from our analysis of WAD. The trains thus excluded make up less than 2 % of the sample, which we consider acceptable, and we still believe that the WAD is a useful measure.

The other method we use to describe the distribution of margins is to look at line sections directly following scheduled stops. Often there are time supplements on these sections, intended to help recover delays that occurred at the station. To measure the extent to which this has been done consistently, as part of a conscious strategy, we have calculated the average size of time supplements directly following scheduled stops. These time supplements and their averages are given in seconds. To aid in visualization, we have rounded off the averages to intervals of 10 seconds each, then used this value as a grouping variable to create aggregates of trains for which we have calculated the degree of punctuality.

Interactions at Stations and on Line Sections

In this paper, we consider and count interactions between trains. We define an interaction as an instance when two trains are at the same place at the same time, which can happen at a station, or on a line section. When on a line section, the trains need to be traveling in the same direction to be counted as an interaction. This happens relatively frequently on double tracks, but is also possible on some single tracks, if there are multiple blocks between two stations.

We are well aware that the number of interactions for trains may not be as easily

influenced by the timetable planner as the size and distribution of margins. We include them in the paper partially because Gorman (2009) showed that interactions have a large influence on punctuality, and mostly because Andersson et al. (2013, 2015) described using scheduled interactions between trains as a trigger for assigning margins.

We count the interaction for both trains, regardless of their order. We do not have any information about the track layout at stations, and we do not use any speed or runtime differentials between trains on a line section in the way that a micro-simulation might, at this stage we simply identify and count the interactions.

Interactions have been identified and counted on the unfiltered dataset, including freight and service trains, passenger trains completing only parts of their journey, and so on.

Interactions can be counted both in the timetable, and in realized traffic. We have tried both, and found the results to be broadly similar. Instead of presenting all of these results, we chose to use the scheduled interactions at stations, and the realized interactions on line sections, where the correlations are the highest.

2.4 Data Analysis

The relationship between punctuality and the studied influencing factors is analyzed using visual analysis of plots, regression analysis and t-tests.

In order to aid the visual analysis, the number of observations of each influencing factor is reduced through an aggregation of observations. In T-SQL, the programming language we use for the most part, we use the “group by” command to aggregate observations according to different values of the influencing variables, and the punctuality is calculated within each of these aggregates.

When creating the aggregates, we perform rounding to create neater groups containing more uniform numbers of observations. For instance, we round off the total size of margins so that each aggregate represents one percentage point of added runtime. Without this rounding, there are many aggregates each representing unique fractions with 14 decimal points, containing uneven numbers of observations. This would make any visualization of the results difficult indeed.

The regression analysis is carried out in Matlab using Weighted Least Squares. Since the focus of the study is to analyze the general trend in the relationship between punctuality and the studied influencing factors, linear regressions are applied. The number of observations in each aggregate is used as weights, to account for the fact that some observations occur much more frequently than others. For example, the Weighted Average Distance used to describe the distribution of time supplements along the journey, took the value of 0.50 in 113 517 cases while the value of 0.90 occurred only 736 times. For this reason, we use the number of observations as weights. We also use the bisquare method to further increase the robustness of the fit against outliers in punctuality, which are otherwise evident from the plots.

T-tests are used in order to study the effect on punctuality by negative time supplements and time supplements directly following station stops.

3 Results with Discussion

3.1 Size of Margins

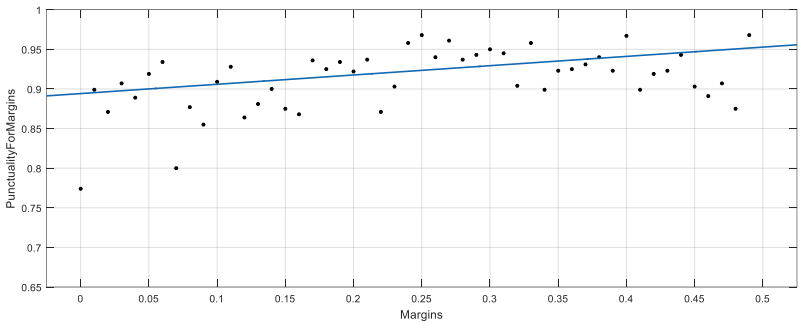

Figure 1 Punctuality and size of margins.

When plotting the size of margins against punctuality, as in Figure 1, we see a positive trend, but also a wide spread. A linear trend line (weighed least squares) is also estimated and results in a positive slope term, see Table 2. The results indicate that margins do seem to increase punctuality, in general, but one should not expect immediate changes in punctuality by slightly tweaking the size of margins in an individual train.

Table 2 illustrates the futility of using the size of margins as the only tool for improving punctuality: to improve punctuality by five percentage points, journey times need to be extended by approximately 40%. This is not acceptable in most cases. One possible way of explaining this difficulty is the behavioral response discussed by Carey (1998). Another is the fact that the distribution of margins also matters, and that if the distribution is inappropriate the time supplements are wasted.

Independent Variable	**Unit**	**Intercept**	**Slope**	**R^2**
Margins	Fraction	0.894	0.117	0.528

Table 2 Summary of linear regression analysis related to size of margins. Punctuality is the dependent variable, expressed as a fraction.

If we only adjust the average size of margins slightly, however, for instance by adding more margins to long distance trains that currently have the lowest margins and punctuality statistics, it is possible to make some gains. The average level of margins across all passenger trains in Sweden is currently 17%. If we increase this by four percentage points to 21%, which is the current average for regional trains, we would expect punctuality to improve by about 0.47 percentage points. While it is not enough of an improvement on its own, it is a start.

Negative Time Supplements

While studying timetables in detail, negative time supplements were identified. These are

instances where the planners have removed time so that the scheduled run time is in fact shorter than the technical minimum. This practice also came up in our interviews, but not in our review of the literature. The timetable planners say that this practice is at the request of the train operating companies, that it is mainly done for local trains on single track lines, that it is accompanied by positive supplements on later line sections where the trains are intended to catch up, and that it has "always" been thus. Several different rationales are cited by the planners:

- It helps timetables fit together, by making it appear as if trains will make it to a station in time for a meeting. It is a way to run more trains than would otherwise be possible and is requested by the train operating companies.
- On some lines, the rules state that arrival times should occur only at whole minutes, which requires rounding. If one always rounds up, this can add several minutes to the journey time, which is often undesirable. For this reason, planners sometimes round down, subtracting seconds.
- On some stations, there are not many passengers, sometimes none, and then the train operating companies think it is a waste for the train to just stand at the platform and wait for the departure time. By having very short or even non-existent dwell times, or using negative time supplements to ensure that the train is artificially late to the station, these "unnecessary" waits can be avoided.
- It has "always" been done this way, despite it being against the rules.

Assessing the effects of this practice by looking at traffic data for a year, we found a t-test, separating those timetables with instances of negative margins from those without, to be more illustrative than a diagram following the count or summed time of such instances. Summary statistics and results of this test are found in Table 3.

Negative Supplements Occur	Count	Punctuality	Variance	P-value
Yes	212 281	0.887	0.100	0E+00
No	316 228	0.928	0.067	

Table 3 Summary of t-test results, negative time supplements.

We estimate a difference in punctuality of 4.1 percentage points, indicating that the practice of using negative time supplements is detrimental to punctuality. With the mix of trains running in Sweden, and the prevalence of this practice, we estimate that by no longer assigning any negative time supplements, overall punctuality would improve by about 1.25 percentage points.

Since the negative supplements reduce scheduled run times below the technical minimum, the reduced runtimes are not "real" and eliminating them from timetables would not be expected to extend actual travel times.

3.2 Distribution of Margins

How margins should be distributed along the journey is an important and somewhat controversial topic, within both research and practice. As we saw in the literature review, there are several possible strategies and little empirical evidence on which works best.

Independent Variable	Unit	Intercept	Slope	R^2
WAD	Fraction	0.856	0.109	0.402
Average supplement following stops	Seconds	0.886	0.001	0.501
Station interactions	Count	0.947	-0.012	0.911
Line interactions	Count	0.913	-0.039	0.974

Table 4 Summary of linear regression analysis related to distribution of margins. Punctuality is the dependent variable, expressed as a fraction.

One of the more often used measures for describing how margins are distributed is the weighted average distance (WAD), which takes a value of 0.5 if margins are perfectly balanced around the middle of the journey, a value close to 0.0 if margins are heavily shifted towards the beginning, and close to 1.0 if they are shifted towards the end. This is perhaps intended more as a descriptive than prescriptive measure, because there are a number of factors that influence where margins are suitable.

All the same, among the planners we interviewed there was an explicit argument of not wanting to shift margins towards the beginning of the journey, since that would risk them being wasted in case a delay occurred later along the journey. Similarly, but coming to the opposite conclusion, both Vromans (2005) and Vekas et al. (2012) performed simulation experiments and concluded that it was better to shift margins slightly towards the beginning of the journey, so that delays that happen early are not carried for too long without being reduced by existing time supplements.

Our dataset of detailed timetables and train observations allows us to add an empirical dimension to the question. In Figure 2, WAD is plotted against punctuality. The result indicates that punctuality, measured at the final destination, increases slightly as time supplements are shifted towards the end of the journey.

We also note that substantial shifts are required for noticeable gains in punctuality. It should be noted that we suspect that this result depends on the method for calculating punctuality. If it were measured in a more holistic way, such as at every station stop, we expect that peak punctuality will be reached at values of WAD closer to 0.5, but have not yet confirmed this.

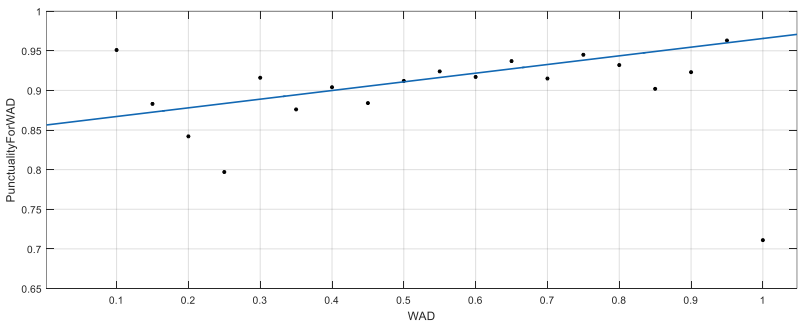

Figure 2 Punctuality and weighted average distance (WAD) of time supplements.

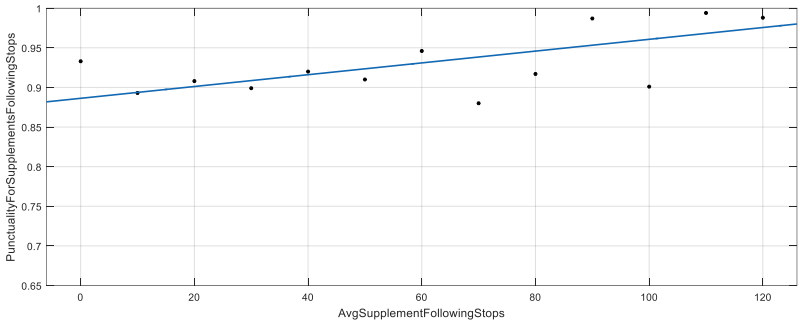

Figure 3 Punctuality and the average time supplement directly following stops.

As it stands, however, it appears that by shifting margins towards the later part of the journey, some punctuality gains can be made. The slope is 0.109, as can be seen in Table 4, and the average WAD across our sample is 0.56. So by shifting this to 0.65, we would expect punctuality to increase by almost a percentage point. Because this is simply shifting the distribution of margins, the total travel time would not increase. For passengers that do not travel the whole distance, however, there may be small differences.

3.3 Supplements directly following stops

One strategy when distributing supplements is to place them directly following scheduled stops. The rationale being that delays often occur at stations, and that it is good to recover delays as soon as possible. This practice was mentioned in a few of our interviews, but only one of the planners mentioned actually doing it.

Analyzing the timetable data, we found supplements following scheduled stops were frequent, and that they were clustered around the “round” numbers of 30, 60, 90 and 120 seconds, although the whole range was used, and we even found some larger supplements. Not having any supplements directly following the stops was the most frequent alternative. Because we are interested in this as a conscious strategy, rather than a chance occurrence, we calculated the average size of these supplements for each train. We then grouped the trains according to their averages, and calculated the punctuality within these groups.

We can identify two key results in Figure 3. Firstly, the positive slope of the trend line (see statistics in Table 4) indicating that larger supplements following stops are associated with higher punctuality. Secondly, punctuality is initially high when the average supplement is 0. These two findings combined suggest that if this strategy is to be used, it should be done with supplements of at least one minute per stop.

Analyzing further with t-tests, we confirm that the punctuality is lower for trains with small supplements, and higher if it is done consistently with larger supplements, and that these differences are statistically significant. Summary statistics of the t-tests are presented in Table 5.

Variable	Values Tested	Count	Punctuality	Variance	P-value
Average time supplement directly following scheduled stops	0 seconds	77 218	0.933	0.062	1
	1-59 seconds	459 416	0.908	0.084	< 0.001
	60+ seconds	28 526	0.943	0.054	< 0.001

Table 5 Summary of t-test results. All values are compared to a baseline with no time supplements following the scheduled stops.

One hypothesis for why the small supplements are associated with a lower punctuality is the argument made by Carey (1998). By allowing more time exiting stations, the train crew takes more time in executing that activity. In this case, they overcompensate. It is possible that this is made worse by the truncation of seconds in some sub-systems in the Swedish railway, so that the driver believes that the supplements given are larger than they really are.

If we for a moment disregard the causes for this phenomenon, and take it as a given, we can estimate that by removing all time supplements following scheduled stops for passenger trains, overall punctuality would improve by about 1.7 percentage points. At the same time, on average 2.3 minutes would be freed up per train, which could be used for margins elsewhere on the journey, or to reduce travel time.

Interactions with Other Trains

Previous research, like Olsson & Haugland (2004) and Gorman (2009), has shown that congestion can be an important cause for delays and reduced punctuality. One measure of congestion is the number of interactions a train has with other trains: if congestion is low this number will be low, but on highly congested lines the number of interactions will be higher.

The number of interactions can also be linked to trains traveling longer distances, being exposed to more traffic as it crosses through various local and regional train systems. Indeed, Harris (1992) showed that distance covered was statistically significant in determining punctuality.

Thus, the number of interactions may help explain both the delays associated with congestion, and the delays associated with long journeys.

Interactions can also be part of a strategy of how to distribute margins, as in Andersson, et al. (2015). For these reasons, we have looked at the link between interactions and punctuality.

We have counted the number of interactions at stations and on line sections, and tracked them separately. Plots of these against punctuality are found in Figure 4 and Figure 5, respectively, while the regression results are found in Table 4.

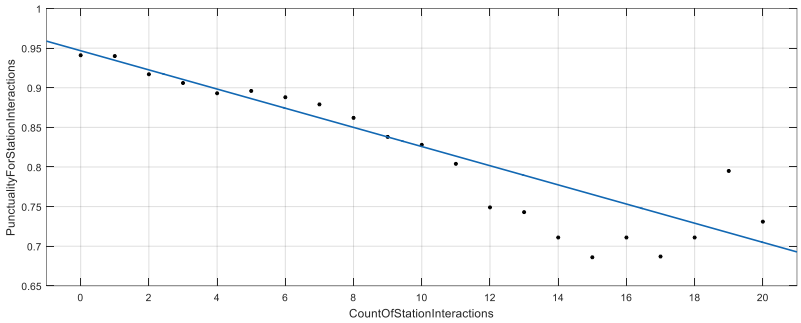

Figure 4 Punctuality and the number of interactions at stations.

At stations, every additional scheduled interaction lowers punctuality by 1.2 percentage points. We have not differentiated between single and double track, considered the layout of the station areas, or any characteristics of the other train, but simply counted the number of times the train is at a station at the same time as at least one other train.

There is a wide distribution of trains along this dimension: the single most common number of interactions at stations is one, which occurs for 25 % of the trains in our sample, more than 95 % of trains are covered by the count of eight, and the average number is 2.89. Thus, the largest improvement can be gained by focusing on trains that have few interactions, and either trying to reduce that number further, or improving the probability of the interactions being successful by adding or shifting margins. For instance, if the average number of interactions at stations could be reduced by 0.50 to 2.39, by multiplying with the slope found in Table 4, we would expect a gain in punctuality across passenger trains of 0.6 percentage points. If we factor in other kinds of trains, the gain in overall punctuality would be about 0.5 percentage points.

Interactions on line sections are much less common in our dataset. Looking at realized traffic, 98% of the trains have no such interactions, and in the timetable, the corresponding number is 97%. When interactions do occur, however, the effect on punctuality is considerable, as shown in Table 4. Each such interaction lowers punctuality by 3.9 percentage points. Again, we have only counted the number of these interactions, without considering the order of the trains or the speed difference. We would expect the effect to be greater for faster trains following slower ones, and will follow up on this in later research.

The low number of these interactions does mean, however, that even if they were to be eliminated entirely, the effect on overall punctuality would be less than 0.1 percentage points. Rather, efforts should be made to ensure that the number of line interactions remains low.

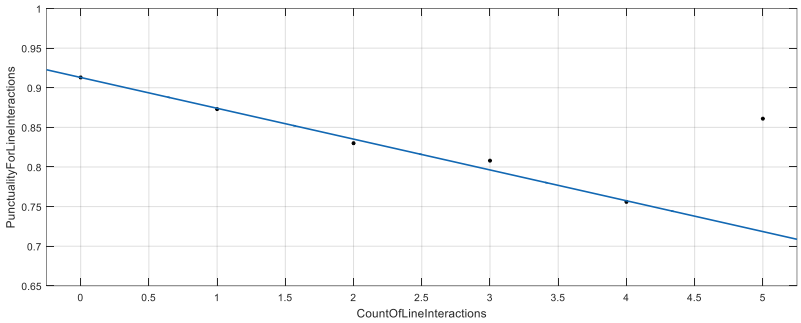

Figure 5 Punctuality and number of interactions on line sections.

4 Conclusions and Recommendations

Based on the results of the literature review, the interviews with planners, and our data analysis, we now give some recommendations on how timetables can be planned in the future to achieve a higher level of punctuality.

4.1 Size of Margins

We have found that increased margins are in fact associated with improved punctuality, but not strongly enough that the size alone can be used to reach the punctuality target of 95 %. Our estimate is that punctuality improves by 0.117 percentage points for each additional percentage point of run time extension. We also note that the average size of margins is about 10 percentage points larger in Swedish practice than in the international literature. This increases the downward variability in run times, makes operations less predictable and causes congestion in some places, consumes capacity, and increases travel times for passengers.

As we have shown, the benefits in terms of punctuality of these margins are rather limited. But since punctuality in Sweden is currently so far from the target, and rapid improvements are required, we still advise a modest increase in the average of about four percentage points, mostly for long distance trains, rather than a decrease. Hopefully this increase can be phased out as improvements from other sources come into effect.

We have found that in practice it is common for timetables to include some negative time supplements, meaning that there are some line sections where the scheduled run time is *shorter* than the technical minimum runtime, and that this is detrimental to punctuality. Our estimate is that by removing all instances of these from a timetable, the punctuality of the affected trains will improve by 4.1 percentage points and overall punctuality by 1.25 percentage points.

4.2 Distribution of Margins

Our results indicate that shifting time supplements towards the end of the journey slightly improves punctuality at the final destination. The effectiveness is about the same as when increasing the size of supplements. The results indicate that an increase in WAD of one

percentage point improves punctuality at the final destination by about 0.1 of a percentage point. We expect that the effectiveness will be lower, if one instead measures punctuality across all scheduled stops. For this reason, we only suggest a modest shift in average WAD from 0.56 to 0.65.

One strategy when distributing time supplements, is to place them directly following station stops. The results of this are mixed, but the conclusion we draw is that there should either be no such supplements, or that they should consistently be of at least one minute. If supplements are lower than 60 seconds, on average, punctuality is 2.5 percentage points lower than if there were no such supplements at all. If the average is 60 or over, however, our estimate is that punctuality improves by one percentage point. This is a significant improvement, but then having time supplements of at least one minute directly after every scheduled stop adds up quickly for most trains. If this is considered too expensive in time, it is better to skip these supplements altogether, since there appears to be a threshold effect.

Finally, in line with other research, interactions between trains are bad for punctuality and should be limited. Our estimate is that every additional interaction at a station lowers punctuality by 1.2 percentage points, while every additional interaction on a line section lowers punctuality by 3.9 percentage points. Thus, gains can be made by reducing the number of interactions, especially on line sections. This would mean ensuring that the first train is able to clear the line section, before the second train enters it, essentially increasing headways at some locations. If an interaction cannot be altogether avoided, holding back one train so that the interaction happens at a station instead of on a line section is expected to improve punctuality by 2.7 percentage points for both trains.

4.3 Estimated Impact of Recommendations

Throughout this paper, we have given a number of examples of how timetable variables can be tweaked in ways that affect punctuality. These changes and their estimated results are summarized in Table 6. Considered as a package, overall punctuality would increase by five percentage points, while the average size of margins is only increased by four percentage points.

Variable Affected	**Recommended Action**	**Gain in Overall Punctuality**
Size of margins	Increase average from 0.17 to 0.21	+0.0047
Negative margins	Eliminate all instances	+0.0125
Weighted Average Distance (WAD) of margins	Go from 0.56 to 0.65	+0.0100
Supplements directly following scheduled stops	Eliminate all those lower than one minute	+0.0170
Station interactions	Reduce average from 2.89 to 2.39	+0.0050
Line interactions	Eliminate all instances	+0.0008
All of the above	*All of the above*	*+0.0500*

Table 6 Estimated effects of recommendations.

Other combinations of adjustments are possible to imagine. While we have proposed a modest increase in margins of four percentage points, it is possible to imagine options where

the average size of margins is unchanged or even decreased down to more internationally comparable levels of 5-7%. We estimate this would lower punctuality by about one percentage point, while travel times would be reduced by 8-10 percentage points from their current levels. Given the need for rapid improvements in punctuality, however, we think a modest increase is more appropriate for the Swedish railway, at the current time.

We would caution against shifting the weighted average distance of margins to more than 0.65, because we suspect that if punctuality is measured at more locations than the final destination, the ideal WAD would be closer to 0.50. Along those lines, one might imagine a scenario where the average WAD is maintained at 0.56 rather than increased. But since punctuality is so often measured at the final destination, we opt for the increase.

The largest gains in punctuality are estimated to come from eliminating the practice of allowing negative margins and supplements lower than one minute following scheduled stops. These policies appear to be mistakes that should not be allowed to persist.

One might further dispute the possibility of decreasing the number of interactions, especially as the capacity utilization and volume of traffic increase. We think some reduction is possible by more conscious planning, but even without the improvements we estimate from these changes, punctuality is set to improve substantially from the changes we propose.

Our contacts and funders at the Swedish Transport Administration wish to comment that they are especially in favor of removing negative margins, slightly shifting the WAD of margins, and experimenting with the balance of dwell times and margins after station stops. This fits well together with other ongoing research and development which they finance.

References

Andersson, E., 2014. “Assessment of Robustness in Railway Traffic Timetables”, Linköping University, Norrköping.

Andersson, E. V., Peterson, A. & Thörnquist Krasemann, J., 2013. “Quantifying railway timetable robustness in critical points”, *Journal of Rail Transport Planning & Management*, issue 3, pp. 95-110.

Andersson, E. V., Peterson, A. & Thörnquist Krasemann, J., 2015. “Reduced railway traffic delays using a MILP approach to increase robustness in Critical Points”, *Journal of Rail Transport Planning & Management*, issue 5, pp. 110-127.

Carey, M., 1998. “Optimizing scheduled times, allowing for behavioral response”, *Transportation Research Part B*, vol. 32, no. 5, pp. 329-342.

Carey, M., 1999. “Ex ante heuristic measures of schedule reliability”, *Transportation Research Part B*, issue 33, pp. 473-494.

Cicerone, S. et al., 2009. “Recoverable robust timetabling for single delay: Complexity and polynomial algorithms for special cases”, *Journal of Combinatorial Optimization*, issue 18, pp. 229-257.

Dewilde, T., Sels, P., Cattrysse, D. & Vansteenwegen, P., 2013. “Robust railway station planning: An interaction between routing, timetabling and platforming”, *Journal of Rail Transport Planning & Management*, issue 3, pp. 68-77.

Gorman, M. F., 2009. “Statistical estimation of railroad congestion delay”, *Transportation Research Part E*, pp. 446-456.

Haldeman, L., 2003. *Automatische Analyse von IST-Fahrplanen*, Zürich.

Hansen, I. A., 2009. “Railway Network Timetabling and Dynamic Traffic Management”, *2nd International Conference on Recent Advances in Railway Engineering (ICRARE-2009)*, pp. 135-144.

Harris, N., 1992. *Planning Passenger Railways: A Handbook*, Transport Publishing Company Limited, London.

International Union of Railways, 2000. “UIC Code 451-1 Timetable Recovery margins to guarantee timekeeping - Recovery margins”, International Union of Railways, Paris.

Nelldal, B.-L., 2009. *Kapacitetsanalys av järnvägsnätet i Sverige. Delrapport 3: Förslag till åtgärder för att öka kapaciteten på kort sikt*, Kungliga Tekniska Högskolan, Stockholm.

Nelldal, B.-L., Lindfeldt, A. & Lindfeldt, O., 2009. *Kapacitetsanalys av järnvägsnätet i Sverige. Delrapport 1: Hur många tåg kan man köra? En analys av teoretisk och praktisk kapacitet*, Kungliga Tekniska Högskolan, Stockholm.

Nie, L. & Hansen, I. A., 2005. ”System analysis of train operations and track occupancy at railway stations”, *European Journal of Transport and Infrastructure Research*, vol. 1, issue 5, pp. 31-54.

Olsson, N. et al., 2015. *Punktlighet i jernbanen - hvert sekund teller*, SINTEF akademisk vorlag, Oslo.

Olsson, N. & Haugland, H., 2004. “Influencing factors on train punctuality - results from some Norwegian studies”, *Transport Policy*, issue 11, pp. 387-397.

Pachl, J., 2002. *Railway Operation and Control*. VTD Rail Publishing, Mountlake Terrace.

Palmqvist, 2014. *Gångtidstillägg för snabbtåg*, Lund University, Lund.

Palmqvist, C., Olsson, N. & Hiselius, L., 2017. “Delays for passenger trains on a regional railway line in Southern Sweden”, *International Journal of Transport Development and Integration*, vol. 1, issue 3, pp. 421-431.

Parbo, J., Nielsen, O. & Prato, C., 2016. “Passenger Perspectives in Railway Timetabling: A Literature Review”, *Transport Reviews*, vol. 36, issue 4, pp. 500-526.

Rietveld, P., Bruinsma, F. & van Vuuren, D., 2001. ”Coping with unreliability in public transport chains: A case study for Netherlands”, *Transportation Research Part A*, vol. 35, pp. 539-559.

Scheepmaker, G. M. & Goverde, R. M., 2015. “The interplay between energy-efficient train control and scheduled running time supplements”, *Journal of Rail Transport Planning & Management*, issue 5, pp. 225-239.

Schöbel, A. & Kratz, A., 2009. “A bicriteria approach for robust timetabling”, *Robust and online large-scale optimization*, pp. 119-144, Springer Berlin Heidelberg.

Trafikanalys, 2016. *Punktlighet på järnväg 2015*, Trafikanalys, Stockholm.

Trafikanalys, 2016. *Railway Transport 2015 quarter 4*, Trafikanalys, Stockholm.

Trafikverket, 2016. *Analysmetod och samhällsekonomiska kalkylvärden för transportsektorn: ASEK 6.0*, Trafikverket, Borlänge.

Trafikverket, 2016. *Noder i Järnvägssystemet T17*, Trafikverket, Borlänge.

Vekas, P., van der Vlerk, M. & Haneveld, W., 2012. ”Optimizing existing railway timetables by means of stochastic programming”, *Stochastic Programming E-Print Series-(SPEPS)*. Humboldt Universität zu Berlin, Berlin.

Wiggenraad, I. P. B., 2001. *Alighting and boarding times of passengers at Dutch railway stations*, Delft.

Vromans, 2005. *Reliability of railway systems*, Rotterdam.

Yuan, J. & Hansen, I. A., 2007. ”Optimizing capacity utilization of stations by estimating knock-on train delays”, *Transportation Research Part B*, vol. 41, pp. 202-217.

Yuan, J. & Hansen, I. A., 2008. ”Closed form expression of optimal buffer times between scheduled trains at railway bottlenecks”, *11th International IEEE Conference on Intelligent Transportation Systems*.