A Mixed Method for Railway Capacity Allocation

Emanuel Broman[1], Jonas Eliasson[2]

[1]*KTH Royal Institute of Technology, Teknikringen 10, SE-114 28 Stockholm, Sweden*
[2]*City of Stockholm, Flemingg. 4, SE-112 26 Stockholm, Sweden*

Abstract

Railways in several countries are now operated through open-access competition. In principal this allows for anyone to run train services at any time, although in practice a process is needed to coordinate and prioritise between requests, especially if capacity is scarce. Central to this process is the national infrastructure agency, which makes priorities based on its own external valuation of operators' potential welfare contribution, typically based on some generalisation. However, such generalisations are sometimes insufficient to differentiate between competitors in an open-access environment, where two or more operators may wish to offer similar services on the same line at the same time of day. More exact methods to calculate welfare are, although desirable, probably unfeasible as they require commercially sensible information such as number of passengers, fares and cost structures.

This paper proposes a method that efficiently allocates capacity between commercial and public operators, as well as between similar operators. Market prices are introduced in a limited form. A mixed method, it uses auctions to allocate pre-defined paths to commercial operators on specified, capacity-constrained lines, while allocating most capacity through a more traditional planned process. To balance between commercial and non-commercial traffic, it uses a novel valuation tool that calculates the social welfare of timetables for public traffic, then compares this to the price offered by commercial firms in an auction.

The result is a feasible method for efficient capacity-distribution between and within market segments, and incentives for more efficient use of infrastructure capacity by operators.

Keywords: railways; capacity allocation; deregulation

1. Introduction

Open-access competition has been introduced on several European passenger railway markets over the last years. Competition seems to be especially beneficial on high-demand lines, although a challenge is to efficiently allocate capacity. When capacity is scarce, an allocation procedure is needed. The infrastructure agency must rearrange trains compared to operators' requests; it moves departure times, inserts extra waiting times at stations or denies requests entirely. The conventional approach to capacity allocation may not be adept to dealing with commercial operators in direct competition; questions are raised regarding its efficiency, transparency and fairness.

Addressing these problems, we propose in this paper a capacity allocation method that deals with commercial and public traffic in a deregulated environment. It takes into account the difficulty that tax-funded operators have of assessing the welfare contribution of marginal changes to timetables. It also addresses the problem of distinguishing between similar commercial operators. This is done through the combination of an auction mechanism for commercial operators with a method to calculate the welfare of tax-funded regional traffic.

1.1. Background

Deregulation of the Swedish railways began in the early 1990s and has progressed ever since. Infrastructure management is now separated from operations, where there is free access for commercial operators. Contracts for unprofitable long-distance lines and commuter train lines are sold through reverse auctions; infrastructure maintenance is purchased from commercial firms; freight transport and profitable passenger operations are both open, largely unregulated markets. Since deregulation, there is now a multitude of operators and of train types.

Traffic volumes have increased as well. This trend started in the early 1990s when new fast trains were introduced on the main lines, which increased the market share of railways. Ten years later, large infrastructure projects financed through public-private partnerships (a connector to Stockholm airport; a bridge across Öresund

* Corresponding author. Tel.: +46 70-237 19 24.
E-mail address: emanuel.broman@abe.kth.se

strait) enhanced the attractiveness of train travel. Meanwhile, regional authorities expanded subsidised traffic, boosting travelling. This rise in traffic volumes has resulted in more congested lines, making an efficient capacity allocation more important.

As efficient capacity allocation has gained in importance, it has also become more complex due to deregulation. Running in convoys to save capacity is for instance less efficient when trains are of different type, as they have different operating speeds and braking and accelerating performances. The separation of infrastructure from operations means the infrastructure agency has less information to base decisions on, especially when the information is commercially sensitive. Incentives have also emerged to 'game' the allocation process, for instance by a dominant actor through making a complete application that is internally conflict-free, in order to make it a natural starting point that others need to adapt to.

The current allocation method seems not to be entirely adapted to a deregulated environment. At its heart are so-called priority criteria, which are a reasonably efficient and well-working measure for solving conflicts between different types of traffic, but inadequate for conflicts between direct competitors. Another important property of the method is negotiation to solve conflicts; this is increasingly difficult when the number of operators increases. The regulatory agency (Transportstyrelsen) has directed criticism against the infrastructure agency (Trafikverket) for a method deemed to be poorly defined and to lack transparency. These concrete problems need to be addressed.

2. Requirements of a capacity allocation method

The prerequisites of a method for allocating access to the railway infrastructure are described in this chapter.

2.1. General prerequisites

The main method for capacity allocation in use in Sweden is the yearly process. It begins with all operators applying for capacity on a fixed date, approximately nine months before the timetable is put to use. When different requests are in conflict, negotiations are organised by the infrastructure agency. If an agreement cannot be reached this way, the infrastructure agency formally declares the line to be capacity constrained, and then unilaterally decides how to solve the conflict. The solution must, according to Swedish law, be such that it maximises social welfare. The infrastructure agency's method to comply with that law is based on so-called priority criteria that give certain types of traffic priority over other, for instance high-speed trains get priority over freight trains.

The tracks are shared by different types of traffic, in this paper divided into long-distance commercial operations and subsidised regional traffic. Freight traffic is not explicitly dealt with in this paper, but it is assumed that enough capacity can be put aside for the short-term market. Fares and freight charges are not regulated; both commercial firms and regional traffic operators set fares freely.

2.2. Different categories of traffic

Long-distance traffic is carried out on commercial terms, either by the state-owned former monopolist or by private firms. These operators are allowed to set prices freely, and they use flexible pricing to achieve their objective to maximise profit. They are assumed to be less flexible in increasing or decreasing capacity, and to have a medium-long planning horizon. There are several operators which are difficult to tell apart from the outside: Incomes and costs are not publicly available, which means their efficiency cannot be assessed; there is no proxy for efficiency or welfare contribution that is good enough at distinguishing between operators that offer similar services; and it is not clear how short-term values such as consumer surplus for a scheduled train service should be weighed against longer-term values such as free and fair competition.

Regional traffic is organised by regional governments. Although traffic is operated under concession, it is the public body that makes decisions on timetables and fares. Forward planning is long. Pricing is comparatively inflexible since fare-structures are typically set yearly for all departures on a political level, and they frequently form central parts of election campaigns on the regional level. Capacity is as inflexible as for long-distance traffic.

There are reasons to believe that the regional traffic authorities are less well suited to judge the value of different timetables, compared to commercial operators. First, they may have the wrong objective. Although arguably any government body should strive to maximise overall welfare, it is possible that they instead aim to maximise the benefits of specific interest groups, or to get maximal attention in media. Secondly, welfare effects are not as immediately visible as revenue is. Commercial operators receive feedback in the form of changing incomes and costs when their service offering changes, but there is no analogous feedback for tax-funded operators. Thirdly, regional authorities do not automatically lose market share if they fail to make good decisions. This is a fundamental difference between firms operating in competitive markets and government entities; commercial firms that efficiently cater to their customers' needs can increase market presence to the detriment of less efficient competitors, but the size of government entities is mostly decided by other factors, such as geographic borders. Fourthly, regional governments have small to non-existent incentives to take account of the cost that they imposes on the national traffic. They have for instance been unwilling to make adjustments to enhance overall capacity utilisation, such as skipping small stations on every other tour, or moving some traffic from train to bus to save capacity for long-distance trains.

2.3. Other prerequisites

A practical limitation is that it takes time for an operator to react to a suggested timetable and formulate a response.

3. The allocation method

The proposed capacity allocation method builds on separate, but linked, mechanisms for commercial and tax-funded operators. The loss imposed on regional traffic operators by altering their timetables, compared to their requests, is calculated with social cost-benefit analysis (social CBA). Commercial operators are allocated train slots through an auction mechanism, and the balance between the two groups of operators is guaranteed by using the loss imposed on regional operators as reserve price in the auction. The reason for this separation is that too little data is available on commercial operations to make a proper CBA, while for public transport there are reasons to believe that an auction mechanism would not produce an efficient outcome (as described more thoroughly in section 2.2).

In summary, the method works as follows: First, regional operators send their applications. Secondly, the infrastructure agency plans corridors for long-distance commercial trains, making adjustments to the regional trains' timetables as necessary. It calculates the loss imposed on regional operators and their passengers in order to find a measure of the externality caused by each commercial slot. Thirdly, those slots are auctioned to commercial operators, with the calculated externality used as reserve price.

3.1. Defining the lines

In the first phase, the infrastructure agency selects the lines it deems suitable for the proposed method. In addition to setting the geographic boundaries, it is decided at which times of day to use the method. An important property of the selected lines is that they are capacity constrained. This means that only parts of the network with high capacity utilisation shall be included, and that they shall only be included at times of day, and on those days of the week, that they are especially heavily utilised. This may for instance be during morning and evening rush-hour. On double-track stretches the time-limit may apply differently for the two tracks, for instance if the outbound track from a city is congested earlier in the morning than the inbound track is.

Operation on selected lines consists at least in part of commercial passenger traffic, as it does on the main lines from Stockholm to Göteborg and Malmö, respectively. At the heart of the method is an auction, where commercial operators bid for departure slots, which implies that it is not useful for those parts of the network that are exclusively used by tax-funded traffic and freight traffic.

3.2. Preparations for the auction

Regional traffic departments are the first to send their timetable requests. It is important that these initial requests are in their view the most ideal traffic and not, as it is sometimes believed to be today, an adaptation to what they believe will be granted.

The infrastructure agency then decides an interval for the number of slots that will be sold to commercial operators. This interval shall be wide enough so that it can be said with certainty that the number of slots sold will fall within it. As with the definition of lines, this will likely be easier with experience. A first version of the regional timetable is made that exactly resembles the wishes of the regional traffic departments. This is then successively adjusted to make space for slots in the commercial timetable (see Figure 1). As the number of commercial slots increases, more intrusive changes need to be made to the regional timetable. When this process is complete, there is one version of the regional timetable and one corresponding version of the commercial timetable for each number of commercial slots within the defined interval. Slots that are defined in one version of the commercial timetable also appear in other versions, so that a version with eleven slots exactly resembles the one with ten slots, except for the one additional departure.

When this process is complete, the social welfare of each version of the regional traffic departments' timetables is assessed (see section 3.3). This method uses an origin/destination matrix for the number of passengers within the network, and then calculates the value of waiting time and in-vehicle time for all passengers. From this we receive a prediction of the socio-economic loss of each version of the timetable that is useful to our purposes, which is to compare the different timetable-versions to one another. For the timetables that correspond to few commercial slots this loss will be small, but as the number of commercial slots increases and the changes to the regional timetables become more severe, the socio-economic loss becomes greater. The change in socio-economic loss between two consecutive versions of the timetable is then used as a reserve price in the auction.

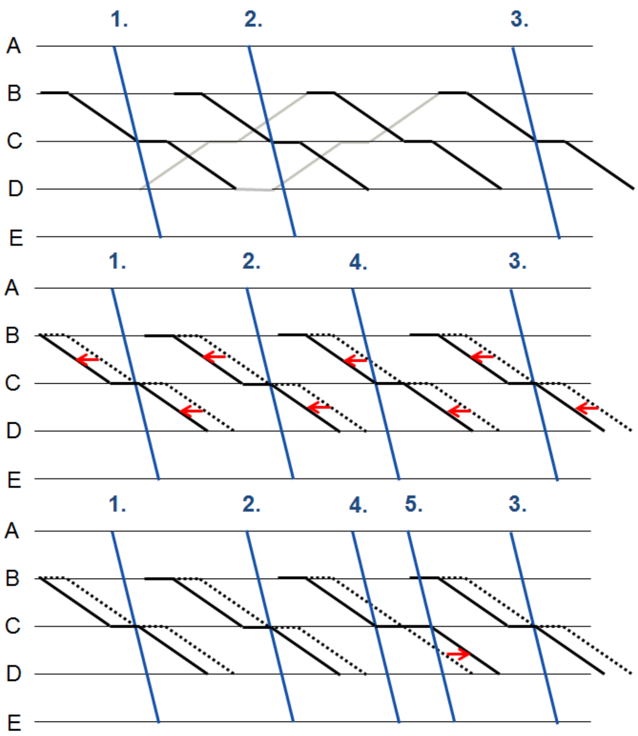

Figure 1

3.3. The welfare of timetables

The method presented in this paper makes use of a model developed by Ait Ali et.al. for calculating the welfare contribution of timetables for public traffic. It is possible with this model to compare the welfare contribution of two alternative timetables. If one of the timetables allows for a commercial train to be scheduled on the same track, the difference between the two timetables can be used as a measure of the commercial train's externality due to crowding.

The model includes an estimation of the number of travellers between any origin-destination (OD) pair in the network at any time of day. This OD matrix is constructed from data on the number of boarding passengers at all times and stations. Destinations are derived from boardings on the return-trip; since passengers on public transport typically used pre-paid electronic tickets, such as monthly travelcards, the same ticket is registered for several trips. The timetabling software RailSys is integrated with the model in order to provide a usable user interface for inserting timetables. RailSys also includes a so called microscopic (i.e. detailed) model of the network, with travel times for all OD pairs.

Based on the OD matric, the network model in RailSys and the inserted timetable, it is possible to calculate waiting times, travel times and a crowding factor for each traffic. Together with public data on the value of travel time, this is used to calculate the consumer surplus. Public data is also used to estimate operating costs.

The sum of changes in consumer surplus and operating costs when replacing one timetable with another is a good-enough estimation of the difference in social welfare between those timetables, for our purposes.

3.4. Auctioning the slots

The auction is organised by the infrastructure agency, and participating are representatives of all commercial operators that plan to have operations on the defined lines during the forthcoming timetable year. There are several auction rounds with up to a day in between, allowing for operators to calculate the effect of the tentative timetable and plan a response.

All slots for all defined lines on the network are auctioned simultaneously. This means that all operators place one bid for each slot in every round. The auction is designed to be slow and steady; bids may be either raised or lowered in every round, and the process continues until no one wishes to change any bid. The purpose of this is for operators to end up with a timetable that is feasible; the number of slots they buy must correspond to what rolling stock they have, timetables must fit the schedules of their on-board staff, and departures must be spread in an economically sound way.

Each departure slot has a reserve price that corresponds to the increased loss it imposes on the region traffic (as described in 1.2). If no one places a bid at least as high as the reserve price for a particular slot, then that slot is not sold. The number of slots sold is therefore decided in the auction, and hence also what version of the regional timetable that will be used. This ensures an efficient allocation of slots between subsidised and commercial traffic.

3.5. Small adjustments

After the timetable for the residual network has been set, a process begins to make small adjustments to the auctioned slots. These changes include adjusting the times at intermediate stops slightly to give adequate time for boarding depending on if there is a connecting train that brings many passengers, and similar small changes. These changes are only allowed if they can be made without major consequences for other traffic. The changes are made after negotiation with the infrastructure agency, and subject to a fee.

4. Is this feasible?

Interviews have been conducted to understand whether the proposed capacity allocation method is feasible. The interviewees are mostly employees of the Swedish infrastructure agency Trafikverket, and all have expert knowledge of various relevant fields, including of current allocation methods and practices, of strategic planning, and of welfare effects of proposed practices. (Westin, 2017) (Nordlöf, 2017) (Svensson, 2017) (Unebrand & Solinen, 2017) (Aronsson, 2017) (Dahlberg, 2017)

An important question is if it is possible to create train paths are that allow operators with different types of rolling stock to make use of speed, acceleration and braking abilities. According to Solinen (2017) this is possible for the two fast train operators on the Stockholm-Göteborg line, SJ and MTR Express. However, the sequence may be important if both operators use slots that are close in time, as SJ’s trains are faster than those of MTR Express. Svensson and Solinen (2017) mention that there is a postal (PostNord) train that departs from Stockholm in rush hour, and that this train is fast enough to use a corridor designed for the passenger trains. As PostNord is a commercial firm, this may be possible.

Unebrand and Svensson (2017) mention that capacity charges, including through auctions, can only be collected after a section is declared to be overloaded. But making such a declaration early in the process is not necessarily a bad idea, according to her. However, Westin (2017) argues that there is a large psychological barrier within the agency against using the overloaded declaration before operators have been consulted, and that this is a serious hindrance to the feasibility of the solution.

Regarding the process, Solinen (2017) claims that it would not be a problem for the regional traffic departments to send their capacity requests earlier than today, as is needed in the proposed process, since they know far ahead how they want to operate their traffic and they do not make large changes from one year to the next. However, both Svensson and Unebrand (2017) argue that the auction process should not be allowed to take too much calendar time, certainly not as much as a month. This is because the date when operators send in their requests is determined through international deals – it is currently the middle of April – and the process must therefore not be much longer than it is today.

One possible downside with the proposed solution concerns the negotiations that take place in today’s process, according to Solinen (2017). Sometimes, the regional traffic operators agree to cancel a departure in favour of one of SJ's fast trains if their passengers are allowed to travel on that train. Such solutions are probably not possible in the proposed solution. However, they are already rather rare.

Westin and Nordlöf (2017) ask if it is possible to 'game' the proposed solution. Since there are naturally few operators – typically only two – that compete on a single line, they may choose to decide in between who gets which slots, and thereby pay too low prices. Or possibly one operator would buy all slots and then choose not to run some of them (unless such behaviour is penalised). Indeed, Nordlöf (2017) argues that the 'natural' outcome of the auction may be that one operator buys all slots, as monopoly is more profitable than competition. If so, market shares may need to be regulated.

5. Conclusions

This paper proposes a method for allocating capacity on the Swedish railway network that combines market prices through auction with a planned process, and a method for estimating the socio-economic value of subsidies regional traffic. It attempts to solve the problems in today's method, including its inflexibility, its opaqueness and its lack of method for solving conflicts between similar competitors. Also, it provides a measure to allocate capacity between commercial and subsidised traffic efficiently.

Interviews have been conducted with staff at the infrastructure agency in order to assess the feasibility of the solution. This has provided insights into the details that need to be addressed before the proposed method can be implemented, as well as some questions of more fundamental nature that need to be addressed.

In all, it seems however that no serious obstacles have been found that should not be possible to overcome. The proposed method thus seems to offer a feasible path to introduce market prices as a means to solve timetabling conflicts on congested stretches of the Swedish railway network. Today's method could thereby be replaced with one that is more efficient and more suited to a deregulated environment.

Bibliography

Aronsson, M. (2017, January 25). PhD, SICS Swedish ICT. (E. Broman, & J. Eliasson, Interviewers)

Dahlberg, H. (2017, January 25). Market & Planning, Business Development, Section for Service Management. (E. Broman, & J. Eliasson, Interviewers)

Nordlöf, P. (2017, February 2). manager Expert Centre, Social CBA. (E. Broman, Interviewer)

Svensson, A. (2017, February 23). Strategic Development. (E. Broman, Interviewer)

Unebrand, J., & Solinen, E. (2017, February 23). planning, Expert Centre, Capacity Central. (E. Broman, Interviewer)

Westin, P.-E. (2017, February 6). Strategic planning, infrastructure charges. (E. Broman, Interviewer)