Economics of Transportation 37 (2024) 100344

Contents lists available at ScienceDirect

Economics of Transportation

journal homepage: www.elsevier.com/locate/ecotra

The economics of regional railway regulation under vertical separation

Antonio Scialà [a,*,1], Francesca Stroffolini [b,1]

[a] Department of Law, Roma Tre University, Via Ostiense 159, 00154 Rome, Italy
[b] Università Federico II di Napoli, Italy

ARTICLE INFO

JEL classification:
L51
L92
L32

Keywords:
Regulation
Railway
Vertical separation
Local public transport
Misalignment of incentives

ABSTRACT

We provide a model of local railway passengers service able to account for the main specific characteristics of the sector under vertical separation. Afterwards, we use this model to carry out both a normative analysis of the operators' investment decisions and an assessment of the welfare effects of simple regulatory instruments. We show that, because of the information asymmetry of train operating company about the productivity of the infrastructure manager, the introduction of a regulatory instrument inducing the former to internalize the effect of her investment on the latter's cost of providing access may be welfare reducing.

1. Introduction

European railway transport is experiencing a long-lasting crisis in many respects. The companies operating in this sector are among the most subsidized public utilities and, compared with other transport modes, the market share is still low in most countries.[2]

Following what happened to other network industries, also in railways the unbundling between network management and service operation has been pursued in several national experiences as well as by the EU regulatory framework (Nash, 2008; European Commission, 1991, 2012). In fact, vertical separation is often considered a necessary step to foster competition in the downstream segment of any public utility industry that requires network infrastructure as an essential input.[3]

As in other network utilities, and given that the EU regulatory framework does not constrain member states to a particular model, more than one organizational model can go under the label of vertical separation. Among the organizational models of the railway system identified by the specialized literature, two have found greater application (Nash, 2008)[4]: the "complete separation" model and the "holding" model.

Very briefly, in the first model, the Infrastructure Manager (hereafter IM) cannot be involved in the operation of the service, while the Train Operating Company (hereafter TOC) cannot absolutely be involved in the management of the infrastructure. In the second model, the IM and the TOC belong to the same holding company but have completely separate functions.[5] Between these two somehow "extreme" models, we can observe a variety of hybrid systems, where differences are also present with respect to the public or private ownership of the operators.

On the empirical ground there is still debate about the costs and benefits of vertical separation as well as the conditions under which vertical separation should be preferred to vertical integration (Smith et al., 2018). Gathon and Pestieau (1995) and Cantos et al. (1999) are among the first studies to predict a positive impact of those reforms introducing vertical separation in railway industry as part of a reform package aimed at liberalizing the market. On the other hand, in a

* Corresponding author.
E-mail address: antonio.sciala@uniroma3.it (A. Scialà).
1 We would like to thank two anonymous referees for valuable comments.
2 See, Pflieger (2014), Arrigo and Di Foggia (2013a,b), European Commission (2016).
3 For a comprehensive view across sector see Newbery (2002). For the railway sector see Friebel et al. (2010).
4 On this point see also Laurino et al. (2015), Mizutani et al. (2015).
5 According to Nash (2008) the "complete separation" model was introduced for the first time in 1989 in Sweden, but it had been pushed to its higher level of disaggregation in UK after the privatization of UK railways. The "holding" model knew its first application in Germany, and it is also applied in other countries like Italy and France.
6 Among the other contributions that claim a negative effect of vertical separation on industry costs see Jensen and Stelling (2007) and Growitsch and Wetzel (2009).

https://doi.org/10.1016/j.ecotra.2024.100344
Received 18 October 2022; Received in revised form 16 January 2024; Accepted 24 January 2024

more recent study (Araújo, 2011), shows that in OECD countries costs have increased between 20% and 40% per cent because of the vertical separation of railway systems.[6]

From this brief summary of the literature on vertical separation in the railway industry, it emerges that one still much debated issue is what may be the sources of the lack of efficiency that might emerge when the system moves away from vertical integration. To the best of our knowledge, until now this problem has not been fully investigated by economic literature through theoretical analysis.

Transportation economics has devoted attention to this issue, highlighting the role of some sector-specific characteristics. In an influential report on the effects of railways privatization in the UK, McNulty (2011)[7] assigns a major role to the misalignment of incentives stemming from the special technological relationship between the state of the rolling stock and the state of the infrastructure, and thereby between the investments in rolling stock maintenance undertaken by the TOC and the investments in the maintenance of the infrastructure undertaken by the IM. This factor has been further investigated by several contributions which agree on claiming that, in the railway industry, misalignment costs contribute considerably more than other transaction costs to the inefficiency of vertical separated systems, and suggest that in vertically separated railways this aspect should be taken into account in the choice and design of regulatory instruments.[8]

The aim of this paper is twofold. First, we intend to provide a stylized theoretical model of local railway passengers service which is able to account for the main specific characteristics of the sector, and in particular for the interactions between operators' decisions on maintenance investment under vertical separation. Second, we will use this theoretical framework to carry out both a normative analysis of the operators' decisions on maintenance investment and an assessment of the welfare effects of simple regulatory instruments.

We place our analysis in the vertical separation case, since the efficiency consequences of incentive misalignment are magnified. Moreover, we think this is particularly appropriate in the case of regional railway transport, where in most cases the local TOCs exploit a national infrastructure (managed by a national IM) which is often shared with national TOCs.[9]

Our assumptions about the ownership of the IM and the TOC aim to develop the analysis in a scenario where the efficiency consequences of incentive misalignment are exacerbated, and where necessary, analytical tractability is preserved. Let us discuss further on this point.

As we focus on local passenger service, also the Public Local Authority (hereafter PLA) is involved in the analysis. In principle, she can either just be the franchiser of the service, or she can have a more active role as owner of the TOC. We consider the extreme case in which the TOC is owned and totally managed by the PLA that maximizes local welfare. It is worth noting that our results would hold *a fortiori* if we assumed a private TOC franchised by the PLA, since the information imperfections would be stronger than under our assumption.

We instead assume that the IM is a national private company. Note that if we assumed that the IM were a state owned enterprise which maximizes social welfare, it would implement the socially optimal level of investment in infrastructure maintenance and the TOC would get perfect knowledge about the level of IM's investment, so that one source of inefficiency would disappear.[10]

We think that, given the specificities of railway industry, an *ad hoc* theoretical framework is needed. In particular, in our model we emphasize the crucial role of investment in rolling stock maintenance. As in other network utilities (e.g. telecommunications and electricity), also in the railway sector the state of the infrastructure has a relevant role, in that it affects both the IM's cost of providing access and, along with the state of the rolling stock, several quality aspects of the service (e.g. punctuality, comfort, safety, any discomfort caused by possible damage either to the infrastructure or to the rolling stock, etc.), which in turn affect service demand. However, an important characteristic that distinguishes railways from other network utilities is that the quality of the infrastructure is not only affected by the IM's investment in the maintenance of the infrastructure, but also by the state of the rolling stock, and thus by the investment in maintenance undertaken by the TOC: well-maintained trains cause less wear and tear of the infrastructure than poorly-maintained ones, and thus lead to a better state of the infrastructure.[11] Our model accounts for this peculiarity.

This characteristic has two important consequences for the effects of the TOC's investment in rolling stock maintenance. First, the TOC's investment in maintenance enhances the quality of the service both directly through the improvement of the rolling stock, and indirectly through the improvement of the infrastructure; second and more importantly, it contributes with the IM's maintenance investment to reducing the IM's cost of providing access. In other words the TOC's investment in rolling stock maintenance generates a positive externality on the IM's cost of providing access and hence on his profit. Because of vertical separation, the TOC does not take into account this latter effect and thus tends to underinvest in rolling stock maintenance with respect to the socially optimal level.

Another important feature differentiating railways from other network utilities which is allowed for by our model is the fact that the revenues raised by the IM by selling access to the network are related to the number of trains employed by the TOC to operate the service, and are thus very weakly related to the final demand expressed in terms of number of passengers (Araújo, 2011). However, through the state of the infrastructure, IM's investment affects the quality of the service and hence its demand. In other words, the IM's investment in infrastructure maintenance generates a positive externality on local welfare. Again, because of vertical separation, the unbundled IM does not take into account this latter effect, which leads to underinvestment in infrastructure maintenance with respect to the socially optimal level.

Finally, a further peculiar characteristic of the railway industry which is considered by our model is that, because of the dependence of local welfare on the IM's investment in infrastructure maintenance, the asymmetric information that the TOC has on the productivity of the IM's investment affects her own investment decisions.[12]

[7] For a more updated review see Laurino et al. (2015).

[8] See, among others, Van de Velde et al. (2012), Merkert et al. (2012), Smith and Nash (2014), Mizutani et al. (2015), Andersson and Hultén (2016).

[9] Recent contributions have shown that from a cost saving perspective vertical separation is preferable to vertical integration when train density is sufficiently low (Mizutani and Uranishi, 2013). Placing our analysis in the vertical separation case, amounts to implicitly assume that in our framework train density belongs exactly in that range where vertical separation is advantageous. This permits us to mildly neglect possible effects of train density on operators' decisions on investment in maintenance, which in our framework is their choice variable.

[10] It should be noted that, especially in Europe, there is the idea that since in a vertical separation scenario the IM should care that fair competition among TOCs is assured, it would be more appropriate for IM to be a state-owned company. However, it should be noted that according to a Public choice approach to the modelling of the public managers' behaviour the management of the state-owned IM would choose the level of investment in infrastructure maintenance which maximizes his private utility, rather than social welfare (Mueller, 2003). This situation resembles the one of a private profit maximizing IM assumed in our analysis.

[11] On this point see, Marschnig (2016), Smith et al. (2017), Lundén and Paulsson (2009).

[12] It is worth pointing out that while the informational advantage of the IM on the value of his productivity with respect to both the regulator and the downstream operators is common to all regulated industries, for other regulated industries this information is much less relevant than in railway for the downstream operators' decisions about the investment in maintenance of their devices.

As for the regulatory framework, we assume that both the private national IM and the TOC are subject to regulation by a National Regulatory Authority (hereafter, NRA) that maximizes social welfare. We consider two regulatory instruments which are widely used by NRAs in many national contexts: the imposition of a lower boundary on the state of the infrastructure along with penalties due by the IM in case of violation, and the setting of the access price. Albeit unable to completely eliminate inefficiency, when compared to optimal regulatory mechanisms these instruments have the advantage of being easily manageable by a NRA and of having parsimonious informational requirements.[13] Our theoretical analysis will show that, because of this asymmetry of information of TOC about the productivity of the IM's investment, the introduction of a regulatory instrument inducing the TOC to internalize the effect of her investment on the IM's cost of providing access may reduce social welfare.

The paper is organized as follows. Section 2 presents a model of regional rail transport. In Section 3 we determine the IM's and TOC's socially optimal investments. Section 4.1 introduces the regulatory framework and the timing of the model. In Section 4.2 the consequences of regulating the IM are analysed. In Sections 4.3 and 4.4 the cases of unregulated and regulated TOC are examined, along with the relative welfare analyses. Section 5 concludes.

2. A stylized model of regional rail transport

We consider a very stylized model of regional rail transport regulated by a NRA.[14] The rail industry consists of three operators:

(1) A national private IM, which provides access to the network and is in charge of the maintenance of the national infrastructure.

(2) A local TOC, which operates the regional service and is in charge of the maintenance of the rolling stock.

(3) A benevolent PLA, which owns the TOC.

We assume that the IM and the TOC are vertically separated operators.

Although it is organized as a separate body with its own balance sheet, we assume that the TOC is totally controlled by the PLA which is responsible for the TOC's investment and financing decisions.

We consider a regional railway of a given length (e.g. 1 km) on which an exogenously given number of trains n operate.[15] The state of the rolling stock w is described by the following function:

$$w(e;h) = r_0 + he, \quad h = \underline{h}, \overline{h}\ ;\ \overline{h} > \underline{h}$$

where r_0 is an exogenous parameter which denotes the initial state of the rolling stock, i.e. when there is no maintenance investment: a higher value of r_0 indicates a better state of the rolling stock; e is the investment in rolling stock maintenance undertaken by the TOC; h denotes the marginal productivity of the TOC's investment in the maintenance of the rolling stock: the higher the value of h, the greater the improvement of the state of the rolling stock for any marginal increase in the TOC's investment.[16] We assume that h reflects the technology needed for the maintenance of the rolling stock, given its specific technical characteristics; this information is private knowledge of the TOC, thereby of the PLA that owns the TOC.[17]

The state of the infrastructure, denoted by i, depends on three elements: (i) its initial state i_0, i.e. the state of the infrastructure when there is no IM maintenance investment: the higher i_0, the better the initial conditions of the infrastructure; *(ii)* the IM's investment in infrastructure maintenance, denoted by m; (iii) the wear and tear caused by the running of the rolling stock, which depends on its state: the better the state of the rolling stock (i.e. the higher the value of w), the lower the wear and tear of the infrastructure. We describe the state of the infrastructure with the following linear form:

$$\begin{aligned} i(m;h,k) &= i_0 + km + w(e;h) = \\ &\quad i_0 + km + r_0 + he \\ k &= \underline{k}, \overline{k}\ ;\ \overline{k} > \underline{k} \end{aligned}$$

where k denotes the marginal productivity of the IM's investment in the maintenance of the infrastructure: the higher the value of k, the greater the marginal improvement of the state of the network for each marginal increase in maintenance investment. We assume that k reflects the technology needed for the maintenance of the local network, given its specific characteristics that are related to hydrogeological and morphological conditions (e.g. hilliness, altitude, rainfall, etc.); this information is IM's private knowledge.

For the sake of simplicity, from hereafter, we normalize to zero both the initial state of the rolling stock r_0 and the initial state of the infrastructure i_0. These assumptions do not affect the results of the paper. The state of the rolling stock and of the state of the infrastructure are redifined as follows:

$$w(e;h) = he \tag{1}$$

$$i(m;h,k) = km + he \tag{2}$$

$$k = \underline{k}, \overline{k} \text{ and } h = \underline{h}, \overline{h}$$

We shall now introduce Q, a function that summarizes various quality dimensions of the service (e.g. punctuality, comfort, safety, any discomfort caused by possible damages to the infrastructure and the rolling stock, etc.): the higher the value of Q, the higher the overall quality of the service. While the state of the infrastructure affects mainly those quality aspects related to safety, the comfort of the service is prevalently affected by the state of the rolling stock. A lower value of w indicates a poorer state of the rolling stock that may lead to a higher probability of rolling stock damages (for example, passenger discomfort due to delay or any other kind of trouble arising in case of train disruption). We assume that the quality of the service is a linear function of the state of the infrastructure and the state of the rolling stock. Therefore, Q takes the following form:

$$\begin{aligned} Q &= \beta i(m;h,k) + w(e;h) = \\ &\quad \beta km + (1+\beta)he \\ \frac{\partial Q}{m} &= \beta k;\ \frac{\partial Q}{\partial e} = (1+\beta)h,\ \beta \in (0,1] \\ \frac{\partial Q}{\partial e} &> \frac{\partial Q}{m} \text{ for } h \geqslant \frac{\beta k}{1+\beta} \end{aligned} \tag{3}$$

Note that, as the state of the rolling stock has a more pervasive role, its betterment would generate two effects on the quality of the service: a

[13] See Laffont and Tirole (1993).

[14] Although our model is tailored on the characteristics of the regional railway service, it is inspired to the broader modelling approach proposed in Chapter 5 of Laffont and Tirole (1993). For a more general view on the new economics of regulation see Laffont (1994). For an insightful economic analysis of utilities privatization in UK, with the lenses of the economics of regulation see Armstrong et al. (1994).

[15] Given the purposes of this paper the latter appears to be a mild assumption. The fact that variations in the number of trains running on the network have small impact on maintenance activities is empirically supported by Johansson and Nilsson (2004).

[16] We are aware that our modelling of investment in both rolling stock and infrastructure maintenance is very basic and tailored on the aims of the paper. Nonetheless, we think that it is able to capture the technological characteristics of railway industry discussed in the Introduction of the paper and relevant for the aim of the paper.

[17] Note that, although the PLA is a public body, it has still incentive to strategically exploit her informational advantage on h so as to increase residents' welfare at the expenses of overall efficiency.

direct (equivalent) effect and an indirect effect measured by parameter β through the wear and tear of the infrastructure. Consequently rolling stock maintenance investment turns out to be more valuable for the quality of the service than infrastructure maintenance investment, even for a range of values of the TOC's investment productivity h lower than the value of the IM's investment productivity k (i.e. in the range $h\epsilon\left[\frac{\beta k}{(1+\beta)}, k\right]$).[18]

Finally, we assume that the quality of the service affects the demand for the final service y, which takes the following linear form:

$$y = \theta + Q(e, m; h, k) - p = \theta + \beta km + (1+\beta)he - p \quad (4)$$
$$\frac{\partial y}{\partial m} = \beta k; \quad \frac{\partial y}{\partial e} = (1+\beta)h$$

where θ is an exogenous preference parameter and p is the price of a single journey.[19] The operators' investments, by increasing the quality of the service, lead to a higher demand for the service.[20] Moreover, concerning the effects of a marginal increase in the TOC's and the IM's maintenance investment on demand, the same considerations formulated for quality Q hold.[21]

2.1. Operators' objective functions

2.1.1. The infrastructure manager's profit function

The IM gains revenues by selling the TOC access to the network for each train employed to run the service; since we have assumed that the number of trains is fixed, access revenues do not depend on service demand (i.e. the number of passengers).[22] The IM's revenues are given by the access price, which is denoted by p_a, paid by the TOC for each train it employs for running the service. The IM incurs two kinds of cost: the cost of maintenance investment, which is a convex function $m^2/2$, and the cost of providing access to each train, denoted by c_a. We assume that the latter is inversely related to the state of the infrastructure: the better the infrastructure conditions are, the lower the cost of providing access is. Taking (2) into account, c_a is expressed by the following linear form:

$$c_a = c_0 - km - he \quad (5)$$
$$c_0 \geqslant km + he$$

where c_0 is a fixed component which accounts for those kind of activities that are independent from the characteristics of the rolling stock (e.g. operation of signalling devices, monitoring devices, presence of support workers, etc.). Both the direct investment in infrastructure maintenance undertaken by the IM and the investment in rolling stock maintenance undertaken by the TOC contribute to a reduction in the access cost. As already pointed out, this last effect is explained by the reduction in the wear and tear of the network arising from a better state of the rolling stock. We assume that the operators' maintenance investments cannot turn the access cost negative.

The IM's profit is given by:

$$\Pi^{IM} = (p_a - c_0 + km + he)n - \frac{m^2}{2}$$

For the sake of simplicity $n = 1$ is assumed in the rest of the paper.[23] Therefore, IM's profit is eventually given by:

$$\Pi^{IM} = p_a - c_0 + km + he - \frac{m^2}{2} \quad (6)$$

2.1.2. The train operating company's profit function

The TOC gains revenues from the tickets purchased by passengers, which can be expressed, taking (4) into account, by

$$R^{TOC} = [\theta + \beta km + (1+\beta)he - p]p$$

As shown in (4), by improving the quality of the service, the operators' investments boost the demand for it, which in turn, increases the TOC's revenues. The TOC incurs two kinds of cost: the access charge p_a paid to the IM for each train employed in the running of the service, and the cost of investment, which is a convex function $e^2/2$. Therefore, the TOC's profit is given by

$$\Pi^{TOC} = [\theta + \beta km + (1+\beta)he - p]p - p_a - \frac{e^2}{2} \quad (7)$$

2.1.3. The public local authority's objective function

We assume that the PLA finances the TOC through two instruments: the price of tickets and a local public subsidy, g. Therefore, the TOC's total gains are given by

$$\Pi^{TOC} + g$$

The local public subsidy is raised through distortionary taxation; this means that each euro of tax paid by local taxpayers involves an additional cost in terms of loss of efficiency; this additional cost is denoted by a parameter λ. As a consequence, for each euro of public subsidy there are both an equivalent gain for the TOC and a cost of $1+\lambda$ for local taxpayers, where λ measures the marginal social cost of public funds.[24]

The PLA's objective function is the local social welfare given by the sum of the net consumer's surplus and the TOC's profit minus the social cost of public funds $(1+\lambda)g$ incurred by local taxpayers. The private gains of the national IM are not included in the local welfare function, since they are external to the local community.[25]

Given the linear demand function (4) the net consumer's surplus CS is defined as

$$CS = \frac{1}{2}[\theta + \beta km + (1+\beta)he - p]^2 \quad (8)$$

Therefore, the PLA's local social welfare function, W^{PLA}, is given by

$$W^{PLA} = CS + \Pi^{TOC} + g - (1+\lambda)g = CS + \Pi^{TOC} - \lambda g \quad (9)$$

[18] Let us point out that in Eq. (3) we have assumed $\beta \in (0, 1]$, so that the case $\beta = 1$ is included in our model (i.e. rolling stock maintenance and infrastructure investment contribute to the same extent to service quality). As it will be clearer later in the paper, we have introduced the parameter β just as a modelling device for tracking the double effect on service quality of an increase in rolling stock maintenance, since the latter affects also the state of the infrastructure.

[19] It should be noted that, since we assume that the number of trains n employed for running the service is exogenously given, this implies that there is a service capacity constraint given by the number of passengers that can be taken on the set of available trains. We assume that y is always below this maximum capacity.

[20] On the other way round, if we consider the inverse demand function, operators' investment in maintenance by increasing the quality of the service will increase customers' willingness to pay.

[21] See, Eq. (3) and the relative comments.

[22] As we have already pointed out in the Introduction, this assumption is based on a well-known characteristic of the railway industry with vertical separation, according to which the revenues from access to the network collected by the IM are very weakly related to the final demand (Araújo, 2011).

[23] Note that, since the focus of the paper is on operators' investment in maintenance choice, and since we are not interested in determining the optimal number of train running on the track, this assumption is without loss of generality.

[24] Notice that we are implicitly assuming that the marginal social cost of local and national taxes is the same. This implies that the marginal social cost of public funds is the same whether the subsidy is directly financed by the PLA through local taxation, or if it is a fiscal transfer from central to local government financed through national tax revenue.

[25] The modelling of the PLA's behaviour is perfectly consistent with the traditional theory of fiscal federalism pioneered by Oates (1972).

3. Benchmark

In order to set a benchmark for the normative analysis carried out in the next sections, we firstly consider welfare maximization by a benevolent social planner who is perfectly informed about the productivity of both the IM's and the TOC's investments (k and h respectively). Social welfare is given by the sum of net consumers surplus, TOC's profit, and IM's profit minus subsidy g weighted by the marginal social cost of public funds λ. This simply amounts to the sum of the PLA social welfare function (9) and the IM's profit:

$$W = CS + \Pi^{TOC} + \Pi^{IM} - \lambda g \tag{10}$$

The social planner's maximization problem is given by:

$$\max_{e,g,m,p,p_a} W = \max_{e,g,m,p,p_a} \left\{ CS + \Pi^{TOC} + \Pi^{IM} - \lambda g \right\}$$
$$s.t. \Pi^{TOC} + g \geqslant 0$$
$$\Pi^{IM} \geqslant 0$$
$$\text{for } k = \underline{k}, \overline{k}, h = \underline{h}, \overline{h}$$

where CS, Π^{TOC} and Π^{IM} are defined in (8), (6) and (7) respectively.

Since W is decreasing in g, it is set so to ensure a non-negative TOC's profit for each value of h and k:

$$-g = \Pi^{TOC} = (\theta + \beta km + (1+\beta)he - p)p - p_a - \frac{e^2}{2} \tag{11}$$

for $k = \underline{k}, \overline{k}, h = \underline{h}, \overline{h}$

Substituting (6), (7), (8), and (11) into (10), the maximization problem becomes:

$$\max_{e,m,p,p_a} W$$
$$s.t.\ \Pi^{IM} = p_a - c_0 + km + he - \frac{m^2}{2} \geqslant 0$$
$$h = \underline{h}, \overline{h}, \quad k = \underline{k}, \overline{k}$$

where:

$$\begin{aligned} W &= \frac{1}{2}(\theta + \beta km + (1+\beta)he - p)^2 \\ &+ (1+\lambda)(\theta + \beta km + (1+\beta)he - p)p \\ &- (1+\lambda)\frac{e^2}{2} - \lambda p_a - c_0 + km + he - \frac{m^2}{2} \end{aligned} \tag{12}$$

Note that the access price reduces welfare proportionally to the social cost of public funds λ. This can be explained by the fact that a marginal increase in the access price involves both an equivalent welfare gain due to the related increase in the IM's profit, and a welfare loss of $1+\lambda$ due to the related decrease in the TOC's profit and increase in public subsidy. The optimal access price p_a^* is thus set to satisfy:

$$\Pi^{IM} = p_a^* - c_0 + km + he - \frac{m^2}{2} = 0$$

which solves[26]:

$$p_a^* = c_0 - km - he + \frac{m^2}{2} \text{ for } h = \underline{h}, \overline{h}, \quad k = \underline{k}, \overline{k} \tag{13}$$

Substituting (13) into (12), the maximization problem shrinks to:

$$\max_{e,m,p} W$$

where

$$\begin{aligned} W &= \frac{1}{2}(\theta + \beta km + (1+\beta)he - p)^2 \\ &+ (1+\lambda)(\theta + \beta km + (1+\beta)he - p)p \\ &- (1+\lambda)\frac{e^2}{2} - (1+\lambda)\left(c_0 - km - he + \frac{m^2}{2}\right) \end{aligned}$$

Denoting $W_p = \frac{\partial W}{\partial p}$, the first order condition w.r.t. p is given by[27]

$$W_p = -(\theta + \beta km + (1+\beta)he - p) + (1+\lambda)(\theta + \beta km + (1+\beta)he - 2p) = 0$$

The first term above represents the welfare loss due to the reduction of consumer surplus; the second term represents the net welfare gain due to the increase in the TOC's profit and the consequent reduction in the public subsidy. Rearranging we obtain:

$$p^* = \frac{\lambda(\theta + \beta km + (1+\beta)he)}{1+2\lambda}$$

Denoting $W_e = \frac{\partial W}{\partial e}$, the first order condition w.r.t. e is given by:

$$\begin{aligned} W_e &= (\theta + \beta km + (1+\beta)he - p)(1+\beta)h \\ &+ (1+\lambda)(1+\beta)hp + (1+\lambda)h \\ &- (1+\lambda)e = 0 \end{aligned} \tag{14}$$

The first term in (14) represents the welfare gain due to the increase in consumer surplus induced by the improvement in the quality of the service; the second term represents the welfare gain related to both the increase in the TOC's revenues and the induced reduction in public subsidies; the third term represents the welfare gain due to the reduction in the cost of access implied by the reduction in the access price[28]; the fourth term represents the welfare loss related to the reduction in the TOC's profit (and the corresponding increase in the public subsidy) which is due to the increase in the cost of investment in rolling stock maintenance.

Substituting p^* into (14), rearranging the formula, and solving by e, we get:

$$e^*(m; h, k) = \frac{(1+\beta)hL_1(\lambda)[\theta + \beta km] + L_2(\lambda)h}{L_2(\lambda) - (1+\beta)^2h^2L_1(\lambda)} \tag{15}$$

where:

$$L_1(\lambda) = 1 + 2\lambda + \lambda^2 \tag{16}$$
$$L_2(\lambda) = 1 + 3\lambda + 2\lambda^2 \tag{17}$$

Finally, denoting $W_m = \frac{\partial W}{\partial m}$, the first order condition w.r.t. m is given by:

$$W_m = (\theta + \beta km + (1+\beta)he - p)\beta k + (1+\lambda)\beta kp + (1+\lambda)(k - m) = 0 \tag{18}$$

The first term is the welfare gain due to the increase in consumer surplus induced by the improvement in service quality; the second term represents the welfare gain related to the increase in TOC's revenues along with the induced reduction in public subsidies; the third effect represents the welfare effects stemming from the net variation of IM's cost which is given by the difference between the reduction in the cost of providing access and the cost of investment in infrastructure maintenance.

Rearranging terms, substituting (15), and following manipulations (see Appendix), we obtain:

$$m^*(h,k) = \frac{k\left[L_2(\lambda) + \beta\theta L_1(\lambda) - (1+\beta)h^2L_1(\lambda)\right]}{\left[L_2(\lambda) - \left((1+\beta)^2h^2 + \beta^2k^2\right)L_1(\lambda)\right]} \tag{19}$$

3.1. Sources of inefficiency

In order to define an appropriate regulatory framework for regional railway service, we need to single out the sources of inefficiency stemming from the characteristics of the operators' behaviour captured by our assumptions.

The following sources of inefficiency arise from the operators' investment decisions. First, by maximizing its profit (see (6)) the private IM does not take into account the effect of investment in infrastructure

[26] Note that given the simplifying assumption $n = 1$ Eq. (13) reflects average cost pricing.

[27] We assume that the second order conditions are always satisfied.

[28] From (13) it is clear that the reduction of the access cost implies an increase in TOC's profit, while letting unchanged IM's profit.

maintenance on the quality of the service, and thereby on service demand, consumers' surplus and TOC's profit. This leads to underinvestment in the maintenance of the infrastructure with respect to the benchmark.

Conversely, the effects of investment in rolling stock maintenance on consumer's surplus are taken into account by the TOC, as we have assumed that she is owned and directly managed by the PLA which maximizes local welfare (9). However, a second source of inefficiency is given by the fact that local welfare does not include the IM's profit – which is a national operator – and thus the TOC's maintenance investment decisions do not consider the effect on the IM's cost of providing access, thereby generating underinvestment with respect to the benchmark.

Nevertheless, another important source of inefficiency arises out of the dependence of local welfare function (9) on the IM's decisions of investment in infrastructure maintenance which, through the improvement of the infrastructure, enhance service quality and demand as well as consumer's surplus. Because of the asymmetry of information of the PLA-TOC[29] with respect to the marginal productivity k of the IM's investment, the PLA-TOC's investment decisions in rolling stock maintenance are taken on the basis of the expected value of k, which leads to a level of investment that is higher than the benchmark for the lowest value of the IM's productivity (i.e. when $k = \underline{k}$). This overinvestment distortion due to the asymmetry of information on k can, in some cases (i.e. for some values of the marginal productivity of the TOC's investment h), more than compensate for the underinvestment caused by the neglecting of the effect on the cost of access. In these cases, we will show how a regulation aimed at internalizing the effect on the cost of access increases the overinvestment distortion and thus possibly leads to a reduction in social welfare.

4. Regulation

4.1. Regulatory framework and timing

Let us now define our regulatory framework. A NRA, which we assume to be in charge of the regulation of railway service, observes both the state of the infrastructure and the state of the rolling stock. However, it is informed only about the probability distribution of productivity k and h of the IM's and TOC's maintenance investment respectively, whose realized values are assumed to be operators' private information.

We consider two very simple regulatory instruments that the NRA can employ to regulate the IM and the TOC: a lower boundary on the observed state of the network and the access price rule. When compared to optimal regulatory mechanisms, and albeit unable to completely eliminate inefficiency, these instruments appear to be more easily manageable by the NRA and to have parsimonious informational requirements.

As for the first instrument, the NRA sets a lower boundary on the observed state of the infrastructure, which the IM must guarantee or else incurs a penalty. The aim of this instrument is to ensure a level of investment in infrastructure maintenance corresponding to the efficient one in the worst state of the world (i.e. $m = m^*(\underline{h}, \underline{k})$).[30]

As for the access price rule, it has to ensure the IM's participation constraint in every state of the world.[31] However, in our framework access price regulation may also be used to pursue a second objective, namely regulating the PLA-TOC so that the effect of her investment on access cost is internalized. We consider two scenarios: the first in which the PLA-TOC is unregulated, and the access price has only to ensure the participation constraint of the IM, taking into account that the NRA does not observe IM's and TOC's marginal productivity; the second, in which access price is also used to regulate the PLA-TOC.

The timing of the model is the following:

- At time 0 the NRA announces:

1. a lower boundary on the state of the network and a penalty to be paid by the IM in case of violation;
2. the rules for setting of the access price.

- At time 1 the IM chooses the amount of investment in infrastructure maintenance that maximizes his profit, and the PLA-TOC chooses the amount of investment in rolling stock maintenance and the price of the service which maximize the expected local welfare.

- At time 2 the NRA observes the state of network and the state the rolling stock, penalties (if any) are imposed on the IM, the access price is paid by the PLA-TOC, the final service is operated, and the public subsidy is paid to the PLA-TOC.

4.2. Regulation of private infrastructure manager

In this section, we will address the design and the effects of the regulation of the private IM imposed by the NRA. We will first analyse the investment decision of an unregulated private IM. Successively, we will investigate the effects of the imposition of a lower boundary on the IM's behaviour. Finally, we will focus on the setting of the access price both in the case of unregulated and regulated PLA-TOC.

As explained in Section 2, the IM is responsible for the investment m in infrastructure maintenance. Since an unregulated private IM will choose the investment that maximizes his private profit as in (6), the IM's maximization problem can be written as follows[32]:

$$\max_m \Pi^{IM} = \max_m \; p_a + h\hat{e}(h,k) + km - \frac{m^2}{2}$$
$$\text{for } k = \underline{k}, \overline{k} \text{ and } h = \underline{h}, \overline{h}$$

where $\hat{e}(h,k)$ is the investment in infrastructure maintenance undertaken by the PLA-TOC. The first order condition of the IM's maximization problem is given by:

$$m = k$$
$$\text{for } k = \underline{k}, \overline{k}$$

[29] Given the assumption that the TOC is owned and managed by the PLA, hereafter we will use the notation PLA-TOC.

[30] This instrument is inspired to measures which can be observed in most railway industries around the world, where – in order to protect the safety of travellers and workers – there are regulatory bodies in charge of verifying that the infrastructure matches strict quality standards. Just to give an example, one can refer to the several health and safety investigation and enforcement powers endowed to the UK regulator (Offrail). Similar powers are assigned to specific national agencies in other European countries, for example in Italy to Agenzia Nazionale per la Sicurezza delle Ferrovie e delle Infrastrutture Stradali e Autostradali (ANSFISA); in France the Etablissement Public de Sécurité Ferroviaire (EPSF); in Germany to Eisenbahn – Bundesamt (EBA); etc..

[31] The access price rules proposed in this paper are an application of the standard theory of price regulation under monopoly with unknown cost. See Baron and Myerson (1982). In the regulatory practice different regulators usually apply different rules, however in most cases the access price is based on an estimation of the cost of providing access and on a non-negativity condition of the IM's profit (that is referred to in the economics of regulation literature as IM's participation constraint).

[32] Note that the IM's choice of investment should take into account the PLA-TOC's choice of investment in rolling stock maintenance $\hat{e}(h,k)$, $h = \underline{h}, \overline{h}$. Besides, because of the asymmetry of information of the PLA-TOC on the productivity parameter h, the level of m should maximize the IM's expected profits with the expectation taken w.r.t. to h. However, because of the linearity of the IM's profit function, neither of these assumptions affects the choice of the IM's profit-maximizing level of investment.

The equation above implies that an unregulated private IM chooses a level of investment such that its marginal cost equates its marginal productivity. Therefore, solving the IM's maximization problem we get:

$$m(k) = k \text{ for } k = \underline{k}, \overline{k}$$

with $m(k) < m^*(h,k)$ for $k = \underline{k}, \overline{k}$ and $h = \underline{h}, \overline{h}$

The underinvestment of $m(k)$ with respect to the optimal level m^* is easily explained by the fact that the IM does not take into account that his own investment, by affecting the state of the network, indirectly affects service demand and thus the PLA-TOC's profit and the consumers' surplus.

Since it observes both the actual state of the network and the state of the rolling stock, and in order to induce the IM to undertake an investment at least equal to $m^*(\underline{h}, \underline{k})$, the NRA imposes an appropriate lower boundary $\underline{i}$ on the state of the network along with a penalty that applies to the IM in case such lower boundary is violated.[33] This regulatory objective can be obtained by setting a lower boundary corresponding to the state of the network that would emerge if the IM undertook the optimal investment $m^*(\underline{h}, \underline{k})$, given the observed state of the rolling stock. The lower boundary takes the form:

$$\underline{i}(h,k) = w(\hat{e}(h,k); h) + \underline{k} m^*(\underline{k}, \underline{h})$$

for $h = \underline{h}, \overline{h}$ and $k = \underline{k}, \overline{k}$

where $w(\hat{e}(h,k); h) = h\hat{e}(h,k)$ is the observed state of the rolling stock. To induce the IM to undertake $m(\underline{k}) \geqslant m^*(\underline{k}, \underline{h})$, the penalty has to be proportional to the difference between the lower boundary and the observed state of the infrastructure, which is given by:

$$i(h,k) = w(\hat{e}(h,k); h) + km(k)$$

for $h = \underline{h}, \overline{h}$ and $k = \underline{k}, \overline{k}$

Denoting by $P^{IM}(k)$ the penalty and by s_I a constant rate, we define:

$$P^{IM}(k) = s_I \left[\underline{i}(h,k) - i(h,k)\right] = s_I \left[\underline{k} m^*(\underline{k}, \underline{h}) - km(k)\right] \quad (20)$$

Since the lower boundary is set on the observed state of the rolling stock, the amount of the penalty does not depend on the investment in rolling stock maintenance. While the penalty is null when the observed state of the network is equal to the lower boundary, it is negative when the former is higher than the latter (i.e. when $i(h,k) > \underline{i}(h,k)$), and it turns out to be a subsidy to the IM. We need to derive the rate s_I^* that induces the IM to undertake $m^*(\underline{h}, \underline{k})$ when $k = \underline{k}$.

Taking into account the introduction of the penalty and Eq. (1), the maximization problem of the regulated IM becomes:

$$\max_m \Pi^{IM}(k, s_I) = p_a + km + h\hat{e}(h,k) - \frac{m^2}{2} - s_I \left[\underline{k} m^*(\underline{k}, \underline{h}) - km(k)\right]$$

The regulated IM's profit-maximizing investment level is defined by the first order condition of the IM's profit maximization problem, and is given by:

$$m^r = k + s_I k$$

where the apex r stands for *regulated*. From the first order condition, it is clear that the introduction of a lower boundary on the state of the network, together with a penalty, increases the marginal private benefit of the IM's investment by factor $s_I k$, since now a marginal increase in the IM's also investment generates the effect of reducing the penalty.

For $k = \underline{k}$ the first order condition leads to:

$$m^r = \underline{k} + s_I \underline{k}$$

Since the NRA knows the value of $\underline{k}$ and $\underline{h}$, as well as the IM's profit function, it is able to derive $m^*(\underline{h}, \underline{k})$. Consequently, to induce the IM to undertake $m^*(\underline{h}, \underline{k})$, the NRA has to set an optimal penalty rate, s_I^*, such that:

$$\underline{k} - m^*(\underline{k}, \underline{h}) + s_I^* \underline{k} = 0 \Longrightarrow s_I^* = \frac{m^*(\underline{k}, \underline{h})}{\underline{k}} - 1 \quad (21)$$

which ensures that:

$$m^r(\underline{k}) = m^*(\underline{k}, \underline{h}) \quad (22)$$

For $k = \overline{k}$, taking into account (21), the profit-maximizing investment level is now defined by the following first order condition and the consequent IM's investment:

$$\overline{k} - m + s_I^* \overline{k} = 0 \Longrightarrow$$

$$m^r(\overline{k}) = \overline{k}(1 + s_I^*) = \frac{\overline{k} m^*(\underline{k}, \underline{h})}{\underline{k}} \quad (23)$$

From (21), (23) and from $\underline{k} < \overline{k}$ it follows that $m^r(\underline{k}) = m^*(\underline{k}, \underline{h}) < \frac{\overline{k} m^*(\underline{k}, \underline{h})}{\underline{k}} = m^r(\overline{k})$. That is, since the IM's maintenance investment is increasing in its productivity k, the imposition of the lower boundary ensures a level of investment in network maintenance higher than $m^*(\underline{k}, \underline{h})$ for $k = \overline{k}$.

Let us now analyse the setting of the access price. The IM's participation constraint in correspondence of the lower boundary $\underline{i}$ for any pairs (h,k) is to be ensured, i.e.:

$$\Pi^{IM}(k,h) = p_a + h\hat{e}(h,k) + km(k,h) - \frac{m(k,h)^2}{2} \geqslant 0$$

for $h = \underline{h}, \overline{h}$ and $k = \underline{k}, \overline{k}$

Because of the asymmetry of information of NRA w.r.t. k, p_a is set on the basis of a fixed-price mechanism. Since the IM's profit is increasing in k and m, and since the level of IM's investment that ensures a state of the network equal to the lower boundary is exactly $m^*(\underline{k}, \underline{h})$, if the IM's participation constraint is satisfied for $k = \underline{k}$ and $m^*(\underline{k}, \underline{h})$, it will be satisfied for any pairs (h,k). This implies that the IM's participation constraint can be written as follows:

$$\Pi^{IM}(k,h) = p_a + h\hat{e}(h,k) + \underline{k} m^*(\underline{k}, \underline{h}) - \frac{m^*(\underline{k}, \underline{h})^2}{2} \geqslant 0$$

for $h = \underline{h}, \overline{h}$ and $k = \underline{k}, \overline{k}$

where the value of the investment in rolling stock maintenance $\hat{e}(h,k)$ chosen by the PLA-TOC depends on hers being regulated or not. Therefore, we analyse the effects of the setting of the access price on the IM's decisions and profit separately in the cases of unregulated and regulated PLA-TOC.

The access price rule in the case of unregulated PLA-TOC

We denote by $e^u(h,k)$ the level of investment in rolling stock maintenance chosen by an unregulated PLA-TOC for each pair (h,k). Since the IM's profit is increasing in both PLA-TOC's investment and productivity, the access price set by the NRA has to satisfy the IM's participation constraint for $h = \underline{h}$, and thereby it is given by:

$$p_a(\underline{h}, k) = -\underline{h} e^u(\underline{h}, k) - \underline{k} m^*(\underline{k}, \underline{h}) + \frac{m^*(\underline{k}, \underline{h})^2}{2} \quad (24)$$

for $k = \underline{k}, \overline{k}$

Note that according to (24), the access price depends on the value of k only through the investment in rolling stock maintenance $e^u(\underline{h}, k)$

[33] Of course, this strategy does not guarantee that the IM would carry out the socially optimal investment in the other states of the world (where either the TOC or the IM or both exhibit a productivity higher than the minimum one). However, if the NRA imposed the optimal investment in other states of the world, the IM's participation constraint would not be assured to hold in the worse state of the world, so that in this case the resulting allocation would not be financially sustainable.

chosen by an unregulated PLA-TOC for any value of k. By substituting (24) into (6), the IM's profit becomes: for $k = \underline{k}$

$$\Pi^{IM}(\underline{k}, h) = \begin{cases} 0, & \text{for } h = \underline{h} \\ \underbrace{\left[\overline{h}e^u(\overline{h}, \underline{k}) - \underline{h}e^u(\underline{h}, \underline{k})\right]}_{A1}, & \text{for } h = \overline{h} \end{cases}$$

and for $k = \overline{k}$

$$\Pi^{IM}(\overline{k}, h) = \begin{cases} \underbrace{s_I^* \left[\frac{\overline{k}^2 m^*(\underline{k}, \underline{h})}{\underline{k}} - \underline{k}m^*(\underline{k}, \underline{h}) \right]}_{B}, & \text{for } h = \underline{h} \\ \underbrace{s_I^* \left[\frac{\overline{k}^2 m^*(\underline{k}, \underline{h})}{\underline{k}} - \underline{k}m^*(\underline{k}, \underline{h}) \right]}_{B} + \underbrace{\left[\overline{h}e^u(\overline{h}, \overline{k}) - \underline{h}e^u(\underline{h}, \overline{k})\right]}_{A2}, & \text{for } h = \overline{h} \end{cases}$$

Note that when the PLA-TOC is not regulated, because of the reduction in the access cost, the IM gains a positive profit in all the states of the world but the worst one. The term B denotes the reduction in the access cost generated by a higher level of IM's productivity and by a higher level of investment in infrastructure maintenance.[34] The terms $A1$ and $A2$ denote the reduction in the access cost generated by a higher level of PLA-TOC's productivity and by a higher level of investment in rolling stock maintenance respectively.[35]

The access price rule in the case of regulated PLA-TOC

If the NRA uses the access price to affect the PLA-TOC's choice of investment level, p_a will be set on the basis of the observed state of the rolling stock, i.e.

$$p_a(h, k) = -he^r(h, k) - \underline{k}m^*(\underline{k}, \underline{h}) + \frac{m^*(\underline{k}, \underline{h})^2}{2} \tag{25}$$

for $h = \underline{h}, \overline{h}$ and $k = \underline{k}, \overline{k}$

where e^r is the investment in rolling stock maintenance chosen by a regulated PLA-TOC.

By substituting (25) into (6), the IM's profit turns out to be

$$\Pi^{IM}(\underline{k}, h) = 0$$

$$\Pi^{IM}(\overline{k}, h) = s_I^* \left[\frac{\overline{k}^2 m^*(\underline{k}, \underline{h})}{\underline{k}} - \underline{k}m^*(\underline{k}, \underline{h}) \right] \quad \text{for } h = \underline{h}, \overline{h}$$

which means that when the access price is also used to regulate the PLA-TOC, the IM obtains a positive profit only for $k = \overline{k}$, due to the reduction in the cost of access arising from higher levels of both investment in the infrastructure and the IM's productivity. However, compared with the case of unregulated PLA-TOC, the IM's no longer benefits from possible reductions in the access cost related either to a higher productivity of the investment in rolling stock maintenance or to a higher level of PLA-TOC's investment.

4.3. Rolling stock maintenance with unregulated local public train operating company

In this section we consider the case of an unregulated PLA-TOC and a private IM regulated as discussed in Section 4.2. In this case, the NRA sets the access price rule according to (24), that is with the sole aim of ensuring non-negative profits for the IM.

The PLA-TOC chooses the price of the railway service p and the investment in rolling stock maintenance e by maximizing the expected local social welfare, with expectation w.r.t. the value of k which is IM's private information. Denoting by EW^{PLA} the local expected social welfare, the PLA-TOC's maximization problem can be written as:

$$\begin{aligned} \max_{p,e} EW^{PLA}(h) = \max_{p,e} & \frac{\upsilon}{2}(\theta + \beta\underline{k}m^r(\underline{k}) + (1+\beta)he - p)^2 \\ & + \frac{(1-\upsilon)}{2}(\theta + \beta\overline{k}m^r(\overline{k}) + (1+\beta)he - p)^2 \\ & + \upsilon\Pi^{TOC}(h, \underline{k}) + (1-\upsilon)\Pi^{TOC}(h, \overline{k}) - \lambda \\ \text{s.t. } \Pi^{TOC} + g \geqslant 0 & \end{aligned}$$

where υ is the probability that $k = \underline{k}$ and $\Pi^{TOC} + g \geqslant 0$ is the participation constraint for the PLA-TOC.

Because of the asymmetry of information on k and the fact that the PLA-TOC's private profit is increasing in $km^r(k)$, the participation constraint of the PLA-TOC should ensure a non-negative profit for $k = \underline{k}$, since:

$$\begin{aligned} & \Pi^{TOC}(h, k) + g \geqslant 0, \text{ for } h = \underline{h}, \overline{h} \text{ and } k = \underline{k}, \overline{k} \\ & \iff \\ & \Pi^{TOC}(h, \underline{k}) + g \geqslant 0, \text{ for } h = \underline{h}, \overline{h} \end{aligned} \tag{26}$$

Now, as the expected local social welfare is decreasing in g and taking into account (26), the value of g which solves the unregulated PLA-TOC's maximization problem is:

$$-g^u(h) = \Pi^{TOC}(h, \underline{k}) \quad \text{for } h = \underline{h}, \overline{h}$$

Substituting the above expression and (7) into EW^{PLA} we get:

$$\begin{aligned} \max_{p,e} EW^{PLA}(h) = \max_{p,e} & \frac{\upsilon}{2}(\theta + \beta\underline{k}m^r(\underline{k}) + (1+\beta)he - p)^2 \\ & + \frac{(1-\upsilon)}{2}(\theta + \beta\overline{k}m^r(\overline{k}) + (1+\beta)he - p)^2 \\ & + \left[(\theta + (1+\beta)he + \beta\widetilde{km^r} - p)p - p_a - \frac{e^2}{2}\right] \\ & + \lambda\left[(\theta + (1+\beta)he + \beta\underline{k}m^r(\underline{k}) - p)p - p_a - \frac{e^2}{2}\right] \\ \text{for } h = \underline{h}, \overline{h}, & \end{aligned}$$

where $\widetilde{km^r} = \upsilon\underline{k}m^r(\underline{k}) + (1-\upsilon)\overline{k}m^r(\overline{k})$.

The price of the service is set to satisfy the following first order condition w.r.t. p:

$$\begin{aligned} \theta + (1+\beta)he + \beta\widetilde{km^r} - p = & \left[\theta + \widetilde{km^r} + (1+\beta)he - 2p\right] \\ & + \lambda\left[\theta + \beta\underline{k}m^r(\underline{k}) + (1+\beta)he - p\right] \end{aligned}$$

On the left hand side of the first order condition there is the PLA-TOC's marginal cost of increasing the price of regional railway service, that is represented by the reduction in the expected consumers' surplus. On the right hand side there are the marginal benefits of increasing p, which are represented by the PLA-TOC's expected marginal private revenues stemming from an increase in p (first term on the r.h.s.), and the PLA-TOC's marginal revenues when $k = \underline{k}$, the latter implying the reduction in the public subsidy g required to satisfy the TOC's participation constraint (second term on the r.h.s.). Notice that the reduction is proportional to the marginal cost of public funds λ: the higher λ, the higher the marginal benefit from a reduction in the public subsidy required to satisfy the participation constraint.

By solving the first order condition w.r.t. p we derive the price of the final service that maximizes the expected local social welfare:

$$p^u(e^u) = \frac{\lambda(\theta + (1+\beta)he^u(h, \widetilde{km^r}) + \beta\underline{k}m^r(\underline{k}))}{1+2\lambda} \quad \text{for } h = \underline{h}, \overline{h}, \tag{27}$$

[34] To be more precise, B denotes the reduction in the access cost generated by a level of IM's investment higher than the optimal one in the worst state of the world (i.e. $m^*(\underline{k}, \underline{h})$).

[35] More precisely, $A2$ denotes the reduction in the access cost generated by a level of PLA-TOC's investment higher than the one that she would have chosen with $h = \underline{h}$ for $k = \underline{k}, \overline{k}$.

where e^u is the level of investment in rolling stock maintenance undertaken by an unregulated PLA-TOC.[36] Consistently with the above discussion of the first order condition, the level of price directly depends on the investment in network maintenance undertaken by a regulated private IM at $k = \underline{k}$, while it depends only indirectly (through e^u) on the expected investment in network maintenance. Besides, since:

$$\frac{\partial p^u(e^u)}{\partial \lambda} = \frac{(\theta + (1+\beta)he^u(h, \widetilde{km^r}) + \beta \underline{k} m^r(\underline{k}))}{(1+2\lambda)^2} > 0$$

the price level increases with the marginal social cost of public funds λ: the higher the marginal cost of public funds, the higher the price set by the PLA-TOC, as this finds more convenient to finance the service through private revenues.

The choice of investment level in rolling stock maintenance must satisfy the first order condition w.r.t. e, which by using (27) is given by

$$EW_e^{PLAu} = (1+\beta)h\left\{\frac{1}{1+2\lambda}\left[(1+\lambda)(\theta+(1+\beta)he^u) - \lambda\beta\underline{k}m^r(\underline{k})\right] + \beta\widetilde{km^r}(k)\right\} + (1+\lambda)\frac{\lambda}{1+2\lambda}\left[(1+\beta)^2h^2e^u + (\theta + \beta\underline{k}m^r(\underline{k}))\right] - (1+\lambda)e^u = 0 \quad (28)$$

for $h = \underline{h}, \overline{h}$

and can be expressed as follows:

$$(1+\beta)h\left\{\frac{1}{1+2\lambda}\left[(1+\lambda)(\theta+(1+\beta)he^u) - \lambda\beta\underline{k}m^r(\underline{k})\right] + \beta\widetilde{km^r}(k)\right\} + (1+\lambda)\frac{\lambda}{1+2\lambda}\left[(1+\beta)^2h^2e^u + (\theta + \beta\underline{k}m^r(\underline{k}))\right] = (1+\lambda)e^u \quad (29)$$

for $h = \underline{h}, \overline{h}$

On the left hand side of (29) there is the PLA-TOC's marginal benefit of investing in rolling stock maintenance, which consists of two elements: the increase in consumers' surplus and PLA-TOC's marginal private revenues from PLA-TOC's investment. Both marginal benefits stem from the fact that an increase in investment in rolling stock maintenance improves the overall quality of the service. This implies an increase in the consumers' willingness to pay, and hence an increase in both the consumers' surplus and the demand for the service (thus in the PLA-TOC's private revenues). On the right hand side, we have the marginal cost of investment in rolling stock maintenance (adjusted for the marginal cost of public funds λ), which stems directly from the quadratic cost function.

Rearranging (29) we obtain:

$$-\left[L_2(\lambda) - (1+\beta)^2h^2(1+2\lambda+\lambda^2)\right]e^u + (1+\beta)hL_1(\lambda)\theta + \beta(1+\beta)h((1+2\lambda)\widetilde{km^r}(k) + \lambda^2\underline{k}m^r(\underline{k})) = 0$$

for $h = \underline{h}, \overline{h}$

where $L_1(\lambda)$ and $L_2(\lambda)$ are given by (16) and (17) respectively.

Finally, by solving (29) w.r.t. e^u, we obtain the level of investment undertaken by an unregulated PLA-TOC:

$$e^u = \frac{(1+\beta)h\left[L_1(\lambda)\theta + \beta((1+2\lambda)\widetilde{km^r}(k) + \lambda^2\beta\underline{k}m^r(k))\right]}{L_2(\lambda) - (1+\beta)^2h^2L_1(\lambda)} \quad (30)$$

for $h = \underline{h}, \overline{h}$ and $k = \underline{k}, \overline{k}$

Now we can determine the value of the social welfare function (10) in each state of the world when the PLA-TOC is unregulated and the private IM is regulated:

$$W(e^u, m^r) = \frac{1}{2}\left[\theta + \beta km^r(k) + (1+\beta)he^u - p^u(e^u, m^r(k))\right]^2 + \Pi^{TOC}(h,k) + \Pi^{IM}(h,k) - \lambda g^u(\underline{k}, h)$$

for $h = \underline{h}, \overline{h}$ and $k = \underline{k}, \overline{k}$

where e^u, p^u, and m^r are defined in (30), (27), and (22)–(23) respectively. By substituting (7), g^u, and (6) evaluated in e^u, we obtain:

$$\begin{aligned} W(e^u, m^r) = {} & \frac{1}{2}(\theta + \beta km^r + (1+\beta)he^u - p^u(e^u))^2 \\ & + (1+\lambda)\left[(\theta + (1+\beta)he^u - p^u(e^u))p^u(e^u) - \frac{e^{u^2}}{2}\right] \\ & + \beta km^r p^u(e^u) + \lambda\beta\underline{k}m^r p^u(e^u) - \lambda p_a(\underline{h}, \underline{k}) \\ & + he^u + km^r - \frac{m^{r^2}}{2} \text{ for } k = \underline{k}, \overline{k}, \\ & \text{for } h = \underline{h}, \overline{h} \end{aligned}$$

where $p_a(\underline{h}, \underline{k})$ is the access price defined in (24).

The above expression makes it clear that, for any level of access price $p_a(\underline{h}, \underline{k})$ and of infrastructure maintenance investment m^r, the welfare loss deriving from delegating the operation of the railway service and the decision about its price to a public local authority strictly depends on the distortion w.r.t. the optimal value e^* of the investment level in rolling stock maintenance e^u undertaken by the local PLA-TOC. This will be the subject of the next section.

4.3.1. Welfare analysis

In this section we want to determine the sign and size of the distortion of the investment level chosen by an unregulated PLA-TOC, e^u, with respect to the social optimal level e^*, for any combination of $\underline{h}, \overline{h}$ and $\underline{k}, \overline{k}$. This can be achieved by determining the sign of the derivative of the social welfare function w.r.t to investment e, evaluated at e^u, since:

$$W_e(e^*) = 0 \text{ and } W_{ee} < 0 \Longrightarrow \quad (31)$$

$$\Longrightarrow W_e(e^u) \gtreqless 0 \Longleftrightarrow e^* - e^u \gtreqless 0 \quad (32)$$

In (31) we have the first order condition and the second order condition of the benchmark case respectively. From those conditions, it follows that if the sign of W_e evaluated at e^u is positive (negative), then e^* is higher (lower) than e^u.

By subtracting the first order condition w.r.t. e (28) of the unregulated PLA-TOC decision problem from the first order condition w.r.t. e (14) of the social welfare function evaluated at e^u, we get

$$\begin{aligned} W_e(e^u) = {} & (1+2\lambda)\beta(1+\beta)h\left[km^*(h,k) - \widetilde{km}^r\right] + \\ & + \lambda^2\beta(1+\beta)h\left[km^*(h,k) - \underline{k}m^r(\underline{k})\right] \\ & + L_2(\lambda)h \end{aligned} \quad (33)$$

For $h = \underline{h}, \overline{h}$ $k = \underline{k}, \overline{k}$

In (33) we can isolate three terms corresponding to the three lines. Each of these three terms is related to a specific source of distortion of e^u w.r.t e^* described in Section 3.1.

Both the first two terms can be explained by the following reasons: (i) the asymmetry of information of the PLA-TOC w.r.t. k on the investment level actually undertaken by a regulated private IM; (ii) the fact that the investment levels $m(k, s_I^*)$ chosen by a regulated private IM are equal to $m^*(\underline{h}, \underline{k})$ for $k = \underline{k}$, and equal to $\frac{\overline{k}m^*(\underline{h},\underline{k})}{\underline{k}}$ for $k = \overline{k}$, for any h, rather than the corresponding benchmark levels $m^*(h,k)$.

More specifically, the first line in (33) captures the distortion of e^u w.r.t e^* due to the fact that, as pointed out in the previous section (see (30)), e^u is set on the expected value of km^r rather than on the optimal level $m^*(h,k)$ which determines e^* (see (15)).

Definition 1. We denote by AI-md direct effect the term:

$$km^*(h,k) - \widetilde{km^r} = km^*(h,k) - \upsilon\underline{k}m^*(\underline{h},\underline{k}) - (1-\upsilon)\overline{k}^2 m^*(\underline{h},\underline{k})/\underline{k}$$

where AI stands for Asymmetry of Information and md refers to the distortion of $m^r(k)$ w.r.t. $m^*(h,k)$.

[36] To simplify the notation, where it is not strictly necessary we omit the dependence of e^u and p^u from exogenous parameters.

As we will show, the sign of this term depends on the state of the world, that is on the particular combination of values of h and k.

The term on the second line in (33) captures the distortion of e^u w.r.t. e^* due to the fact that the retail price p^u is set on the basis of $\underline{k}m^*(\underline{h},\underline{k})$ for any couple of values of h and k, since only this value matters for the participation constraint when there is asymmetry of information on k (see (27)). In other words, this term indicates the distortion of e^u w.r.t. e^* arising from the distortion of p^u w.r.t. p^*, when both are evaluated at e^u.

Definition 2. We denote by price effect of the asymmetry of information the term:

$$km^*(h,k) - \underline{k}m^*(\underline{h},\underline{k}) \text{ for } h = \underline{h}, \bar{h}$$

Since this term is never negative and is null at $(\underline{h},\underline{k})$, from (32) it follows that by the sole *price effect* of the asymmetry of information, the PLA-TOC would realize a level of investment lower than the optimal one.

Finally, the term on the third line of (33) measures the distortion of e^u with respect to e^* due to the fact that the PLA-TOC does not take into account the reduction in the cost of access generated by her own investment in the rolling stock maintenance which is given by h.

Definition 3. We denote by externality effect the term:

$$L_2(\lambda)h$$

The *externality effect* is always positive, so that it leads to underinvestment in rolling stock maintenance.

From the previous analysis it follows that as both the price effect and the externality effect are positive, to determine the sign of (33) and thereby (using (32)) the direction and the size of the distortion of the investment level e^u w.r.t e^*, we need to determine the sign of the *AI-md direct effect* for any combination of values of h and k. Whenever this sign is positive, all effects go in the same direction, which implies $e^u < e^*$. Instead, if the sign of this term is negative, by the sole *AI-md direct effect* the PLA-TOC would realize a level of investment higher than $e^*(h,k)$; in this case the final result of the interaction of the three effects will depend on the size of the overinvestment due to the sole *AI-md direct effect* and the size of the under-investment due to both the *price effect* and the *externality effect*.

We summarize the results in the following Proposition:

Proposition 1. *With an unregulated PLA-TOC and a regulated private IM,*

1. *when $k = \bar{k}$ there is underinvestment in the maintenance of the rolling stock;*
2. *when $k = \underline{k}$ we can have also overinvestment, moreover*
   - (a) *underinvestment in the maintenance of the rolling stock in state of the world $(\underline{h},\underline{k})$ implies underinvestment also in state of the world $(\bar{h},\underline{k})$;*
   - (b) *overinvestment in the maintenance of the rolling stock in state of the world $(\underline{h},\underline{k})$ may imply either underinvestment or overinvestment in state of the world $(\bar{h},\underline{k})$.*

Two remarks are worthwhile. First, in the case $k = \bar{k}$ the *AI-md direct effect* becomes:

$$\bar{k}m^*(h,\bar{k}) - \upsilon\underline{k}m^*(\underline{h},\underline{k}) - (1-\upsilon)\bar{k}^2 m^*(\underline{h},\underline{k})/\underline{k} > 0 \text{ for } h = \underline{h}, \bar{h} \tag{34}$$

which implies that the *AI-md direct effect* contributes to inducing the PLA-TOC to undertake a level of investment lower than the optimal one, i.e. (33) is positive in the states of the world $(\underline{h},\bar{k})$ and $(\bar{h},\bar{k})$.

Second, note that the size of the overinvestment is higher the lower the a-priori probability υ that the state $\underline{k}$ occurs is. Since in the case $(\underline{h},\underline{k})$ the *price effect* is equal to zero, it follows that the sign and the size of the distortion of e^u w.r.t e^* (i.e. the sign and the absolute value of (33) evaluated at $h = \underline{h}$ and $k = \underline{k}$) depend on which effect will prevail between the overinvestment generated by the *AI-md direct effect* and the underinvestment generated by the *externality effect* which is equal to $\underline{h}$.

As we will see in the next section, this conclusion has relevant consequences for the effects of PLA-TOC regulation, inasmuch as a regulated PLA-TOC may also widen the magnitude of the distortion.

4.4. Rolling stock maintenance with regulated local public train operating company

The decision of a NRA to regulate a PLA-TOC can be motivated by the following two aims:

(1) to ensure an amount of investment in rolling stock maintenance which is not lower than $e^*(\underline{h},\underline{k})$ in state of the world $(\underline{h},\underline{k})$ and thereby in each state of the world;

(2) to induce the PLA-TOC to internalize the reduction in the cost of access generated by its investment in the rolling stock maintenance, i.e. to internalize what we have called *externality effect*.

From the analysis of the previous section, it follows that the achievement of this second objective automatically implies the achievement of the first one, because the distortion ascribable to the sole asymmetry of information of the PLA-TOC on k, which we have named *AI-md direct effect*, leads to a level of investment higher than $e^*(\underline{h},\underline{k})$, in state of the world $\underline{h},\underline{k}$ (see (42)).[37]

As we have highlighted in Section 2, in our framework the NRA may employ the access price to regulate the PLA-TOC; this consists in setting the access price on the basis of the observed state of the rolling stock $w(h,e^r) = he^r$ (condition (25)), which depends on the level of investment undertaken by the PLA-TOC

$$p_a(\underline{k},e^r) = -he^r - \underline{k}m^*(\underline{k},\underline{h}) + \frac{m^*(\underline{k},\underline{h})^2}{2} \tag{35}$$

where e^r stands for the investment undertaken by a *regulated* PLA-TOC.

The PLA-TOC's maximization problem is the same as the one formalized in Section 4.3 obtained substituting p_a with $p_a(\underline{k},e^r)$:

$$\begin{aligned}\max_{p,e} EW^{PLAr}(h) = \max_{p,e}\ & \frac{\upsilon}{2}(\theta + \beta\underline{k}m^r(\underline{k}) + (1+\beta)he - p)^2 \\ & + \frac{(1-\upsilon)}{2}(\theta + \beta\bar{k}m^r(\bar{k}) + (1+\beta)he - p)^2 \\ & + (1+\lambda)\left[(\theta + (1+\beta)he - p)p - p_a - \frac{e^2}{2}\right] \\ & + \beta\left(\lambda\underline{k}m^r(\underline{k}) + \widetilde{km^r}\right)p\end{aligned}$$

For $h = \underline{h}, \bar{h}$

The value of e^r solves

$$\begin{aligned}EW_e^{PLAr} = & \lambda(1+\beta)hp^u(e^r) - \left[1 + \lambda - (1+\beta)^2h^2\right]e^r \\ & + \left[\theta + \beta\widetilde{km^r}(1+\beta) + (1+\lambda)\right]h = 0\end{aligned}$$

By substituting p^u as defined in (27) we obtain

$$EW_e^{PLAr}(h) = EW_e^{PLAu}(h) + L_2(\lambda)h = 0 \tag{36}$$

where $EW_e^{PLAu}(h)$ – defined in (28) – is the first order condition w.r.t. e when the PLA-TOC is not regulated (i.e. it is determined by (24)) and the additional term in (36) is exactly the *externality effect* (see (33)).

From (28) and (36) we derive

$$\begin{aligned}e^r = & \frac{(1+\beta)hL_1(\lambda)\theta + \beta(1+\beta)h((1+2\lambda)\widetilde{km^r} + \lambda^2\underline{k}m^*(\underline{k},\underline{h}))}{L_2\lambda - (1+\beta)^2h^2L_1(\lambda)} + \\ & + \frac{L_2(\lambda)h}{L_2(\lambda) - (1+\beta)^2h^2L_1(\lambda)}\end{aligned} \tag{37}$$

[37] Remember that in state of the world $(\underline{h},\underline{k})$ the *price effect* is null.

$h = \underline{h}, \overline{h}$

Condition (36) can be easily explained. Since the access price is a cost for the PLA-TOC, the regulation of its value on the basis of the observed state of the rolling stock induces the PLA-TOC to take into account the effect of her own investment on the access price; this is equivalent to internalizing the effect of investment in the rolling stock maintenance on the cost of access, i.e. internalizing the *externality effect*.[38] As a consequence, the distortion of e^r w.r.t. e^* will only be due to the asymmetry of information of the PLA-TOC with respect to k, i.e. the *direct Ai-md effect* and the *price effect* of asymmetric information, both disentangled in (33).

Proposition 2. *With a regulated private IM, when $k = \overline{k}$, the PLA-TOC's regulation aimed at internalizing the externality effect leads to underinvestment in the maintenance of the rolling stock. Instead, when $k = \underline{k}$, it leads to overinvestment when $h = \underline{h}$, while it can lead to either underinvestment or overinvestment when $h = \overline{h}$.*

Note that, for the case $(\overline{h}, \underline{k})$, the implications of PLA-TOC regulation in terms of under or overinvestment are ambiguous, in that in this state of the world the *price effect of AI* is positive and the *Ai-md effect* is negative, and hence the internalization of the externality induced by the regulation of the PLA-TOC may lead either to (i) a reduction in underinvestment, (ii) an increase in over-investment, or (iii) a shift from under to overinvestment in rolling stock maintenance. In the first case we would have a reduced distortion, in the second case an increased distortion, while in the third case it would not be possible to establish *a-priori* if the distortion increases or decreases.

4.4.1. Welfare effects of PLA-TOC regulation

In what follows we want to assess the welfare effects of regulating a PLA-TOC by adopting the access price rule (35) instead of rule (24). This can be done by disentangling both a *direct* welfare effect, which concerns the impact of the two rules on the value of the access price, and an *indirect welfare effect* which captures the impact of removing the *externality effect* on sign and size of the distortion of the investment level e^u undertaken by an unregulated PLA-TOC (w.r.t. e^* for any pair (h, k)).

As far as the *direct welfare effect* is concerned, it is easy to verify that since $p_a(h, k) < p_a(\underline{h}, k)$ for $k = \underline{k}, \overline{k}$, the removal of the *externality effect* raises the investment level undertaken by a PLA-TOC for any h, i.e. $e^r > e^u$. Because the access price has a social cost of λ,[39] the adoption of the rule $p_a(h, k)$ instead of $p_a(\underline{h}, k)$ generates a direct expected welfare gain, denoted by ΔEW^{dir}, which is equal to

$$\Delta EW^{dir}(h) = \lambda \left[p_a(\underline{h}, \underline{k}) - r p_a(\underline{h}, \underline{k}) - (1-r) p_a(\overline{h}, \underline{k}) \right] =$$
$$\lambda \left[r \underline{h} e^r(\underline{h}, \widetilde{km^r}) + (1-r) \overline{h} e^r(\overline{h}, \widetilde{km^r}) - \underline{h} e^u(\underline{h}, \widetilde{km^r}) \right]$$

where r is the probability of $h = \underline{h}$. By substituting (30) and (37), the above expression reduces to

$$\Delta EW^{dir} = \lambda(1-r) \left[\overline{h} e^u(\overline{h}, \widetilde{km^r}) - \underline{h} e^u(\underline{h}, \widetilde{km^r}) \right] + \lambda L_2(\lambda) \left[\frac{r \underline{h}^2}{Soc_e(\underline{h})} + \frac{(1-r)\overline{h}^2}{Soc_e(\overline{h})} \right] > 0$$

where $Soc_e(h) = L_2(\lambda) - (1+\beta)^2 h^2 L_1(\lambda)$

As far as the *indirect welfare effect* is concerned, we can state that the removal of the *externality effect* increases welfare in that state of the world where it reduces the size of the distortion of e^u w.r.t. e^*. This unambiguously occurs when the other two effects disentangled in (33), i.e the direct *Ai-md effect* and the *price effect* of asymmetric information, contribute with the *externality effect* to generating a level of investment e^u lower than e^*. As stated in Proposition 1 this is the case when $k = \overline{k}$.

On the other hand, when $k = \underline{k}$ we can have either underinvestment or overinvestment in the unregulated case, given that for $h = \underline{h}$, the *Ai-md effect* alone would lead to $e^u > e^*$, while for $h = \overline{h}$ the direction of the *Ai-md effect* is ambiguous. This, in turn, implies that the internalization of the *externality effect* leads unambiguously to overinvestment in state of the world $(\underline{h}, \underline{k})$ (see Proposition 1), while again the result is in principle ambiguous in state of the world $(\overline{h}, \underline{k})$.

To sum up, the welfare effect of the internalization of the *externality effect* depends on which outcome occurs in the unregulated case: overinvestment or underinvestment. The following Proposition holds:

Proposition 3. *A regulatory access price rule that induces the PLA-TOC to internalize the effect of the investment in rolling stock maintenance on the access cost increases the welfare in the states of the world where $k = \overline{k}$. When $k = \underline{k}$ it can decrease it, but this is more likely to occur just in the case where it is more likely that this welfare loss is lower than the expected welfare gains generated in the other states of the word. Therefore regulation is likely to be socially valuable.*

The consequence of Proposition 3 is quite interesting: regulating the PLA-TOC may lead to welfare losses in situations where the realization of the worst state of the world is more likely to occur.

From the obtained results, we can draw the conclusion that the adoption of an access price rule which induces the PLA-TOC to internalize the effect of the investment in rolling stock maintenance on the access cost affects the expected welfare through three channels:

(i) it reduces the access price which increases the expected welfare in proportion λ.

(ii) It reduces the distortion of the investment in rolling stock maintenance with respect to the optimal values just in the states of the world in which $k = \overline{k}$.

(iii) In the worst state of the world (i.e. $(\underline{h}, \underline{k})$), PLA-TOC regulation may decrease welfare; this is more likely to occur the higher the a priori probability associated to $\underline{k}$. However this is just the case where the welfare gains generated by the regulation of the PLA-TOC in the other states of the word are not likely to overcompensate the welfare loss in state $(\underline{h}, \underline{k})$, so that the expected welfare increase.

5. Conclusions

In this paper, we have provided a first theoretical analysis of regional railway passengers service accounting from the main characteristics of the sector. In a framework in which we assume that the train operating company is owned and managed by the local government, and there is a national infrastructure manager our model permits to highlight the misalignment of incentives to investment in the maintenance of the rolling stock and of the network. In particular, three sources of inefficiency are identified. The first two sources of inefficiency are quite standard and are related to the positive externalities generated by the investment in the maintenance of the infrastructure and rolling stock. Thus, due to these externalities there is underinvestment both in infrastructure and rolling stock maintenance.

The third source of inefficiency is more peculiar to the railway sector, and to the best of our knowledge, it has never been pointed out by the theoretical economic literature on railway sector. This is related to the fact that the productivity of the investment in infrastructure maintenance is private information of the infrastructure manager, i.e. there is asymmetry of information of the train operating company with respect to the marginal productivity of investment in infrastructure maintenance. We show that this kind of inefficiency counterbalance the first two, and it may lead also to overinvestment in rolling stock maintenance.

[38] Here the role of the access charge largely resembles that of pigouvian taxation in internalizing positive externalities.

[39] A marginal increase of the access price involves an equivalent welfare gain due to the related increase in IM's profit, and a welfare loss of $1 + \lambda$ due to the related decrease of TOC's profit and so to an increase of public subsidies.

Then, we investigate the consequences of these results for regulation. In a regulatory framework in which:

- the infrastructure manager is induced to implement the optimal investment in the worst situation;
- the train operating company is regulated through the access price, so that the externality generated by the investment in rolling stock maintenance is internalized;

We show that investment decisions may be less efficient then in a regulatory framework where the train operating company is unregulated, and most importantly this is more likely to occur the higher the probability that the worst conditions realize (i.e. low IM's and TOC's marginal productivity of maintenance investment).

CRediT authorship contribution statement

Antonio Scialà: Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft, Writing – review & editing. **Francesca Stroffolini:** Conceptualization, Formal analysis, Investigation, Methodology, Writing – original draft.

Appendix

Proof of Eq. (19). Solving (18) w.r.t. m, substituting (15), and multiplying both sides by $(-W_{ee})(-W_{mm})$ we get (see the subsequent proof):

$$m^*(h,k)(-W_{ee})(-W_{mm}) = k(-W_{ee})L_2(\lambda) + \beta k\theta L_1(\lambda)(-W_{ee}) + \beta(1+\beta)^2 kh^2\theta L_1(\lambda)^2 + \beta^2k^2(1+\beta)^2h^2L_1(\lambda)^2 m^*(h,k) + \beta k(1+\beta)L_1(\lambda)L_2(\lambda)h^2$$

solving by $m^*(h,k)$ and using the explicit expression of $-W_{ee}$ we get:

$$m^*(h,k) = \frac{k\left[L_2(\lambda)-(1+\beta)^2h^2L_1(\lambda)\right]L_2(\lambda)}{D(h,k)} + \frac{\beta k\theta L_1(\lambda)L_2(\lambda)+\beta k(1+\beta)L_1(\lambda)L_2(\lambda)h^2}{D(h,k)} \tag{38}$$

where:

$$D(h,k) = W_{mm}W_{ee} - W_{em}^2 = L_2(\lambda)^2 - \left[(1+\beta)^2h^2+\beta^2k^2\right]L_1(\lambda)L_2(\lambda)$$

From the assumption that second order condition of the social planner maximization problem are satisfied we have that $D(h,k) > 0$

We can simplifies (38) to obtain

$$m^*(h,k) = kL_2(\lambda)\frac{\left[L_2(\lambda)-(1+\beta)^2h^2L_1(\lambda)+\beta\theta L_1(\lambda)+\beta(1+\beta)h^2L_1(\lambda)\right]}{L_2(\lambda)\left[L_2(\lambda)-\left((1+\beta)^2h^2+\beta^2k^2\right)L_1(\lambda)\right]} = \frac{k\left[L_2(\lambda)+\beta\theta L_1(\lambda)-(1+\beta)h^2L_1(\lambda)\right]}{\left[L_2(\lambda)-\left((1+\beta)^2h^2+\beta^2k^2\right)L_1(\lambda)\right]} \quad \blacksquare$$

Proof of Eq. (38). Substituting (15) in (38) gets

$$m^*(h,k)(-Soc_e(h))(-Soc_m(k)) = k(-Soc_e(h))L_2(\lambda) + \beta k\theta L_1(\lambda)(-Soc_e(h)) + \beta k(1+\beta)^2h^2L_1(\lambda)^2\theta + \beta^2k^2(1+\beta)^2h^2L_1(\lambda)^2m^*(h,k) + \beta k(1+\beta)hL_1(\lambda)L_2(\lambda)h \tag{39}$$

where

$$-Soc_e(h) = L_2(\lambda) - (1+\beta)^2h^2L_1(\lambda)$$
$$-Soc_m(k) = L_2(\lambda) - \beta^2k^2L_1(\lambda)$$

$$m^*(h,k)\left[(-Soc_e(h))(-Soc_m(k)) - \beta^2k^2(1+\beta)^2h^2L_1(\lambda)^2\right] = k(-Soc_e(h))L_2(\lambda) + \beta k\theta L_1(\lambda)(-Soc_e(h)) + \beta k(1+\beta)^2h^2L_1(\lambda)^2\theta + \beta k(1+\beta)hL_1(\lambda)L_2(\lambda)h$$

with

$$D(h,k) = \left[(-Soc_e(h))(-Soc_m(k)) - \beta^2k^2(1+\beta)^2h^2L_1(\lambda)^2\right] = L_2(\lambda)^2 - (1+\beta)^2h^2L_1(\lambda)L_2(\lambda) - \beta^2k^2L_1(\lambda)L_2(\lambda) > 0$$

By substituting $-Soc_e(h)$ in (39) the result follows. ■

Proof of Proposition 1.

1. For $k = \overline{k}$ and $h = \underline{h}, \overline{h}$ the *AI-md effect* turns out to be:

$$km^*(h,k) - \widetilde{km}(k,s_I^*) = \overline{k}m^*(h,\overline{k}) - \upsilon\underline{k}m^*(\underline{h},\underline{k}) - (1-\upsilon)\overline{k}^2m^*(\underline{h},\underline{k})/\underline{k} = \overline{k}m^*(h,\overline{k}) - m^*(\underline{h},\underline{k})\frac{\upsilon(\underline{k}^2-\overline{k}^2)+\overline{k}}{\underline{k}} \tag{40}$$

where the latest inequality follows from the assumption that $\underline{k} < \overline{k}$, noticing that $m^*(h,\overline{k}) > m^*(\underline{h},\underline{k})$, and that $\upsilon \in [0,1]$. From (33) evaluated at $(\overline{k},h)$ we get:

$$W_e(\overline{k},h,e^{AL}(h,.)) > 0 \Longleftrightarrow e^{AL}(h,\widetilde{km}(k,s_I^*)) < e^*(h,\overline{k}) \text{ for } h = \underline{h},\overline{h} \tag{41}$$

2. for $k = \underline{k}$, $h = \underline{h}$. The *AI-md direct effect* becomes

$$\underline{k}m^*(\underline{h},\underline{k}) - \upsilon\underline{k}m^r(\underline{k}) - (1-\upsilon)\overline{k}m(\overline{k}) = -\frac{(1-\upsilon)(\overline{k}^2-\underline{k}^2)m^*(\underline{h},\underline{k})}{\underline{k}} < 0 \tag{42}$$

which implies that by the sole *AI-md direct effect*, the TOC would realize a level of investment higher than e^*.[40] For $k = \underline{k}$, $h = \overline{h}$. The *AI-md direct effect* becomes

$$\underline{k}m^*(\overline{h},\underline{k}) - \widetilde{km^r} = \underline{k}m^*(\overline{h},\underline{k}) - \upsilon\underline{k}m^*(\underline{h},\underline{k}) - (1-\upsilon)\frac{\overline{k}^2m^*(\underline{h},\underline{k})}{\underline{k}}$$

which can be written as

$$\underline{k}m^*(\overline{h},\underline{k}) - \widetilde{km^r} = -(1-\upsilon)\left[\frac{\overline{k}^2m^*(\underline{h},\underline{k})}{\underline{k}} - \underline{k}m^*(\underline{h},\underline{k})\right] + \underline{k}\left[m^*(\overline{h},\underline{k}) - m^*(\underline{h},\underline{k})\right] \tag{43}$$

where the first term is the *AI-md direct effect* in state of the world $(\underline{h},\underline{k})$ and is negative (condition (42)) while the second term is positive. Therefore, in this case, the *AI-md direct effect* can turn out to be either positive or negative, pushing the PLA-TOC towards underinvestment or overinvestment respectively.

(a) The analysis carried out above shows that the sign of the overall effect of the asymmetric information of the PLA-TOC about k on the investment in rolling stock maintenance cannot be univocally determined. However, since from (43) it is clear that the magnitude of the

[40] It is informative to break down the above effect into the one due to asymmetry of information and the one due to the distortion w.r.t. the benchmark of the investment level chosen by the IM. To do this, let us add and subtract $(1-\upsilon)\overline{k}m^*(\underline{h},\overline{k})$ in (42), which can be written as

$$-(1-\upsilon)\left[\overline{k}m^*(\underline{h},\overline{k}) - \underline{k}m^*(\underline{h},\underline{k})\right] + (1-\upsilon)\left[\overline{k}m(\underline{h},\overline{k}) - m(\overline{k},s_I^*)\right]$$

where the first term captures the effect of the asymmetry of information when the levels of investment chosen by the IM are equal to the benchmark level for any value of k, whereas the second term captures the effect of the distortion in the investment level of IM for $k = \overline{k}$. Since the first term is negative while the second is positive, it follows that the over-investment effect in (42) is due to the effect of asymmetry of information which more than compensates for the underinvestment due to the distortion in the investment level of IM.

AI-md direct effect is greater in state of the world $(\underline{h}, \underline{k})$ than in the state of the world $(\bar{h}, \underline{k})$, we are able to state that when underinvestment in rolling stock maintenance emerges in state of the world $(\underline{h}, \underline{k})$, it will emerge even more strongly in state of the world $(\bar{h}, \underline{k})$.[41]

(b) On the other hand, when overinvestment in rolling stock maintenance emerges in state of the world $(\underline{h}, \underline{k})$ either overinvestment (even though lower than in the case $(\underline{h}, \underline{k})$) or even underinvestment may emerge in state of the world $(\bar{h}, \underline{k})$. ■

Proof of Proposition 2. By substituting (33) into (36) and subtracting this latter from (14) evaluated at e^r we obtain:

$$\begin{aligned} W_e(e^r) &= W_e(e^u) - L_2(\lambda)h = \\ &(1+2\lambda)\beta(1+\beta)h\left[km^*(h,k) - \upsilon\underline{k}m^*(\underline{k},\underline{h}) - (1-\upsilon)\frac{\bar{k}^2 m^*(\underline{k},\underline{h})}{\underline{k}}\right] \\ &+ \lambda^2\beta(1+\beta)h(km^*(h,k) - \underline{k}m^*(\underline{k},\underline{h})) \end{aligned}$$

where the second line refers to the *direct Ai-md effect* and the third line to the *price effect of AI*. From (34), (42), and taking into account that the *price effect* is null for $k = \underline{k}, h = \underline{h}$ and negative in all the other cases, it follows that

$$\begin{aligned} &W_e(e^r) > 0 \iff e^r < e^* \\ &\quad \text{for } k = \bar{k},\ h = \underline{h}, \bar{h} \\ &W_e(e^r) < 0 \iff e^r > e^* \\ &\quad \text{for } k = \underline{k} \text{ and } h = \underline{h} \end{aligned}$$ ■

Proof of Proposition 3. If $e^u > e^*$, the removal of the *externality effect* is welfare reducing because it increases the overinvestment distortion of an amount equal to $e^r - e^u$. Note that this result is more likely to arise the lower the probability that just the state $\underline{k}$ occurs is.

On the other hand, if $e^u < e^*$, the removal of the *externality effect* generates two opposite welfare effects: a welfare gain to the extent that it eliminates the underinvestment equal to $e^* - e^u$, and a welfare loss to the extent that it leads to an overinvestment equal to $e^r - e^*$.

Let us analyse these two cases for state of the world $(\underline{h}, \underline{k})$, which is of special relevance from a regulatory viewpoint as it is the worst state of the world.

In the case of $e^u < e^* < e^r$, the regulation of a PLA-TOC is socially valuate if the expected loss arising from the elimination of the *externality effect* (measured by $L_2(\lambda)\underline{h}$) in state of the world $(\underline{h}, \underline{k})$ is lower than the expected gain occurring in the other states of the world. This is obviously more likely to occur the lower the probability that state of the world $(\underline{h}, \underline{k})$ occurs is. ■

References

Andersson, M., Hultén, S., 2016. Transaction and transition costs during the deregulation of the Swedish Railway market. Res. Transp. Econ. 59, 349–357.

Araújo, S., 2011. Has Deregulation Increased Investment in Infrastructure?. OECD Economics Department Working Papers, vol. 892, OECD Publishing, Paris.

Armstrong, M., Cowan, S., Vickers, J., 1994. Regulatory Reform: Economic Analysis and British Experience. vol. 20, MIT Press, Cambridge (MA).

Arrigo, U., Di Foggia, G., 2013a. Assessing the level of state aid to the railway sector in selected European countries: Insights from the Italian case. Eur. J. Bus. Econ. 8 (4), 1–9.

Arrigo, U., Di Foggia, G., 2013b. Schemes and levels of state aid to rail industry in Europe: evidences from a cross-country comparison. Eur. J. Bus. Econ. 8 (3), 14–20.

Baron, D.P., Myerson, R.B., 1982. Regulating a monopolist with unknown costs. Econometrica 911–930.

Cantos, P., Pastor, J.M., Serrano, L., 1999. Productivity, efficiency and technical change in the European railways: a non-parametric approach. Transportation 26 (4), 337–357.

European Commission, 1991. Council directive of 29 July 1991 on the development of community's railways (91/440/EEC). http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1416170084502&uri=CELEX:32014R0269.

European Commission, 2012. Council directive of 21 November 2012 establishing a single European railway area (2012/34/EU). http://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1416170084502&uri=CELEX:32014R0269.

European Commission, 2016. Fifth Report on the Monitoring development of the rail market. https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31991L0440.

Friebel, G., Ivaldi, M., Vibes, C., 2010. Railway (de) regulation: a European efficiency comparison. economica 77 (305), 77–91.

Gathon, H.-J., Pestieau, P., 1995. Decomposing efficiency into its managerial and its regulatory components: The case of European railways. European J. Oper. Res. 80 (3), 500–507.

Growitsch, C., Wetzel, H., 2009. Testing for economies of scope in European railways: an efficiency analysis. J. Transp. Econ. Policy 43 (1), 1–24.

Jensen, A., Stelling, P., 2007. Economic impacts of Swedish railway deregulation: A longitudinal study. Transp. Res. E: Logist. Transp. Rev. 43 (5), 516–534.

Johansson, P., Nilsson, J.-E., 2004. An economic analysis of track maintenance costs. Transp. policy 11 (3), 277–286.

Laffont, J.-J., 1994. The new economics of regulation ten years after. Econometrica 507–537.

Laffont, J.-J., Tirole, J., 1993. A Theory of Incentives in Procurement and Regulation. MIT Press, Cambridge (MA).

Laurino, A., Ramella, F., Beria, P., 2015. The economic regulation of railway networks: A worldwide survey. Transp. Res. A: Policy Pract. 77, 202–212.

Lundén, R., Paulsson, B., 2009. Introduction to wheel–rail interface research. In: Wheel–Rail Interface Handbook. Elsevier, pp. 3–33.

Marschnig, S., 2016. Innovative track access charges. Transp. Res. Procedia 14, 1884–1893.

McNulty, R., 2011. Realising the Potential of GB Rail: Final Independent Report of the Rail Value for Money Study. Department for Transport and Office of Rail Regulation, London.

Merkert, R., Smith, A.S., Nash, C.A., 2012. The measurement of transaction costs—evidence from European railways. J. Transp. Econ. Policy 46 (3), 349–365.

Mizutani, F., Smith, A., Nash, C., Uranishi, S., 2015. Comparing the costs of vertical separation, integration, and intermediate organisational structures in European and East Asian railways. J. Transp. Econ. Policy 49 (3), 496–515.

Mizutani, F., Uranishi, S., 2013. Does vertical separation reduce cost? An empirical analysis of the rail industry in European and East Asian OECD Countries. J. Regul. Econ. 43, 31–59.

Mueller, D.C., 2003. Public Choice III. Cambridge University Press, Cambridge (UK).

Nash, C., 2008. Passenger railway reform in the last 20 years–European experience reconsidered. Res. Transp. Econ. 22 (1), 61–70.

Newbery, D.M., 2002. Privatization, Restructuring, and Regulation of Network Utilities. vol. 2, MIT Press, Cambridge (MA).

Oates, W.E., 1972. Fiscal Federalism. Harcourt Brace Jovanovich, New York.

Pflieger, G., 2014. What "regulatory state"? Explaining the stability of public spending and redistribution functions after regulatory reforms of electricity and rail services in the United Kingdom and Germany. Law Policy 36 (2), 195–221.

Smith, A.S., Benedetto, V., Nash, C., 2018. The impact of economic regulation on the efficiency of European railway systems. J. Transp. Econ. Policy (JTEP) 52 (2), 113–136.

Smith, A., Iwnicki, S., Kaushal, A., Odolinski, K., Wheat, P., 2017. Estimating the relative cost of track damage mechanisms: combining economic and engineering approaches. Proc. Inst. Mech. Eng., F: J. Rail Rapid Trans. 231 (5), 620–636.

Smith, A.S., Nash, C., 2014. Rail Efficiency: Cost Research and Its Implications for Policy. OECD, International Transport Forum Discussion Paper, 2014/22.

Van de Velde, D., Nash, C., Smith, A., Mizutani, F., Uranishi, S., Lijesen, M., Zschoche, F., 2012. EVES-Rail: Economic Effects of Vertical Separation in the Railway Sector. Technical report, Community of European Railways, Brussels.

[41] By the same token, when overinvestment emerge in the state of the world $(\bar{h}, \underline{k})$, then it will emerge even more in the state of the world $(\underline{h}, \underline{k})$.