TOWARDS A COMPREHENSIVE ECONOMIC THEORY OF RAIL CONGESTION

BRUNEL Julien, Réseau Ferré de France, Paris, France julien.brunel@rff.fr
MARLOT Grégoire, Réseau Ferré de France, Paris, France gregoire.marlot@rff.fr
PEREZ Maria, Réseau Ferré de France, Paris, France maria.perez-herrero@rff.fr

ABSTRACT

As the European rail markets progressively open up to competition, there is a growing concern for optimizing the infrastructure manager production processes (investments, timetabling and path allocation). Line capacity is essentially what the infrastructure managers have to sell as their final service. Nevertheless capacity is a complex issue with numerous meanings and not a standard definition. The capacity of a railway line depends on how it is used and how the different parameters are combined (infrastructure factors, traffic features or operating requirements).

In railway sector, the capacity shortage has traditionally been considered as the inability for a train operator to obtain the desired train path (scarcity). However, this capacity perception seems restrictive. A lack of capacity can be experienced before scarcity, as unexpected transmitted delays are positive related to the density of traffic (congestion). In order to reduce capacity shortage, the infrastructure manager cans activate different levers.

On the short-run, the infrastructure manager should look for an optimal scheduling. It exposes to a trade-off between resilience and capacity. This trade-off may involve several efficient price signals in order to reflect the costs of the different options. On the long-run, the infrastructure manager is looking for an optimal trade-off between the total costs of the capacity shortage (expectable delays, unexpected delays and the part of demand that is not satisfied, whether it has been rejected or deferred) and the cost of increasing the capacity of the network (building and operating). Short-term levers must necessarily be linked to the long-term perspective of capacity investments by regulation in order to guarantee the welfare maximization.

This paper proposes few facts for the definition of capacity shortage in rail sector and the potential levers that an infrastructure manager can develop. This research relies on data from the French rail infrastructure manager, concerning delays, traffic, investments costs, and on a detailed analysis of timetabling and path allocation processes.

Keywords: Capacity, congestion, investment, marginal cost, scarcity, scheduling, price signal, shortage, rail transport, resilience.

1. INTRODUCTION

Nowadays there is a growing concern for providing efficient price signals through network access charges. Whereas it is sometimes very difficult to assess the marginal costs of rail traffic, it is even more difficult to take into account capacity shortage. As a result, access charges are often too low in highly congested areas or during peak hours, increasing the need for public funding, whereas they can be too high in others cases, deterring the use of the network. More generally, economic appraisal of capacity investments is rather difficult. Cost-benefit analyses value time savings, not capacity increases. This can lead to over- or underinvestment. Moreover, there is no such thing as economic optimisation of timetabling and path allocation.

The problem is far easier for vertically integrated operators: capacity shortages and delays translate directly into a loss in commercial revenues. Congestion is not an externality, and thus dealing with capacity shortages is a matter of maximizing commercial revenues under technical constraints. There exists a considerable literature of both analytical and simulation-based methods, which have been used to study delays and capacity assessment in railroad line haul networks. The propagation of unexpected delays, resulting from incidents, has been less studied, even if infrastructure managers often make provision for large headways in the schedule to deal with such delays.

These methods could be useful for an infrastructure manager aimed at providing efficient price signals to competing train operating companies, and optimizing its own production processes (investments, timetabling, path allocation), regarding not only commercial revenues but also welfare impacts. Nevertheless, it is necessary to give an economic value to congestion, which is not an easy task as soon as one cannot observe congestion on a “free access” network. One can only observe the demand and the congestion resulting from given timetabling, path allocation process and access charges. The problem is the same for railroads.

Moreover, there are few academic papers dealing with railroad congestion pricing, all of them considering congestion and scarcity as different phenomenon, calling for distinct valuation and internalization process. The aim of this paper is to provide a few hints for the development of a comprehensive economic theory of rail congestion, which would be a useful tool for the implementation of a rail congestion-pricing scheme, but also for the optimisation of timetabling, path allocation, and capacity investments.

The structure of the paper is as follows. Section 2 reviews the academic literature dealing with road and airport congestion. Section 3 tries to characterize railroad congestion from an empirical point of view. Section 4 studies the relationship between congestion, scarcity and the need for capacity investments from an economic point of view. Section 5 proposes a comprehensive method for the economic appraisal of capacity shortage, which could be used for pricing as well as optimisation of timetabling and path allocation, and economic appraisal of capacity investments. At last, section 6 offers concluding remarks.

2. RELATED LITERATURE

a) The literature regarding road congestion pricing

A wide literature deals with congestion in the road sector, from classic contributions such as Pigou (1920), Walters (1961) or Vickrey (1963) to more recent works from Arnott, De Palma and Lindsey (1993), Chu (1995) or Verhoef (2001). From an economic perspective, congestion is basically a standard externality problem. When car users decide to make an additional trip, they impose additional costs on themselves, but also on other users. Traffic engineer's speed flow curve has well established the existence of a relationship between traffic density and speed. When a new car enters the road, traffic density increases, speed drops and, therefore, travel time lengthens. The economic issue of road congestion lies in the fact that drivers do not perceive the cost they incur for other users. It results in an excessive consumption of road traffic and congested roads during peak hours.

Governments traditionally curtail congestion on the supply side, with the expansion and the improvement of road networks. This solution is restricted by fiscal, physical or environmental constraints. However, economists explain that the problem of congestion can be addressed on the demand side by pricing or regulation. Academic literature shows that peak/off-peak pricing is an efficient solution to tackle congestion, and allows users to internalize the external costs generated and reallocate the traffic demand during the day (Vickrey, 1963). Furthermore, the literature also demonstrates that, given certain circumstances, congestion pricing covers the construction costs of roads (Mohring and Harwitz, 1962, Arnott et al., 1993). Hau (1998) proposes a rigorous non-mathematical interpretation of this literature and relaxes some of the assumptions of the cost recovery problem. It includes the constant return of scale and the perfect divisibility of the investment hypotheses.

b) The literature regarding airport congestion

As Quinet (1997) pointed out, congestion does not appear only on roads but also in other transportation modes, even where traffic is scheduled in advance. By contrast with road sector, congestion has received less attention in these sectors. A sizeable literature nevertheless has studied airport congestion that happens in the neighbourhood of large airports due to runways or traffic control saturation.

In a seminal paper, Carlin and Park (1970) estimate the marginal cost of delays in New York's LaGuardia airport. The congestion cost is defined as the additional delay imposed on the following planes in the queue during the busy period. One of the main results is that landing fees are inefficient, because the congestion costs are significantly higher than what is paid by air carriers. This paper explores the possibility of imposing a congestion toll. A contemporary paper of Levine (1969) also advocates more differentiated fees depending on the time of the day in order to reflect congestion during peak hours.

A substantial literature follows these seminal articles. Some of them propose to assess empirically the cost of congestion. This is notably the case of Morrison and Winston (1989). This paper intends to estimate econometrically the relationship between airport activity and arrival and departure delays using US data. It clearly exhibits that an increasing level of activity causes an increase in average delays. In other words, when capacity is used to its fullest, an additional slot increases the probability of delays due to a reduction in the ability to recover from an incident. Another interesting contribution to this literature is given by De Rus and Nombela Merchan (2006), which propose a desegregated analysis of airport delays in Madrid Barajas.

Brueckner (2002) points out that, when an air carrier is dominant in an airport, an optimal pricing rule should charge only the cost of delays it imposed on other companies, because the dominant company internalizes the cost of delays that is imposed on its own flights. Nevertheless, Morrison and Winston (2007) quantifies the welfare gains from modifying congestion pricing to take into account the internalization of dominant air carriers at hubs, and finds a small difference between the net benefits generated by the optimal congestion pricing suggested by Brueckner and more traditional congestion tolls, not taking into account the dominance of a company in a hub.

3. RAILROAD CONGESTION FROM AN EMPIRICAL POINT OF VIEW

a) Two types of congestion

As for road or air transport, congestion in rail transport means that the total travel time increases with the traffic density. In the rail transport case, the increase in travel time is mainly due to the delays. The importance of rail congestion was noticed by the High Level Group on infrastructure charging (Nash, 1999) which noticed that, when traffic approaches capacity, delays are expected to be more frequent.

In rail transport, there are two types of delays. The first one is generated by the operational constraints of a railroad network: delays for meets with opposing rail traffic on single-track lines, and for following and overtaking slower rail traffic moving in the same direction, for example. This is why railroad traffic needs to be scheduled, just like air traffic, beyond obvious safety reasons: if traffic is badly scheduled, there will be a lot of unexpected delays. Nevertheless, efficient scheduling does not imply no delay compared to the "free flow" situation, since it will drastically reduce the capacity of the network: it rather involves a trade-off between speed, expectable delays and capacity, as soon as the infrastructure manager is able to assess the impact of an additional train on the travel times of all the existing trains.

The second type of delays is originated by an incident (failure of the rolling stock, failure of the infrastructure, inadequate behaviour of the crew, etc.). This incident generates delays to the following trains, and given the complexity of the network a lot of trains can be affected, even on different sections of the network. These transmitted delays are obviously unexpected. The likelihood of delays increases as capacity utilization rises, because heavy traffic reduces the network manager's ability to resolve the incident, and the delay is transmitted to more trains, with a snowballing effect.

b) Capacity and expectable delays

Scheduling requires estimating the travel times and delays in the network for each train, which depends on the capacity of each section of the network. Nevertheless, the actual capacity is neither easily defined nor quantified. The capacity of a railway network and the delay across it are closely related. If "delay" is defined as the difference between the actual running time and the free running time (i.e. the time the train takes to traverse the network, when traveling at its maximum allowable speed and not experiencing delays due to other trains), the delays encountered by trains under different operating assumptions can be used to evaluate the capacity of a section of a network. Capacity can be defined as the number of trains that can safely coexist in a network, or a portion of it, when interference delays are taken into consideration.

Capacity analysis in railway transportation is dependent on various operational aspects. The first aspect is the track configuration. The network can consist of single, double, triple or even more track, there can be more or less junctions, and the signalling system can allow for more or less trains. The second aspect is the characteristic of each train, such as train length, speed, acceleration rate and deceleration rate (they need to be considered in order to increase or reduce speed without violating the speed limit), and priority (to cross a junction or seize a track, the train with the lower priority should wait and stop until the train with higher priority passes). The third aspect is the speed limits on the different track segments and junctions. Sometimes trains cannot be dispatched at their maximum speed; different trains can have different speeds limits even though their paths may use the same tracks, etc.

There exists a considerable literature of both analytical and simulation-based methods, which have been used to study delays and capacity assessment in railroad line haul networks with specific configurations. Frank (1966) studied delay on a single track rail line with unidirectional and bidirectional traffic. The author estimated the number of trains that could travel on the network by considering only one train on each link between sidings and using single train speeds, and assuming deterministic travel times. This work was later extended by Petersen (1974) to accommodate for two different train speeds, while assuming independent and uniformly distributed departure times, equally spaced sidings and a constant delay for each encounter between two trains.

More recently, Chen and Harker (1990) extended this model to calculate delays for different types of trains over a specified single track section as a function of the schedules of the trains and the dispatching policies. They assumed a constant probability of delay between trains. Greenberg and al. (1988) proposes a queuing model on single track, low speed rail lines, assuming that trains departures follow a Poisson process. Higgins and Kozan (1998) present a model in urban networks and quantify the expected delays for passenger trains on a complex multitrack rail network. This paper also investigates the influence of modifying scheduled slack time on expected delays. It suggests that, although large reductions in expected delays are achievable with small amount of slack time, little improvements are observed when slack time is increased further (e. g. from 8% to 16%). Dessouky and Leachman (1995) used a simulation modelling methodology to analyse the capacity of tracks and delay to trains in a complex rail network. Krueger (1999) used simulation to develop a regression model to define the relationship between train delay and traffic volume. Yuan and Hansen (2007) proposed probability models that provide an estimate of delays and the use of track capacity. Murali et al. (2010) presented a simulation-based technique to generate delay estimates over track segments as a function of traffic conditions, as well as network topology to facilitate routing and scheduling freight trains.

The infrastructure manager internalizes the expected delays when he designs the path of trains. This does not signify the absence of delays in the network. Residual delays appear since we observe a second type of delays: the unexpected delay.

c) Capacity and unexpected delays

To a certain extent, the unexpected delays are also taken into account in the scheduling process: the infrastructure manager designs the paths with a slack time (the headway is larger than required by safety issues). The smaller is the slack time, the greater is the number of paths offered, but also, the lower is the resilience of the timetable, and the greater is the probability of delays: on the one hand, large headways reduce the transmission of delays, but on the other hand, they also reduce capacity. The slack time can be very significant: it represents 20% of the capacity on the French high-speed lines, for instance.

Here again, scheduling involves a trade-off between delays and capacity, so that infrastructure managers cannot increase the headways enough to supress all transmission of delays. The headways are generally design more by trial and error than through intensive modelling and computing. In order to understand the effects of a marginal train on the transmitted delays, and thus improving the scheduling, it is interesting to measure the increase in unexpected transmitted delays due to an increase in traffic.

As expressed previously, the idea of a relationship between traffic density and unexpected delays is quite familiar in airport economics. In comparison, there are very few papers dealing with this issue in the railway sector. Gibson et al. (2002) develop a regression model to define a correlation between capacity utilization and unexpected delays for the British network. In this study, an exponential form was chosen to estimate for the relationship between capacity utilization (C_{it}) and reactionary delay (D_{it}) across the network.

The regression analysis was performed for 24 strategic routes of the British network .The results of the analysis show that there is a positive relationship between capacity and unexpected transmitted delays.

Similarly, an extensive econometric analysis has been conducted for the French railway network, with comparable results (Brunel, Marlot and Perez, 2013). This study focuses on 42 lines of the French railway network, with 3 measuring points for each line. It shows a positive econometric relationship between traffic and unreliability rate or the length of delay. According to the line and its features (allowed speed, number of tracks, signaling…), it shows a positive econometric relationship between traffic and unreliability rate: an additional train on the line increases the probability of delays, for it and for the other trains. The marginal congestion cost is made up of a direct effect which is internalized by the supplementary train and of an indirect effect that generates an external cost on next users.

For example, according to this study, when traffic equals six trains per hour, in a high speed line, an additional train causes around 1 extra minute delay on the next trains. In the case of an intercity line with high traffic density, the direct effect of an additional train when traffic equals 12 trains per hour, an additional train causes around 8 extra minute delays on the next trains. The varied traffic lines density emphasizes that congestion would not emerge with the same intensity in the entire network.

4. THE WELFARE VALUE OF CONGESTION AND SCARCITY

a) The need for a welfare value of railway congestion

Our analysis of the scheduling process suggests that it involves a trade-off between capacity (i.e. the number of trains allowed to access the network), on the one hand, and delays, both expectable or not, on the other hand: if there are too many trains on the network, their expected travel time will be long, with either queues forming at the bottlenecks or a reduced speed; moreover, the likelihood that a delay originated by an incident propagates to a lot of trains will be very high. In order to optimize the scheduling process, it would be necessary to consider the cost of capacity shortage. Nevertheless, it is not an easy task to value this cost.

The analytical tools described in the literature following Franck (1966) could be useful for an infrastructure manager aimed at providing efficient price signals to competing train operating companies, and optimizing its own production processes (investments, scheduling, path allocation), regarding not only commercial revenues but also welfare impacts. Nevertheless, in the context of a vertically separated railroad system, with an independent infrastructure manager and different train operators and activities, or if the commercial revenues do not reflect the value of the services for the travellers (for example if the train services are heavily subsidized, which is often the case for urban and suburban services), these methods are pointless without an economic value of congestion.

Giving an economic value to congestion is not an easy task as soon as one cannot observe congestion on a “free access” network, as it is the case for road congestion. In the railroad case, one can only observe the congestion resulting from given scheduling process, path allocation process and access charges. From a train operator point of view, as soon as the expectable delays are internalized through the scheduling process, the capacity shortage takes two different forms:

- congestion, i.e. unexpected transmitted delays related to the density of traffic;
- scarcity, i.e. the inability for a train operator to obtain the desired train path.

The distinction between pure congestion and scarcity is frequently made in the literature dealing with railway congestion pricing. In particular, the High Level Group on infrastructure charging (Nash, 1999) specifies the case of rail congestion noticing this distinction between congestion and scarcity. Quinet (2003) explains that rail infrastructure charges should include external costs, including the costs of delays due to heavy traffic, and that the cost of scarcity should be revealed using auctions or priority rules. Quinet (2003) concludes that given the actual oligopolistic structure of rail market a combination of central planning and action seems to be the best solution to reveal the value of slots for the rail operators. Nash & Matthews (2005) do not specifically address the difference between congestion and scarcity. Their paper is more interested in the issue of scarcity, and exhibits several methods to measure this cost.

The following sections propose to review several methods to value capacity shortage: scarcity and congestion in rail transport.

b) Measuring the welfare value of delays

The delay or the rejection of a train has no welfare value in itself. It is a social cost only insofar as there are people who are experiencing delays, or feeling uncomfortable because of the overcrowding of the trains, or who are not able to travel at the desired time, or not able to travel at all. The welfare value of the capacity shortage is mainly the value of the welfare losses due to overcrowding, shifts in travel time or impossibility to travel (linked to the scarcity of paths and the expected delays internalized through the scheduling process), and delays (linked to the trade-off between capacity and timetable resilience).

In order to value congestion, it is necessary to know the marginal cost (delays) of adding a supplementary train in the rail system. Firstly, it is necessary to define the total delay in a railway line due to an increase of one traffic unit (the marginal delay). If the consequences of adding a new train are given in minutes, travel time can be converted to a monetary basis using a value of time as a shadow price for the train user. This valuation should be adjusted in order to take into account the value of reliability, which measures the willingness to pay for the reduction in the variability of travel time.

This parameter has received a recent interest in the academic literature. According to it, reliability should be considered using a reliability multiplier. This multiplier values one minute of unexpected delay relative to scheduled travel time. In a recent paper, Börjesson and Eliasson (2011) have made a comparison of the reliability multiplier's values obtained in transports economics literature: Wardman (2001) finds an average value of 7.4. In a recent paper, Abrantes and Wardman (2011) point out an average value of 6.4 with a standard deviation of 3.8. Even so, Rietveld et al. (2011) propose a value of 2.4.

c) Measuring the welfare value of scarcity

Nash & Matthews (2005) explains that scarcity cost can be revealed using a market based approach and particularly, auctions. This solution is in theory the best solution to reveal these values but, in reality, actions are too difficult to settle, given the number and the complexity of slots that should be allocated. An alternative method would be to design an auction with pre-packed slots, or to allocate rail capacity to several potential rail operators through priority rules, as suggested by Quinet (2003). Nevertheless, pre-packed slots and priority rules internalize, to a certain extent, the value of scarcity, so that such processes will not reveal the full scarcity cost. These are second best solutions.

An alternative approach should be to value the opportunity cost of any particular allocation of slot. It can refer to a cost-benefit approach, which values consumer cost to travel later or earlier from the desired time. The information required for this method is arduous. It is necessary to know what the passengers do when the capacity is constrained; to what extent they shift their departure time and take different trains, or simply cancel their journey.

Moreover, in urban areas, the cost of transport congestion is not the cost to travel later or earlier but the cost of travelling in an overcrowded environment (standing, standing packed). Various studies have been made in order to value this cost, generally thanks to stated preferences surveys. Wardman and Whelan (2011) purpose a meta-analysis of crowding studies in the UK, exclusively SP surveys. These studies generally conclude that overcrowding induces time multipliers, generally around 1.5-2.

A final source of complexity for the estimation of the optimal rail congestion cost is that there are few operators, the market being oligopolistic. The dominant rail operator can be interested in rising prices, limiting supply in several markets. This strategy of the dominant operator generates a social loss of welfare. There are fewer passengers that what would be socially optimal. This loss of welfare is hard to value as we have mentioned before since it refers to the opportunity cost of the slots that are not supplied. Furthermore, the valuation of this cost raises many questions: What would have been the strategy of an operator under perfect competition? What prices would the operator propose? What would be the induced demand? These questions are crucial in order to determine the real cost of capacity shortage in rail market. These questions do not have stills answers .They call for further researches.

To conclude, we can observe that the cost of capacity shortage can be computed from several methods. The valuation of this cost (scarcity and congestion) provides a good benchmark in order to determine an optimal scheduling. In what follows, we propose to enlarge this analysis in order to show that the value of congestion also interests pricing or investment policy.

5. OPTIMAL PRICING, SCHEDULING, AND CAPACITY INVESTMENTS

a) Optimal scheduling

The network manager must offer an optimal number of paths and realize an optimal trade-off between resilience and capacity. As explained above, the congestion costs internalized through the scheduling process must be taken into account, along with the congestion and scarcity costs. It must value the cost of the capacity that is not provided to train operators, but allocated to the time margin designed to increase the resilience of the timetable, and the cost of expected delays.

Assessing the optimality of scheduling through a cost benefit analysis is probably very difficult, because there is probably no analytical solution, and it is not possible to compute, at the scale of the whole network, every possible schedule, and every possible allocation of the paths to the different train operators.

Nevertheless, simulation models allow the testing of different schedules on a part of the network; assumptions can be made to simplify the problem; and finally central planning is probably necessary to a certain extent, for example to define priority rules, or to define some of the services. These questions call for further researches.

b) Optimal pricing

We have shown above that despite some slack time taken into account on the scheduling process, we can observe residual, unexpected delays. These delays, due to congestion, have a negative impact on social welfare, and this effect can be valued. This phenomenon can be understood as a standard externality problem, as explained above. The literature on road congestion is extensive on this topic. In rail transport, the problem is basically the same. A train operator deciding to use the network during peak-hours generates an external cost on other users. In order to realize an efficient use of the network, it is necessary to make the train operators pay for this external cost through adequate track access charges.

However, this congestion fee would only be optimal under several conditions. We can identify at least two of them. First, the infrastructure must offer an optimal number of paths and realize an optimal trade-off between resilience and capacity. In the same way, the infrastructure manager must increase the capacity when the welfare benefit of the additional capacity is superior to its cost (building and operating). It means that the short term congestion fee must necessarily be linked to the long term perspective of capacity investments. .

c) *Optimal capacity investments*

If congestion is the result of a trade-off between capacity and delays in the scheduling process, then one can expect that, from a long-term point of view, the optimal level of congestion is the result of an adequate investments policy. The infrastructure manager can reduce the congestion when he invests in additional capacities. In particular, one may believe that a benevolent infrastructure manager should expand capacity until the marginal benefit equals the marginal cost of building it. On the contrary, the infrastructure manager could limit the capacity or the number of paths offered, in order to raise unjustified profits by limiting the capacity. This is why the natural monopoly must be regulated. The regulator must compare the congestion cost (and, if needed, the scarcity cost) to the cost of an increase in capacity. It must also assess the optimality of scheduling through a cost benefit analysis.

From the infrastructure point of view, scheduling and capacity investments are only different ways to tackle a capacity shortage, short term and long term. One must be aware that, if the scheduling process internalizes some of the congestion, which should not be taken into account in the optimal track access charges, the value of this internalized congestion should be taken into account in the assessment of capacity investments.

On the short-term, the infrastructure manager is looking for an optimal trade-off between total travel time (including the expectable delays associated with the interference of the trains on the network, and the probability of unexpected delays) and capacity (the number of paths offered, i.e. the number of trains that can actually access the network). It needs to balance the costs of expectable and unexpected delays *versus* the cost of restraining the number of paths (rejecting some of the trains).

On the long-term, the infrastructure manager is looking for an optimal trade-off between the total costs of the capacity shortage (expectable delays, unexpected delays and the part of demand that is not satisfied, whether it has been rejected or deferred) and the cost of increasing the capacity of the network.

As a result, from an economic point of view, there must be a relationship between the costs of expectable delays, which are already internalized through the scheduling and path allocation process, the costs of unexpected delays, the costs of rejecting some of the trains (scarcity costs), and the costs of increasing the capacity of the network. Nevertheless, the analytical solution to this problem remains to be defined.

6. CONCLUSION

The economic literature suggests that congestion can appear in sectors where traffic is scheduled by advance. In the railway sector, most of the congestion can be internalized through the scheduling process, leading to reduce either speed or capacity. Nevertheless, all the congestion cannot be internalized through the scheduling process: the likelihood of delays increases as capacity utilization rises, because heavy traffic reduces the network manager's ability to resolve the incident, and the delay is transmitted to more trains, with a snowballing effect. Econometric analyses afford evidences of this congestion, at least in Great Britain and in France.

From an economic point of view, this observation supports the implementation of a congestion charge in order to provide incentives for an efficient use of the network. Estimates of congestion costs suggest that these costs are generally low. Congestion is limited to a very little number of lines and nodes of the network. Nevertheless, in these cases, congestion costs could be very high.

Moreover, congestion costs, even internalized, must be taken into account in the scheduling process and in the economic appraisal of capacity investments. Nevertheless, estimating the welfare costs of the congestion internalized through the scheduling process is arduous. One can observe that in Europe there are very few cases of scarcity as the directive 2001/14 defines this concept. Until now, rail operators are rarely competing for the same paths. This could be explained by the fact that liberalization of rail in Europe is recent and unachieved. In a monopolistic market, train operators are internalizing a part of the congestion, through their own trade-off between different trains and activities, and pricing.

7. REFERENCES

Abrantes, P.A.L., Wardman, M.L., (2011) Meta-analysis of UK values of travel time: an update. *Transportation Research Part A 45 (1), 1–17.*

Arnott, R., De Palma, A., Lindsey, R. (1993) A structural Model of Peak-Period Congestion: A Traffic Bottleneck with Elastic Demand, *American Economic Review,* 83 (1), pp. 161-179.

Börjesson, M., Eliasson, J. (2011), On the use of “average delay” as a measure of train reliability, *Transportation Research Part A,* 45, pp. 171-184.

Brueckner, J. K. (2002) Airport congestion when carriers have market power, *American Economic Review*, 92 (5), pp. 1357–1375.

Brunel J., Marlot G., Perez M. (2013) Measuring congestions in rail sector : the French experience.

Carlin, A., Park, R. E. (1970) Marginal cost pricing of airport runway capacity, *American Economic Review,* 60 (3), pp. 310-319.

Chen, B., Harker, P.T., (1990). Two-moment estimation of the delay on a single-track rail line with scheduled traffic. Transportation Science 24 (4), 261–275.

Chu, X. 1995. Endogenous trip scheduling: The Henderson approach reformulated and compared with the Vickrey approach. Journal of Urban Economics 37: 324-343.

Dessouky, M.M., Leachman, R.C., (1995). A simulation modeling methodology for analyzing large complex rail networks. Simulation 65 (2), 131–142.

Doll C (2002) Transport User Cost and Benefit Case Studies. UNITE (UNIfication of accounts and marginal costs for Transport Efficiency) Deliverable 7. Funded by 5th Framework RTD Programme. ITS, University of Leeds, Karlsruhe, November 2002

Frank, O., (1966). Two-way traffic in a single line of railway. Operations Research 14, 801–811.

Gibson, S., Copper, G., Ball, B. (2002) Developments in transport policy: the evolution of capacity charges in the UK network, *Journal of transport economics and policy*, 32 (2), pp. 341-354.

Greenberg, B.S., Leachman, R.C., Wolff, R.W., (1988). Predicting dispatching delays on a low speed single track railroad. Transportation Science 22 (1), 31–38.

Harker, P.T., Hong, S., (1990). Two moments estimation of the delay on a partially double-track rail line with scheduled traffic. Journal of Transportation Research Forum 31 (1), 38–49.

Hau, Timothy D. (1998) Congestion Pricing and Road Investment. In Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility, ed. Kenneth J. Button and Erik T. Verhoef. Cheltenham, UK: Edward Elgar, 39-78.

Higgins, A., Kozan, E., (1998) Modeling train delays in urban networks. Transportation Science 32 (4), 251–356.

Krueger, H., (1999) Parametric modelling in rail capacity planing. In: Proceedings of the 1999 Winter Simulation Conference, pp. 1194–2000.

Levine, M. E. (1969) Landing fees and the airport congestion problem, *Journal of Law and Economics*, 12, pp. 79–108.

Mohring, H., Harwitz, M. (1962) *Highway Benefits: An Analytical Framework*, Evanston, Illinois: Northwestern University Press.

Morrison, S., Winston, C. (1989) Enhancing performance of the deregulated air transportation system, *Booking paper on economics affairs, microeconomics*, pp. 61-123.

Morrison, S., Winston, C. (2007) Another Look at Airport Congestion Pricing", *American Economic Review*, 97 (5), pp. 1970-1977.

Nash, C. (1999) *Calculating transport congestion and scarcity costs, Final report of the expert advisors to the High level group on infrastructure charging (working group 2)*.

Nash, C., Matthews, B. (2005) *Rail infrastructure charges – the issue of scarcity,* Working paper, University of Leeds.

Nombela Merchán, G., De Rus, G. (2006) Analisis economico de la congestion en los aeropuertos europeos in De Rus, G. (ed.) *La politica de transporte europea : el papel del analisis economico*, Fundacion BBVA.

Pigou, A.C. (1920), *The Economics of Welfare*, London : MacMillan.

Petersen, E.R., (1974.) Over the road transit time for a single track railway. Transportation Science 8, 65–74.

Petersen, E.R., Taylor, A.J., (1982) A structured model for rail line simulation and optimization. Transportation Science 16, 192–206.

Quinet, E. (1997) Full social cost of transportation in Europe, in Greene, D. L. et al. (ed.) *The full costs and benefits of transportation: Contribution to theory, methods and measurement*, Springer.

Quinet, E. (2003) Short term adjustment in rail activity, *Transport Policy*. 10 (1), pp. 73-79.

Van Loon, R., P. Rietveld, and M. Brons.(2011) Travel-time Reliability Impacts on Railway 13 Passenger Demand: a Revealed Preference Analysis. Journal of Transport Geography, 14 Vol. 19, 2011, pp. 917-925.

Verhoef, E.T. (2001) "An integrated dynamic model of road traffic congestion based on simple car-following theory: exploring hypercongestion" Journal of Urban Economics 49 505-542

Vickrey, W. S. (1963) Pricing in urban and suburban transport, *American Economic Review* 53 (2), pp. 452-65.

Vickrey, W. S. (1969), Congestion Theory and Transport Investment, *American Economic Review*, 59 (2), pp. 251-260.

Walters, A. A. (1961) "The theory and measurement of private and social cost of highway congestion", Econometrica, 29(4):676-697

Yuan, J., and Hansen, I. A. Optimizing capacity utilization of stations by estimating knock-on train delays. Transportation Research Part B 41 (2007), 202-217