Appraising the benefits of bottleneck removal in rail transport: a simplified CBA approach and a case study in north-eastern Italy

Paolo Beria [1*], Raffaele Grimaldi [2]

[1] DAStU – Politecnico di Milano, 20133, Milano, Italy
[2] DAStU – Politecnico di Milano, 20133, Milano, Italy

Abstract

The removal of infrastructure bottlenecks is widely considered among the most profitable interventions, in socio-economic terms, and rail transport is not an exception. However, the measurement of the related benefits is difficult and no specific manuals indications seem to exist.

The aim of this paper is to propose a simplified approach to estimate the effects of a capacity constraint for a simple rail network, and assess its removal through a CBA.

In the first part, we describe the costs and benefits involved in bottleneck removal projects, suggesting possible sources and references. Then we briefly analyse the transport economics literature on the evaluation of rail bottleneck removals. In the following we introduce the proposed methodology, based on the use of a standard logit model. The model is specified initially for a single link and then extended to a more complex network. We also discuss the effect of regulation in the distribution of calculated surplus variations.

We complete the paper with a numerical case study about the appraisal of capacity investments in Friuli-Venezia Giulia, a north-eastern Italian region whose main rail network is part of two European TEN-T corridors (Mediterranean and Baltic – Adriatic).

JEL classification: D61, R41, R42
Keywords: Cost Benefit Analysis; bottleneck; rail.

1. Introduction[1]

There are many situations, especially in the early stages of the appraisal and planning process, in which indications on the amount of expected benefits of a new project is needed and a full multimodal transport model capable of simulating the effects is not available.

This paper presents a simplified approach to the calculation of bottleneck removal benefits, to be consistently used in a Cost Benefit Analysis (CBA).

The aim is to provide a method to help analysts face the lack of the huge amount of data and simulations needed for a CBA of a bottleneck. In particular, the method is suitable to be applied without a full transport simulation model describing the upstream and downstream network and the complete origin-destination (OD) matrix of traffic. Instead, the methodology is based on two assumptions: the availability of a relatively

* Corresponding author: Paolo Beria (paolo.beria@polimi.it; tel: +39.02.2399.5424)

[1] The present paper comes from the joint work of both authors. It is not possible to attribute single parts to the authors.

simple calibrated bimodal logit model – capable to simulate the modal shares of traffic in the bottlenecked section only – and the applicability of the so called "rule of half", to consistently calculate the user surplus variations in a CBA without explicitly knowing the demand function.

The paper is structured as follows. We start by describing the costs and benefits to be assessed to properly evaluate rail bottleneck removal projects (Section 2). In the following section (3) we outline how known literature deals with the benefits of bottleneck removal into CBAs, in particular with consumers' surplus. Section 4 describes the core of the methodology, applied to the simplest case of single segment network. The following Section 5 extends the base case to a generic network, in which one or more segments may reach saturation. Section 6 briefly outlines a parallel but crucial aspect: how market conditions and capacity allocation mechanisms influence the users and the social costs of bottleneck. Finally, Section 7 presents a numerical case study about the appraisal of capacity investments in the Italian region Friuli-Venezia Giulia. Section 8concludes.

2. A CBA of bottleneck removal projects

The assessment of a bottleneck removal project involves numerous effects to be quantified. In the following we will refer specifically to rail projects, but this method is generally suitable for scheduled transport modes.

In Table 2 we list the costs and benefits typically related to this kind of projects. Cost items are common to most infrastructure project. Benefits instead, must include the evaluation of scarcity and congestion removal (Nilsson, 2012). Both can be divided into sub-effects.

Costs	*Benefits*
1. Investment costs;	**5.** Removal of scarcity problems:
2. New infrastructure operating and maintenance costs;	- Surplus gain for users that would be otherwise excluded from the saturated infrastructure;
3. Residual value of the investment.	- Saved net external costs due to users that would otherwise shift to other modes or longer paths;
4. Possible impacts during construction (with respect to existing services, other infrastructures and the environment)	- Possible wider economic effects.
	6. Removal of congestion problems:
	- improvement in service performances (regularity, speed, crowding, etc.);

Source: our elaborations;
Note: we assume here that the new infrastructure built to solve the bottleneck has the same characteristics (speed, maximum axial load, etc.) of the former one. Clearly, if present, these further benefits must be added.

Table 1 - General list of costs and benefits to be considered in the appraisal of a bottleneck removal

Scarcity problems (Point 5 in Table 1) generate, as a direct effect, firstly surplus losses to users that might be excluded from the saturated infrastructure and in some ways might impact also users that continue travelling on it: we will discuss this issue in depth in the following section.

Secondly, as an indirect effect, the shift of excluded users to potentially more impactful modes or longer paths after the saturation of the existing infrastructure, will also generate external costs (more accidents, congestion, air and noise pollution, climate change, etc.), if not correctly internalised, to the society. The amount is the difference between gross external costs generated by all the shifting users, before and after the project, and specific taxes associated to generated transport (that is, fuel duty but not generic VAT[2]).

It is worth noticing that, while users' surplus variation depends on the difference in costs in the saturated section only (as we will discuss later), the net external costs must be accounted for the entire path followed by the shifting user, that is usually much longer than the saturated section only. Reference values for unit marginal external costs in the EU can be found for example in Maibach et al. (2008), while transport specific taxes can be estimated referring to Hyelén et al. (2013) for tax rates and the British *Transport Appraisal Guidance* (UK DfT, 2013: Unit 3.5.6) for fuel consumption formulas.

Thirdly, another class of potential benefits caused by scarcity problem removals is wider economic effects. Wider economic effects represent all those benefits that are not captured by direct benefits to users in a well constructed standard CBA, after allowing for external costs (Vickerman, 2007): literature includes agglomeration economies, increased competition, productivity of firms and network effects. These second order effects – that can be considered to be negligible in many transport projects – might potentially be significant in bottleneck removal projects because of the sharp increase in perceived generalised costs of transport due to the reaching of capacity. Unfortunately, the evaluation of wider economic effects usually requires complex economic models. An alternative way is the one suggested in the British *Transport Appraisal Guidelines* (UK DfT 2013: Unit 3.5.14), but it still needs a lot of detailed and geographically disaggregated data. A parametric estimation of what should be the WEE capable of justifying an otherwise negative CBA of bottleneck removal is thus a more manageable way to give policy maker an answer.

We have to finally consider also possible congestion problems (Point 6 in Table 1), particularly if the chosen maximum rail capacity value is not conservative enough. Regardless what criterion has been used to determine this value, the performance of rail services will gradually worsen when the number of trains will get closer to this value, in terms of less regularity of the service (more delays) and possible speed reductions. The increase in maximum capacity will thus provide also benefits to the users of those services, in particular passengers, and these should be included in the final Net Present Value (NPV).

We will discuss more in details in the following section how scarcity and congestion removal are treated in literature.

3. The core benefits of bottleneck removal: scarcity and congestion

Nilsson (2012) explains how congestion, in the way it is usually understood in road transport, takes the form of two different components in rail transport:

- a scarcity problem; and,

[2] Actually, VAT applied on fuel duty represent a transport specific tax and should be considered.

- a congestion problem.

The problem of scarcity rises when it becomes difficult, and gradually impossible, to build a timetable capable of guarantee all the requested slots. The problem of congestion instead appears after the timetable has been designed, since the planned timetable might be disturbed by delays and irregularities.

With this respect, Nash and Samson (1999) outlined how *"the main consequence of full utilisation of capacity is that users simply cannot get the capacity they want when they want it; they have to run their trains at times and possibly speeds different to their preferred alternative, or to give up the journey."*

The way bottleneck removal benefits should be assessed in socio-economic CBAs seems quite neglected in the literature and guidelines, quite surprisingly considering the importance and frequency of the issue.

The *Railway Project Appraisal Guidance* (RailPAG: EC-EIB, 2004) states that *"a bottleneck will typically be associated with congestion. Small deviations to train plans (train delays) will result in the need to readjust train schedules significantly. A removal of the bottleneck will hence often have significant repercussions on reliability. The improvement in reliability will affect the costs of the rail operators, and these savings could be significant but difficult to trace and measure. Improvement in reliability will also lead to shorter travel times and so to an improvement in the welfare of travellers over and above average time savings. These improvements will have to be valued separately and accounted for in the analysis, otherwise the real benefits of the project will not be covered. On the other hand attention must be paid not to double-count time savings under the "reliability" label."* The guidance focuses on effects for users remaining on the rail line, but does not consider what happens to the others.

Saturation and the consequent mode shift is instead considered in the Strategic Business Case of the British "New Lines Program " (Network Rail, 2009), in which freight benefits due to a release of freight paths on the conventional West Coast Main Line after the building of the planned parallel High Speed 2 line, have been calculated using the so called "sensitive lorry miles avoided" approach (SRA, 2003): they attach to each kilometre (mile) that is expected to shift from road to rail transport after the bottleneck will be removed a net external cost saving (i.e. the difference between generated gross external costs – in terms of accidents, congestion and environmental impacts – and paid fuel taxes). Again, it seems that no surplus loss has been assessed for trains excluded from the congested rail line without the investment.

In the following we will focus on the scarcity problem, leaving the decrease in remaining service performances to other contributions like Eliasson and Borjesson (2012), who discuss about how rail travel time-capacity relationships, that would allow estimating the congestion benefits of a bottleneck removal, is a field that would require further research. The authors demonstrate how timetable assumptions influence the results of railway investment appraisals.

4. Base case: single line and single relationship

4.1 Problem description

When a multimodal transport simulation model for the study area is available, the effect of a bottleneck in terms of mode and path shifts can be simulated and this will provide all data needed for a CBA. In particular, the transport model allows to calculate the generalised cost (hereinafter "GC") for every type of user and for every OD couple, both with and without the constraint. For example, a model can estimate the GC of rail and road transport between any origin and destination involved, in absence of capacity constraints. In case of bottleneck, say in the rail corridor, the GC of rail increases, while GC of road option remains the same. The model is then capable of calculating the shares of the two modes, with and without bottleneck. Moreover, if the basic rail path is saturated, the model should be able to calculate also the GC of alternative rail paths, allowing a correct distribution of demand on the network. The model is then providing the CBA with the traffic flows and with the GCs in all scenarios.

However, suitable regional multimodal transport models are often not available. Consequently, any CBA not capable to correctly take into account saturation effects (including up- and downstream paths) result biased. Our approach, tries to practically evaluate the socio-economic scarcity effects of a bottleneck, starting from its practical consequence, that is an artificially constrained modal share, trying to simplify and overcome the difficulties related to the construction of such a relationship without a model. This method is suitable to be applied at the early stages of the planning evaluation process, when only general inputs exist.

Let's consider a relation A to B, where two transport modes compete, for example "rail" and "road". An example (Figure 1) can help us in understanding the situation.

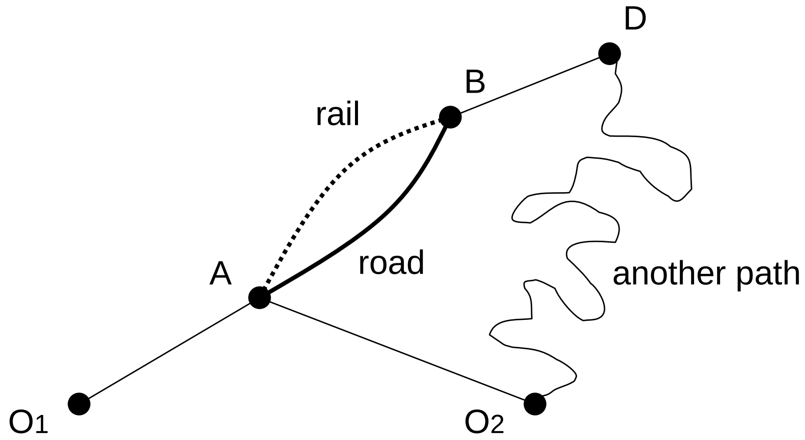


Figure 1 - Example of network

Users of the section A-B come from the two origins O_1 and O_2 and all go to the destination D. Without a model, we do not exactly know their individual GCs, neither where they exactly come from. We just know that their paths O_1ABD and O_2ABD pass through section A-B, divided into the two available modes, "rail" and "road". Actually, their choices depend on many individual characteristics, like their origin, their upstream and downstream paths and modes, their value of time (in its different components), their transport costs (both depending on what they carry), etc, all unknown.

Let's consider the case in which traffic exogenously increases and a bottleneck in the A-B "rail" segment occurs.[3] In reality users will individually evaluate their best possible

[3] We do not deal here with the issue of how rail capacity can be assessed and which values should be chosen: it is however clear that this is a crucial issue, as its value – and the consequent timetable

substitute options, choosing the best one in terms of GCs. For example, some of them (say those from O_1) could prefer to switch to the "road" mode, because the extra-cost for them is lower than the extra-cost of the saturated rail line and of any other known alternative. To the contrary, others (say those from O_2) could radically change path and now prefer the link "another path".

The scarcity cost of the bottleneck (more precisely, its shadow cost) is then dependent on:

1. the OD matrix;
2. the network characteristics;
3. the users' characteristics.

Those are the typical outputs of a transport model. In the following sections we will describe how to make an evaluation having only relatively limited information.

4.2 Generalised costs and traffic under saturation conditions

Random utility models (typically logit) applied to a specific network section are commonly used to describe the behaviour of heterogeneous users starting from the average costs or utilities (Cascetta, 2001) and assuming a particular distribution of the singular costs (or utilities), not included in the deterministic measure.

So, through a logit model it is possible to estimate a modal share due to average conditions, given a certain distribution of individual behaviours, without knowing the behaviour of individual users. In this example, we will refer to a simple binomial logit, like the one in Equation 1, where p_{rail} probability of choice (which translates into the mode share) of rail transport, GC_{rail} and GC_{road} are the respective average generalised costs, κ and λ are the calibration parameters of the model (we use κ as a rail specific constant, but other formulations are possible) and t is the considered year.

$$p_{rail,t} = \frac{\exp(\lambda \cdot GC_{rail,t} + k)}{\exp(\lambda \cdot GC_{rail,t} + k) + \exp(\lambda \cdot GC_{road,t})} \qquad \text{Equation 1}$$

The following Figure 2 depicts the concept, where $Q_{rail,t} = p_{rail,t} \cdot Q_{tot,t}$ is the rail traffic at year t.

assumptions (Eliasson and Borjesson, 2012) – determines how much scarcity or congestion will rise, thus changing the results of the CBA.

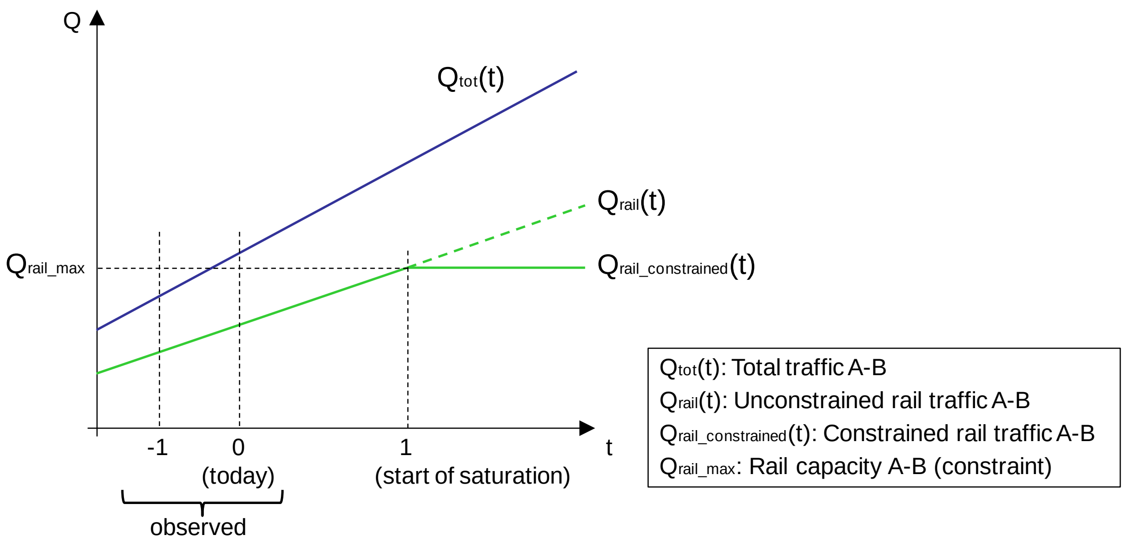


Figure 2 - Exogenous and constrained traffic

Rail, road and the consequent total traffic on the section A-B would exogenously rise. However, at a certain moment, rail increase is capped as the section reaches its capacity as described before: exogenous increase is no more possible on that mode and part of traffic shifts to its second best alternative; in this simplified case the parallel road.

We assume that:

a. the traffic and the modal split are known for some years in the past, under different conditions and for different origin-destination pairs. For example, in years with different fuel costs or with a different network configuration. Independent observations are used to estimate the calibration parameters.
b. it is possible to make an estimation of average generalised costs for the users of the two modes passing through section A-B, possibly divided in homogeneous O-D groups.[4]

In this case, we build one (or more, if A-B traffic is too heterogeneous to be modelled singularly) calibrated logit capable of simulating the modal split in section A-B in function of GCs.

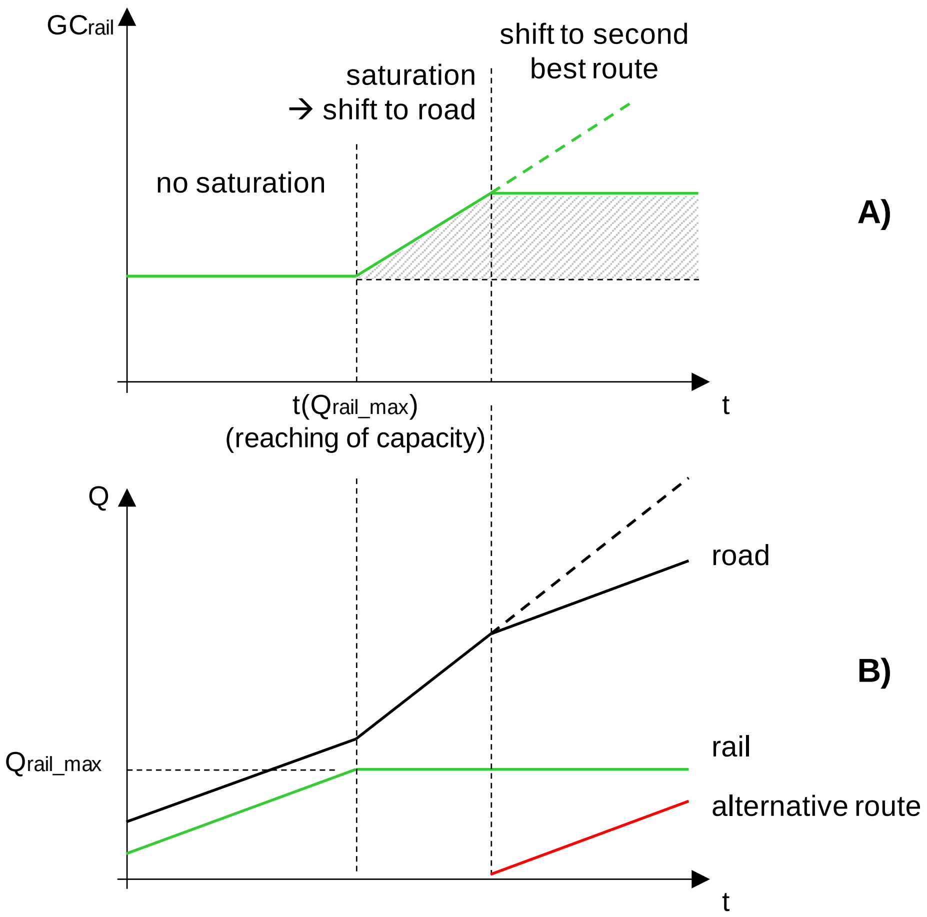
The calibrated model can then be applied yearly, since the moment of reaching capacity. The idea is to simulate which are the average GCs of the constrained mode that would generate, in future years, the same modal share forced by the constraint.

By substituting the constrained rail share $p_{rail\_constrained}$ and properly turning the equation, we can obtain the needed function of GC with respect to the constrained modal share, as in Equation 2.

[4] The more are the users groups for which we can estimate average GCs (point b), the more observations (point a) we will need to calibrate the logit.

$$GC_{rail\_constrained,t} = \frac{1}{\lambda} \cdot \left\{ \ln \left[\frac{p_{rail\_constrained,t}}{1 - p_{rail\_constrained,t}} \cdot \exp\left(\lambda \cdot GC_{road,t}\right) \right] \right\} - k \qquad \text{Equation 2}$$

The extra-cost of the saturation is the difference between $GC_{rail\_constrained}$ and GC_{rail}. We could say that this is the extra-cost capable of shifting the former marginal user, i.e. the user with the less costly alternatives or the user with the lower utility from passing through the network section under consideration.[5]



[5] It is worth mentioning a paradox that might rise in the application of the logit. When we calculate the costs of the capacity constraint on the less used mode ("rail", accounting for, say, 30%) we see that the average generalised cost of the other mode ("road", accounting for, say, 70%) is already lower (otherwise "road" would be less used). But, when we force the modal shift to simulate the effect of saturation, the difference in generalised costs (GC_{road} - GC_{rail}) would be positive, resulting in a benefit of the constraint. This paradoxical result is because the existing users of the "rail" already individually perceive lower generalised costs with respect to average ones and consequently use the mode that has the lower cost for them, even if not lower on average. This paradox is however not a problem if the logit is correctly calibrated, typically introducing a modal constant or assuming (as a parameter) that part of the traffic using the dominant mode is captive, both determining generalised costs coherent with previous calculations.

Figure 3 - Generalised costs under saturation conditions (A) and traffic (B)

Figure 3 summarises the simulated effect. The imposing of a capacity constraint to the mode "rail" (Q_{rail_max}) translates into an increase in the GC of rail[6] (graph A). Consequently, rail traffic stops increasing and levels to the maximum capacity (graph B). This situation is however not infinite (even assuming infinite capacity of "road"), because at a certain point a second best "alternative route" (for example another longer rail path) could become favourable to shifters (graph B). The effect is that the generalised cost increase is capped. The resulting shaded area in graph A[7] represents the shadow cost of the constraint. The constrained rail GC is then defined by Equation 3.

$$GC_{rail\_constrained,t} = \min\left\{\begin{array}{c}\frac{1}{\lambda}\cdot\left\{\ln\left[\frac{\exp(\lambda\cdot GC_{road})}{p_{rail\_constrained,t}} - \exp(\lambda\cdot GC_{road})\right]\right\} \\ GC_{rail\_alternative\_path}\end{array}\right. \qquad \text{Equation 3}$$

However, since it is unknown both which users/goods will actually shift and their up- and downstream origins and destinations, it would be impossible to estimate their new costs without making further assumptions. What is known is that, coming up to the saturation of one mode (say, the rail), the users that have the less expensive alternative or lower utility will be those shifting/disappearing first,[8] starting from the former marginal user.

This is summarised in the so called "rule of half", which is the way commonly used in CBAs to overcome the ignorance about the form of the demand function (Abelson and Hensher, 2001; Kidokoro, 2004; the World Bank, 2005; Maffii and Parolin, 2013; Grimaldi and Beria, 2013). It assumes that all users are linearly distributed between the former marginal user in the initial situation and the new marginal user in the final situation.

In conclusion, the logit tells us what is the average generalised cost of transport in both situations and the "rule of half" subsumes the distribution of the non-marginal users.[9]

A similar approach is used by Jorge and de Rus (2004) to evaluate capacity expansion projects in airports.

4.3 The scarcity cost of a bottleneck and the benefit from its removal: users' surplus

Our assumptions allows to calculate yearly which is the extra-cost associated to a congested network segment that, if imposed to the users, would cause a modal split equal to the one forced by the capacity constraint.

The difference between the average generalised cost (GC_{rail}) and the one that simulates the effect of the capacity constraint ($GC_{rail\_constrained}$), is the extra cost for users, associated to the reaching of capacity. This extra-cost should be applied as a whole to

[6] We will discuss later what this increase actually represents, according to market conditions.

[7] The area is triangular if no alternative path exist in the obtained range of GC, or trapezoidal if a third alternative path exist.

[8] We assume here that no grandfather's rights for capacity exist: the track operator will price differently in case of saturation and users will modify their behaviour consequently, irrespectively of their previous use of the infrastructure. We will discuss further this hypothesis in Section 5.

[9] Clearly, both hypotheses are strict. However, for early appraisals we can obtain a consistent estimation in absence of a multimodal transport model to describe the costs of single users.

the traffic that remains on the saturated segment, and through the above introduced "rule of half" to the traffic that shifts to the other mode (Figure 5). The “rule of half” subsumes that we don't know how much the shift will cost them, but we know that it will cost something in between the difference between the GC of the rail before and after reaching the capacity (otherwise they would not have shifted, see Grimaldi and Beria (2013)).

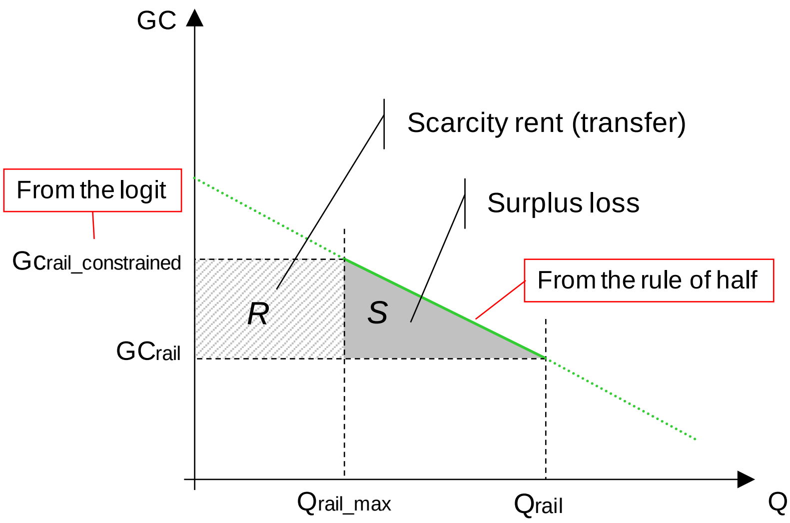


Figure 4 - Shadow cost of bottleneck (at year *t*)

The cost of constraint for the users at year *t* is then:

$$\Delta S_{users,t} = \boldsymbol{R} + \boldsymbol{S} = Q_{rail\_max} \cdot (GC_{rail,t} - GC_{rail\_constrained,t}) + \tfrac{1}{2}\,(Q_{rail,t} - Q_{rail\_max}) \cdot (GC_{rail,t} - GC_{rail\_constrained,t})$$

However this extra-cost does not match with the social cost of the bottleneck. We can divide it into two components: the rectangular part (area “R”, the shaded one in Figure 5) is the cost that remaining users “pay” to stay on the saturated line (as a real extra-toll or as a scarcity rent). This, however, is a transfer between the users (more precisely: goods or passengers transported on the line) and another subject (the train operating company, the network manager or the government, according to regulation. See Section 5).

The second component (area “S”, the grey triangle in Figure 5) is instead the net surplus loss associated to the users that renounce to travel on the segment (shifting to another path or giving up on travelling).

The social surplus loss of the bottleneck at year *t* is then:

$$\Delta S_t = \boldsymbol{S} = \tfrac{1}{2}\,(Q_{rail,t} - Q_{rail\_max}) \cdot (GC_{rail,t} - GC_{rail\_constrained,t})$$

This will be one of the benefits when performing the CBA (see Section 6). It is to be noticed that the over- or under-estimation of bottleneck removal benefits resulting from this method depends on the quality of the calibration of the logit (but this is true for any model) and on the distribution of O-D pairs (the more it is homogeneous, the more the “rule of half” is correct).

5. Extended case: networks

The same methodology described above for a single segment, can be extended to a simple network. In this case every segment is to be treated singularly as above, but the way the different traffic components are allocated to the network must follow an optimisation criterion to be decided and that can introduce some complexity in the calculation.

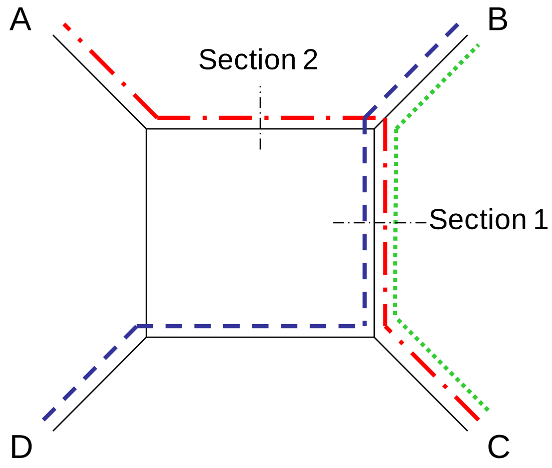


Figure 5 - Example of network structure and traffic components' paths.

In the left part of example Figure 6 we see a network made of four segments and connecting four origins/destinations. The OD matrix is made of 16 cells and possible paths are two per OD pair. In the right part of Figure 6 an example with just three active relationships (A-C, B-C, and B-D) is depicted, each one following a path involving some segments of the network. In particular, one segment is interested by all three OD pairs and two segments are interested by two OD pairs each.

Three different situations may rise.

The first is when one single section (say Section 1 of Figure 6) is saturated by the three traffic components. This case is however trivial, as it must be treated exactly as described in previous Section 3, with the sole difference that the same extra-cost (to be calculated) is applied to three traffic components (via three different logit models) under one capacity constraint.

The second and third cases occur when two sections (say Section 1 and Section 2 of Figure 6) reach the capacity during the analysed period. In both cases the shadow cost of the bottleneck can be calculated as follows:

a. Each OD pair has an uncongested average generalised cost $GC_{o,d}$, to which one must add the extra-costs (K_x) to be calculated, one per each section of the network (equal to 0 if no saturation rise), every year.

$$GC_{o,d\_constrained} = GC_{o,d} + K_1 + K_2 + \ldots$$

b. setting all $K_x = 0$ determines the rise of saturation for the two segments, when the logit (one per OD pair) are performed.

c. Starting from the most saturated section[10] (say Section 1), the corresponding extra-cost K_1 is calculated in order to have 100% use of capacity ($Q_{o,d} = Q_{rail\_max,o,d}$).

In the second case, this extra cost of Section 1 causes a reduction of traffic such as also the other Section 2 falls below saturation. It must be noticed that the model “automatically” determines the new mix of the traffic components (A-C, B-C, B-D in

[10] This algorithm is the simplest. Other optimisation algorithms may be applied.

the example), according to their overall generalised cost, once the cost of constraint is included. So it might also happen that all the shifted traffic comes from one single OD pair or any other distribution depending on relative GCs.

In the third case the extra-cost K_1 is not sufficient to solve all upstream and downstream saturation problems, including that of Section 2. In this case a second extra-cost K_2 must be calculated and applied to the involved traffic components of sections (only A-C in this example). As the process described is static, this will however lead to a further decrease of flows on Section 1 (now below 100%). Stopping here would clearly provide an overestimation of the cost of the bottlenecks. So, the process can be repeated iteratively (step 3 will decrease the K_1 previously calculated), until equilibrium is reached with all $Q_{o,d} = Q_{rail\_max,o,d}$.

6. The effect of market conditions on users surplus

As previously mentioned in Paragraph 1.3, the distribution of the extra-cost calculated above depends on the market conditions. In particular, it depends on the presence of "grandfathers' rights" in the allocation of capacity and on the way access is priced to the final users.

case	*1.1*	*1.2*	*2*	*3*
Cost of scarcity for final users	-R-S	worse than -R-S	-R	-R
Extra revenues for TOCs	+R			
Extra revenues for the IM			+R	
Extra revenues for the Regulator				+R
ΔS	-S	worse than -S	-S	-S

Table 2 - Distributional effect of regulation (sign + is a benefit)

Naming the two components of users' surplus as in Figure 5 (***R*** and ***S***), we distinguish into 4 cases, summarised in Table 2:

1) There is no access regulation and the Infrastructure Manager ("IM") does not exploit the cost of scarcity and thus loses the willingness to pay of users, pricing slots at the same price of unsaturated conditions (this happens for example when applying the "grandfathers' rights" principle). In this case it might happen that:
   - 1.1) Train Operating Companies ("TOCs") are able to discriminate their users raising the price. In this case final users (the passengers or the transported goods) pay and the TOCs benefit of the same amount as a scarcity rent.
   - 1.2) Also TOCs do not discriminate users with higher willingness to pay (WTP) and go on with their "historical" customers applying "historical" prices up to the reaching of saturation, whatever is the WTP of lost demand. From the social point of view this is the worst case, but a quantification of surplus loss is not possible as the WTP of included and excluded demand is not known.
2) There is no access regulation, but the IM is able to exploit the scarcity rent, by pricing differently the scarce slots. The TOCs and their customers will pay and the IM will get the extra revenues, generating monopoly extra-profits.

3) The Regulator applies scarcity prices for the access to the slots, in order to reflect the calculated cost of scarcity: in this case the demand with higher WTP (through the TOCs) will pay for the extra charge. The remaining demand, with lower WTP from passing in the saturated section, will give up or will find another route. The regulator will re-invest the revenues (not necessarily to solve this bottleneck, that might not be the most socially viable option). This situation is not common, but the most adherent to the effect accounted theoretically by CBA. It is worth noticing that the regulator revenues are not a net benefit, because another component of the market (the users) is paying for it.

It is clear that in all cases what changes is the distribution of the "R" component among the actors of the society according to regulatory conditions. However, in all cases there is a constant surplus loss due to scarcity, that is the component "S" and that represents the surplus loss for the excluded demand. Only in case 1.2 the surplus loss is higher than S, because excluded customers are not necessarily the ones with lower WTP (as happens in other cases). This situation is however very unrealistic, as at least the TOCs will be probably able to recognise that some of the customers that they do not serve are willing to pay more than the "historical" ones.
In conclusion, these different conditions affect the subject that will "pay" the extra-cost of bottleneck. However, this – if the hypothesis of standard CBA theory holds – has no effect on the net social surplus, with the relevant exception of the worst case 1.2.

7. An application: Friuli Venezia Giulia rail network

This methodology was developed and used to evaluate the social cost associated to the possible saturation of the main rail network of the Friuli Venezia Giulia region, in north-eastern Italy, and the consequent justifiable investments aimed at increasing capacity. This evaluation was made at the very early planning stage, in order to give decision makers an order of magnitude and help them identify possible strategic directions for rail transport. Consequently available information was quite limited and many used inputs are parametric or taken from the literature.

Three hypotheses have been made:

1. The demand growth is exogenous and not dependent from possible projects, that is no significant performance improvements are forecasted due to new capacity expanding investments;
2. Capacity on alternative rail paths is sufficient. This is particularly needed for the Brenner rail, which represent the natural rail alternative for the considered area and for which a capacity expansion is already planned and partially under construction (Grimaldi, 2012);
3. Forecasted increase in passenger services was given the priority and considered as a constraint (that is, those services do not have to compete for slots with freight trains).

In order to simplify the problem, the main rail network – part of two European TEN-T corridors (Mediterranean and Baltic–Adriatic) – has been grouped into five main sections, as shown in Figure 7. In our opinion these simplifications are acceptable because the analysis is focused on freight traffic, and different routes within the same main sections are substantially indifferent. Each section is crossed by different flows, that is traffic having common origins and destinations. Freight flows were modelled into six relations, using the network as described inTable 3.

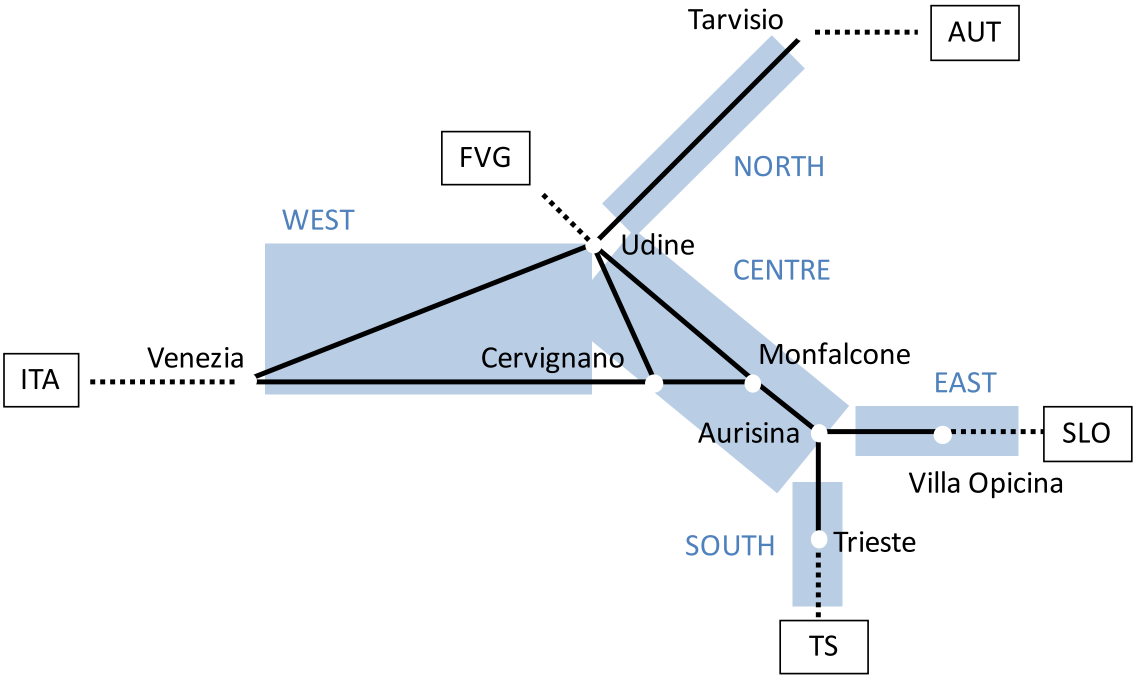


Figure 6 – Schematic representation of the main rail network in Friuli Venezia Giulia.

Flow	Description	Sections
Italy-Austria (ITA-AUT)	Italian and western origin/destinations (excluding FVG) to/from Austria and north-eastern ones	WEST, NORTH
Trieste port – Austria (TS-AUT)	Trieste port to/from Austria and north-eastern origin/destinations	SOUTH, CENTRE, NORTH
Friuli Venezia Giulia – Austria (FVG-AUT)	Origins/destinations inside Friuli Venezia Giulia region to/from Austria and north-eastern origin/destinations	NORTH
Italy-Slovenia (ITA-SLO)	Italian and western origin/destinations (excluding FVG) to/from Slovenia an eastern ones	WEST, CENTRE, EAST
Trieste port – Slovenia (TS-SLO)	Trieste port to/from Slovenia and eastern origin/destinations	SOUTH, EAST
Friuli Venezia Giulia – Slovenia (FVG-SLO)	Origins/destinations inside Friuli Venezia Giulia region to/from Slovenia and eastern origin/destinations	CENTRE, EAST

Table 3 – Description of considered flows and travelled network sections.

Different growth trends were estimated for the different flows, up to 2050: in some cases those trends saturated the capacity of existing rail sections. In the base case, saturation appears only in the CENTRE section in 2030 and freight slot requests on this section would be 141% of estimated freight capacity in 2050 (that is, 114 trains per day on the Monfalcone-Aurisina route).

Figure 7 presents estimated traffic flows and associated average rail generalised costs using the presented logit model for each traffic flow, with and without the capacity constraint on the three flows passing through the CENTRE section (TS-AUT, ITA-SLO and FVG-SLO). As one can see they would have different behaviour in case of saturation. FVG-SLO flows would suffer from a significant increase in their average generalised costs from 2030 onwards, resulting in an even sharper reduction of expected rail flows (also because of the short average distance, see Table 4). The effect on the TS-AUT is relatively limited, because of the availability of an alternative rail path via Slovenia having low extra costs. The ITA-SLO flow is the largest and the constraint would substantially cap its growth to the values of 2030-2035.

	Increase in rail GCs	*Reduction in rail flow*	*Average distance*	*Traffic type*
TS-AUT	+4.0%	-15.6%	500	maritime
ITA-SLO	+16.3%	-25.9%	1,000	land
FVG-SLO	+42.9%	-75.7%	300	land

Table 4 - Estimated increase in rail generalised costs and reduction in rail traffic in 2050.

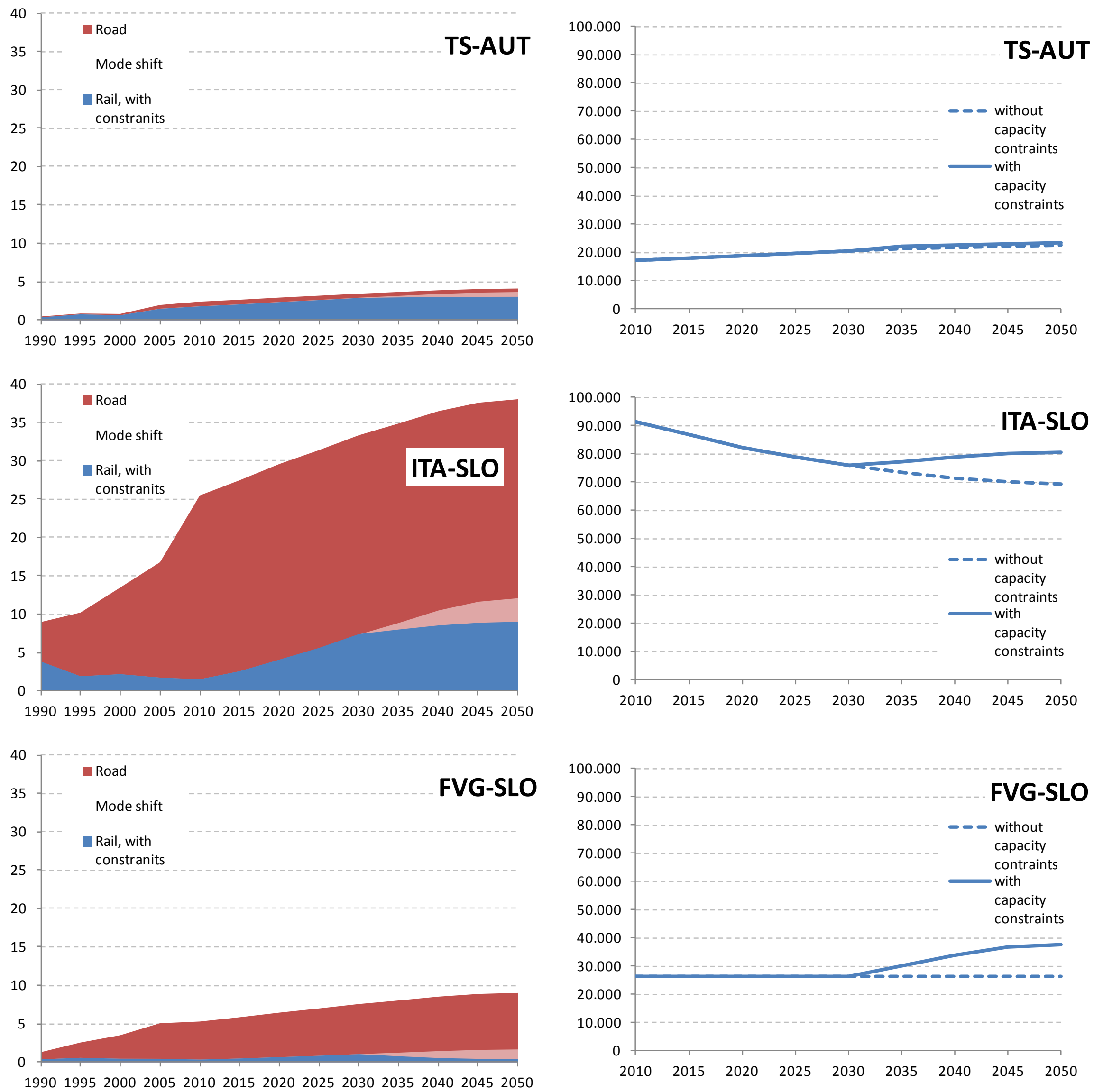


Figure 7 - Estimated traffic flows (left row, in million tons per year) and associated average rail generalised costs (right row, in Euro per train)

The estimated increase in generalised costs for those traffic flows allowed us to estimate the consumer surplus losses associated to the saturation of the Monfalcone-Aurisina rail section from 2030 on. Additionally, we evaluated the net external costs (that is, the difference between gross external costs and fuel taxes) generated by the mode shiftto road transport and longer paths.

This being an early evaluation, no increase in the performances of existing services (also because we judged used capacity values – provided by RFI, the Italian rail infrastructure manager – to be conservative enough) and no possible wider economic effects were estimated. In Table 6 we report the estimated social cost of the constraint. Using typical operating and maintenance costs and residual values, we suggested that capacity increase measures, up to a cost of 638 million Euros in 2025, would be justified.

	Present Value	
Surplus loss (shifted users)	-186,790,192	$€_{2013}$
Generated gross external costs	-590,626,561	$€_{2013}$
Generated fuel taxes	371,173,740	$€_{2013}$
Social cost of the constraint	-406,243,013	$€_{2013}$

Table 5 – Estimated social cost of the constraint

We estimated also a potential scarcity rent of 1,061,714,641 $€_{2013}$ (the sum of the *R* areas in Figure 5 in each year) which, as we discussed in section 6, represents a transfer within the society but might trigger significant second order effects.

8. Concluding remarks

Bottleneck removals generate two family of benefits, related to the removal of scarcity problems (surplus gains for users otherwise excluded, change in generated net external costs due to avoided mode or path shifts and possible wider economic effects) and of congestion problems (improvement in service performance).

Transport evaluation literature does not give indications about how to properly calculate bottleneck removal benefits in scheduled transport modes. This fact is particularly problematic when making early appraisals without having a multimodal transport model.

In this paper we propose a simplified approach for the evaluation of consumers' surplus based on a simplified binomial logit model related only to the saturated section and thus needing little information. The aim is to provide a method to help analysts face the lack of the huge amount of data and simulations needed for a CBA of a bottleneck. The model is used to calculate, using a standard logit model, the extra cost which would determined a modal share equal to the one forced by the capacity constraint. This extra-cost is used to determine the shadow cost of the bottleneck.

Indications about how to evaluate the other possible benefits are then provided.

Bibliographical references

Abelson, P., & Hensher., P. (2001). Induced travel and user benefits: clarifying definitions and measurement for urban road infrastructure. In Hensher, D., Button, K. (Eds.), *Handbook of transport systems traffic control*, Handbooks in Transport 3. Elsevier, Pergamon.

Cascetta, E. (2001). *Transportation systems engineering: theory and methods*(Vol. 49). Springer.

Maibach, M., Schreyer, C., Sutter, D., Van Essen, H. P., Boon, B. H., Smokers, R., ... & Bak, M. (2008). Handbook on estimation of external costs in the transport sector. *Internalisation Measures and Policies for All external Cost of Transport (IMPACT). CE Delft*.

EC-EIB (2005). *RailPAG: Railway Project Appraisal Guidance*. European Commission and European Investment Bank.

Eliasson, J., & Borjesson, M. (2012). On timetable assumptions in railway investment appraisal. In *Transportation Research Board 91st Annual Meeting*(No. 12-2124).

Grimaldi, R. (2012). La galleria di base del Brennero: considerazioni su costi e benefici. In Grimaldi, R. (ed.), *C'è luce in fondo al tunnel? Analisi e spunti sulle politiche*

infrastrutturali ferroviarie alpine, Maggioli Editore, Sant'Arcangelo di Romagna (Italy).
Grimaldi, R., & Beria, P. (2013). Open issues in the practice of cost benefit analysis of transport projects. *13th World Conference on Transport Reasearch*, Selected Proceeedings, July 15-18, 2013 – Rio de Janeiro, Brazil.
Hylén, B., Kauppila, J., & Chong, E. (2013). Road Haulage Charges and Taxes: Summary analysis and data tables 1998-2012. In *International Transport Forum Discussion Papers* (No. 2013/8). OECD Publishing.
Jorge, J. D., & de Rus, G. (2004). Cost–benefit analysis of investments in airport infrastructure: a practical approach. *Journal of Air Transport Management*, *10*(5), 311-326.
Kidokoro, Y. (2004). Cost-benefit analysis for transport networks: Theory and application. *Journal of Transport Economics and Policy (JTEP)*, *38*(2), 275-307.
Maffii, S., & Parolin, R. (2013). Estimating Benefits for Modal Shifters: a Methodological Remark. *13th World Conference on Transport Reasearch*, July 15-18, 2013 – Rio de Janeiro, Brazil.
Nash, C., & Sansom, T. (1999). *Calculating transport congestion and scarcity costs*. Final Report of the Expert Advisers to the High Level Group on Infrastructure Charging, Commission of European Communities, Brussels.
Network Rail (2009). *New Lines Program: Strategic Business Case*. Planning and Regulation, Route Planning, London (UK).
Nilsson, J. E. (2012). *Congestion and scarcity in scheduled transport modes.* No. 2012: 25. CTS-Centre for Transport Studies Stockholm (KTH and VTI).
The World Bank (2005). Transport Note No. TRN-11: treatment of induced traffic. *Notes on the Economic Evaluation of Transport Projects.* Transport Notes, Transport Economics, Policy and Poverty Thematic Group, Washington, DC (USA).
UK DfT (2013). *Transport Analysis Guidance.* Department for Transport, London (UK). Website: http://www.dft.gov.uk/webtag
Vickerman, R. (2007). *Recent Evolution of Research into the Wider Economic Benefits of Transport Infrastructure Investments*. Discussion Paper No. 2007-9, ITF-OECD Joint Research Centre.