MEASURING CONGESTION IN RAIL SECTOR: THE FRENCH EXPERIENCE

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

In the rail sector, there are few academic papers dealing with congestion. Most papers focus on scarcity of train paths, or, following Petersen (1974), the relationship between scarcity and speed. Gibson and al. (2002) examine another relationship, between traffic and delays: one can expect delays to increase as capacity utilization rises. Heavy traffic reduces the network manager's ability to resolve an incident and in consequence we observe a snowball effect in the transmission of the delay.

This paper presents an econometric analysis of the relationship between traffic density (i.e. number of trains per hour), reliability rate and average delay can be a relevant method to assess congestion. In this research, a database from the French rail infrastructure manager is used, which records traffic information in the French network, and notably the delay at each measuring point. These data give a precise account of the delays and the traffic density. We focus on 42 lines of the French railway network, with 3 measuring points for each line. The dataset includes 6.4 million trains.

The econometric analysis shows a positive econometric relationship between traffic and unreliability rate: an additional train on the line increasing the probability of being late, we are able to compute a marginal cost of congestion, which can be used for pricing and cost benefit analysis of capacity investments. This study may support the implementation of a congestion charge in the French network. It also provides à guideline to the network manager to invest in new capacity.

Keywords: congestion, rail transport, reliability, delay, capacity, resilience, externality, marginal cost.

1. INTRODUCTION

In its statement on the network reference document of 2012, the French rail regulator (Autorité de Régulation des Activités Ferroviaires, ARAF) states that congestion is a crucial issue for the railway system and that track access charges shall provide incentives for a better use of the network (statement 2011-002 of the 2nd February 2011).

Rail capacity is one of the most pressing issues facing the French network. Demand for rail transport is growing in metropolitan areas. Expansion of rail capacity faces a range of obstacles and financial challenges. Nevertheless, there are few academic papers dealing with congestion in rail transport. Congestion is traditionally supposed to occur when demand on an infrastructure exceeds the available capacity. Things can be a little more subtle: most of the economic literature about road congestion relies either on a static speed/flow relationship, which can be observed through road statistics or on a dynamic peak-load approach (bottleneck model, etc.). This literature is quite extensive. Nevertheless, literature about congestion is less extensive in sectors where traffic is scheduled in advance. Since the seminal papers of Levine (1969) and Carlin et al. (1970), some papers estimated marginal congestion cost in air transport. In the rail sector, very few academic papers deal with congestion. In this context, Réseau Ferré de France undertakes various studies in order to objectify congestion on the national rail network. The present study aims to investigate the influence of a traffic density on reliability rate and average delay on a given rail track. This analysis has been realized by the economic consulting firm Microeconomix.

The structure of the paper is as follows. Section 2 briefly reviews the academic literature dealing with the economics of congestion in transport, and notably in rail transport. It presents the intuition behind this paper. This section underlines that, in the presence of a high traffic density, the railway infrastructure is more exposed than a low traffic track, resulting in a higher probability of delays. This intuition is empirically confirmed by Gibson et al. (2002). Section 3 provides a detailed description of the methodological framework. Section 4 describes the data set used in the present paper. Section 5 exposes the results of our analysis. At last, section 6 offers concluding remarks.

2. RELATED LITERATURE

Many networks suffer from peak-load demand problems. In general, congestion refers to the existence of limited capacity networks whose demand varies periodically. The economic literature concerning congestion is quite extensive in the road sector. When cars users decide to make an additional trip, they impose additional costs on themselves, on the infrastructure provider and on other users. The literature shows that pricing congestion allows users to internalize the external costs generated and reallocate the traffic demand during the day (Vickrey, 1963). The literature also shows that, given certain circumstances, congestion pricing covers the construction costs of highways (Mohring and Harwitz, 1962, Arnott and al., 1993, Hau, 1998).

Congestion also appears in sectors where traffic is scheduled in advance. A sizeable literature has studied airport congestion that happens in the neighborhood of large airports due to runways or traffic control saturation following the papers of Levine (1969) and Carlin et al. (1970). This literature includes several papers which empirically estimates the marginal congestion cost in air transport (e. g. UNITE, 2002, Morrisson and Winston, 1989, Nombela Merchan and de Rus, 2006). These papers verify a relationship between traffic density and the probability of delays in the airport industry: when capacity is highly used, an additional slot increases the probability of delays due to a reduction in the ability to recover from an incident.

Few academic papers consider congestion in rail transport. Some notable exceptions are the High Level Group on infrastructure charging (Nash, 1999) and papers of Quinet (2003) and Nash and Matthews (2005). These papers specify the case of pricing railway congestion from a theoretical point of view.

From an empirical point of view, there are relatively few papers which estimate rail congestion. These papers have estimated two types of congestions. The first one is the expected congestion. This type of congestion refers to the delays 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. There exists a considerable literature of both analytical and simulation-based methods, which study delays and capacity assessment in railroad line networks with specific configurations following Frank (1966) or Petersen (1974).

A second type of delays is originated by an incident (failure of the rolling stock, failure of the infrastructure, inadequate behavior of the crew, etc.). An 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 delays are obviously unexpected. They increase 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. This idea is quite familiar in airport economics as stated previously. It is also intuited by the papers of Quinet (2003) and Nash and Matthews (2005) for rail transport.

These delays can be measured with an adequate monitoring system. For instance, it has been empirically studied in the British rail network by Gibson et al. (2002). In this paper, a regression analysis confirms the existence of a relationship between capacity utilization and delays. 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. This relationship is given by:

$$D_{it} = A_i \exp(\beta C_{it})$$

where A_i is a route section specific constant and β is a route specific constant.

The regression analysis was performed for 24 strategic routes on British network using 1998 and 1999 financial year data on reactionary delay per train and a measure of capacity utilization based on the winter 1998 timetable.

The results of the regression show that β is statistically significant for 20 out of the 24 routes. It means that it exists a positive relationship between capacity and reactionary delays. This relationship justifies the congestion charge implemented since 2001 by Network Rail. An additional path increases the probability of delays and, therefore, its monetary cost in a performance regime framework.

In what follows, we propose to investigate the existence of this relationship between traffic and delays in the case of the French network, using an internal dataset of the French rail infrastructure manager.

3. THE MODEL AND THE ECONOMETRIC STRATEGY

The present section proposes a mathematical framework in order to estimate empirically the marginal congestion cost in railways. This mathematical framework enables to isolate the marginal effect of a train on the total delays.

In this section, we notice R_i^* the deviation between the real time and the scheduled time of a train for a given traffic density Q_i. The train can be on time ($R_i^* = 0$), early ($R_i^* < 0$), or late ($R_i^* > 0$).

We define the variable R_i representing the delay of train. One can therefore notice:

$$R_i = \begin{cases} 0 \quad si\ R_i^* \leq 0 \\ R_i\ si\ R_i^* > 0 \end{cases}$$

The expected delay of train for a given traffic density is:

$$E(R_i) = p(R_i^* \leq 0) \cdot E(R_i | R_i^* \leq 0) + p(R_i^* > 0) \cdot E(R_i | R_i^* > 0)$$

As the expected delay is null when the train is on time or early ($p(R_i^* \leq 0) \cdot E(R_i | R_i^* \leq 0) = 0$), this equation can be written:

$$E(R_i) = p(R_i^* > 0) \cdot E(R_i | R_i^* > 0) \quad (1)$$

This equation indicates that the expected delay of train for a given traffic density is equal to the product of the expected delay of delayed trains and the number of trains delayed.

The total amount delays of trains for a given traffic is, by definition, the expected delay of train multiplied by the number of trains, i. e. $Q_i \cdot E(R_i)$. Therefore, it follows that the marginal delay imposed by an additional train is the derivative of the total amount of delay function with respect to the level of traffic.

It can also be written as:

$$\frac{\partial (Q_i.E(R_i))}{\partial Q_i} = Q_i \left(\frac{\partial E(R_i)}{\partial Q_i}\right) + E(R_i) \quad (2)$$

In this equation, the right hand term is the expected delay of the additional train given the traffic density: this is a direct effect, internalized by the train. The direct effect is equal to the expected delays for a given traffic density. This term, expressed by equation (1), can be directly computed from the data set.

The left hand term of equation (2) represents the marginal delay imposed by the additional train on the following trains. It is an indirect effect which corresponds to the pure externality effect of congestion. The indirect effect, cannot be computed directly and needs and econometrical analysis in order to be estimated.

Using equation (1), the indirect effect can be rewritten as:

$$\frac{\partial E\ \ R_i\ \ Q_i}{\partial Q_i} = \frac{\partial}{\partial Q_i}\ p\ \ R_i^* > 0\ \ \cdot E\ \ R_i\ R_i^* > 0$$

$$= \frac{\partial p\ \ R_i^* > 0}{\partial Q_i} E\ \ R_i\ R_i^* > 0\ \ +\ \ R_i^* > 0\ \ \cdot \frac{\partial E\ \ R_i\ R_i^* > 0}{\partial Q_i} \qquad (3)$$

In this expression, the expected delay $E\ \ R_i\ R_i^* > 0$ and the probability of being late $p\ \ R_i^* > 0$ are known from the data set. Nevertheless, the two derivatives should be estimated.

The first term ($\frac{\partial p\ \ R_i^*>0}{\partial Q_i}$) (a) describes the marginal effect of an additional train on the probability of being late. The second one ($\frac{\partial E\ \ R_i\ R_i^*>0}{\partial Q_i}$) (b) represents de marginal effect of an additional train on the expected delay.

In what follows, we propose to estimate the first effect using a probit model. For each train, the dependent variable equals 1 if the train is delayed ($R_i\ si\ R_i^* > 0$) and 0 if not. The marginal effect on the probability of being late is directly calculated using the results from the probit regression for each level of traffic.

The second effect is estimated using a linear regression:

$$y_{ij} = \alpha_j + \beta x_{ij} + \varepsilon_{ij}$$

where y_{ij} is the delay for train *i* in the measuring point *j*, x_{ij} is the traffic density associated at each observation and ε_{ij} is the error term. In this expression, α_j is supposed to be specific for each measuring point. Both the probit and the linear models regressions include fixed effects in order to consider the delay heterogeneity associated to the different measuring points, direction and train use (passenger, freight, etc.).

4. THE DATA SET

In this empirical research, we use data from an internal database of Réseau ferré de France in order to estimate the previous parameters. This internal database records traffic information in the French network, and notably the delays at each measuring point. The data provided by this database allow us to know precisely the performance (reliability rate and delay) of each line at each level of traffic.

The data is recorded by an automatic system which detects the train circulation and registers the traffic details concerning the train. These automatic measuring points are associated to the measuring points which are utilized for the construction of the schedule. The system allows obtaining, for each train which crosses a measuring point, the data presented in the following table.

Table I: Summary variables

VARIABLE	DESCRIPTION
Internal circulation number	Specific and unique number associated at each train
Circulation number	Number associated to a specific stopping pattern
Date/Hour:	Date et real hour when the train crosses the measuring point
Week day	-
Timetable type	Determines the kind of stop: Origin, Passage, Arrival or Departure (for a stop in a train station) or Terminus
Time deviation	It is the deviation between the real time and the scheduled time (delay)
Statistical category	Informs about the train activity (HSL, regional activity, national activity, freight, etc.) and if the train is loaded or empty

However, railway lines have different characteristics. They have diverse uses (passenger trains or freight trains), different traffic densities (lines with heavy traffic or lines with low traffic) and varied levels of performance. For that reason, we have subdivided the French network in several groups of lines with similar characteristics

In this classification, the network is divided in 4 categories depending on uses (freight, regional, national) and speed levels (high speed lines or not):

- High speed lines: routes with a speed higher than 250kph
- Intercity lines: routes between population centers mainly used by freight and passenger long distance trains.
- Regional lines: routes between suburbs, towns and cities, without special speed requirements, and mainly used by regional and commuters trains.
- Only freight lines: freight specific routes with no mixed traffic, and generally low traffic density.

At the same time, these categories are subdivided in subcategories depending on the traffic density (trains per weekday per route): high, medium or low traffic density. The traffic is highly concentrated around several nodes of the networks. For example, we can observe lines with 15 trains per hour during the peak-hours period in some regional railway lines near Paris. By contrast, some local lines can only have one train per hour during the peak-hours. The varied traffic lines density emphasizes that congestion would not emerge with the same intensity in the entire network.

In the present study, we have focused our analysis on 42 lines of the French railway network, with 3 measuring points for each line. The lines belong to these different groups of lines presented above. The dataset includes 6.4 million trains. These lines have been assembled in 9 subgroups using the strategic segmentation. The dataset used in this research contains all train circulations in these lines during 2011.

5. RESULTS

In this analysis, the variable traffic has to be defined. For each observation (each train recorded), we have obtained a level of traffic which equals the number of train scheduled in the same line and direction during the previous hour. Then, an econometrical analysis is pursued to measure the additional delay (in minutes) in a railway route due to an increase of one traffic unit (the marginal delay). As mentioned above, an additional train is likely to be delayed and to impose an additional delay on the next trains. The consequences of an additional train (direct and indirect effect) have been considered separately in our analysis, in order to assess the effect that an additional train generates on other trains. The indirect effect is the pure externality from an economist point of view whereas the indirect effect is internalized by the additional train.

Some of the parameters are directly computed using the data set. Some others are estimated with the econometric analysis, as described above. Two econometrical regressions are conducted in order the estimate the marginal cost of congestion (indirect effect) (in minutes) : the probit model which estimates the marginal effect of traffic on the probability of being late, and the linear model which estimates the marginal effect of an additional train on the expected delay.

The results of the econometric analysis are presented table II. The regressions have been estimated separately for the 9 groups of lines. Table II presents the results of the two regressions. The first column represents the average marginal effect of an additional train on the probability of being late. It is the parameter($\frac{\partial p \;\; R_i^*>0}{\partial Q_i}$) in equation (3). The second column represents the marginal effect of an additional train on the expected delay. It corresponds to the parameter ($\frac{\partial E \;\; R_i \; R_i^*>0}{\partial Q_i}$) in equation (3).

Table II: Regressions results

Strategic Classification	Type of line	Probit	Linear regression
G1	High Speed	0.0096*** (0.0024)	0.020** (0.017)
G2	Intercity lines	0.020*** (0.00042)	0.49** (0.12)
G3	Intercity/Regional lines	0.013*** (0.000054)	0.10** (0.018)
G4	Intercity lines high traffic density	0.022*** (0.00024)	0.67** (0.073)
G5	Intercity lines low traffic density	0.018*** (0.00057)	0.67 (0.30)
G6	Intercity lines medium traffic density	0.010*** (0.0011)	0.19 (0.14)
G7	Regional lines high traffic density	0.025*** (0.00024)	0.14** (0.024)
G8	Regional lines low traffic density	0.056*** (0.0064)	0.67** (0.31)
G9	Regional lines medium traffic density	-0.025*** (0.0024)	1.05 (1.10)

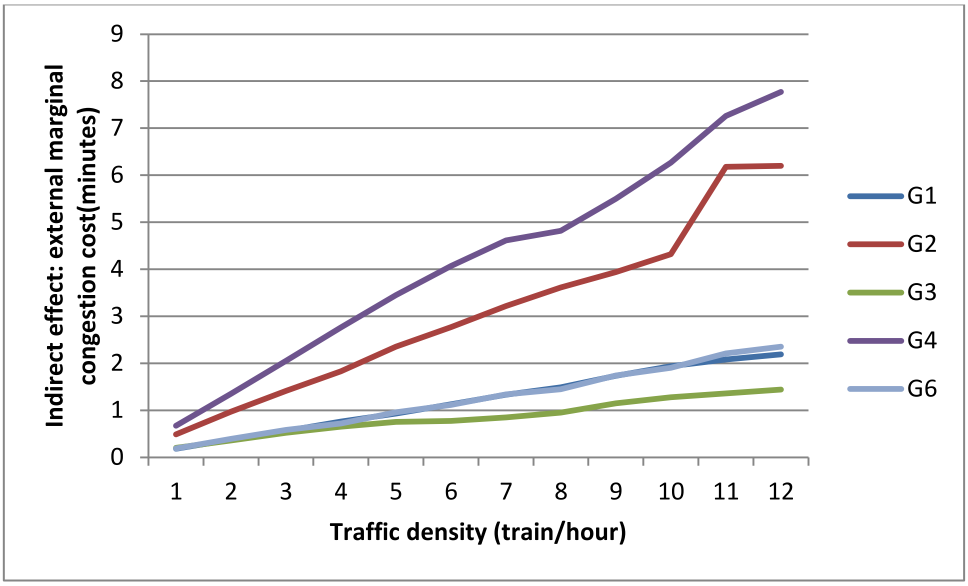
*Standard error in parentheses. *p<0.10, **p<0.05, ***p<0.001*

These results can be interpreted as follows: for high speed lines, an additional train increases the probability of being late by 0.96 points and increases the expected delay by 0.20 minutes for the following trains. Moreover, these results show that for certain types of lines, the congestion is not statistically significant. It is the case of intercity lines. It not surprising since this group corresponds to low traffic group of lines.

Once these two regressions have been estimated, it is possible to compute the average direct effect, given by equation (1), and the indirect effect by group, in accordance with equation (3). The results of these computations are presented in table III. This table can be interpreted as follows: for G4, an extra train has an average expected delay of 7.9 minutes (direct effect) and it adds 0.68 minutes in average for each following trains (indirect effect). One should notice that these effects are not constant with the traffic density. Some parameters used to compute the direct or indirect effect are varying in accordance with the level of traffic. For instance, the parameter $p(R_i^* > 0)$ which represents the probability of being late is computed for each level of traffic, and varies according the density of traffic. Therefore, these effects should be calculated by traffic density in order to know the total marginal congestion cost due to an additional train in each group.

Table III: Congestion marginal cost

Strategic Classification	Direct Effect	Indirect Effect
G1	4.45	0.19
G2	5.57	0.47
G3	2.73	0.13
G4	7.90	0.68
G5	7.35	0.59
G6	4.21	0.18
G7	2.01	0.19
G8	2.91	0.67
G9	3.24	0.30



In order to check the robustness of these results, some tests have been realized. A first test is realized in order to verify the existence of the relationship with another definition of delay. The previous results considered a train delayed if delay was superior to zero. Nevertheless, the data shows that many trains have in fact little delays (less than 5 minutes). A little delay associated to a train could be a measure error in some points, so we decided to test our results using a different delay definition. Two tests have been done considering only delays superior to three and five minutes respectively. In both cases, the estimated relationships are significant.

Some regressions analyses are also conducted for several specific points. We have considered that marginal effects are homogeneous between measuring points or lines in the same group. The test shows that there exist some differences between measuring points and lines. In some measuring points the congestion effects are higher than in others sections of the network, but the effect remains significant from a statistical point of view.

These results therefore provide strong evidence of our intuitive idea: an additional train increases the probability of late trains. It means that there is a form of unexpected congestion in the railways. The direct effect is internalized by the supplementary train, but the indirect effect generates an external cost on other users.

Delays increase the travelling time for passengers. It has a negative impact on social welfare. From an economic perspective, this phenomenon can be understood as a standard externality problem. High traffic density during peak-hours generates an external cost on other users. Track access charges can send to train operators the correct signal of the marginal social cost of adding a train. This pricing rule would allocate demand in an efficient way during the periods of the day. It would reflect the external negative effects that an additional train generates on travelling time for passengers. A congestion fee would permit to internalize the external cost imposed by the additional train on others when rail operators decide the number of paths supplied.

6. CONCLUSION

The present paper investigates one form of rail congestion. It shows that the economics of congestion in rail transport is nevertheless quite small in comparison with road transport for instance. An economic analysis of congestion in rail transport is proposed. This paper presents an econometrical analysis of delays on the French rail network, which establishes a relationship between delays and heavy traffic in several points of the French network. This study may support the implementation of a congestion charge and the improvement of cost benefit analysis methods. Nevertheless, the observed reliability rates in this study depend on the features of the line, the trade-off between capacity and resilience for the design of the train paths, and on the way train paths are allocated between different trains. In order to determine the optimal level of congestion and the optimal capacity policies, including congestion pricing, it is necessary to consider and analyze all these choices.

REFERENCES

- 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.

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

- 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.

- 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.

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

- Microeconomix (2012) Etude économétrique de l'effet de la congestion ferroviaire sur la régularité des trains à l'attention de RFF.

- 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*.*

- 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.

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

- Quinet, E. (2003) Short term adjustment in rail activity, *Transport Policy*. 10 (1), pp. 73-79.
- Vickrey, W. S. (1963) Pricing in urban and suburban transport, *American Economic Review* 53 (2), pp. 452-65.