https://doi.org/10.1016/j.tra.2019.11.011

High-speed rail networks, capacity investments and social welfare

Valentina Bracaglia [a], Tiziana D'Alfonso [b,*], Alberto Nastasi [b], Dian Sheng [c], Yulai Wan [d], Aming Zhang [e]

[a] Autostrade per l'Italia S.p.A., Via Bergamini 50, Roma
[b] Department of Computer, Control and Management Engineering, Sapienza Università di Roma, Via Ariosto 25, Roma
[c] School of Management, Huazhong University of Science and Technology, Wuhan 430074, China
[d] Department of Logistics and Maritime Studies, Hong Kong Polytechnic University, Hong Kong, China
[e] Sauder School of Business, University of British Columbia, 2053 Main Mall, Vancouver, Canada

Abstract

In this paper, we analytically study the performance of two topologies of high-speed rail (HSR) networks: isolated-corridors and grid networks. We evaluate how HSR configuration affects capacity of newly developed infrastructure, profits and social welfare by considering a number of factors, namely economies of traffic density, market size, operating cost and cost of capital. Our investigations focus on a social welfare-maximizing entity which provides HSR train services as well as develops infrastructure. We find that although the grid network allows for more new markets, the isolated-corridors network may perform better in terms of social welfare when the cost of capital is high, the demand in the new markets established by the isolated-corridors network is high, the operating cost (excluding density effect) is high and the traffic density effect is weak. We also identify cases where the optimal network configuration in terms of social welfare may not be optimal in other aspects, such as capacity, consumer surplus and profits. For example, when the density effect is strong, grid network is likely to be socially optimal, but it faces more difficulty in recovering the invest cost comparing with the alterative network.

Keywords: High-speed rail, network configuration, traffic density effect, capacity, social welfare

*__Corresponding author__
Tel: +39 06 77274 105
Fax: +39 06 77274 074
E-mail address: dalfonso@dis.uniroma1.it (Tiziana D'Alfonso)

1. Introduction

High-speed rail (HSR) has been a focal element of transport policies in many countries (Zhang et al., 2019). According to the International Union of Railways (UIC, 2018a), in 2016 HSR systems have carried over 715 billion passenger-km (see also Table 1). Up to October 2018, there are 9,230 km of HSR lines in operation in Europe, 1,697 km under construction and 11,067 km being planned. Major developments of HSR networks will take place in Spain and France, which already operate the widest network in the European continent, i.e., 2,852 km and 2,814 km of lines respectively (UIC, 2018b). In Asia, the leading investors in HSR are China and Japan. The Chinese HSR network (the largest in the world) covers 27,684 km and is planned to cover an additional 11,551 km, with 10,026 km of lines already under construction (UIC, 2018b). Japan, a pioneer in HSR investment with 3,041 km of HSR lines in operation, is now constructing 402 km and plans to build additional 194 km by 2046 (UIC, 2018b).

*** *Table 1* ***

Three strategic models for HSR infrastructure development have been observed around the world (Perl and Goetz, 2015): (i) exclusive corridors between megacities of more than 10 million inhabitants, (ii) hybrid networks composed by a mix of newly developed HSR links and existing conventional rail lines, which easily extend HSR services into smaller cities at a lower cost but a lower average service speed, and (iii) comprehensive national grid networks mainly made up of a few new and exclusive HSR lines which intersect with each other to link all major and middle-sized communities across the country. The Japanese Tokaido Line can be considered as a dedicated HSR corridor at its initial development. It originally linked Tokyo and Osaka with dedicated HSR tracks (512 km), but it has been later expanded to serve smaller cities. Another example of dedicated corridor is the Northeast Corridor (Boston - New York City - Washington DC) in the USA served by the Amtrak's Acela Express trains. On the other hand, China provides a pivotal example of comprehensive grid network with a backbone of four major north–south and four major east–west passenger dedicated lines. Another example of an HSR comprehensive grid is provided by Spain, which has been developing a radial network centered on Madrid.

Obviously, the extension of an HSR infrastructure involves considering various tradeoffs between service coverage, capital and operating costs. For instance, extension of an HSR infrastructure to form a grid network provides greater service coverage due to the large number of city pairs connected but incurs higher capital and operating costs, whereas a single isolated corridor has lower capital and operating costs but offers less service coverage. On the one hand, network effects arising in grid networks generate positive externalities which pivot on traffic density and the number of origin-destination pairs served. On the other hand, investment costs as well as operating costs increase as the number of markets served by HSR (and traffic) rises. These trade-offs shape consequently incentives of capacity investment, profits and social welfare.

In this paper, we develop an analytical framework to study performance of two alternative topologies of HSR networks: isolated-corridors and grid networks. We consider a set of cities, one existing HSR corridor and a chance to build one new corridor. There are two options in developing the new corridor, either having the new corridor intersect with the existing corridor to form a grid network, or making the new corridor isolated from the existing corridor. We evaluate and compare capacity investment levels, profits and social welfare of these two options and examine the role of various factors, namely economies of traffic density, market size, operating cost and cost of capital. We finally investigate the potential conflicts in determining the network structure based on social welfare and other concerns, such as capacity level, consumer surplus and profits.

Our investigations focus on a social welfare-maximizing entity which not only provides the HSR services but also invests in the infrastructure. Indeed, this circumstance is observed in several major HSR projects. For instance, China State Railway Group (CR), previously known as China Railway Corporation, is responsible for not only operating China's passenger and freight train services, both conventional and high-speed, but also planning, developing, financing and managing the national railway network. CRC used to be part of the now-defunct Ministry of Railways and was converted to an enterprise wholly-owned by the State in 2013. It is financed solely by the Ministry of Finance and reports directly to the State Council (The World Bank, 2017, pp. 394). In France, SNCF is the state-owned French National Railway Company. SNCF Mobilités is the transport division of SNCF which is responsible for the provision of transport services, i.e., freight and passenger services including the *Train à Grande Vitesse* (meaning "High-speed train"). SNCF Réseau is the infrastructure division of SNCF and is responsible for infrastructure development, operations and maintenance.[1]

[1] Prior to 1997, SNCF managed both rail infrastructure and train operations. SNCF was restructured in 1997 to satisfy the legal order of the European Union for railways, which required vertical separation of the accounts for rail infrastructure and train operations. Ownership of the railway network was transferred to a separate company, named Réseau Ferré de France (RFF). RFF focused on track improvement and development, network investment and financing, and contracted with SNCF to undertake the maintenance and operation of railway infrastructure. On 1 January 2015, RFF and SNCF were once again restructured and combined into the SNCF Group. (The World Bank, 2017, pp. 529).

This paper relates to two strands of literature.[2] The first strand discusses qualitatively HSR performance according to the level of integration with the rest of the network, speed and type of services (Givoni, 2006; Campos and De Rus, 2009; Perl and Goetz, 2015). Givoni (2006) concludes that HSR is best designed to substitute conventional railway services on routes where more capacity is needed, to reduce travel time to further improve the railway service, or to substitute other modes, i.e., air. Campos and De Rus (2009) discuss the economic definition of HSR, showing that it is not speed but the network exploitation model that determines this concept. The authors identify four models: (i) exclusive exploitation (either high-speed trains on high-speed tracks or conventional trains on conventional tracks), (ii) mixed high-speed (high-speed trains on high-speed and conventional tracks), (iii) mixed conventional (conventional trains on high-speed or conventional tracks), and (iv) fully mixed (high-speed and conventional trains on high-speed or conventional tracks). The exclusive exploitation and the mixed high-speed models, for example, allow a more intensive usage of HSR infrastructure and thus higher traffic capacity. The other models, however, must accommodate (with the exception of multiple-track sections of the line) slower trains which frequently occupy a larger number of slots and reduce the possibilities for providing HSR services. The exclusive exploitation model has been adopted by the Japanese Tokaido Line which used standard gauge at its first introduction in 1964, whilst the narrow gauge was used by a conventional line. Chinese HSR started with the mixed HSR model by upgrading conventional lines, but then converted to an exclusive exploitation model after developing the majority of the network with new lines dedicated to HSR services. Perl and Goetz (2015) find HSR being generally profitable in heavily travelled corridors of 300–800 km between major population centers (i.e., Tokyo-Osaka line), while longer distance HSR operations involving smaller population centers can struggle to cover their costs.[3]

[2] A vast branch of literature focuses on the intermodality between HSR and air transport, namely airline-HSR competition (e.g., Yang and Zhang, 2012; D'Alfonso et al., 2015, 2016; Jiang and Zhang, 2016; Wan et al., 2016) and airline-HSR cooperation (e.g., Avenali et al., 2018; Jiang and Zhang, 2014; Jiang et al., 2017; Socorro and Viecens, 2013, Xia and Zhang, 2016; 2017; Xia et al; 2018). In our paper, we abstract away from the discussion on the effects of HSR operations over the competitiveness of air transport, in order to focus on the impact of network configuration on HSR investment levels, profit and social welfare. Our analysis could inform about the pattern of HSR expansion; thus stakeholders including substitute air transport operators can be better prepared for the challenges due to intermodality.

[3] For instance, the Japanese Tokyo-Osaka line has been regarded as the most successful HSR operation in the world, but the extension of the line to cities located in less densely populated areas lead to financial losses for Japan National Railways. The expanded high-speed train network in China has started to make a profit in the populated east (China.com, 2016). In 2015, six high-speed rail lines made a profit, with the Beijing-Shanghai route topping the list at a net profit of Y6.58 billion ($990 million), all connecting mega cities in populated areas with strong economies such as Beijing, Tianjin, Shanghai, Hangzhou, Ningbo, Shenzhen and Guangzhou. Except for the Beijing-Tianjin HSR, the other five lines have managed to turn from deficit to profit within five years of operations. Conversely, services running through the vast central and western regions are still far from breaking even. Performance is closely related to the population and economic strength of cities HSR serves.

The second strand of literature addresses the design problem of a rail transit line. In principle, the basic parameters to be determined in planning a rail transit line include the rail line length, number and locations of stations and fare. The design of these parameters depends very much on the population density in the planning area. This is because the urban population density directly influences the level of passenger demand. Li et al. (2012) provides a literature review of various analytical models of transit service design and highlights main characteristics with respect to: (i) decision variables, such as the location or spacing of transit routes and/or stations, and fare; (2) the objective function, which is typically the minimization of the total system cost (i.e., the sum of operator and user costs); (3) the geometry of transit lines, such as a linear structure or rectangular grid; and (4) demand characteristics, including fixed or elastic demand. Such previous studies (e.g., Liu et al., 1996; Wirasinghe and Seneviratne, 1986; Li et al., 2011) mainly focus on how to optimize the design variables of the transit services approximately and/or numerically, given a specified network configuration.

None of these papers examines analytically how different network configurations affect the economic performance of the transport system. Our paper is the first attempt in the literature to provide an analytical framework to study the performance of alternative topologies of HSR networks. We aim at providing guidance to shape future investments in HSR infrastructure by identifying the relevant factors for the success of alternative topologies. In particular, we isolate relevant constellations of parameters where a given network topology performs better in bringing higher capacity investments, profits and social welfare.

The paper is organized as follows. Section 2 describes the model. Section 3 derives analytical results for the case of grid networks and isolated-corridors networks. Section 4 derives policy implications and presents numerical simulations on the impact of network configuration on traffic, capacity investments, consumer surplus and social welfare. Section 5 concludes.

2. The model

Let consider a set of seven cities $\mathcal{C} = \{A, B, C, D, H, E, F\}$. An HSR operator provides transport services on one existing HSR corridor CD which consists of two links, CH and DH (Figure 1). The existing corridor serves three origin-destination (OD) markets, $\mathcal{OD}^e = \{CD, CH, DH\}$. Let k be the capacity of the existing links, i.e., the maximum number of HSR rides that can be provided on links CH and DH. We assume k is sufficiently large, such that capacity constraints on existing links are not binding.

*** Figure 1 ***

We evaluate two alternative scenarios:

1) building a new corridor AB that connects cities A, H and B and forms a grid with the existing corridor (Section 2.1), and
2) building a new corridor EF that connects the cities E, B and F and forms two isolated corridors (Section 2.2).

In both scenarios, the same HSR operator operates all the existing and new links in the network and maximizes social welfare. We assume that the HSR operator is also the infrastructure developer. As a result, it chooses capacity on the new corridor, incurring related investment costs, and by determines the traffic volume in each OD markets.

2.1 Grid network

The grid network is made up of four links (Figure 2), $\mathcal{L}^g = \{AH, BH, CH, DH\}$. Each link has unitary length. The network serves 10 OD markets, $\mathcal{OD}^g = \{AB, CD, AD, AC, BD, BC, BH, CH, DH, AH\}$. Compared to the case of one single CD corridor (Figure 1), the grid network generates seven new markets.

*** Figure 2 ***

Since CD corridor already exists, the capacity investment costs are only relevant to AH and BH links but not the CH and DH links. Then, equilibrium traffic in each OD market depends on whether the new or existing links are involved. Therefore, let q_n and p_n (q_e and p_e) indicate traffic and price in markets involving one new (existing) link, i.e., markets AH and BH (markets CH and DH). Similarly, let q_{2n} and p_{2n} (q_{2e} and p_{2e}) indicate traffic and price in markets involving two new (existing) links, i.e., market AB (market CD). Note that q_{2n} is along two high-cost new links, while q_{2e} is along two low-cost existing links. Finally, let q_{2m} and p_{2m} indicate traffic and price in markets involving one existing link and one new link, i.e., markets AD, AC, BD and BC. The subscript *2m* stands for the *mixed* usage of one low-cost link and one high-cost link.

Following Singh and Vives (1984), we assume the representative consumer has a quadratic utility function which is determined by the number of trips taken in all OD markets. This approach has been used in transport literature (Jiang and Zhang, 2014; D'Alfonso et al., 2015, 2016). We assume all markets are independent and hence there is no substitutability or complementarity among OD markets. Based on this assumption and symmetry, we can write the utility function as:

$$U^g = \alpha\,(q_{2n} + q_{2e} + 4\,q_{2m} + 2\,q_n + 2\,q_e) - \frac{1}{2}({q_{2n}}^2 + {q_{2e}}^2 + 4\,{q_{2m}}^2 + 2\,{q_n}^2 + 2\,{q_e}^2) \qquad (1)$$

where the superscript g stands for the case of the grid network. The representative traveler's marginal utility in market i $(i = n, 2n, e, 2e, 2m)$ is $\alpha - q_i$ and hence the marginal utility at zero trips is the same across markets and indicated by $\alpha \geq 0$.

Solving the representative passenger's problem (i.e., maximizing (1) subject to the budget constraint), we can obtain the inverse demand curve in each market *i*:

$$p_i = \alpha - q_i, \ \forall i = n, 2n, e, 2e, 2m \qquad (2)$$

Then, the parameter α can be interpreted as the strength of demand for travel as the demand increases at any price level when α increases. We now turn to the supply side. Let $k^g \geq 0$ be the capacity of the links AH and BH, i.e., the maximum number of HSR rides that can be provided on links AH and BH. Let R_i be the total number of rides offered by the HSR operator in OD market *i*. We have $q_i = R_i \times S \times LF$, where S represents the number of train seats per ride and LF represents the load factor. We assume that train size and load factor are fixed and hence $S \times LF$ is constant and can be normalized to 1. Then, number of rides is equivalent to number of trips, and price per seat, price per passenger and price per HSR ride are equivalent. We assume that the HSR technology exhibits economies of traffic density[4]. Thus, the operating cost on each link decreases with the traffic density Q_l^g of the link $l \in \mathcal{L}^g$ and for newly-built and existing links respectively:

$$Q_n^g = q_n + q_{2n} + 2\,q_{2m} \qquad (3)$$
$$Q_e^g = q_e + q_{2e} + 2\,q_{2m}$$

Following Brueckner and Spiller (1991), Zhang (1996), Brueckner (2001) and Bilotkach (2007) and Jiang and Zhang (2014) we define the HSR cost function as follows:

$$C^g = 2\,\left(c\,Q_e^g - d\,{Q_e^g}^2\right) + 2\,\left(c\,Q_n^g - d\,{Q_n^g}^2\right) + 2\,r\,k^g \qquad (4)$$

where $r \geq 0$ is the marginal (capital) cost of building one unit of capacity on link AH or BH. $d \geq 0$ is the density parameter and $c \geq 0$ is the cost of serving one unit of traffic (one ride, one passenger or one seat) on one link. The above setting leads to the following social welfare, profit and consumer surplus functions:

[4] Strong economies of traffic density have been identified in the railway industry (see Jiang and Zhang 2014).

$$SW^g = \alpha\,(q_{2n} + q_{2e} + 4\,q_{2m} + 2\,q_n + 2\,q_e) - \frac{1}{2}({q_{2n}}^2 + {q_{2e}}^2 + 4\,{q_{2m}}^2 + 2\,{q_n}^2 + 2\,{q_e}^2) - 2\,\left(c\,Q_n^g\, - d\,{Q_n^g}^2\right) - 2\,\left(c\,Q_e^g\, - d\,{Q_e^g}^2\right) - 2\,r\,k^g$$

$$\pi_{op}^g = p_{2n}\,q_{2n} + p_{2e}\,q_{2e} + 4\,p_{2m}\,q_{2m} + 2\,p_n\,q_n + 2\,p_e\,q_e - 2\,\left(c\,Q_e^g\, - d\,{Q_e^g}^2\right) - 2\,\left(c\,Q_n^g\, - d\,{Q_n^g}^2\right) \tag{5}$$

$$\pi^g = p_{2n}\,q_{2n} + p_{2e}\,q_{2e} + 4\,p_{2m}\,q_{2m} + 2\,p_n\,q_n + 2\,p_e\,q_e - 2\,\left(c\,Q_e^g\, - d\,{Q_e^g}^2\right) - 2\,\left(c\,Q_n^g\, - d\,{Q_n^g}^2\right) - 2\,r\,k^g$$

$$CS^g = SW^g - \pi^g$$

where social welfare (SW^g) is defined as the difference between passenger utility and the total costs, and π^g (π_{op}^g) is profit after (before) deducting the investment cost. We will refer to π_{op}^g as operating profit. CS^g is the consumer surplus.

The HSR operator decides the capacity, k^g, of the links AH and BH and the amount of traffic $\boldsymbol{q} = (q_{2n}, q_{2e}, q_{2m}, q_n, q_e)$, subject to the capacity constraint on each link. That is, it solves the following problem:

$$\begin{aligned} &\max_{k^g,\boldsymbol{q}}\; SW^g \\ &s.t.\quad Q_n^g \le k^g \\ &\qquad\;\; Q_e^g \le k \end{aligned} \tag{6}$$

For the sake of tractability, we focus on the case where $Q_e^g \le k$ is not binding. The optimal quantities supplied in each market and optimal capacity are reported below.

$$\begin{aligned} q_{2n}^{g*} &= \frac{(1-6d)(\alpha - 2c + 2\alpha d) - 2(1-10d)r}{(1-14d)(1-6d)} \\ q_{2e}^{g*} &= \frac{(1-6d)(\alpha - 2c + 2\alpha d) - 8dr}{(1-14d)(1-6d)} \\ q_{2m}^{g*} &= \frac{\alpha - 2c + 2\alpha d - r}{1-14d} \end{aligned} \tag{7}$$

$$q_n^{g*} = \frac{(1-6d)(\alpha - c - 6\alpha d) - (1-10d)r}{(1-14d)(1-6d)}$$

$$q_e^{g*} = \frac{(1-6d)(\alpha - c - 6\alpha d) - 4dr}{(1-14d)(1-6d)}$$

$$k^{g*} = Q_n^{g*} = \frac{(1-6d)(4\alpha - 7c) - (5-42d)r}{(1-14d)(1-6d)}$$

Note that the capacity constraint on the newly built corridor is always binding at the equilibrium, i.e., $k^{g*} = Q_n^{g*}$ always hold at the optimum. Let SW^{g*}, π^{g*}, π_{op}^{g*} and CS^{g*} denote the equilibrium social welfare, profit, operating profit and consumer surplus in the grid network obtained by plugging solution (7) into (5).

We remark that the equilibrium outcomes should satisfy a broader range of conditions, namely the non-negativity conditions for optimal traffic volumes and ticket prices, the second-order concavity conditions for the maximization problem, and the non-arbitrage conditions on prices, that is $p_{2e} \leq 2p_e$, $p_{2n} \leq 2p_n$ and $p_{2m} \leq p_e + p_n$. It is easy to prove that the non-arbitrage conditions are always satisfied. The second-order concavity conditions hold when $d < 1/14$, and therefore all the following analysis is based on $d \in [0, \frac{1}{14})$.

Lemma 1 clarifies some analytical properties of equilibrium quantities in each market. All proofs are available in the Appendix.

Lemma 1. When a welfare-maximizing HSR operator invests in an HSR corridor to form a grid network:

(i) Let $\theta = \frac{r}{1-6d}$. At equilibrium, it results:

   a. $q_{2e}^{g*} \geq q_{2m}^{g*} \geq q_{2n}^{g*}$, with $q_{2e}^{g*} - q_{2n}^{g*} = 2\theta$ and $q_{2m}^{g*} - q_{2n}^{g*} = \theta$

   b. $q_e^{g*} \geq q_n^{g*} > \frac{q_{2n}^{g*}}{2}$, with $q_n^{g*} = \frac{q_{2n}^{g*}}{2} + \frac{\alpha}{2}$ and $q_e^{g*} = q_n^{g*} + \theta = \frac{q_{2n}^{g*}}{2} + \frac{\alpha}{2} + \theta$

   c. $q_{2e}^{g*} - q_{2m}^{g*} = q_{2m}^{g*} - q_{2n}^{g*} = q_e^{g*} - q_n^{g*} = \theta$

(ii) Non-negative condition for quantities are satisfied if and only if $q_{2n}^{g*} > 0$ holds. This is equivalent to $r < \frac{(1-6d)(\alpha-2c+2\alpha d)}{2(1-10d)}$.

(iii) Quantity difference within two-link markets or within one-link markets (θ) increases in r and d.

(iv) When $r = 0$, $q_{2e}^{g*} = q_{2m}^{g*} = q_{2n}^{g*} = \frac{\alpha-2c+2\alpha d}{1-14d}$ and $q_e^{g*} = q_n^{g*} = \frac{\alpha-c-6\alpha d}{1-14d}$.

Lemma 1 highlights the asymmetry introduced by capacity costs. Since capacity costs are present on AH and BH links, but not on the CH and DH links, $q_e^{g*} \geq q_n^{g*}$. Moreover, q_{2n} involves two high-cost links, while q_{2e} involves two low-cost links and q_{2m} involves one low-cost link and one high-cost link, which is reflected by $q_{2e}^{g*} \geq q_{2m}^{g*} \geq q_{2n}^{g*}$. When the marginal cost of capacity is zero (equivalently, when we abstract away capacity investment on the new corridor), equilibrium traffic is the same within one-link markets and within two-link markets.

Based on Lemma 1, Lemma 2 shows the impact of density parameter on traffic in each market.

Lemma 2. When a welfare-maximizing HSR operator invests in an HSR corridor to form a grid network, equilibrium quantity supplied in each market increases as the density parameter d increases.

Intuitively, Lemma 2 holds because the operating cost of serving the same amount of traffic reduces as the density effect becomes stronger.

2.2 Isolated-corridors network

The isolated-corridors case (Figure 3) also involves four links, each with unitary length, $\mathcal{L}^c = \{CH, DH, EB, FB\}$. However, there are only six OD markets, $\mathcal{OD}^c = \{CD, CH, DH, EB, FB, EF\}$. That is, only three new markets are generated after adding the EF corridor.

*** Figure 3 ***

Similar to the grid network case, only newly added corridor EF incurs capacity investment costs. Then, the equilibrium traffic depends on the involvement of newly-added links and the number of links used to serve the market. Therefore, following the previous notation, q_n and p_n (q_e and p_e) denote traffic and price in new (existing) one-link markets, i.e., markets EB and FB (markets CH and DH). Let q_{2n} and p_{2n} (q_{2e} and p_{2e}) indicate traffic and price in new (existing) two-link markets, i.e., market EF (market CH).[5] The marginal utility derived by the representative traveler at zero traffic is $\beta \geq 0$ in both one-link and two-link markets along corridor EF. This parameter β can be considered as a measure of the strength of demand for travel on one-link and two-link markets along corridor EF as it shifts the demand curve. Then, the utility function is specified as:

$$U^c = \alpha\ (2\ q_e + q_{2e}) + \beta\ (2\ q_n + q_{2n}) - \frac{1}{2}({q_{2n}}^2 + {q_{2e}}^2 + 2\ {q_n}^2 + 2\ {q_e}^2) \quad (8)$$

[5] To save on notation, we shall use the same symbols q_n and p_n as well as q_e and p_e (q_{2n} and p_{2n} as well as q_{2e} and p_{2e}) as used in section 2.1. We will use different symbols for equilibrium values in the two scenarios.

The total cost function is:

$$C^c = 2\left(c\, Q_e^c\ - d\, Q_e^{c\,2}\right) + 2\left(c\, Q_n^c\ - d\, Q_n^{c\,2}\right) + 2\, r\, k^c \tag{9}$$

where

$Q_n^c = q_n + q_{2n}$

$Q_e^c = q_e + q_{2e}$

Note that Q_n^c and Q_e^c measure, respectively, traffic density on new links (EB or FB) and existing links (CH or DH). $k^c \geq 0$ is the capacity of the links EB or FB. The superscript c stands for the case of the isolated-corridors network. We may then write down social welfare, profits and consumer surplus:

$$SW^c = \alpha\,(2\, q_e + q_{2e}) + \beta\,(2\, q_n + q_{2n}) - \frac{1}{2}({q_{2n}}^2 + {q_{2e}}^2 + 2\,{q_n}^2 + 2\,{q_e}^2)$$
$$- 2\left(c\, Q_e^c\ - d\, Q_e^{c2}\right) - 2\left(c\, Q_n^c\ - d\, Q_n^{c2}\right) - 2\, r\, k^c$$

$$\pi_{op}^c = p_{2n}\, q_{2n} + p_{2e}\, q_{2e} + 2\, p_n\, q_n + 2\, p_e\, q_e - 2\left(c\, Q_e^c\ - d\, Q_e^{c2}\right) - 2\left(c\, Q_n^c\ - d\, Q_n^{c\,2}\right) \tag{10}$$

$$\pi^c = p_{2n}\, q_{2n} + p_{2e}\, q_{2e} + 2\, p_n\, q_n + 2\, p_e\, q_e - 2\left(c\, Q_e^c\ - d\, Q_e^{c2}\right) - 2\left(c\, Q_n^c\ - d\, Q_n^{c\,2}\right)$$
$$- 2\, r\, k^g$$

$$CS^c = SW^c - \pi^c$$

The welfare-maximizing HSR operator decides the capacity, k^c, and the traffic volume in each market $\boldsymbol{q} \coloneqq (q_{2n}, q_{2e}, q_n, q_e)$ subject to the capacity constraint on each link. Thus, it solves the following maximization problem:

$$\max_{k^c,\boldsymbol{q}}\ SW^c$$
$$s.t.\quad Q_n^c \leq k^c \tag{11}$$
$$Q_e^c \leq k$$

Again, we focus on the case where $Q_e^c \leq k$ is not binding. The equilibrium quantity supplied in each market is reported below:

$$q_{2n}^{c*} = \frac{\beta - 2c + 2\beta d - 2r}{1 - 6d} \tag{12}$$

$$q_{2e}^{c*} = \frac{\alpha - 2c + 2\alpha d}{1 - 6d}$$
$$q_n^{c*} = \frac{\beta - c - 2\beta d - r}{1 - 6d}$$
$$q_e^{c*} = \frac{\alpha - c - 2\alpha d}{1 - 6d}$$
$$k^{c*} = Q_n^{c*} = \frac{2\beta - 3(c + r)}{1 - 6d}$$

Again, the non-negativity conditions, second-order concavity conditions and non-arbitrage conditions need to be satisfied.[6] The capacity constraint on the newly-built corridor is always binding at the equilibrium, i.e., $k^{c*} = Q_n^{c*}$ always hold at the optimum. Denote SW^{c*}, π^{c*}, π_{op}^{c*} and CS^{c*} as the equilibrium social welfare, profit, operating profit and consumer surplus in the isolated-corridors network obtained by plugging (12) into (10).

Lemma 3 shows the impact of density parameter, d, on equilibrium traffic in each market.

Lemma 3. When a welfare-maximizing HSR operator invests in an HSR corridor to form an isolated-corridors network, equilibrium quantity supplied in each market increases as the density parameter d increases.

3. Impact of network configuration on traffic, capacity and social welfare

In this section, we compare the performance of grid and isolated-corridors networks in terms on traffic and capacity. We first consider the case where capacity investment costs are suppressed, i.e., $r = 0$. Lemma 4 compares equilibrium traffic in the two alternative networks.

Lemma 4. When capacity costs are suppressed, i.e., $r = 0$, at the equilibrium,

(i) $q_e^{c*} \leq q_e^{g*}$, $q_{2e}^{c*} \leq q_{2e}^{g*} = q_{2m}^{g*}$. The equal signs of the inequalities hold when d = 0.

(ii) $k^{g*} < k^{c*}$ if and only if $\beta > \tilde{\beta} = \frac{2(\alpha - c - 6\alpha d)}{1 - 14d}$; otherwise, $k^{g*} \geq k^{c*}$ and the equal sign holds when $\beta = \tilde{\beta}$.

Lemma 4 suggests that in each market involving at least one existing link, the grid network tends to generate more traffic than isolated-corridors network. However, despite that the grid network generates more new markets, the newly developed capacity may be higher in the isolated-corridors

[6] The non-arbitrage conditions are automatically hold. The second-order concavity conditions hold when d<1/6.

network as long as the market size of these new markets is far much larger in isolated corridor network than the grid network. Proposition 1 compares social welfares in the two networks when capacity investment costs are suppressed.

Proposition 1. When capacity investment costs are suppressed, i.e., $r = 0$, at the equilibrium,

(i) when $\alpha \geq \beta$, $SW^{g*} > SW^{c*}$;

(ii) when $\alpha < \beta$, $SW^{g*} > SW^{c*}$ if and only if $\alpha < \beta < \hat{\beta} = \frac{4c}{3-2d} + \sqrt{\frac{21\alpha^2-72\alpha c+64c^2-98\alpha^2 d+288\alpha cd-256c^2 d+188\alpha d^2-160\alpha cd^2-88\alpha^2 d^3}{(3-2d)^2(1-14d)}}$;

(iii) when $\beta = \tilde{\beta}$, $SW^{g*} < SW^{c*}$.

When capacity investment costs are suppressed, and $\alpha \geq \beta$, the inferiority of the isolated-corridors network relative to the grid network can be seen immediately, since the grid model generates more new markets. However, if the markets along the newly-built corridor EF is sufficiently larger than those along the corridor AB, the result may be reversed. Moreover, when the two networks produce the same capacity level on the newly-built corridors, the grid network generates less social welfare than the isolated-corridors network. In the next section, Proposition 6 identifies, for any feasible (β, r) combination, the behaviour of social welfare indifference curve.

We now consider the case where capacity investment costs play a role, i.e., $r > 0$. Proposition 2 compares traffic in the two networks.

Proposition 2. At the equilibrium, when $r > 0$, we have:

(i) $q_{2e}^{g*} \geq q_{2e}^{c*}$ and $q_e^{g*} \geq q_e^{c*}$, and the equal signs hold when $d = 0$.

(ii) When $\alpha = \beta$, $q_{2n}^{g*} \geq q_{2n}^{c*}$ and $q_n^{g*} \geq q_n^{c*}$, and the equal signs hold when $d = 0$.

Proposition 2 suggests that in markets involving only existing links, a grid network tends to generate higher traffic than isolated-corridors network. Such traffic difference vanishes when economies of traffic density no longer plays a role. When the newly developed markets have the same size as the existing markets, the above observation applies to newly developed markets as well.

We now compare equilibrium capacities of the two networks. Proposition 3 clarifies the impact of c, r and d on capacity.

Proposition 3. Let $\tilde{\beta}(r)$ be the cut-off value of β which makes $k^{g*} = k^{c*}$. Then, $\tilde{\beta}$ decreases in c and r, while it increases in d. That is, at equilibrium, the isolated-corridors network may generate more capacity than the grid network with large β when: (i) the cost of capital r or operating cost c is at the high end, or (ii) density effect d is weak.

The idea of Proposition 3 can be shown graphically (Figure 4) by writing the equilibrium capacity as a function of capital cost: $k^{g*}(r) = \frac{\alpha(4-24d)+42cd-7c}{(1-14d)(1-6d)} - \frac{(5-42d)r}{(1-14d)(1-6d)}$ and $k^{c*}(r) = \frac{2\beta-3c-3r}{1-6d}$.

*** Figure 4 ***

By comparing the slopes of $k^{g*}(r)$ and $k^{c*}(r)$, one can prove that $k^{g*}(r)$ is steeper than $k^{c*}(r)$, i.e., $\frac{(5-42d)}{(1-14d)(1-6d)} - \frac{3}{1-6d} = \frac{2}{(1-14d)(1-6d)} \geq 0$. In other words, capacity investment in grid network is more sensitive to capital cost than in the case of isolated-corridors network. When β increases, $k^{c*}(r)$ shifts upward while keeping $k^{g*}(r)$ unchanged. Since $k^{c*}(r)$ has a negative but smaller slope in magnitude, the intersection with $k^{g*}(r)$ will occur at larger *r* values first. That is, with a sufficiently large β, isolated-corridors network may generate more capacity than grid network when the cost of capital r is at the high end. Applying the similar logic, one can show that with a sufficiently large β, isolated-corridors network may generate more capacity than grid network when the operating cost c is at the high end, or the density parameter d is at the low end.

We now compare the performance of grid and isolated-corridors networks in terms of social welfare. Since social welfare is maximized at equilibrium and the capacity constraints of newly-added corridors are always binding, following the envelope theorem, we have $\frac{dSW^*}{dx} = \sum_i \frac{\partial SW^*}{\partial q_i} \frac{dq_i{}^*}{dx} + \frac{\partial SW^*}{\partial x} = \frac{\partial SW^*}{\partial x}$, for $x = c, d$. Propositions 4 and 5 discuss the impact of *r*, *c* and *d* on social welfares.

Proposition 4. At the equilibrium, $\frac{dSW^{g*}}{dr} < 0$ and $\frac{dSW^{c*}}{dr} < 0$. Moreover, as capital cost r increases, the social welfare tends to decrease faster in the grid network than the isolated-corridors network if and only if the grid network generates larger capacity, i.e. $\frac{dSW^{g*}}{dr} - \frac{dSW^{c*}}{dr} < 0$ if and only if $k^{g*} > k^{c*}$.

Proposition 5. At the equilibrium, in both networks, $\frac{dSW^{g*}}{dc} < 0$ and $\frac{dSW^{c*}}{dc} < 0$, but $\frac{dSW^{g*}}{dd} > 0$ and $\frac{dSW^{c*}}{dd} > 0$. Moreover, as operating cost *c* (density parameter *d*) increases, the social welfare tends to decrease (increase) faster in the grid network than the isolated-corridors network unless the capacity of the isolated corridor network is substantially larger than that of the grid network.

Proposition 5 states that, as operating cost, c, increases, the grid network tends to become less preferred than the isolated-corridors network in terms of social welfare due to the fact that the grid network serves more OD markets. However, if the isolated-corridors network generates substantially larger capacity and hence substantially higher traffic than grid network, an increase in operating cost will disadvantage isolated-corridors network. The impacts of density parameter, d, on social welfares are opposite to the impacts of an increase in operating cost.

4. Discussion and policy implications

In this section, to better understand the difference of these two networks and provide further policy implications, we analyze the indifference curves of social welfare, capacity, profit and consumer surplus. In particular, let SWdiff be the indifference curve composed of all the (*β, r*) combinations which make $SW^{g*} = SW^{c*}$ hold. Similarly, let Kdiff, πdiff1, πdiff2 and CSdiff be the indifference curves which shows all feasible (*β, r*) combinations when both networks result in the same equilibrium capacity, operating profit, profit and consumer surplus, i.e. $k^{g*} = k^{c*}$, $\pi_{op}^{g*} = \pi_{op}^{c*}$, $\pi^{g*} = \pi^{c*}$, and $CS^{g*} = CS^{c*}$, respectively.

Figure 5 illustrates the shape of a typical SWdiff curve. Indeed, an increase in *β* (moving from *point 1* to *point 2*) improves utility and traffic density of the isolated-corridors network, while having no impact on the grid network. Moreover, as the cost of capital, *r*, increases (from *point 1* to *point 3*), both networks generate lower utility and less operating cost, but grid network's utility drops faster while cost reduces slower. As a result, SWdiff is downward sloping, leading to Observation 1.

*** *Figure 5* ***

Observation 1. The grid network generates higher social welfare than the isolated-corridors network when the cost of capital, *r*, is small and the size of new markets in isolated-corridors network, *β*, are small, or one of them is small.

Observation 1 can be easily seen from Propositions 3 and 4. That is, two factors may cause the isolated-corridors network to outperform the grid network in social welfare. First, when grid network has larger capacity, the social welfare of the grid network decreases faster than isolated-corridors network as the capital cost *r* increases (Proposition 4). Second, as capital cost increases, it makes the isolated-corridors network more likely to generate higher capacity at a given level of *β*, as the threshold $\tilde{\beta}(r)$ decreases in r (Proposition 3). These two factors suggest that isolated-corridors are more likely to generate higher social welfare when capital cost is high.

Note that the indifference curve of capacity, Kdiff, is also downward sloping, which can be easily proved. Based on Lemma 4(ii) and Proposition 1(iii), we can prove the following key relationship between SWdiff and Kdiff as stated in Proposition 6 and illustrated in Figure 6.

Proposition 6. For any feasible (*β, r*) combination, Kdiff is always to the right of SWdiff.

As illustrated by the shaded area in Figure 6(a), Proposition 6 implies that at any given level of capital cost, r, there is a range of median level β in which the grid network generates more capacity investment but lower social welfare. This is a special parameter range where the investment cost of grid network outweighs the benefit of larger capacity and hence higher traffic. However, as long as the isolated-corridors network generates more capacity, it also produces higher social welfare.

**** Figure 6 ****

Note that Proposition 6 holds when one-link markets and two-link markets have the same sizes, i.e. market size indicator *α* applies to all markets in the grid network and *β* applies to all the markets along the EF corridor in the isolated-corridors network. However, as demand for travel can reduce as travel distance increases, one may assume that one-link markets tend to have a larger size than the two-link market. Then, Kdiff and SWdiff may have an intersection in certain constellations when operating cost *c* is very high. Figure 6(b) illustrates such cases. Note that there is a small parameter range (dark shaded area) where the isolated-corridors network generates more capacity investment but lower social welfare.

We observe the following relationship between SWdiff and the other indifference curves (Figure 7). In general, CSdiff and πdiff2 seem to be located to the right of SWdiff. Note that the grid network generates higher profit than isolated-corridors in the area on the right-hand of πdiff2. Therefore, the area in-between SWdiff and πdiff2 represents the cases where isolated-corridors are preferred in terms of both social welfare and profit. However, this never happens to the grid network. That is, when grid network is preferred in terms of social welfare (on the left-hand side of SWdiff), the isolated-corridors network is preferred in terms of profit; when grid network is preferred in terms of profit (on the right-hand side of πdiff2), the isolated-corridors network is preferred in terms of social welfare. The comparison between SWdiff and CSdiff suggests that network choice based on consumer surplus tends to be consistent with the choice based on social welfare. There only exists a small area between SWdiff and CSdiff in which grid network generates lower social welfare but higher consumer surplus. The indifference curve of operating profit before deducting of investment costs (πdiff1) has two branches.[7] The grid network generates higher operating profit in the area above the left branch or below the right branch. Observation 2 summarizes the potential inconsistency in preferred network structure when aspects other than social welfare are taken into account.

*** *Figure 7* ***

Observation 2. There exist areas of (*β, r*) combinations in which the following cases occur:

(a) $SW^{g*} < SW^{c*}$ but $K^{g*} > K^{c*}$

(b) $SW^{g*} < SW^{c*}$ but $CS^{g*} > CS^{c*}$

(c) $SW^{g*} > SW^{c*}$ but $\pi^{g*} < \pi^{c*}$, and $SW^{g*} < SW^{c*}$ but $\pi^{g*} > \pi^{c*}$

(d) $SW^{g*} > SW^{c*}$ but $\pi_{op}^{g*} < \pi_{op}^{c*}$, and $SW^{g*} < SW^{c*}$ but $\pi_{op}^{g*} > \pi_{op}^{c*}$

Observation 2 implies that there exist a wide parameter range where the network choice based on social welfare is different from the network choice based on consumer surplus, profit or operating profit.

We then discuss the impact of the operating cost *c* by presenting simulation outcomes with $\alpha = 1000$, $d = 0.01$ and varying levels of *c*.[8] Figure 8 suggests that both SWdiff and Kdiff shift to the left as *c* increases. The former is proved by Proposition 3 and the latter can be immediately seen from Propositions 4 and 6. As both indifference curves fall into cases where $k^{c*} - k^{g*} \leq 0$ (Proposition 6), this leads to $\frac{dSW^{g*}}{dc} - \frac{dSW^{c*}}{dc} < 0$ (Proposition 4). In a word, the parameter range of $SW^{g*} > SW^{c*}$ shrinks as *c* increases. One may observe that Kdiff shifts faster than SWdiff. The same rationale applies to the indifference curves of consumer surplus (CSdiff) and profit (πdiff2) as well and these two curves move closer to SWdiff as *c* increases.

---

[7] The right branch may not appear in all the figures presented in this paper, since only a limited range of β is shown in the figures. The right branch will appear as long as the the *β*-axis is extended.

[8] To save space, we only present a typical case, the observations presented hold qualitatively with a wide range of *α* and *d* parameters.

*** *Figure 8* ***

Based on the above discussion, Observations 3 and 4 state a few main observations from the numerical simulation and the ideas are summarized in Figure 8.

Observation 3. When the operating cost, *c*, increases, the following tend to occur:

(i) The grid network becomes *less* likely to generate higher social welfare, consumer surplus and capacity than isolated-corridors network.

(ii) The grid network becomes *more* likely to generate higher profit (after deducting the investment cost) than the isolated-corridors network.

Observation 4. The (*β, r*) parameter ranges of the following inconsistent cases *shrink* as *c* increases: (a) $SW^{g*} < SW^{c*}$ but $K^{g*} > K^{c*}$, (b) $SW^{g*} < SW^{c*}$ but $CS^{g*} > CS^{c*}$, and (c) $SW^{g*} > SW^{c*}$ but $\pi^{g*} < \pi^{c*}$. The (*β, r*) parameter range of $SW^{g*} < SW^{c*}$ but $\pi^{g*} > \pi^{c*}$ *expands* as *c* increases.

Observations 3 and 4 suggest that although an increase in operating cost helps the isolated corridor network outperform in social welfare, it may be disadvantaged in terms of profit. Although the inconsistency between social welfare and consumer surplus reduces as *c* increases, the chance of having grid network generate higher social welfare but lower profit increases at a high level of *β*. This implies that when the operating cost is high, the policy maker may end up choosing the isolated-corridors case due to the consideration of profit and investment cost recovery even though the grid network generates higher social welfare and is more beneficial to consumers. This finding seems consistent with the real-life difference in China and the USA in network choice. China (grid network with lower operating costs) seems to care about social welfare and consumer surplus more than cost recovery while the USA (isolated-corridors network with higher operating costs) has a strong concern on cost recovery.

We now turn to the impact of the density parameter *d* based on simulation analysis which assumes *α* = 1000, *c = 400* and varying *d* (Figure 9). Figure 9 suggests that although SWdiff, Kdiff, CSdiff and πdiff2 all shift to the right as *d* increases, SWdiff shifts slower than the other indifference curves, enlarging the gap between SWdiff and the other curves. These shifts of indifference curves lead to Observations 5-6.

*** *Figure 9* ***

Observation 5. When the density parameter, *d*, increases, the following tend to occur:

(i) The grid network becomes *more* likely to generate higher social welfare, consumer surplus and capacity than isolated-corridors network.

(ii) The grid network becomes *less* likely to generate higher profit (after deducting investment cost) than isolated-corridors network.

Observation 6. The (*β, r*) parameter ranges of the following inconsistent cases *expand* as *d* increases: (a) $SW^{g*} < SW^{c*}$ but $K^{g*} > K^{c*}$, (b) $SW^{g*} < SW^{c*}$ but $CS^{g*} > CS^{c*}$, and (c) $SW^{g*} > SW^{c*}$ but $\pi^{g*} < \pi^{c*}$. The (*β, r*) parameter range of $SW^{g*} < SW^{c*}$ but $\pi^{g*} > \pi^{c*}$ *shrinks* as *d* increases.

In general, the impact of increasing density effect is opposite to the impact of increasing operating cost.

5. Concluding remarks

In this paper, we have developed an analytical framework to study investment levels and performance of two alternative topologies of HSR networks: grid and isolated-corridors networks. We evaluate how network configuration affects capacity investment levels, profits and social welfare and how the traffic density effect, operating cost, market size and cost of capital play a role.

Our main findings are as follows. The difference in capacity investment is mainly driven by the relative market size. If the market size of OD markets along the isolated (EF) corridor is sufficiently large, the isolated-corridors network may generate more capacity than the grid network. Indeed, an increase in the size of markets along the isolated corridor improves utility and traffic density of the isolated-corridors network while having no impact on the grid network. Moreover, high capital cost, high operating cost and a weak density effect may also contribute to higher capacity in the isolated-corridors network than the grid network. Similarly, the isolated-corridors network becomes more likely to outperform the grid network in terms of social welfare if capital cost increases, operating cost increases, or the density effect decreases. These results hinge on tradeoffs between service coverage and investment, and between the density effect and operating costs faced by individual network structures. Adding new HSR links to form a grid network provides greater service coverage (hence higher traffic density) but incurs higher capital and operating costs. On the other hand, a single isolated corridor requires lower capital and operating costs given similar market size but offers less service coverage.

We further investigate whether the network generating higher social welfare requires higher capacity investment and generates higher consumer surplus and profit. In most of the cases, more capacity investment is consistent with higher social welfare. However, the exceptional case may happen when the market size along the isolated corridor is at medium level, and then it is possible to have grid network generate higher capacity but lower social welfare, leading to a case of overinvestment. We also observe parameter ranges where network choice based on social welfare would conflict with the choice based on consumer surplus or profit. Contrary to the impact of increasing operating cost, an increase in the density effect may in general increase the chance of these conflicts, calling for more extra government policies to balance the interests of different stakeholders.

Our investigation is based on a social welfare-maximizing entity who operates HSR services and invests in HSR infrastructure. Indeed, this circumstance is observed in some major developments of the HSR network (e.g., China, France, Spain). In general, further developments of this work may investigate whether the above results hold qualitatively with a profit-maximizing decision maker. Finally, in this paper, the passenger transport market is monopolistic. However, when the rail transport market is fully liberalized, competition occurs on some city-pairs, among different operators (e.g., in Italy, Trenitalia, the incumbent, and Nuovo Trasporto Viaggiatori, the new entrant, provide HSR services). Then, further developments of this work may investigate whether the above results hold qualitatively with different market structures in the passenger transport market.

Acknowledgments

We wish to thank the Editor, two anonymous referees, the participants at the 2018 and 2019 meetings of the International Transport Economic Association (ITEA) and at the 2018 meeting of the Air Transport Research Society (ATRS), where we presented earlier versions of this paper. This research has been partly undertaken at the CTS – Center for Transportation Studies at University of British Columbia, which is thankfully acknowledged, and it has been developed as a result of a mobility stay funded by the Erasmus Mundus Programme of the European Commission under the Transatlantic Partnership for Excellence in Engineering – TEE Project. Financial support from Sapienza University of Rome (Grants No. C26N1339HR and C26A14YJL5), National Natural Science Foundation of China (Grant No. 71801099) and the General Research Fund from Research Grants Council of Hong Kong (Grant No. RGC/PolyU 152195/17E) are also gratefully acknowledged.

References

Avenali, A., Bracaglia, V., D'Alfonso, T., Reverberi, P., 2018. Strategic formation and welfare effects of airline-high speed rail agreements. Transportation Research Part B: Methodological, 117, 393-411.

Campos, J., De Rus, G., 2009. Some stylized facts about high-speed rail: A review of HSR experiences around the world. Transport policy, 16(1), 19-28.

China.com (2016). China high speed rail: lucrative in east, loss-making in center and west. Available at: http://english.china.com/news/business/56/20160802/722066.html. Last retrieved: 17/10/2018

D'Alfonso, T., Jiang, C., Bracaglia, V., 2015. Would competition between air transport and high-speed rail benefit environment and social welfare? Transportation research part B: methodological, 74, 118-137.

D'Alfonso, T., Jiang, C., Bracaglia, V., 2016. Air transport and high-speed rail competition: Environmental implications and mitigation strategies. Transportation Research Part A: Policy and Practice, 92, 261-276.

Givoni, M., 2006. Development and Impact of the Modern High-speed Train: A Review. Transport reviews, 26(5), 593-611.

Jiang, C., Zhang, A., 2014. Effects of high-speed rail and airline cooperation under hub airport capacity constraint. Transportation Research Part B: Methodological, 60, 33-49.

Jiang, C., Zhang, A., 2016. Airline network choice and market coverage under high-speed rail competition. Transportation Research Part A: Policy and Practice, 92, 248-260.

Jiang C., T. D'Alfonso, Y. Wan, 2017. Air-rail cooperation: Partnership level, market structure and welfare implications. Transportation Research Part B: Methodological. DOI: 10.1016/j.trb.2017.01.006.

Li, Z. C., Lam, W. H., Wong, S. C., 2011. On the allocation of new lines in a competitive transit network with uncertain demand and scale economies. Journal of Advanced Transportation, 45(4), 233-251.

Li, Z. C., Lam, W. H., Wong, S. C., Sumalee, A., 2012. Design of a rail transit line for profit maximization in a linear transportation corridor. Transportation Research Part E: Logistics and Transportation Review, 48(1), 50-70.

Liu, G., Quain, G.J., Wirasinghe, S.C., 1996. Rail line length in a cross-town corridor with many to many demand. Journal of Advanced Transportation 30 (1), 95–114.

Perl, A. D., Goetz, A. R., 2015. Corridors, hybrids and networks: three global development strategies for high speed rail. Journal of Transport Geography, 42, 134-144.

Singh, N., Vives, X., 1984. Price and quantity competition in a differentiated duopoly. RAND Journal of Economics, 15 (4), 546–554.

Socorro, M., Viecens, M., 2013. The effects of airline and high-speed train integration. Transportation Research Part A: Policy and Practice, 49, 160–177.

The World Bank, 2017. Railway reform: Toolkit for improving rail sector performance. Available at: http://documents.worldbank.org/curated/en/529921469672181559/Railway-reform-Toolkit-for-improving-rail-sector-performance. Last retrieved: 17/10/2018.

UIC - International union of railways, 2018a. Data of high-speed passenger-km. Updated 30 November 2017. Available at: https://uic.org/IMG/pdf/high_speed_passenger-km_20171130_.pdf. Last retrieved: 17/10/2018.

UIC - International union of railways, 2018b. High speed lines in the world. Updated 1 October 2018. Available at: https://uic.org/IMG/pdf/20181001-high-speed-lines-in-the-world.pdf. Last retrieved: 17/10/2018.

Wan, Y., Ha, H. K., Yoshida, Y., Zhang, A., 2016. Airlines' reaction to high-speed rail entries: Empirical study of the Northeast Asian market. Transportation Research Part A: Policy and Practice, 94, 532-557.

Wirasinghe, S.C., Seneviratne, P.N., 1986. Rail line length in an urban transportation corridor. Transportation Science 20 (4), 237–245.

Xia, W., Zhang, A., 2016. High-speed rail and air transport competition and cooperation: A vertical differentiation approach. Transportation Research Part B: Methodological, 94, 456-481.

Xia, W., Zhang, A., 2017. Effects of air and high-speed rail transport integration on profits and welfare: the case of air-rail connecting time. Journal of Air Transport Management, 65, 181-190.

Xia, W., Jiang, C. Wang, K., Zhang, A., 2018. Xia, W., Jiang, C., Wang, K., & Zhang, A. (2019). Air-rail revenue sharing in a multi-airport system: Effects on traffic and social welfare. Transportation Research Part B: Methodological, 121, 304-319.

Yang, H., Zhang, A., 2012. Effects of high-speed rail and air transport competition on prices, profits and welfare. Transportation Research Part B: Methodological, 46(10), 1322-1333.

Zhang, A., Wan, Y., & Yang, H. (2019). Impacts of high-speed rail on airlines, airports and regional economies: A survey of recent research. Transport Policy, 81, A1-A19.

Figures

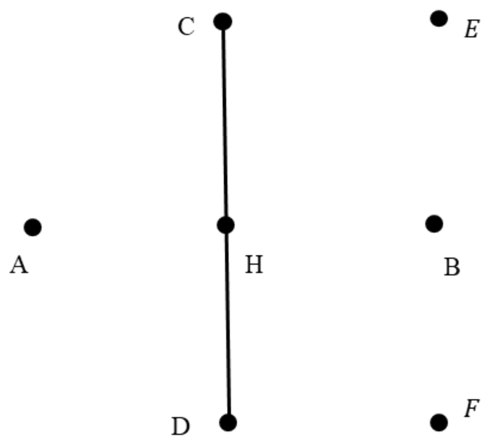

Figure 1 Network structure: existing corridor

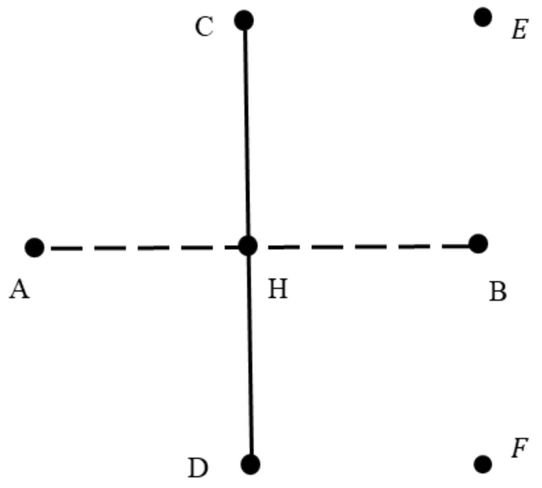

Figure 2 Grid network. The dashed line represents the new corridor.

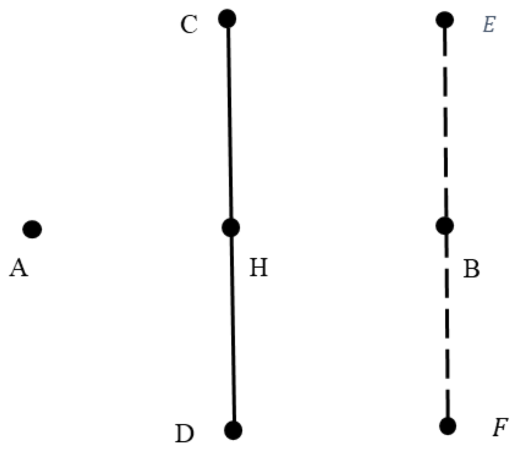

Figure 3 Isolated-corridors. The dotted line represents the new corridor.

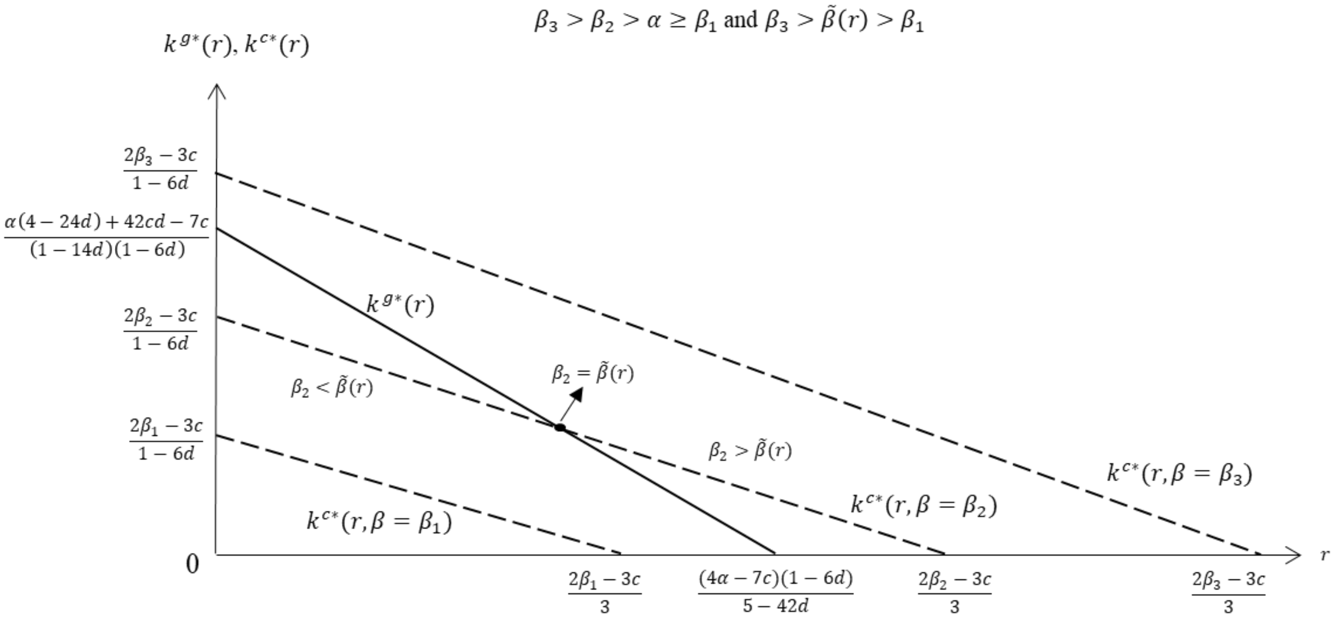

Figure 4 Equilibrium capacity in grid and isolated-corridors networks

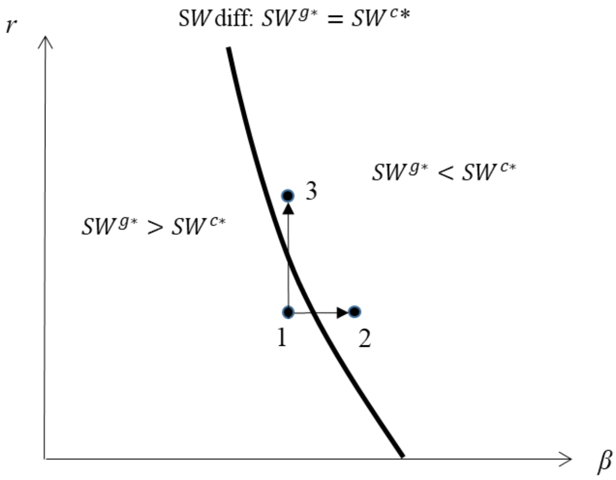

Figure 5 Illustration of the social welfare indifference curve

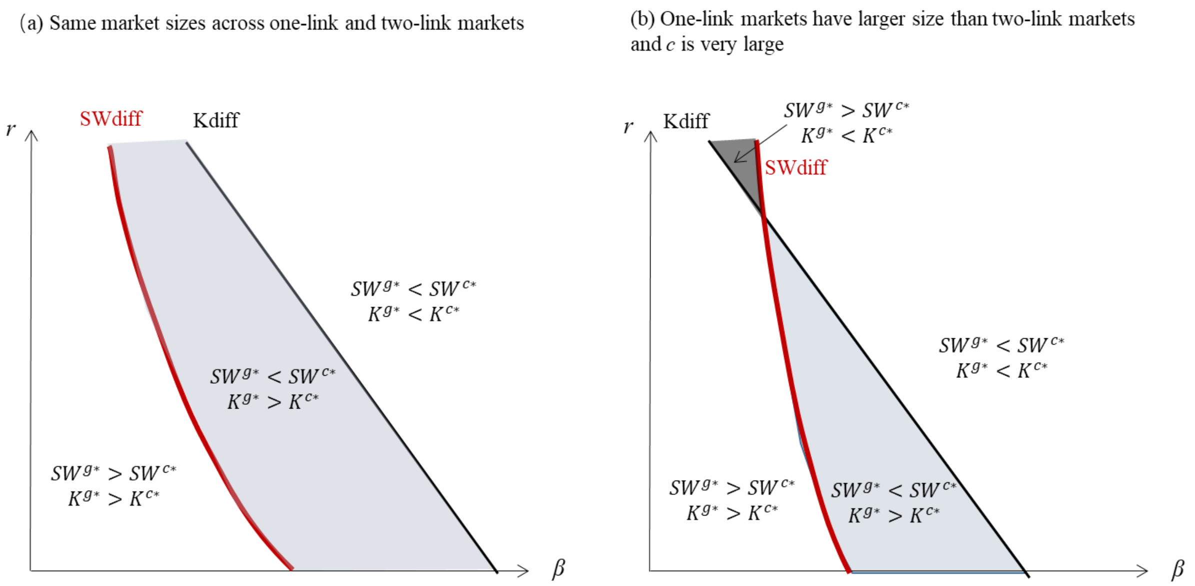

Figure 6 Parameter range where higher investment leads to lower social welfare

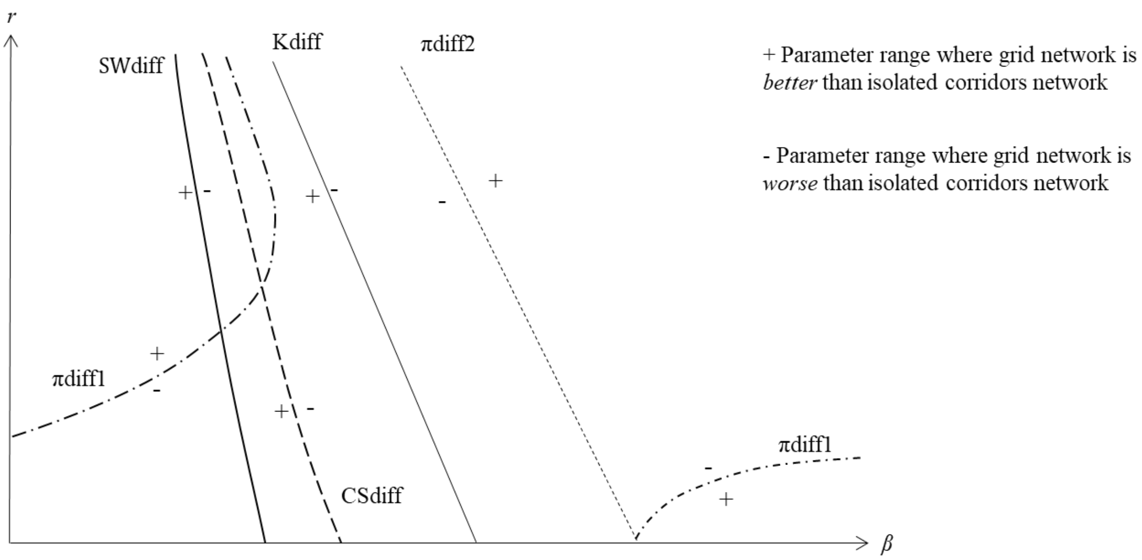

Figure 7 Illustration of social welfare, capacity, consumer surplus and profit indifference curves

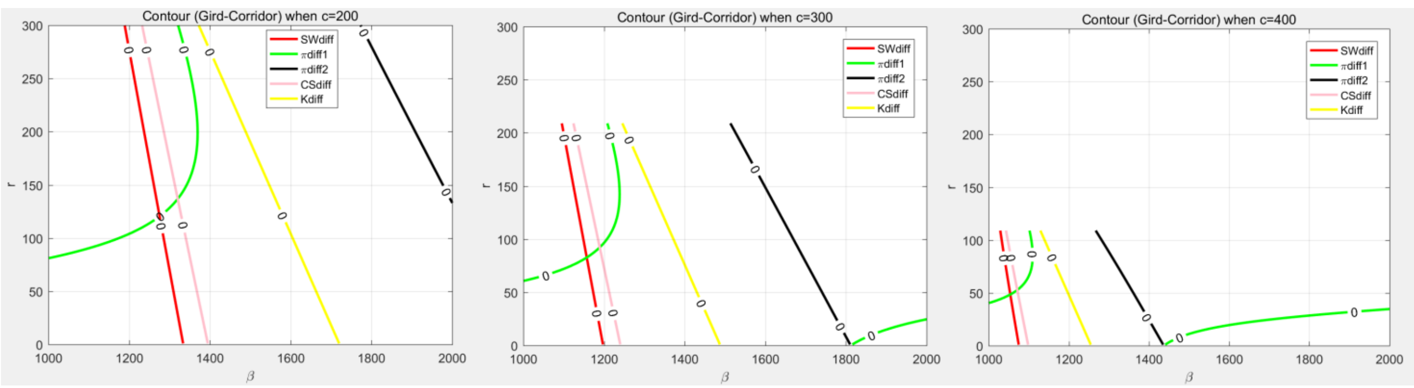

Figure 8 The impact of c on social welfare, capacity, consumer surplus and profit indifference curves (α = 1000, d = 0.01, c = 200, 300, 400). The feasible range of r and hence the indifference curves shrink as c increases.

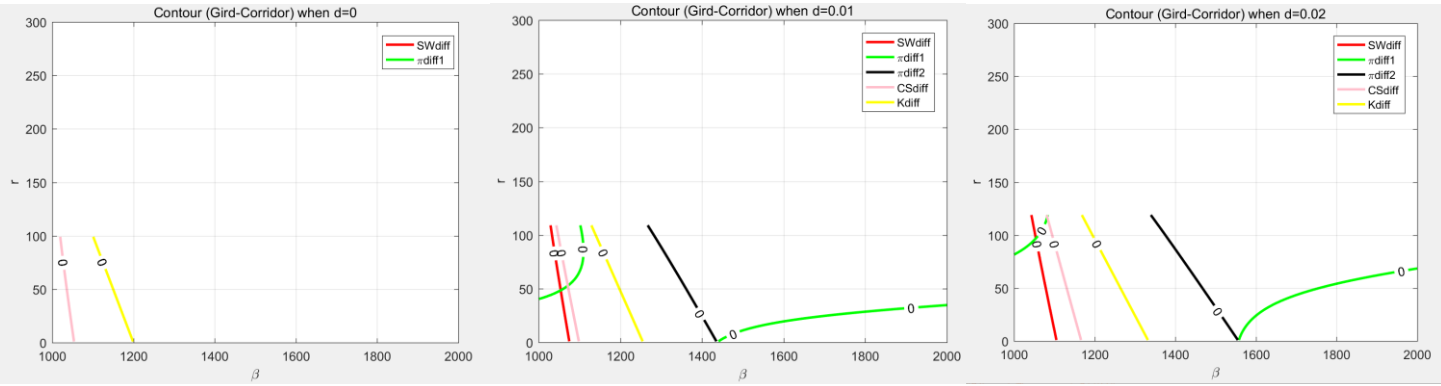

Figure 9 The impact of *d* on social welfare, capacity, consumer surplus and profit indifference curves (α = 1000, c = 400, d = 0, 0.01, 0.02). Note that when d = 0, SWdiff overlaps CSdiff, πdiff2 overlaps Kdiff, and πdiff1, does not exist as the operating profits are both zero for the two networks.

Table

Table 1 Data of high-speed rail passenger-km (billion passenger-km). Source: UIC (2018a)

Year	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**	**2016**
China (Chinese Railways)	46.3	105.8	144.6	214.1	282.5	386.3	464.1
Japan (JR group)	76.9	79.56	84.2	87.4	89.2	97.4	98.6
Korea (Korail)	11.0	13.6	14.1	14.5	14.4	15.1	16.3
Taiwan High Speed Rail Corp.	7.5	8.1	8.6	8.6	8.6	9.7	10.5
Italy (Trenitalia)	11.6	12.3	12.3	12.8	12.8	12.8	12.8
Spain (Renfe Operadora)	11.7	11.2	11.2	12.7	12.8	14.1	15.1
Germany (DB AG)	23.9	23.3	24.8	25.2	24.3	25.3	27.2
France (SNCF)	51.9	52.0	51.1	50.8	50.7	50.0	49.1
Other European Companies	7.3	10.5	14.8	15.2	18.2	20.0	22.0
Total	*248.2*	*316.6*	*365.7*	*441.3*	*630.6*	*241.2*	*717.7*

Appendix

Proof of Lemma 1

Since $2q_n^{g*} - q_{2n}^{g*} = \alpha > 0$, we have $q_n^{g*} > \frac{q_{2n}^{g*}}{2}$.

Since $q_{2e}^{g*} - q_{2n}^{g*} = \frac{2r}{1-6d} \geq 0$, we have $q_{2e}^{g*} \geq q_{2n}^{g*}$ and the equal sign holds only when r = 0.

Since $q_e^{g*} - q_n^{g*} = \frac{r}{1-6d} \geq 0$ and the equal sign holds only when r = 0, we have $q_e^{g*} \geq q_n^{g*} > \frac{q_{2n}^{g*}}{2}$.

Since $q_{2m}^{g*} - q_{2n}^{g*} = \frac{r}{1-6d} \geq 0$ and the equal sign holds only when r = 0, we have $q_{2m}^{g*} \geq q_{2n}^{g*}$. Moreover, it is straightforward to see that $q_{2e}^{g*} - q_{2m}^{g*} = \frac{r}{1-6d} \geq 0$.

Therefore, as long as $q_{2n}^{g*} \geq 0$, all the other quantities are non-negative. Then, items (i)-(ii) and (iv) are proved.

Item (iii) is straight-forwarder since $\frac{\partial\theta}{\partial d} = \frac{6r}{(1-6d)^2} > 0$, $\frac{\partial\theta}{\partial r} = \frac{1}{(1-6d)^2} > 0$.

Q.E.D.

Proof of Lemma 2

In the AB market:

$$\frac{\partial q_{2n}^{g*}}{\partial d} = \frac{4(4\alpha - 7c)}{(1-14d)^2} - \frac{4(5 - 84d + 420d^2)r}{(1-14d)^2(1-6d)^2}$$

$$= \frac{4\{(1-6d)^2(4\alpha - 7c) - [(5-42d)(1-6d) - 12d(1-14d)]r\}}{(1-14d)^2(1-6d)^2}$$

$$= \frac{4[(1-6d)^2(4\alpha - 7c) - (5-42d)(1-6d)r + 12d(1-14d)r]}{(1-14d)^2(1-6d)^2}$$

$$= \frac{4k^{g*}}{(1-14d)} + \frac{48d(1-14d)r}{(1-14d)^2(1-6d)^2} > 0$$

Then, from Lemma 1(i), Lemma 1(iii) and the above inequality, we have $\frac{\partial q_{2e}^{g*}}{\partial d} = \frac{\partial q_{2n}^{g*}}{\partial d} + 2\frac{\partial\theta}{\partial d} > 0$, $\frac{\partial q_{2m}^{g*}}{\partial d} = \frac{\partial q_{2n}^{g*}}{\partial d} + \frac{\partial\theta}{\partial d} > 0$, $\frac{\partial q_n^{g*}}{\partial d} = \frac{1}{2}\frac{\partial q_{2n}^{g*}}{\partial d} > 0$, $\frac{\partial q_e^{g*}}{\partial d} = \frac{1}{2}\frac{\partial q_{2n}^{g*}}{\partial d} + \frac{\partial\theta}{\partial d} > 0$.

Q.E.D.

Proof of Lemma 3

In CD and CH (DH) markets it results:

$\frac{\partial q_{2e}^{c*}}{\partial d} = \frac{4(2\alpha-3c)}{(1-6d)^2} > 0$

$\frac{\partial q_{e}^{c,*}}{\partial d} = \frac{2(2\alpha-3c)}{(1-6d)^2} > 0.$

In EF market, as non-negativity condition of q_{2n}^{c*} implies $\beta - 2c + 2\beta d - r > 0$ and $d < 1/14$, we have $4\big(2\beta - 3(c+r)\big) = 8\beta - 12c - 12r > \beta - 2c + 2\beta d - r > 0$. Then, $\frac{\partial q_{2n}^{c*}}{\partial d} = \frac{4(2\beta-3(c+r))}{(1-6d)^2} > 0$. In the EB and FB market, it results $\frac{\partial q_{n}^{c*}}{\partial d} = \frac{2(2\beta-3(c+r))}{(1-6d)^2} > 0$.

Q.E.D.

Proof of Lemma 4

(i) When $r = 0$, it results $q_e^{c*} - q_e^{g*} = -\frac{4d(\alpha-2c+2\alpha d)}{(1-14d)(1-6d)} = -\frac{4d}{1-6d} q_{2e}^{g*} \leq 0$ since $0 \leq d < 1/14$ and q_{2e}^{g*} is non-negative. Moreover, from Lemma 1(iv), $q_{2e}^{c*} - q_{2e}^{g*} = q_{2e}^{c*} - q_{2m}^{g*} = -\frac{8d(\alpha-2c+2\alpha d)}{(1-14d)(1-6d)} = 2\big(q_e^{c*} - q_e^{g*}\big) \leq 0.$

(ii) When $r = 0$, $k^{g*} - k^{c*} = \frac{(4\alpha-7c)(1-6d)-(2\beta-3c)(1-14d)}{(1-14d)(1-6d)} < 0$ if and only if $\beta > \frac{2(\alpha-c-6\alpha d)}{1-14d} = \tilde{\beta}$. $\tilde{\beta} - \alpha = q_{2n}^{g*} > 0.$

Q.E.D.

Proof of Proposition 1

Items (i) and (ii):

When r = 0, $SW^{g*} - SW^{c*} = \frac{-7\alpha+24\alpha c-16c^2+28\alpha d-80\alpha cd-44\alpha^2 d^2-\beta(8c-112cd)+\beta^2(3-44d+28d^2)}{2(1-14d)(1-6d)}$. The sign of $SW^{g*} - SW^{c*}$ depends on the sign of the numerator which is positive if and only if $\beta < \frac{4c}{3-2d} + \sqrt{\frac{21\alpha^2-72\alpha c+64c^2-98\alpha^2 d+288\alpha cd-256c^2 d+188\alpha d^2-160\alpha cd^2-88\alpha^2 d^3}{(3-2d)^2(1-14d)}} = \hat{\beta}$. When $\beta = \alpha$, $SW^{g*} - SW^{c*} = 2q_{2e}^{c*} q_{2e}^{g*} > 0$ and therefore $\hat{\beta} > \alpha$ always holds. Then, $SW^{g*} - SW^{c*} > 0$ holds if and only if $\beta < \hat{\beta}$ (including $\beta \leq \alpha$).

Item (iii):

When r = 0, replace β with $\tilde{\beta}$ obtained from Lemma 4(ii) into q_{n}^{c*} and $q_{2\mathrm{n}}^{c*}$ so as to obtain the quantities expressed as a function of α when $k^{g,*} = k^{c,*}$ holds. Then, one can show:

$$SW^{c,*}|_{r=0,\beta=\tilde{\beta}} - SW^{g,*}|_{r=0,\beta=\tilde{\beta}} = \frac{(\alpha+2\alpha d-2c)[(5-36d+164d^2)\alpha-2(3-2d)c]}{2(1-14d)^2(1-6d)}$$

As $q_{2\mathrm{n}}^{g*} = \frac{\alpha+2\alpha d-2c}{1-14d} > 0$, $\alpha > \frac{7c}{4}$ (since $k^{g*}|_{r=0} > 0$) and $5 - 36d + 164d^2 > 0$ hold, one can show:

$$SW^{c*}|_{r=0,\beta=\tilde{\beta}} - SW^{g*}|_{r=0,\beta=\tilde{\beta}} > \frac{q_{2\mathrm{n}}^{g*}(11-82d)c}{8(1-6d)} > 0, \text{ since } d < 1/14.$$

Therefore, when r = 0 and $\beta = \tilde{\beta}$, $SW^{c*} > SW^{g*}$.

Q.E.D.

Proof of Proposition 2

Item (i): $q_{2e}^{g*} - q_{2e}^{c*} = \frac{8d[\alpha-2c+2\alpha d-r]}{(1-14d)(1-6d)} = \frac{8dq_{2m}^{g*}}{(1-6d)} \geq 0$ and the equal sign holds when $d = 0$. Moreover, $q_e^{g*} - q_e^{c*} = \frac{4d[\alpha-2c+2\alpha d-r]}{(1-14d)(1-6d)} = \frac{4dq_{2m}^{g*}}{(1-6d)} \geq 0$ and the equal sign holds when $d = 0$.

Item (ii): Let $\beta = \alpha$, and then $q_{2n}^{g*} - q_{2n}^{c*} = \frac{16\alpha d^2+8d(\alpha-2c-r)}{(1-14d)(1-6d)} \geq 0$ since $\alpha > 2c + r$ and the equal sign holds when $d = 0$.

Q.E.D.

Proof of Proposition 3

We can solve for cut-off $\tilde{\beta}(r) = \frac{2\alpha-2c-12\alpha d-r}{1-14d}$ which makes $k^{g*} = k^{c*}$. Then, taking derivatives of $\tilde{\beta}$, we have $\frac{\partial\tilde{\beta}}{\partial c} = -\frac{2}{(1-14d)} < 0$, $\frac{\partial\tilde{\beta}}{\partial r} = -\frac{1}{(1-14d)} < 0$. Regarding the impact of d on $\tilde{\beta}$, we have $\frac{\partial\tilde{\beta}}{\partial d} = \frac{2(8\alpha-7(2c+r))}{(1-14d)^2}$. Since $\alpha - 2c - r > 0$, then $8\alpha - 14c - 7r > 8\alpha - 16c - 8r = 2(\alpha - 2c - r) > 0$, that is $\frac{\partial\tilde{\beta}}{\partial d} > 0$.

Q.E.D.

Proof of Proposition 4

Taking derivatives of equilibrium social welfares w.r.t. *r*, we get:

$\frac{dSW^{g*}}{dr} = -2Q_n^{g*} < 0$ and $\frac{dSW^{c*}}{dr} = -2Q_n^{c*} < 0$

$\frac{dSW^{g*}}{dr} - \frac{dSW^{c*}}{dr} = -2(k^{g*} - k^{c*}) < 0$ if and only if $k^{g*} > k^{c*}$.

Q.E.D.

Proof of Proposition 5

Taking derivatives of equilibrium social welfares w.r.t. *c*, we get:

$$\frac{dSW^{g*}}{dc} = -2\left(q_e^{g*} + q_{2e}^{g*} + 2q_m^{g*}\right) - 2\left(q_n^{g*} + q_{2n}^{g*} + 2q_{2m}^{g*}\right) = -2(Q_e^{g*} + Q_n^{g*}) < 0$$

$$\frac{dSW^{c*}}{dc} = -2(q_e^{c*} + q_{2e}^{c*}) - 2(q_n^{c*} + q_{2n}^{c*}) = -2(Q_e^{c*} + Q_n^{c*}) < 0$$

From Lemma 1(i), we have:

$$\begin{aligned}\frac{dSW^{g*}}{dc} - \frac{dSW^{c*}}{dc} &= -2\left(q_e^{g*} + q_{2e}^{g*} + 2q_m^{g*}\right) - 2(q_e^{c*} + q_{2e}^{c*}) - 2(k^{g*} - k^{c*}) \\ &= -2\left(\frac{8dq_{2m}^{g*}}{(1-6d)} + \frac{4dq_{2m}^{g*}}{(1-6d)} + 2q_{2m}^{g*}\right) - 2(k^{g*} - k^{c*}) \\ &= -2\left(2 + \frac{12d}{(1-6d)}\right)q_{2m}^{g*} - 2(k^{g*} - k^{c*})\end{aligned}$$

That is, $\frac{dSW^{g*}}{dc} - \frac{dSW^{c*}}{dc} > 0$ if and only if $k^{c*} - k^{g*} > \frac{2}{(1-6d)}q_{2m}^{g*}$.

Taking derivatives of equilibrium social welfares w.r.t. *d*, we get:

$$\frac{dSW^{g*}}{dd} = 2\left(q_{2n}^{g*} + q_n^{g*} + 2q_{2m}^{g*}\right)^2 + 2\left(q_{2e}^{g*} + q_e^{g*} + 2q_{2m}^{g*}\right)^2 = 2\left(Q_n^{g*2} + Q_e^{g*2}\right) > 0$$

$$\frac{dSW^{c*}}{dd} = 2(q_n^{c*} + q_{2n}^{c*})^2 + 2(q_e^{c*} + q_{2e}^{c*})^2 = 2\left(Q_n^{c*2} + Q_e^{c*2}\right) > 0$$

Then, it results:

$$\begin{aligned}\frac{dSW^{g*}}{dd} - \frac{dSW^{c*}}{dd} &= 2\left[\left(Q_n^{g*2} + Q_e^{g*2}\right) - \left(Q_n^{c*2} + Q_e^{c*2}\right)\right] = 2\left[\left(Q_e^{g*2} + k^{g*2}\right) - \left(Q_e^{c*2} + k^{c*2}\right)\right] \\ &= 2\left[\left(Q_e^{g*2} - Q_e^{c*2}\right) + \left(k^{g*2} - k^{c*2}\right)\right]\end{aligned}$$

That is, $\frac{dSW^{g*}}{dd} - \frac{dSW^{c*}}{dd} < 0$ if and only if $k^{c*2} > \left(Q_e^{g*2} - Q_e^{c*2}\right) + k^{g*2} > k^{g*2}$, since $Q_e^{g*} > Q_e^{c*}$.

Q.E.D.

Proof of Proposition 6

We first prove that there exists no feasible pair of (β, r) such that both $SW^{g*} - SW^{c*} = 0$ and $k^{g*} - k^{c*} = 0$ hold, together with the satisfaction of the following concavity and non-negativity conditions:

(C1) SW^{g} is concave while the newly built capacity constraint is binding. That is, its Hessian matrix is negative definite, and it is equivalent to have $d < 1/14$.

(C2) All quantities are non-negative. From the proof of Lemma 1 and $q_{2n}^{c*} < 2q_{\mathrm{n}}^{c*}$ and $q_{2e}^{c*} < 2q_{\mathrm{e}}^{c*}$, we know that as long as $q_{2n}^{g*} \geq 0$, $q_{2e}^{c*} \geq 0$ and $q_{2n}^{c*} \geq 0$, all the other quantities will be non-negative.

Suppose there exists (β, r) such that both $SW^{g*} - SW^{c*} = 0$ and $k^{g*} - k^{c*} = 0$ hold. Solving these two equations for (β, r), we can get the following two sets of solutions:

$$\begin{cases} r = (1+2d)\alpha - 2c \\ \beta = \alpha \end{cases} \quad \text{and} \quad \begin{cases} r = \frac{(164d^2-36d+5)\alpha-(6-4d)c}{7-58d} \\ \beta = \frac{(9-38d)\alpha-8c}{7-58d} \end{cases}$$

With the first set of solution, we can rewrite $q_{2n}^{c*} = \frac{-r}{1-6d} < 0$. Thus, the first set of solution is outside of the feasible range.

With the second set of solution, substituting the solution into expressions of quantities, we can get:

$$q_{2\mathrm{e}}^{c*} = \frac{(1+2d)\alpha-2c}{1-6d},\ q_{2\mathrm{n}}^{c*} = \frac{(52d-1-404d^2)\alpha-(10-92d)c}{(7-58d)(1-6d)},\ q_{2\mathrm{n}}^{g*} = \frac{(44d-3-284d^2)\alpha-(2-44d)c}{(7-58d)(1-6d)}$$

$q_{2\mathrm{e}}^{c*} - q_{2\mathrm{n}}^{c*} = \frac{4[2(1-6d)\alpha-c]}{7-58d}$, where $7-58d > 0$ holds as (C1) holds. Suppose $q_{\mathrm{e}}^{c*} = \frac{(1-2d)\alpha-c}{1-6d} > 0$ holds. Then, since $1-6d > 0$ due to (C1), we can show $2(1-6d)\alpha - c > (1-2d)\alpha - c > 0$. That is, $q_{2\mathrm{e}}^{c*} > q_{2\mathrm{n}}^{c*}$ and hence $q_{2\mathrm{e}}^{c*} > 0$ holds iff $q_{2\mathrm{n}}^{c*} > 0$. Then, the main focus is to check the non-negativity conditions of $q_{2\mathrm{n}}^{c*}$ and $q_{2\mathrm{n}}^{\mathrm{g}*}$.

As $q_{2\mathrm{n}}^{c*} - q_{2\mathrm{n}}^{g*} = \frac{2[(1+10d)\alpha-4c]}{7-58d}$, there are two cases:

Case 1: Suppose $q_{2\mathrm{n}}^{c*} - q_{2\mathrm{n}}^{g*} \geq 0$, it implies $(1+10d)\alpha - 4c \geq 0$. That is, $\alpha \geq \frac{4c}{1+10d}$. Moreover, it is easy to show that $44d - 3 - 284d^2 < 0$ as (C1) holds. Then, we have $q_{2\mathrm{n}}^{g*} \leq \frac{(44d-3-284d^2)4c-(1+10d)(2-44d)c}{(7-58d)(1-6d)(1+10d)} = \frac{-2c}{(1+10d)} < 0$.

Case 2: If $q_{2\mathrm{n}}^{c*} - q_{2\mathrm{n}}^{g*} < 0$, it implies $(1+10d)\alpha - 4c < 0$. That is, $c > \frac{(1+10d)\alpha}{4}$. Moreover, it is easy to show that $10 - 92d > 0$ as (C1) holds. Then, we have $q_{2\mathrm{n}}^{c*} < \frac{4(52d-1-404d^2)\alpha-(10-92d)(1+10d)\alpha}{4(7-58d)(1-6d)} = \frac{-\alpha}{2} < 0$.

In both Case 1 and Case 2, at least one quantity will be negative and (C2) fails to hold. Thus, the second set of solution is also outside of the feasible range. As a result, there exists no (β, r) such that SWdiff and Kdiff intersects within the feasible range.

Then, the only remaining step is to prove that when $k^{g*} = k^{c*}$, one of the points on the Kdiff curve is on the right-hand side of SWdiff curve, i.e. $SW^{c*} > SW^{g*}$ holds at that point. Since r = 0 is within in the feasible range, we focus on the point when r = 0 and $k^{g*} = k^{c*}$. From Lemma 4(ii) and Proposition 1(iii), we know that when r = 0 and $\beta = \tilde{\beta}$, both $k^{g*} = k^{c*}$ and $SW^{c*} > SW^{g*}$ hold.

Q.E.D.