Centre for Transport Studies
STOCKHOLM

On timetable assumptions in railway investment appraisal

Jonas Eliasson
Maria Börjesson

Centre for Transport Studies, Royal Institute of Technology

CTS Working Paper 2013:31

Abstract
The benefits captured in an appraisal of a railway investment are determined by what timetables the analyst assumes in the scenarios with and without the investment. Without an explicit, objective and verifiable principle for which timetables to assume, the appraisal outcome is virtually arbitrary. This means that appraisals of railway investments cannot be compared to each other, and opens the door for strategic behaviour by stakeholders conducting seemingly objective cost-benefit analysis. We explain and illustrate the nature and extent of the problem, and discuss the practical consequences for appraisal comparability and conscious or unconscious misrepresentation. Finally, we discuss possible objective principles for appraisal timetable construction and contrast this with current practice, which is shown to be likely to exaggerate investment benefits in an appraisal.

Keywords: Cost-benefit analysis, appraisal, railway investments, timetables.

JEL Codes: R42, R48.

Centre for Transport Studies
SE-100 44 Stockholm
Sweden
www.cts.kth.se

vti

KTH VETENSKAP OCH KONST

1 INTRODUCTION

Railway investments are substantial financial commitments. In most countries, developing and maintaining the railroad network is a public responsibility. As with all public transport policy decisions, benefits can be weighed against costs and the cost-efficiency of different investments can then be compared. This makes well-developed and well-structured cost-benefit analysis (CBA) methods indispensable. The increasing interest for investments in high-speed railways makes this even more important, considering the huge costs at stake.

The outcome of a railway investment appraisal depends on the assumed timetables with and without the investment. It is meaningless to speak of the benefits of a railway investment as such – it is the use of the investment, i.e. the timetable, that decides what benefits a railway investment will generate. Railway investments are different from most other transport investments in that its use is planned and decided by society to a large extent, through capacity allocation, regulations, track access charges, subsidies, public transport provision etc., and by commercial operators, who strive to maximize profits, often with some monopolistic power. This stands in contrast to most other transport investments, such as roads or bicycle paths, where the use of the physical infrastructure is mainly decided by individual decisions by large numbers of travellers.

The importance of timetable assumptions becomes particularly evident when appraising capacity improvements, since an increase in capacity can be used in any combination of increased frequencies, decreased travel times, and reduced delays. Depending on how the new capacity is used, an appraisal will give different results. The appraisal outcome will also depend on what is assumed about the use of the existing capacity in the do-nothing case.

This means that the analyst responsible for carrying out an appraisal can influence how large benefits an investment will generate by choosing more or less socially efficient timetables. This influence can be conscious or subconscious. It is usually difficult or impossible to gauge the social efficiency of a timetable for outside observers and decision-makers, and sometimes even for the analyst. Clearly, this means that CBA results may become misleading and to some extent even arbitrary. Even worse, if the analyst represents a stakeholder with an interest in discrediting or promoting certain investments, there is an opportunity for strategic behaviour, with very little possibility for outsiders to discover this.

The purpose of this paper is to explain and illustrate the crucial importance of timetable assumptions in railway investment appraisal, why explicit principles for appraisal timetable construction are necessary, discuss the potential problems that the lack of such principles creates, discuss different such principles, and explain why current practice will tend to exaggerate investment benefits. The key insight is that without an explicit principle governing the choice of appraisal timetables both with and without an investment, the social benefit of a railway investment is not defined.

The intended audience for this paper is not primarily railway planners but people working with appraisal methodology. Appraisal methodology covers all sorts of inputs, valuations and processes, from future fuel prices to the value of a statistical life. To ensure that CBAs for different investments can be compared and ranked consistently, there are often national guidelines for appraisal methods, inputs and parameters. In contrast, principles for timetable construction are surprisingly neglected. The ultimate objective of this paper is to start a discussion about such principles.

In Section 2, a simple model is formulated, which captures the main features of applied transport CBA. The model is a simplified version of current Swedish appraisal practice and guidelines. Section 3 shows, by way of examples, why the social benefits of a railway investment cannot be defined without a principle for choosing timetables before and after the investment, and that this opens the door for both honest confusion and strategic behaviour by the analyst. Section 4 discusses advantages and disadvantages of different timetable principles. Section 5 discusses the (implicit) principles for timetable construction in current appraisal practice, and what consequences these principles are likely to have. Section 6 concludes.

1.1 Literature and practice

Despite the obvious importance of timetables in appraisal, we have failed to find any mentioning in appraisal guidelines of how analysts should construct timetables. For example, the extensive guideline RAILPAG (Railway Project Appraisal Guidelines), intended to be ”a common framework for the appraisal of railway projects across the EU” does not mention how timetables should be designed, what stakeholder should have the responsibility, or how the appraisal result might be affected by it. Neither of the national appraisal guidelines of Sweden, Norway, UK or the Netherlands mentions the topic, despite having detailed formal appraisal guidelines covering virtually all other parameters and dimensions.

It seems that the practice in most railway appraisals is to treat the timetable as an exogenous variable outside the analyst’s control or responsibility, much like the future oil price or population growth. This is strange for two reasons. First, in order to make appraisals of different investments comparable, appraisal guidelines devote substantial efforts to forecast and harmonize assumptions regarding virtually all variables entering the CBA, from future oil prices and population growth to transit fares and vehicle operating costs. The timetable, on the other hand, is left to the analyst’s discretion. Second, the timetable is not really an exogenous variable: it is to a considerable extent a policy instrument under at least partial public control through capacity allocation, track charges, and public provision or subsidies of public transport. Thinking about the timetables with and without an investment is an integral part of designing and evaluating the investment, and this should be made explicit.

The research literature on the subject is scarce. In most published railway appraisals, the timetable seems to arrive by fiat. One of few examples investigating the importance of the timetable is Adler, Pels & Nash (2010), who study how the competition between airlines and high-speed rail may affect service levels, and how this affects the benefits of high-speed rail investments. Bristow, Preston & Nash (1998), discussing appraisal principles in the partially deregulated British railway industry, touch upon related issues, such as whether investment CBAs should include user benefits or only profits and non-user benefits. Nash & Preston (1991) discuss what principles regarding future use of railway investments should be assumed, in the context of a discussion about whether financial or social investment criteria should be used in railway investment decisions. Nash (1992) discusses the interactions between investments in the railway system, both between different track investments (which may be complements or substitutes) and between track investments and investments in rolling stock.

Errors in demand forecasting are often caused by wrong assumptions about future supply, and future timetable assumptions seem to be a large culprit. Tegnér (2001) concludes that the optimistic timetable (high frequencies, low fares, short travel times) used in the demand forecast for the Ostkustbanan railway line was a major reason for a fourfold overprediction of increase in demand. SIKA (2000) found that the actual number of travellers on three studied Swedish railway lines was overpredicted with between 26% and 72%, quoting wrong timetable assumptions as one of the reasons for the demand shortfall. Olsson (2006) underlines the integration of timetable assumptions and investments, concluding that the estimates for travel time, frequency, punctuality and number of travellers that an investment will generate relies on the realized timetable. Mackie & Preston (1998) list twenty-one sources of errors in appraisal, among them “Incorrect definition of the base and investment cases”, “Errors in planning assumptions” and “Transport inputs incorrect”, all of which can be viewed as examples of wrong assumptions of the future timetable.

The subject of this paper is timetable principles for investment appraisal. A related issue is construction of socially efficient timetables. Somewhat surprisingly, the literature and methodology on this topic is also scarce. There is a developing literature concerning different aspects of railway operations optimization, such as crew and vehicle scheduling and train routing. This literature typically uses engineering objectives such as minimizing total operations cost. Caprara, Kroon, Monaci, Peeters & Toth (2007) give an overview of the rail operations research literature, including problems such as line planning, timetabling, platforming, rolling stock circulation, shunting and crew planning. Kroon & Peters (2003) present an algorithm to account for variable trip times in the timetable design problem. Vansteenwegen & Oudheusden (2006) present an algorithm to construct optimal buffer times in a general railway network with connections; the central trade-off is that low buffer times increase the risk of missed connections, while high buffer times increase travel times. Claessens, Van Dijk & Zwaneveld (1998) consider the problem of cost optimal railway line allocation, minimizing operating costs while meeting constraints on passenger flows and service capacity. There are almost no studies focusing on optimization of strategic timetable design using social welfare as the objective function. A rare example of an economical objective function is Brännlund, Lindberg, Nõu & Nilsson (1998) present an algorithm that schedules a set of trains to obtain a profit-maximizing timetable while not violating track capacity constraints.

The literature on strategic appraisal of investments and policy measures, including fields such as demand modelling and cost-benefit analysis, is virtually disjoint from the railway operations research literature. Compared to the welfare economics-oriented appraisal tradition, the railway operations research tradition usually handles demand as exogenous or with simplistic models, and if passenger welfare is taken into account into the problem, it is handled in a comparatively crude and heuristic fashion (such as maximizing the number of passengers able to reach their destination without an interchange), compared to the detailed welfare measures usually used in economically and behaviourally based appraisal. The strategic appraisal literature, on the other hand, seldom deals with the details of capacity or timetables. Our hope is that the strait between those two traditions can be bridged.

2 THE MODEL

In this section, we set up a simple framework for cost-benefit analysis with a demand and a supply side, capturing the essential features of a transport CBA. The framework and its parameters is a simplified version of current Swedish guidelines, so the examples are perfectly realistic. The simplifications are made to avoid obscuring the key points of the analysis. Only consumer and producer surplus are included, omitting external effects, taxes and charges. The examples are as homogeneous as possible: they consider only one class of travellers and one ticket price.

Starting with the demand side of the model, the timetables are simplified down to the essentials: train frequencies and travel times. This is how timetables are usually characterized in strategic demand modelling and appraisal: frequencies and travel times are assumed to be homogeneous during a certain time period S that is analysed (e.g. rush hours or a weekday). More detailed demand modelling and appraisal uses schedule-based demand modelling, rather than frequency-based, but for our purposes, this would just add complexity to the exposition without adding further insights.

Consider a line segment between two stations of length L. Let n_i be the total number of trains of train type i in both directions traversing this line segment during the analysed time period S. Symmetry is assumed, implying that the number of trains in each direction is $n_i/2$. All trains of type i have travel time t_i and headway $2S/n_i$. For simplicity, the fare p_i is assumed to be constant and exogenous[1]. The generalized travel cost is defined as $c_i = p_i + \alpha t_i + \beta * 2S/n_i$,

[1] This is consistent with appraisal practice, although often not realistic. In this paper we will assume a demand function with constant elasticity ε: $D(c) = D^0 \left(\frac{c}{c^0}\right)^{\varepsilon}$. Then the optimal price becomes p* = $(\gamma_2 t+\gamma_3 L)\varepsilon/(1+\varepsilon)$, so it is independent of both the number of trains n (as long as n is independent of t) and initial demand D^0. With for example a linear demand function, this is no longer true.

where α and β are monetary values of travel time and waiting time, respectively. Travel demand between the stations is D_i = $D(\mathbf{c})$. The consumer surplus (CS) of a change in generalized costs from $\{c_i^0\}$ to $\{c_i^1\}$ is by definition:

$$CS = \sum_i \int_{c_i^0}^{c_i^1} D(c)dc \approx \sum_i ½[D(c_i^0) + D(c_i^0)](c_i^0 - c_i^1)\,. \qquad (1)$$

In the numerical examples, we will use the rule of a half approximation, which is valid for small changes $(c_i^0 - c_i^1)$.

The producer surplus (PS) is defined as fare revenues minus operations costs. Operations costs increase with the number of trains, the number of passengers, travel distance and travel time. We apply a cost function used in Swedish appraisal, where operations costs are assumed to be linear in train operating time ($n_i t_i$), passenger time ($D_i t_i$) and passenger distance ($D_i L$). The producer surplus of a change of travel times and the number of trains from situation 0 to situation 1 is then:

$$PS = \sum_i (D_i^1 - D_i^0)p_i + \gamma_1(n_i^0 t_i^0 - n_i^1 t_i^1) + \gamma_2(D_i^0 t_i^0 - D_i^1 t_i^1) + \gamma_3(D_i^0 - D_i^1)L \qquad (2)$$

where γ_1-γ_3 are parameters. Total social benefits are defined as CS + PS.

On the supply side, the travel time is a function of the minimal travel time, infrastructure capacity and the number of trains. The capacity relationships can be modelled in different ways, depending on the context and the need for details. In one extreme, real, detailed timetables are constructed, which are then analysed graphically or through simulation. In the other extreme, aggregate, analytical relationships based on the capacity relationships are used[2]. Just as the demand side, the aggregate approach is usually used for strategic planning and appraisal.

We use a capacity relationship used by the Swedish Transport Administration for strategic planning and appraisal[3], similar in structure to capacity relationships used in other countries. The consumed capacity C is the fraction of the analysed time period S when the line segment is occupied by trains. Each line segment is divided into homogenous track segments k, which are the smallest building blocks of the network. C is calculated for each track segment on the line segment, and the highest C is applied to the whole line segment. For single tracks, C is defined as

$$C = \sum_i (\max_k (n_i T_{ik} + \rho n_i M_i)/S), \qquad (3)$$

[2] The same trade-off between detail/realism and ease-of-use/analytical tractability exists for road capacity relationships, where there are also both dynamic and static volume-delay relationships.

[3] This relationship is in fact not very realistic: it is too generous in the sense that it underestimates how travel times increase at high capacity utilization. Nevertheless, it illustrates the points we wish to make. Besides, it is of practical importance, since it is the relationship actually used in Swedish practice.

where M_{ik} is a time addition for meetings, depending on train type i, T_{ik} is the minimal travel time of track segment k for train type i and ρ is the probability that two trains meet. The consumed capacity is thus the fraction of the time period S during which a train occupy the segment, including the waiting time for meetings. The scheduled travel time for trains of type i on the line segment, t_i, is then computed from the minimal travel time of the line segment, T_i, track length L and C:

$$t_i = T_i + max[(0.2*C - 0.06)L,0], \quad (4)$$

where L is given in kilometres and T_i is given in minutes.

3 NUMERICAL EXAMPLES

3.1 One train type

Consider a 100 km single track line segment between two stations, divided into 10 track segments of equal lengths, with meeting stations between each track segment. Parameter values are adapted from the Swedish appraisal guidelines and given in Table 1.

Table 1. Parameters of the CBA framework used in the examples.

Value of in-vehicle time α	70 SEK/h
Value of waiting time β	70 SEK/h*
γ_1	89.1 SEK/min/train
γ_2	0.266 SEK/min/pass
γ_3	0.102 SEK/km/pass
Demand elasticity	-0.7 $[D(c) = D^0\left(\frac{c}{c^0}\right)^{-0.7}$ where D^0 and c^0 is the initial situation]

**According to this guideline, the value of waiting time is the same as the value of in-vehicle time for waiting times under 15 minutes, and lower for longer waiting times.*

Assume that there is only one train type, with a minimum travel time of 40 minutes on the line segment. Assume that there is initially n=4 trains/hour passing between the two stations (2 trains/hour in each direction), carrying 1000 passengers/hour. The probability that a train meets another train at any given track segment is[4] 0.27. With M=5 minutes, we have C=0.36, implying a total timetable travel time of 41.11 minutes (using (3) and (4))[5].

It is suggested that the frequency should be increased to 7 trains/hour to improve the service and increase demand. To avoid increased travel times, the capacity on the track must be increased. It is thus suggested that the number of meeting stations is doubled, making the track segments 5 km long. The net social benefit of this capacity increase, given a frequency of 7 trains/hour, turns out to be nearly 23 000 SEK/hour (details in Table 2).

[4] $\rho = 4 \cdot \frac{\frac{40}{60}}{10} = 0.27$

[5] $C = \max_k \frac{n \cdot T_k + \rho \cdot n \cdot M}{T} = \frac{4 \cdot 4 + 0.27 \cdot 4 \cdot 5}{60} = 0.36$ and $t = T_i + max[(0.2 \cdot C - 0.06)L, 0] = 40 + max[(0.2 \cdot 0.36 - 0.06) \cdot 100, 0] = 41.11$ minutes.

Another analyst, however, suggests that the relevant do-nothing scenario is the current frequency of 4 trains/hour, and that this should be compared to the investment scenario with 7 trains/hour. In this analysis, the investment only yields around 7 000 SEK/hour, less than a third of the social benefits in the first analysis (details in Table 3). A third analyst suggests that the investment should be evaluated keeping the frequency fixed at 4 trains/hour, arguing that this is the pure investment benefit. This reduces social benefits by almost a half to less than 4 000 SEK/hour.

Table 2. Example 1, appraisal A.

	Base	Investment	Difference
Trains/hour	7	7	0
Travel time (minutes)	48.8	41.4	7
Passengers/hour	985	1085	-100
Generalized cost (minutes)	67	58	9
Consumer surplus (SEK/h)			7300
Producer surplus (SEK/h)			15484
Total social benefits (SEK/h)			22784

Table 3. Example 1, appraisal B.

	Base	Investment	Difference
Trains/hour	4	7	-3
Travel time (minutes)	41.11	41.0	0
Passengers/hour	1000	1085	-85
Generalized cost (minutes)	65	58	7
Consumer surplus (SEK/h)			7429
Producer surplus (SEK/h)			4421
Total social benefits (SEK/h)			11851

The question of which of these analyses is correct cannot be settled without a principle determining the timetables with and without the investment. Two other possible principles would be to assume that frequencies should be set to maximize social benefits or operator profits in both scenarios.

Table 4 summarises investment benefits under different timetable assumptions. Appraisal A in the first row assumes that train frequencies are increased regardless of whether the investment is built. This is common in applied appraisal practice: a planned target timetable is assumed, and investments are evaluated assuming this timetable. Appraisal B in the second row compares the current situation with the planned situation. This is another reasonable alternative, although it potentially adds the effects of the investment as such with the effect of the timetable change as such. Appraisal C evaluates the investment at the current frequency, thus separating the frequency effect from the investment effect, although obviously not using the new capacity to its full extent. Appraisal D and E assume that train frequencies are set to maximize operator profits and social benefits, respectively, in both scenarios.

Table 4. Investment benefits assuming different timetables.

		CS	PS	CS+PS	Principle
A.	n^0=7, n^1=7	7300	15484	22784	Planned frequency
B.	n^0=4, n^1=7	7480	-332	7148	Current vs. planned frequencies
C.	n^0=4, n^1=4	1305	2513	3818	Current frequency
D.	n^0=3.5, n^1=5.7	7992	3691	11682	Profit-maximizing frequencies
E.	n^0=4.0, n^1=6.0	7429	4421	11851	Welfare-maximizing frequencies

We will explain in some detail why so different conclusions can be obtained. Figure 1 shows consumer surplus (CS), producer surplus (PS) and total social benefits (CS+PS) as a function of the number of trains in the initial situation without investment. All benefits are compared to the reference situation with *n*=4 trains/h.

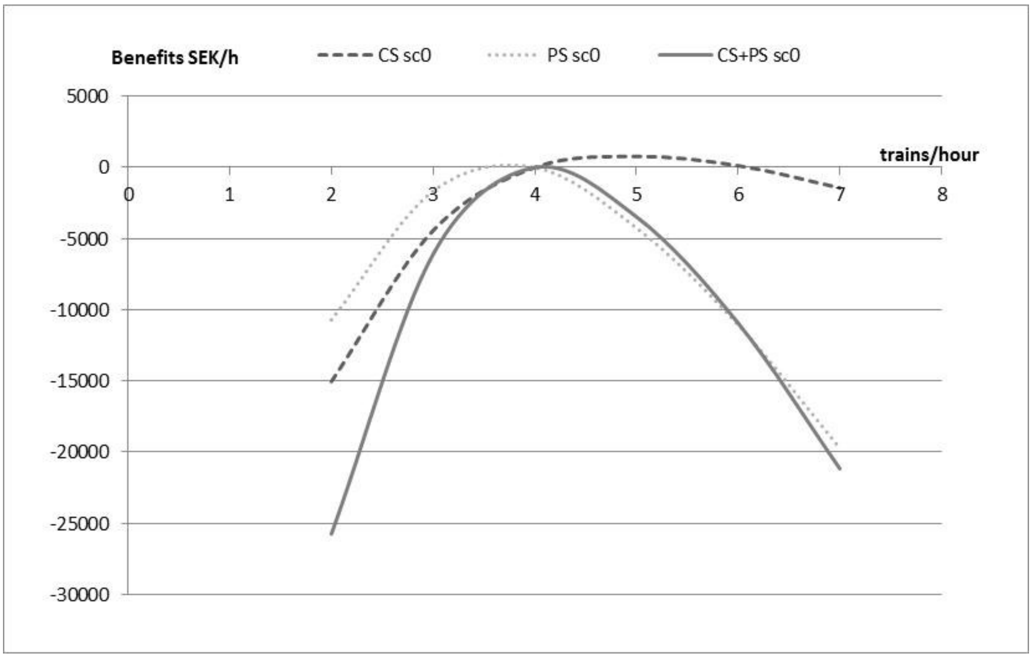


Figure 1. Example 2: Consumer surplus, producer surplus and total benefits as functions of the number of trains, initial situation without investment

The CS increases with the number of trains up to 5 trains/h, driven by declining waiting time. Above 5 trains/h, the CS declines with increasing number of trains because the travel time increases due to capacity constraints. The PS has a maximum at 3.5 trains/hour; the fare revenues increase slower than operating costs after that point. Total benefits PS+CS attain a maximum at 4 trains/h. We can note in passing that the profit-maximizing optimum is different from the social optimum; a commercial operator trying to maximize profits will run fewer trains than what is socially optimal. This is one of the arguments for society providing, regulating or subsidizing public transport.

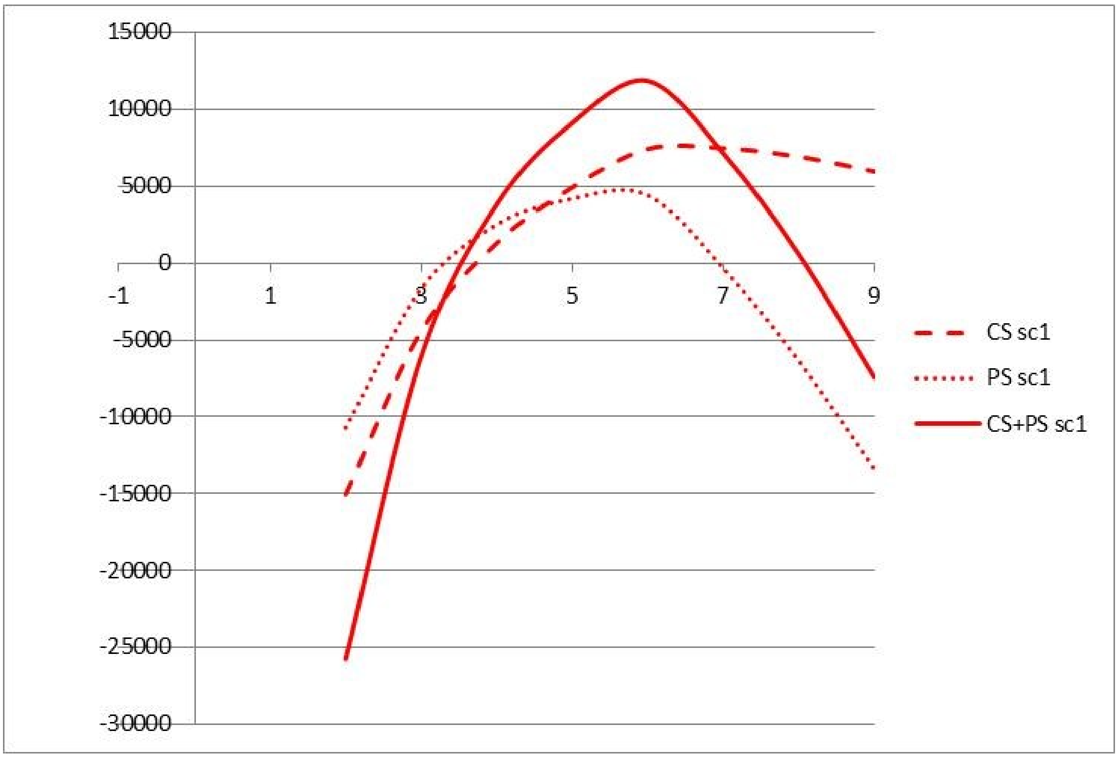
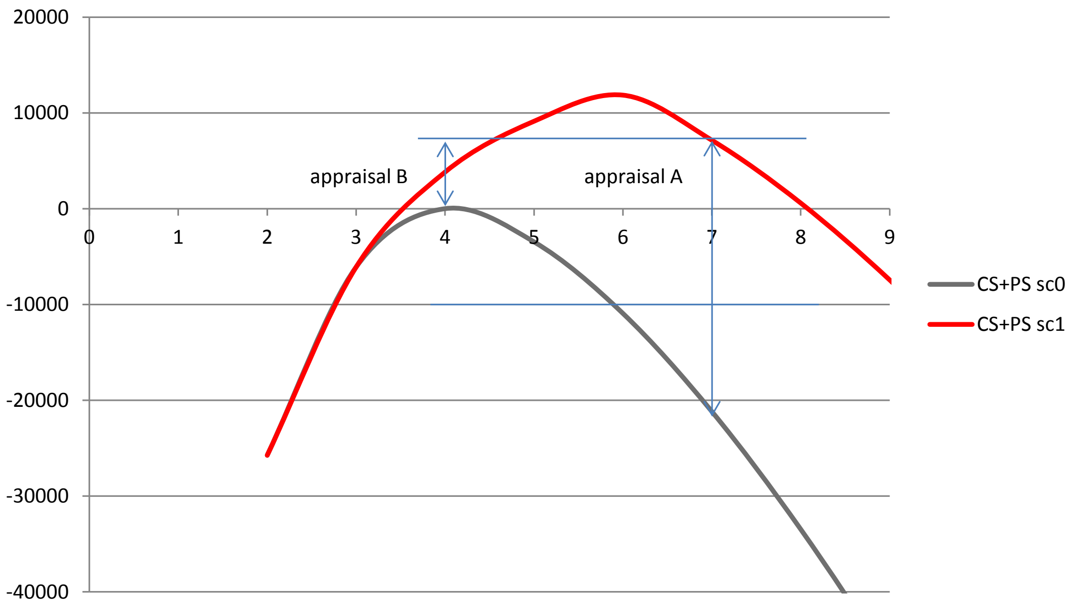


Figure 2. Example 2: Total benefits as functions of the number of trains, with and without investment; the two arrows show two CBAs.

Figure 2 corresponds to Figure 1, but shows the benefits in the investment scenario. The PS has its maximum at 5.7 trains/hour and the total benefits CS+PS has a maximum at 6.0 trains/hour. The investment increases both CS and PS compared to the reference situation.

Figure 2 shows total benefits with and without the investment, showing Appraisal A and Appraisal B as arrows. The key point is that the net benefit of the investment depends on what is assumed about the number of trains before and after the investment, n^0 and n^1. In Appraisal A, the frequency in the do-nothing scenario is far from the social optimum, causing the appraisal to show large benefits. In Appraisal B, the timetable happens to be close to socially optimal in the do-nothing scenario, hence shows much smaller benefits than Appraisal A.

In this simplistic example, this is intuitively obvious. But in a real example, these effects are effectively impossible to see through, especially for a third-party observer. An appraisal might even show negative benefits, if the timetable is further from the optimal in the investment scenario than in the do-nothing scenario. For example, this happens if one assumes 4 trains/hour without the investment and anything above 8 trains/hour with the investment. We are aware of such phenomena happening in practice, confusing the analysts involved.



The appropriate timetable principle depends on the institutional context. If there is an institutional setting when the number of trains should be determined based on welfare maximization, it seems apparent that these frequencies should be applied when evaluating the investment (4 trains/hour in the do-nothing scenario and 6 trains/hour in the investment scenario). If there is an institutional setting where the producer benefits are maximized, the number of trains that optimize the producer surplus should be applied in the appraisal. Appraisals D and E happen to be close to each other and to Appraisal B, but this needs not be the case.

It should be emphasized that it is impossible for an outside observer, and often also for the analyst, to see through these dependencies when the appraisal is presented with just one set of timetable assumptions in the do-nothing and the investment scenarios.

Capacity utilization does not only affect travel times but also delays, since high capacity utilization increases knock-on delay effects. In practice, this is partly compensated for by adding buffer times to the timetable travel times. We have not included delay effects in the example, but analytically we can interpret the travel times in the example as if they include buffer times and average delays (valued by an appropriate multiplier). In the above example we have assumed capacity constraints, but it is easy to show that the analysis is sensitive to timetable assumptions also where there are no capacity constraints, and the investment only reduces travel time. For capacity improvements the analysis is, however, particularly sensitive since the added capacity can allow for different improvements: travel times, frequencies or reliability.

3.2 Two types of trains

Next, assume that there are two operators with different objectives, and infrastructure is funded by the government. This example aims at illustrating how the real timetables result from the process of negotiation between actors within a given institutional settings, and that the operators may have incentives to state a future timetable different from their intended timetable. Based on this example we will then discuss what principle for timetable assumptions in appraisal that is appropriate, depending on the institutional setting.

Assume that there are two train types competing for capacity: regional trains and long-distance trains. The two demand functions for the train types are completely separable; we can imagine that the two trains serve separate markets, and only happen to share this particular stretch of track. Regional trains are operated by a public operator that strives to maximize consumer surplus (the regional government covers operating deficits), while long-distance trains are run by a commercial operator trying to maximize profits. The institutional setting is such that the regional train operator first decides the number of trains it will run, and then the long-distance train operator decides the number trains it will run, given the stated intentions of the regional train operator. The government funds infrastructure investments, but does not take into account that the timetables assumed for the do-nothing and the investment scenarios affect the outcome of the appraisal. The operators, however, are fully aware of the importance of the timetables used in the appraisal scenarios.

Assume that regional trains are slower, taking 45 minutes to traverse this track if there are no capacity constraints, while long-distance only need 25 minutes. In the current situation, there are 3 regional trains per hour with a total of 1000 passengers, and 3 long-distance trains per hour, also with 1000 passengers. The other parameters are unchanged from the example with a single train type. However, in this example the longest track segment on the 100 km single track line segment between two stations is 15 km.

Now, the regional train operator wants to increase capacity by increasing the number of meeting stations such that the longest track segment decreases from 15 km to 3 km, and extend the timetable to 6 trains/hour. Given 6 regional trains/h, the long-distance operator would maximize profits by running 5 trains/hour in the new situation. The first row of Table 5 shows the benefits of the investment with the timetables that will be the actual outcome in the do-nothing and the investment scenarios: 87 kSEK/h.

However, since the government does not take into account the impact of the timetables in the appraisal, the operators have both incentive and opportunity to act strategically when stating their intended timetables, which are inputs to the appraisal. Since the consumer surplus and the producer surplus both increase when the investment is built, the operators support the capacity investment and have incentives to increase the CBA benefits, increasing the probability that the government decides to build the investment.

The operators know what timetable that gives the maximal social welfare in the investment scenario (3 regional and 5 long distance trains), so they can act strategically by stating that they intend to run this number of trains if the investment is built. This yields a benefit of 163 kSEK/h (second row of Table 3), almost twice the benefit of the actual outcome. In the investment scenario the timetable maximizes the social welfare, but in the do-nothing scenario the train frequency is higher than the optimal.

The operators can also act strategically by stating (truthfully) their intentions to run 6 regional trains/h and 5 long-distance trains/h with the investment. However, they state that they also intend to run this timetable even if the investment is not built. Then the benefit would increase further to 306 kSEK/h.

The resulting benefit obviously depends on the institutional setting. Another possible setting can be that there is one profit maximizing operator that runs both train types. The realized timetable in the do-nothing and investment scenarios would then be the ones that maximize the total PS, yielding an invesment benefit of 115 kSEK/h. A third possible setting is that both train types are run by a public authority, maximizing social welfare. This would give 2 regional and 3 long distance trains in the do-nothing scenario, and 3 regional trains and 5 long distance trains in the investment scenario, yielding an investment benefit of 101 SEK/h. In any of these cases, the operator(s) can easily act strategically by stating intended timetables that increases investment benefits.

Table 5. An example with two types of trains (regional and long-distance).

	Without investment	With investment	Total benefits (kSEK)
	(R=regional trains, L = long-distance trains)		
Actual outcome: the regional operator maximizes CS first. The long-distance operator maximizes PS, given number of regional trains.	3R,3L	6R,5L	87
Operators state that they will run the timetable that is socially optimal in the investment scenario.	3R,3L	3R,5L	163
Actual timetable outcome in the investment scenario applied in both the do-nothing and investment scenarios.	6R,5L	6R,5L	306
Examples of principles for choice of timetables			
Welfare maximizing	2R,3L	3R,5L	101
Profit maximizing	1R,3L	2R,6L	115

4 WHICH PRINCIPLE SHOULD BE USED?

The above examples illustrate that the benefit of an investment is not defined without a principle defining timetables with and without the investment. The example also illustrates the problem of assuming timetables based on the stakeholder's stated intentions. Even if it was possible to use binding contracts over 10-20 years, forcing operators to run the planned timetable in the forecast year, the problem of assuming a timetable for the do-nothing scenario remains. The same problem arises when there is a future target timetable, which is the case in some countries. The target timetable is designed to be consistent with the investment scenario, but much less effort is usually spent on the timetable in the do-nothing scenario. Hence, the benefit of the investment is still somewhat arbitrary.

An explicit principle for choosing timetables in the do-nothing and the investment scenarios is also necessary to ensure comparability of different suggested investments. The timetables must be derived taking into account demand and supply as well as the prevailing institutional settings and capacity constraints. Like all cost-benefit calculations this principle should be transparent and possible to verify by a third party.

4.1 Current practice

The process of timetable construction in appraisal varies between countries using cost benefit analysis for transport investment decisions, but there are also many similarities. To give an example, the current practice in Sweden can roughly be summarized as follows. Starting from current timetables, the plans and wishes from operators regarding traffic changes are collected, and needs for capacity improvements are identified. Based on a judgment of what capacity improvements are reasonable, operators' plans are modified, and a future overall timetable is constructed, assuming these capacity improvements. This is (somewhat confusingly) called the 'baseline timetable'. In most investment appraisals, the baseline timetable represents the investment scenario, while the do-nothing scenario is constructed by removing the investment, keeping the number of trains in the baseline timetable fixed, and increasing their travel times to meet capacity constraints.

Sometimes the current timetable is used in the do-nothing scenario and a planned timetable in the investment scenario. Although this seems like a better practice than to use the same timetable in the two scenario, this too, opens up for the possibility for conscious or unconscious strategic behaviour from regions, operators or other stakeholders wanting to promote certain investments and discredit others, perhaps in competition for national investment funding. The stakeholder may state (and perhaps honestly believe) that given the investment, the plan is to run a timetable that will maximize social benefits, although this timetable will not actually be the outcome because of the prevailing institutional setting. This will also exaggerate benefits in the investment scenario, since the current timetable is usually not socially optimal – especially as evaluated by the formal CBA framework

In current practice, it is seldom or never analysed how total benefits differ if the increased capacity is used to improve travel times, frequencies or reliability. Usually, it is up to the analyst to decide how the timetables should differ between the do-nothing and the investment scenarios. An example: during the preparation of the recent National Transport Investment Plan (covering 2010-2021), 12 capacity-improving investments were analysed in the Greater

Stockholm area. The added capacity was used to improve frequencies in two cases, to improve travel times in three cases, to improve both frequency and travel times in three cases, and to improve reliability in four cases. (As to reliability improvements, there are no quantitative relationships at all, so the size of these effects is left to the analyst's judgment.) In none of these cases was it analysed how benefits would have differed if the capacity improvement had been used differently.

4.2 Which principle?

So what transparent and verifiable principles for choosing the timetables in appraisal are appropriate in different institutional contexts? The institutional context varies between countries; the example in Section 3.2 is inspired by the Swedish context. Regional trains (commuter trains) are run by the regional governments (through public transit operators), long-distance trains are operated by a profit-maximizing but government-owned near-monopoly operator, and freight trains by several profit-maximizing companies (with one dominating firm). The track allocation process is far from transparent but it is clear that the regional train operators have a strong bargaining power. The most appropriate may then be to assume a simplified model of the track allocation process, such as the one in the example, and assume the calculated actual outcome timetables in the do-nothing and in the investment scenario. However, track capacity allocation is actually supposed to be based on maximizing social welfare, albeit this principle has been difficult to operationalize. This could motivate assuming that train frequencies are determined to optimize social welfare, under the constraint that long-distance and freight operators will not run more trains than what maximizes the profit.

In countries where the railway is vertically integrated, infrastructure investments and timetable are more integrated decisions. In settings like this the timetable in the investment scenario is probably well planned and close to the social optimum. Hence, the principle to optimize the social benefit seems appropriate in this institutional setting. This principle may also be the most appropriate approximation in many countries where the railway system is regulated and subsidised by society in several ways, including public provision or subsidisation of public transport, subsidised track access charges, public control over capacity allocation etc. The underlying idea would then be that the investment appraisal should capture the maximal social benefit that the infrastructure can generate and that society is trying to achieve maximal social welfare through its combination of regulation, subsidies and traffic provision. If traffic supply is decided mainly by profit-maximizing train operators, with monopolistic power[6], it would be natural to let the timetable be decided by maximal producer surplus.

[6] Choosing profit maximization as the principle with which to construct timetables makes the practice to assume constant and exogenous fares even more suspicious than usually, so this should then be modified.

4.3 Discrepancy with the realized timetable

There will as always be a discrepancy between the model and the reality. A criticism of applying a model for timetable assumption is therefore that since the CBA-model is a simplified representation of reality, omitting or misrepresenting certain effects to some extent, the model computed timetable will not necessarily be identical to the later realized timetables. This is, however, not unique to the timetable assumptions, but to the entire cost-benefit analysis. When carrying out appraisals, we are working under the assumption that the CBA and model assumptions are sufficiently accurate representations of reality, in the sense that the CBA-ranking of investments or policy measure corresponds to the true ranking. Hence, even if the timetable computed according to the most appropriate principle, given the institutional setting, does not correspond to the actual outcome, it is necessary to stick to a single predefined principle.

If not sticking to a principle, this opens up a particularly insidious comparison: comparing the actual, current timetable (that can be observed), with the timetable that gives the highest benefits after the investment (in case the principle to maximize welfare is applied). Few people would then argue against such a comparison – after all, it seems natural to use the actual timetable as a reference. But even if the current timetable is well-designed in reality (for the sake of argument), it may not be optimal when evaluated in the formal CBA framework. Then, if the analyst chooses a timetable in the investment scenario which is optimal when evaluated by the CBA framework – and why shouldn't she? – the benefits will be exaggerated. Even an investment that does not improve anything will show a benefit according to such a CBA. This shows that it is necessary to stick to the same CBA framework and timetable principle both in the investment and the do-nothing scenarios.

5 CONCLUSIONS

The social benefit of a railway investment cannot be defined without specifying timetables with and without the investment. Without an explicit principle for designing these timetables, there is a risk for arbitrariness, conscious or unconscious strategic behaviour when conducting the appraisal, counterintuitive appraisal outcomes, and a lack of comparability between suggested investments. This applies to all types of railway investments, but it is especially conspicuous and potentially insidious for capacity improvements, since the added capacity can be used in several ways – improved travel times, frequencies or reliability.

The conclusion that the timetables are important is in one sense almost trivial, but the awareness about the extent of the issue seems to be limited. Going through research literature and guidelines for railway appraisal, we have found no more than mentioning *en passant* of this issue or what principles should govern the design of timetables. Sometimes it is mentioned that it is important to have a good timetable for the planned investment to make the analysis reveal the highest possible benefit, but not that the timetable in the do-nothing scenario is equally important for the analysis.

Current Swedish practice, which we believe is fairly typical, is to collect operators' plans and wishes for future traffic, specify the investments needed to accommodate these plans (or most of them), and construct a timetable based on this. The same number of trains is usually assumed both with and without the investment, but travel times in the investment scenario are longer. Different timetables are seldom or never tested. In other counties, where the railway is more vertically integrated, the infrastructure and timetables are decided together but the same basic problem remains: a much greater effort is usually spent finding the appropriate timetable in the investment scenario than in the do-nothing scenario.

We have illustrated how the appropriate timetable principle depends on the institutional setting. One principle is to forecast the timetable that will be the outcome of the process of negotiation between operators and the state, taking into account demand and supply as well as the prevailing institutional settings and capacity constraints. In some contexts, assuming socially optimal timetables might be a reasonable approximation of what society tries to achieve through regulations, policies, public transport provision, capacity allocation, pricing, subsidies etc. In other contexts, profit-maximizing timetables may be a more realistic approximation. There are also institutional context where the appropriate principle is a mix of the above two.

Whatever principle is chosen, a process is needed to systematically explore the social benefits of different timetables. To do this, current appraisal and modelling methods have to incorporate railway capacity analysis from the railway operations research tradition. At a minimum, one should explore how different timetables (in the do-nothing and the investment scenario) affect the total benefits of an investment.

Since the cost-benefit analysis is a simplified representation of reality, omitting or misrepresenting certain effects to some extent, the CBA-optimal timetable will not necessarily be the reality-optimal timetable. But when carrying out appraisals, we are working under the assumption that the CBA is a sufficiently accurate representation of reality, in the sense that the CBA-ranking of investments or policy measure corresponds to the true ranking. A similar assumption must be made for the evaluation of timetables. If it turns out that the socially optimal timetable is clearly unrealistic, this would mean that we are working with an unrealistic CBA framework, that is not a sufficiently accurate representation of reality to base judgments or ranking on.

Railway investments often represent huge public spending commitments, and hence it should be uncontroversial that decision-makers need reliable appraisal methods for comparing suggested investments against each other and against other uses of public resources. In this paper, we have argued that having an explicit, verifiable principle for timetable construction is necessary for CBA to be meaningful. Without such a principle, and a way of verifying that it is applied, CBA for railway investments will be a pointless exercise. In worst case, it becomes an instrument for lobbyist wolves dressed in the sheepish clothes of transport economists.

6 ACKNOWLEDGMENTS

Thanks to Lena Wieweg for fruitful discussions, and to Chris Nash and Jan-Eric Nilsson for useful comments on an earlier draft of the paper. Lennart Lennefors and Magnus Wahlborg at the Transport Administration helped explaining the capacity relationships used by the Transport Administration. Partial funding from the Transport Administration is gratefully acknowledged.

7 REFERENCES

Adler, N., Pels, E., & Nash, C. (2010). High-speed rail and air transport competition: Game engineering as tool for cost-benefit analysis. *Transportation Research Part B: Methodological*, *44*(7), 812–833.

Bristow, A. L., Preston, J. M., & Nash, C. A. (1998). Investment planning and appraisal issues in the privatized railway — the British experience. *Transport Reviews*, *18*(4), 353–362.

Brännlund, U., Lindberg, P. O., Nõu, A., & Nilsson, J.-E. (1998). Railway timetabling using Lagrangian relaxation. *Transportation Science*, *32*(4), 358–369.

Caprara, A., Kroon, L., Monaci, M., Peeters, M., & Toth, P. (2007). Passenger Railway Optimization. *Barnhart and Laporte (eds): Handbooks in Operations Research and Management Science* (Vol. Volume 14, pp. 129–187). Elsevier.

Claessens, M. T., Van Dijk, N. M., & Zwaneveld, P. J. (1998). Cost optimal allocation of rail passenger lines. *European Journal of Operational Research*, *110*(3), 474–489.

Kroon, L. G., & Peeters, L. W. P. (2003). A Variable Trip Time Model for Cyclic Railway Timetabling. *Transportation Science*, *37*(2), 198–212.

Mackie, P., & Preston, J. (1998). Twenty-one sources of error and bias in transport project appraisal. *Transport Policy*, *5*(1), 1–7.

Nash, C. (1992). Appraisal of rail projects. *Project Appraisal*, *7*(4), 211–218.

Nash, C., & Preston, J. (1991). Appraisal of rail investment projects: recent British experience. *Transport Reviews*, *11*, 295–309.

Olsson, N. O. E. (2006). Impact Analysis of Railway Projects in a Flexibility Perspective. *Transport Reviews*, *26*(5), 557.

SIKA. (2000). *Uppföljning av investeringar [Ex-post assessment of investments.]* (No. 2000:11). SIKA report. SIKA. Retrieved from http://www.sika-institute.se/

Tegnér, G. (2001). *Ostkustbanan – modell och verklighet. [The Ostkustbanan railroad - forecast and reality.]* (No. 2001-04-30). Transek report. Transek AB.

Vansteenwegen, P., & Oudheusden, D. V. (2006). Developing railway timetables which guarantee a better service. *European Journal of Operational Research*, *173*(1), 337–350.

8 APPENDIX: FINDING OPTIMAL TIMETABLES

In the most detailed, realistic case, finding a socially optimal timetable is extremely difficult, primarily because of the complicated nature of capacity relationships. But in the context of strategic appraisal, things are a little simpler. First, the timetables are usually characterized only by travel times and frequencies between each pair of stations, and these are also organized along lines. Second, it is usually enough to study variations of the baseline timetable, where only a few lines are affected – finding a global optimum may not be necessary for our purposes.

The difficult part of the problem is to specify a useable relationship between capacity utilization and travel times (and delays, if these are to be treated explicitly rather than through buffer times or similar proxies). This problem is beyond the scope of this paper. There are several possible ways to combine analytical and simulation-based methods to specify such relationships, especially if they can be tailored to the application at hand, but this is clearly a field for future research.

Given a technical capacity relationship, the problem of finding a socially optimal timetable is simple to formulate. To simplify notation, we will pretend there is only one pair of stations, one type of trains, and one type of travellers. The extension to several pairs of stations and to several types of travellers and trains is trivial. We also, just as before, omit taxes, charges and external effects – this is also trivial to add. Given the initial frequency n^0, the problem is to find the optimal frequency n. The problem can then be written:

$$\max_n \left\{ \int_{c^0}^{c} D(s)ds + Q - Q^0 \right\}$$

Technical constraints:

$c^0 = b^0 + \alpha t^0 + \frac{\beta}{2n^0}$	Generalised costs
$c = b + \alpha t + \frac{\beta}{2n}$	
$t^0 = \mathrm{G}(n^0)$	Capacity relationships
$t = \mathrm{G}(n)$	
$D^0 = D(c^0)$	Demand functions
$D = D(c)$	
$Q^0 = b^0D^0 - \gamma_1 n^0 t^0 - \gamma_2 D^0 t^0 - \gamma_3 D^0 L$	Fare revenues – operational costs
$Q = bD - \gamma_1 nt - \gamma_2 Dt - \gamma_3 DL$	

The demand function D(c) may be of different kinds, from simple elasticity relationships to large-scale multimodal demand models. In some cases it may be sufficiently realistic to assume constant demand. Demand modelling is a well-developed area, so even if there are practical difficulties with acquiring data and striking a balance between detail and ease-of-use, there is no principal difficulties to construct a working demand relationship. It should be noted that for the problem to work, the demand can only depend on the entire generalized cost, not on its individual components. In other words, there must be consistency between the relative valuations used in the demand model, and the valuations used for appraisal. We stress this since such consistency is usually not the case in applied appraisal practice.

The capacity relationship $t = \mathrm{G}(n)$ would in a real case relate all travel times to all frequencies affecting the same line. This relationship may be very complicated, but it is often possible to limit the number of affected tracks and lines to keep this manageable. In any case, such a relationship must be available in some form if the analyst is to be able to decide the effects of a suggested capacity improvement. But this is clearly a field requiring more research.