Transportation Research Part B 82 (2015) 91–113

Contents lists available at ScienceDirect

Transportation Research Part B

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

Airport congestion pricing and terminal investment: Effects of terminal congestion, passenger types, and concessions

Yulai Wan[a,*], Changmin Jiang[b,c], Anming Zhang[d,e]

[a] Department of Logistics and Maritime Studies, Hong Kong Polytechnic University, Hong Kong, China
[b] Asper School of Business, University of Manitoba, Winnipeg, MB, Canada
[c] Transport Institute, University of Manitoba, Winnipeg, MB, Canada
[d] Sauder School of Business, University of British Columbia, Vancouver, BC, Canada
[e] China Academy of Financial Research, Shanghai Jiao Tong University, China

ARTICLE INFO

Article history:
Received 14 November 2014
Revised 15 October 2015
Accepted 15 October 2015
Available online 11 November 2015

Keywords:
Airport pricing
Terminal congestion
Runway congestion
Passenger types
Terminal concessions
Terminal investment

ABSTRACT

None of the airport-pricing studies have differentiated the congestion incurred in the terminals from the congestion incurred on the runways. This paper models and connects the two kinds of congestion in one joint model. This is done by adopting a deterministic bottleneck model for the terminal to describe passengers' behavior, and a simpler static congestion model for the runway. We find that different from the results obtained in the literature, uniform airfare *does not* yield the first-best outcome when terminal congestion is explicitly taken into account. In particular, business passengers are at first-best charged a higher fare than leisure passengers if and only if their relative schedule-delay cost is higher. We further identify circumstances under which passengers are, given a uniform airport charge scheme, under- or over-charged with respect to the terminal charge. Furthermore, when concession surplus is added to the analysis, the airport may raise (rather than reduce) the airport charge in order to induce more business passengers who in turn will lengthen leisure passengers' dwell time and hence increase their chance of purchasing concession goods. Finally, the impacts of terminal capacity expansion and time-varying terminal fine toll are discussed.

1. Introduction

Airport congestion has become an important and growing phenomenon in the airline industry. As a potential solution, imposition of congestion tolls has been proposed and widely discussed in the literature. Various factors that influence the design of airport congestion tolls have been examined. For example, airline market structure needs to be taken into account since airlines with market power may internalize the congestion cost they impose to their own flights (e.g., Daniel, 1995; Brueckner, 2002; Pels and Verhoef, 2004; Zhang and Zhang, 2006; Basso, 2008; Brueckner and Van Dender, 2008; Rupp, 2009; Flores-Fillol, 2010; Ater, 2012; Lin, 2013). Airport concessions should also be part of the picture, given that the number of passengers plays a different role in contributing to the airport's congestion level and concession revenues (e.g., Oum, et al. 2004; Czerny, 2006; Yang and Zhang, 2011; Bracaglia et al., 2014). The types of passengers may matter as well, because passengers of different types may have distinct responses to congestion tolls due to their different values of time (Czerny and Zhang, 2011, 2014b). D'Alfonso et al. (2013)

* Corresponding author. Tel.: +852 2766 7868.
E-mail address: sarah.wan@polyu.edu.hk (Y. Wan).

http://dx.doi.org/10.1016/j.trb.2015.10.006

investigated the interaction among these factors while incorporating the connection between congestion delays and concession consumption. They derived corresponding optimal airport charges based on the assumption that an increase in congestion delays will increase passengers' dwell time, and hence induce higher probability of their purchasing concession goods. This assumption does not reflect the fact that airport congestion may occur either in the terminals or on the runways.[1]

Interestingly, none of the previous studies have differentiated the congestion incurred in the terminals from the congestion incurred on the runways, despite of the fact that these two kinds of congestion clearly have different implications to airlines and airports in the above-mentioned contexts. In particular, terminal congestion seems to be less of a concern to the airlines' operations, but it will likely affect passenger behavior and airport concession activity to a large extent. On the other hand, runway congestion is more of an issue to the airlines but has less to do with airport concessions. In other words, airport concessions and passenger types are related more closely with terminal congestion than with runway congestion. For instance, passengers cannot control the time to take-off, but can choose when to arrive at the airport. Passengers can consume concessions during their dwell time at the terminal, but they are required to stay in aircraft and wait for their turns to take-off when the runway is congested and as a consequence, they are unable to purchase any concession goods during the wait. Therefore, separating these two kinds of airport congestion may help clarify and deepen our understanding of the interactions between different factors in designing an optimal airport charge. Furthermore, these two kinds of airport congestion show different characteristics, with terminal congestion being totally "atomistic" while runway congestion shows a certain degree of "internalization."[2]

Taken together, these observations suggest that it is important to investigate the impacts of separating terminal congestion from runway congestion, in order to improve our understanding of airport congestion and the design of an optimal airport toll. In this paper we treat terminal congestion and runway congestion as two different kinds of congestion, and consider both congestion types simultaneously in a single airport pricing model. To capture the difference between terminal and runway congestions, we adopt a deterministic bottleneck model for the terminal and a simpler static congestion model for the runways (the latter is standard in the recent airport pricing literature). The bottleneck model was introduced by Vickrey (1969) to account for congestion dynamics, in particular individual drivers' decision on when to depart their home during the morning rush hours. Arnott et al. (1990) derived the optimal coarse toll and optimal road capacity based on a bottleneck model with a fixed number of identical travelers, i.e. inelastic travel demand. Arnott et al. (1993) modified the previous studies by incorporating elastic travel demand. The model was then extended to cases where travelers are heterogeneous (e.g., Arnott et al., 1994; van den Berg and Verhoef, 2011). Compared with the conventional static congestion model which assumes a constant level of congestion over a period of time for all the travelers, the bottleneck model can capture such features as varying congestion over time and travelers' response to congestion tolls by adjusting departure time. However, except for the case of a single dominant player, the Vickrey-type bottleneck model would only fit the cases where users of a congestible facility are all "atomistic" (such as cars on highways). Airport terminals face individual passengers who are by definition atomistic and so won't take into account other passengers when making decisions. As a consequence, it appears to be a perfect context for the usage of the bottleneck model. Runways, on the other hand, face airplanes operated by a few large airlines that potentially internalize the congestion they impose on their own flights. Therefore, the bottleneck model may not be a good fit for runway congestion.[3] Our integrated model is related to D'Alfonso et al. (2013) but here we treat the two kinds of congestion explicitly and separately. Instead of assuming all passengers incur the same amount of terminal dwell time, which is roughly determined by the overall passenger volume at the airport, this new modeling approach enables the terminal dwell time to be elicited by individual passengers' airport arrival behavior, which is affected by both the number of passengers and the passenger types. Another difference with D'Alfonso et al. (2013) is that the present paper no longer assumes away airline price discrimination from the analysis.

We find that contrary to Czerny and Zhang (2011, 2014b), when terminal congestion is taken into account, uniform airfare *does not* yield the first-best outcome. This is because some passengers may prefer arriving at the airport far in advance to avoid long queues at check-in and security screening, while others may prefer arriving at the airport relatively late to avoid long dwell times before boarding. Intuitively, adding a passenger who arrives at the airport at a particular time may increase the queuing time for those arriving after that passenger and the airside dwell time for those arriving before by pushing their arrival times forward. As a result, different passengers cause different externalities on the others. Thus, the amount of terminal externality to

---

[1] See Zhang and Czerny (2012) and Basso and Zhang (2007) for recent surveys of these and other studies. For early congestion pricing studies see, for example, Levine (1969), Carlin and Park (1970), Morrison (1983), and Mohring (1999).

[2] A number of early papers have focused only on runway congestion, arising from the "all the flights leave at 8 a.m." phenomenon and/or the flight banking phenomenon (see further discussion in Section 2). Such clustering schedule behavior by airlines could however also lead to congestion at the pre-departure procedures at the terminal. In addition, the stringent new security requirements that were implemented as a direct result of the September 11th terrorist attacks have made such pre-departure procedures more cumbersome and time-consuming (an effect often referred to as the "hassle factor," see Ito and Lee 2005). Together, they make terminal congestion often a non-negligible matter (e.g., Appold and Kasarda, 2007; Saraswati and Hanaoka, 2012; Lin and Chen, 2013).

[3] The Vickrey bottleneck model was recently successfully applied to runway congestion with airlines having market power (e.g. Daniel, 1995, 2009; Silva et al., 2014). Such model setting includes only one dominant airline plus one or more atomistic fringe airline(s). Nevertheless, this model has difficulties in describing airports' dynamic congestion when two or more airlines have significant market shares – but such an oligopoly market structure is quite common in the industry (e.g., Morrison and Winston, 1989; Brander and Zhang, 1990, 1993; Zhang and Czerny, 2012). As indicated by Silva et al. (2014), the Vickrey model may not possess pure-strategy Nash equilibrium when it is applied in the Cournot oligopoly case. Silva, Lindsey, de Palma and van den Berg (2014) confirmed, in a duopoly setting, that with the Vickrey congestion technology either a pure strategy equilibrium does not exist or it exits while no queue occurs. Thus, the Vickrey bottleneck model may not be a good fit for runway congestion when the airport is served by several significant airlines. Verhoef and Silva (2015) describe the dynamic congestion pattern at airports served by two non-atomistic airlines by replacing bottleneck congestion with Chu (1995)'s flow congestion which ignores interactions among travellers departing at different time instants. Verhoef and Silva's work provides a possible alternative in modeling dynamic runway congestion in the future. More remarks on the relevance of road bottleneck models to the present setting will be given in the text.

be internalized varies across passenger types. However, an additional passenger leads to the same amount of runway congestion regardless the passenger types. Our analysis shows that when both types of passengers are levied a uniform airport charge, an increase in airport charge will reduce the number of leisure passengers and the total number of passengers, but may reduce or increase the number of business passengers. Furthermore, various conditions are identified under which the terminal charges (conditional on an optimal uniform airport charge) are above or under the externality that a certain passenger imposes on the others. Finally, when comparing the airport pricing rule with the one found in the literature (e.g., D'Alfonso et al., 2013) we find that if the volume of business passengers increases in airport charge, the airport will raise rather than reduce the airport charge in order to lengthen leisure passengers' dwell time and hence increase their chance of purchasing concession goods.

To sum up, the main contribution of this paper are to: (i) highlight the difference between terminal congestion and runway congestion that has been overlooked in the airport pricing literature, (ii) propose a modeling approach to incorporate this difference, (iii) show the potential problem of treating terminal congestion the same way as runway congestion, and (iv) show the structure and distortion of the optimal uniform airport charge, as well as the underlying reasons. Although we are not aiming at providing a toll scheme that can lead to first-best outcome, our analysis may provide some other insights. For example, instead of relying only on a pricing mechanism, the airport can provide appropriate facilities to minimize the difference between the relative schedule delay costs of different passenger groups, so that the uniform airport charges can still internalize the right amount of terminal externality.

The structure of the paper is as follows. Section 2 sets up the model. Section 3 models passengers' equilibrium arrival pattern at the terminal and characterizes the demand functions. Section 4 derives the first-best outcomes, and Sections 5 examines airlines' and the airport's equilibrium behaviors, with airport concessions being assumed away in both sections. Section 6 examines the first-best outcomes and optimal airport charges by incorporating concessions. Section 7 provides further discussion on policy implications and issues which can be studied with this modeling approach in the future. Finally, Section 8 contains the concluding remarks.

2. The model

Consider a congestible airport served by n identical airlines. There are two types of passengers departing from the airport[4]: business and leisure. Business passengers (denoted as B) have a higher value of time than leisure passengers (denoted as L), which is represented with $v_B > v_L$ (e.g., Morrison, 1987; Morrison and Winston, 1989; USDOT, 1997; Pels et al., 2003). Denote q_h^i as the number of "type h" passengers flying with airline i, $i = 1, 2 \ldots n$, and $Q_h = \sum_{i=1}^{n} q_h^i$, $h = B, L$. The (inverse) demand function of type-h passengers is $\rho_h(Q_h)$, which is two-times differentiable with $\rho_h'(Q_h) < 0$ and $\rho_L''Q_h + 2\rho' < 0$ for any (Q_B, Q_L) pair in the relevant range of our analysis. The latter implies that the first-order conditions lead to the global maximum of the revenues summed across all airlines when there are no user costs, and is a standard assumption in modeling oligopoly rivalry (Tirole, 1988).

Following Brueckner (2002), Basso and Zhang (2008) and Czerny and Zhang (2014a), we consider two periods at the airport with (i) all passengers preferring departing during the peak period to departing during the off-peak period, and (ii) runway congestion occurring only at the peak period. To simplify the analysis, we further assume that all the departing flights in concern are scheduled at time $\hat{t}$ of the peak period (and the gates will close at this time as well). In other words, $\hat{t}$ is considered as the preferred departure time of the peak period by all airlines. Borenstein and Netz (1999) show that schedule competition among airlines may result in airlines locating their flights closely together in time (the so-called "all the flights leave at 8 a.m." phenomenon).[5] Fig. 1 illustrates the activity timeline for a particular departing passenger who completes pre-departure procedures prior to $\hat{t}$. The pre-departure procedures at the terminal – which include check-in, security screening and passport control (if any) – create queues, and will be considered collectively as a bottleneck.

Consider, more specifically, the departing passenger in Fig. 1. She arrives at the airport at time t and has to wait in line for $T_v(t)$ units of time before being processed. The check-in and screening procedures take T_f units of time per passenger and have a capacity of s passengers per unit of time. Defining $T(t) = T_v(t) + T_f$, the passenger completes the pre-departure procedure at time $a(t)$, or equivalently $t + T(t)$, and then proceeds to the departure lounge to wait for boarding at $\hat{t}$. We assume that being late and missing the flight is never an acceptable option for passengers, i.e. $a(t) \le \hat{t}$. Once passengers are boarded on the aircraft, as all the flights are scheduled and ready for take-off at $\hat{t}$, they will form a queue and incur runway delays.

We assume that the waiting time in the terminal queue and in the runway queue incurs the same time cost, given by v_h ($h = B, L$). In addition, if a passenger finishes the pre-departure procedures earlier than $\hat{t}$, she will incur an early

[4] Our model will focus on departure passengers. Arrival and departure do not share the same terminal procedures although they share runways. Besides, arrival passengers' concession activity is less likely to be correlated with terminal congestion. Note also that here, departure passengers refer to originating passengers rather than connecting passengers. In most of the cases, connecting passengers do not follow the same terminal procedures as originating passengers. For example, most connecting passengers will get their boarding passes when checking-in for the first flight at their very origins; and in the U.S. they do not need to go through security screening again at the transiting airport as long as they stay in the airside of the terminal. When passengers transit between an international leg and a domestic leg, however, substantial delays due to the crowded terminal can occur. As flights are scheduled by airlines (and so connecting passengers have less control on their airport arrival time), the behavior of connecting passengers is not modeled here.

[5] Alternatively, this assumption may be analogous to flight banking behavior at hub airports found in the literature (e.g. Daniel, 1995; Mayer and Sinai, 2003) with $\hat{t}$ as the preferred departure time of one representative bank.

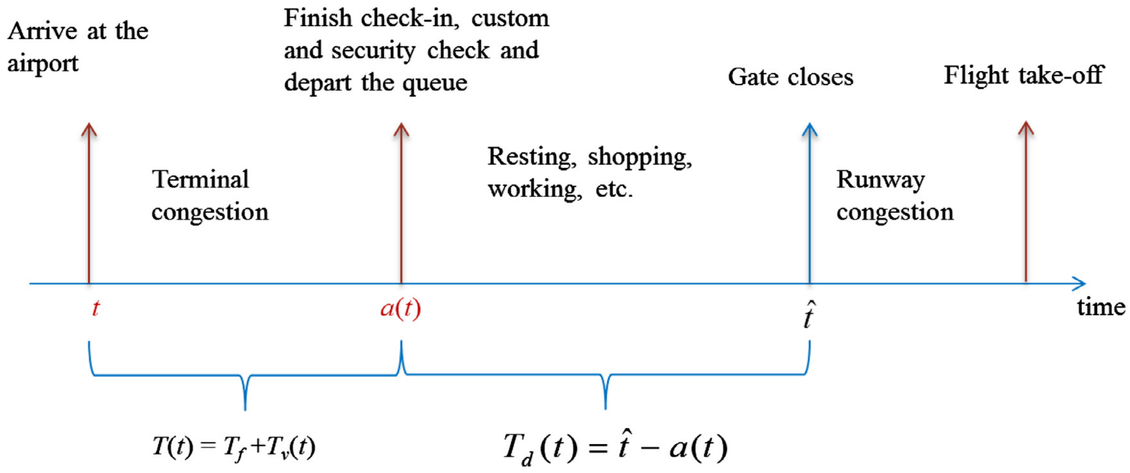

Fig. 1. Activity timeline of a departing passenger.

"schedule delay cost."[6] This cost is given by $v_h\beta_h \cdot (\hat{t} - a(t))$, where term $v_h\beta_h$ denotes the unit early schedule delay cost of type-h passengers and parameter β_h is the relative cost of early schedule delay to time delay.[7] Here, we assume $0 < \beta_h < 1$, i.e., the early schedule cost is less than the pure waiting time cost, v_h. This is because passengers feel more comfortable and can perform more productive activities when waiting in the lounge during their "dwell time" than lining up in a queue (e.g., Thomas, 1997; Graham, 2009; Lin and Chen, 2013). Note that while the empirical results suggest $v_B > v_L$, it is not clear whether $\beta_B > \beta_L$, or $\beta_B = \beta_L$, or $\beta_B < \beta_L$. As a consequence, it is not clear whether $v_B\beta_B > v_L\beta_L$ (or $v_B\beta_B = v_L\beta_L$, or $v_B\beta_B < v_L\beta_L$), i.e., business passengers' time valuation is higher (or the same, or lower) than leisure passengers' time valuation when the early schedule delay is considered. As to be seen below, the relative magnitude of β_B and β_L is a key factor influencing some of the paper's main results.

Following the common simplification in the literature (e.g., Arnott et al., 1990, 1993, 1994) the fixed "free flow" processing time T_f is taken to be constant and is, without loss of generality, further normalized to zero. Thus, the terminal cost of a type-h passenger arriving at the airport at time t is:

$$c^h(t) \equiv v_h T(t) + \beta_h v_h \cdot (\hat{t} - a(t)) = v_h T(t) + \beta_h v_h \cdot (\hat{t} - t - T(t)) \tag{1}$$

where the passenger's dwell time is defined as $T_d(t) = \hat{t} - a(t)$. Note that the first term on the right-hand side (RHS) of (1) is the waiting (queuing) cost, whereas the second term is the schedule-delay (dwell time) cost.

The runway delay is, as is common in the literature, denoted as $D(Q)$, where $Q = Q_B + Q_L$ is the total number of passengers served by all carriers, with $D'(Q) > 0$ and $D''(Q) \geq 0$.[8] In particular, if we assume that all flights will form a random queue and the runway capacity is K passengers per unit of time, the first take-off will be at $\hat{t}$ and the last one will be at $\hat{t} + (Q/K)$. Consequently, the expected runway delay is:

$$D(Q, K) = \int_0^{\frac{Q}{K}} z\frac{K}{Q}dz = \frac{Q}{2K} = \theta Q,$$

where $\theta = 1/(2K))$. The above is an example of deriving the congestion delay at the runway; in particular, this linear delay function has been used by, among others, De Borger and Van Dender (2006), Basso (2008), Basso and Zhang (2008) and Yang and Zhang (2011). The remainder of the paper, however, will not impose a particular functional form on runway congestion; rather, we will consider D as a general function of Q while suppressing the notation of K. A type-h passenger who arrives at the airport at time t incurs a generalized cost that is equal to the sum of airfare, runway congestion cost and terminal cost:

$$f_h(t) \equiv p_h + v_h D(Q) + c^h(t), \tag{2}$$

where p_h is the ticket price paid by the type-h passenger and $c^h(t)$ is given by (1).

The airport and airline behavior is modeled as a two-stage game: in the first stage, the airport chooses its charges to maximize social welfare. In the second stage, airlines simultaneously determine output levels to maximize their respective profits. Here, each airline may make separate decisions on the number of business passengers and the number of leisure passengers, implying that airlines can price discriminate between the two passenger types (e.g., Czerny and Zhang, 2014b, 2015; van der Weijde, 2014). For simplicity, we normalize all the other costs incurred by airlines and the airport to zero.

[6] This is in accord to the term used in the conventional bottleneck model. Note that schedule delays at the runway measure the absolute difference between the passengers' most preferred and their actual departure times (e.g., Douglas and Miller, 1974; Panzar, 1979; Basso, 2008; Flores-Fillol, 2010).

[7] Qian and Zhang (2013) applied a similar approach to model schedule delay costs in bottleneck models.

[8] See the survey paper by Zhang and Czerny (2012) and references cited therein for the treatment of runway congestion.

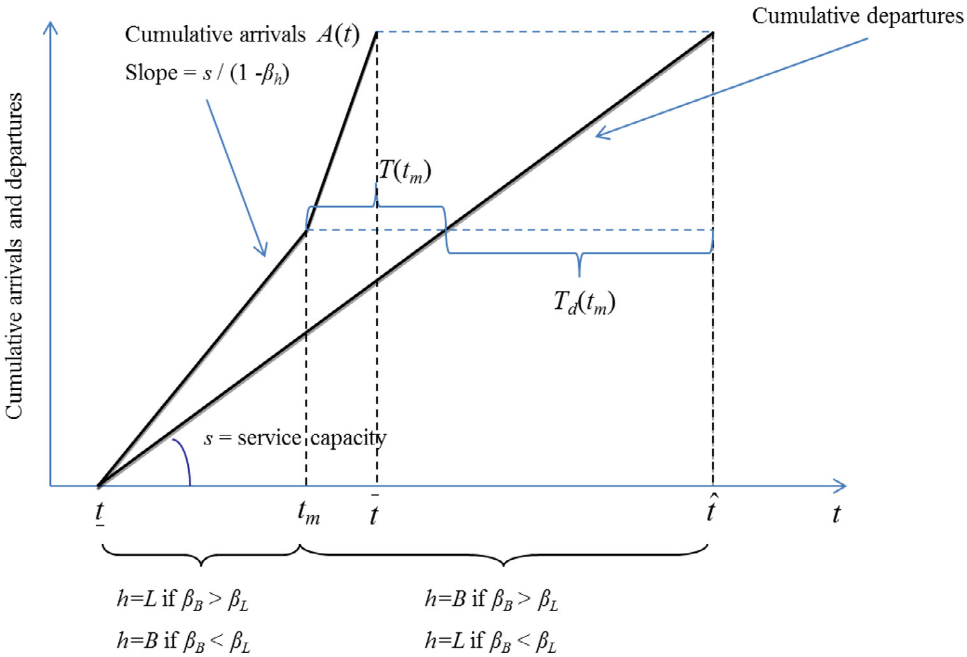

Fig. 2. Cumulative arrivals at the airport and departures from pre-departure processing sites at the terminal.

3. Passenger equilibrium and demand functions

Before analyzing the subgame perfect Nash equilibrium of the two-stage game, we characterize passenger equilibrium and demand functions. Given airport and airline decisions, individual passengers determine whether to take the flights leaving at time $\hat{t}$ and if so, when to arrive at the airport. Consider the timing decision first. According to the literature (e.g., Arnott et al., 1994; Arnott and Kraus, 2003; van den Berg and Verhoef, 2011), at the equilibrium the cumulative numbers of passengers arriving at the airport and departing from the bottleneck (i.e. completing the pre-departure procedures) follow the patterns shown in Fig. 2.[9] Specifically, all passengers will arrive at the airport during the time interval $[\underline{t}, \bar{t}]$. A passenger who arrives at $\underline{t}$ incurs only the early schedule-delay cost (i.e. zero queuing cost). On the other extreme, a passenger who arrives at $\bar{t}$ incurs only the queuing cost and she leaves the bottleneck exactly at time $\hat{t}$, thereby incurring zero early schedule-delay cost. Any passenger arriving between $\underline{t}$ and $\bar{t}$ incurs both the queuing and schedule-delay costs. The bottleneck will always operate at its capacity until the last passenger is processed. Thus the rate of leaving the bottleneck is capacity s, which defines the slope of the "cumulative departures" curve in Fig. 2.

As passengers minimize their individual terminal costs by choosing their arrival time, we take the first-order condition of Eq. (1) with respect to t and obtain:

$$T'(t) = \frac{\beta_h}{1 - \beta_h}.$$

Note[10] that the relationship between $A(t)$ – the total number of passengers that arrive by t – and $T(t)$ is: $T(t) = (A(t) - s(t - \underline{t}))/s$. Therefore, it is easy to find that the slopes of the "cumulative arrivals" curve are equal to:

$$A'(t) = \frac{s}{1 - \beta_h},$$

indicating the arrival rates for type-h passengers, $h = B, L$. If $\beta_B > \beta_L$, i.e. the relative schedule-delay cost of business passengers is higher than that of leisure passengers, then business passengers are more sensitive to early schedule delay and hence are willing to accept a longer waiting time in return for a shorter dwell time (e.g., Arnott et al., 1994; Arnott and Kraus, 2003; van den Berg and Verhoef, 2011). As a result, leisure passengers arrive during the interval $[\underline{t}, t_m]$, which is then followed by business passengers who arrive during $[t_m, \bar{t}]$. Further, the arrival rate of the former is lower than the arrival rate of the latter, suggesting that the left

[9] As indicated in conventional bottleneck models, it is assumed here that passengers are informed of the congestion at the airport. Although this assumption is perhaps a little bit strong, it is still sensible for two reasons: first, frequent travelers usually have a good knowledge about the airport's congestion level. Second, while occasional travelers have less experience with, and knowledge about, the airport than frequent travelers, it is easy to learn about on-time performance of an airport or even a specific flight from the internet as long as the passengers care about such information.

[10] In the classic bottleneck model, the notation β_h refers to the per-unit early schedule delay cost. In the present paper, however, it refers to the ratio between the per-unit early schedule delay cost and the per-unit waiting time cost.

segment of the cumulative-arrivals curve is flatter than the right segment. Likewise, if $\beta_B < \beta_L$, business passengers will arrive earlier at a lower rate.[11]

At equilibrium, passengers of the same type must face the same terminal cost; otherwise, passengers with a higher cost will adjust their arrival times to those associated with a lower cost. Based on this equilibrium condition, terminal costs can be derived as follows:

$$\begin{aligned} &c^B(t) = c^B(\bar{t}) = v_B T(\bar{t}),\, c^L(t) = c^L(\underline{t}) = v_L \beta_L (\hat{t} - \underline{t}) \text{ if } \beta_B > \beta_L;\\ &c^B(t) = c^B(\underline{t}) = v_B \beta_B (\hat{t} - \underline{t}),\, c^L(t) = c^L(\bar{t}) = v_L T(\bar{t}) \text{ if } \beta_B < \beta_L. \end{aligned} \tag{3}$$

Because the queue starts at $\underline{t}$ and ends at $\hat{t}$, we have $\hat{t} - \underline{t} = (Q_B + Q_L)/s.$. As all passengers arrive at the airport by $\bar{t}$, it follows that

$$\bar{t} - \underline{t} = \frac{Q_B(1-\beta_B)}{s} + \frac{Q_L(1-\beta_L)}{s}.$$

The time spent by the last passenger waiting before being processed is equal to

$$T(\bar{t}) = \frac{1}{s}\left[Q_B + Q_L - (\bar{t} - \underline{t})s\right] = \frac{Q_B\beta_B + Q_L\beta_L}{s}.$$

We can then rewrite Eq. (3) into the following:

$$\begin{aligned} &c^B(t) = v_B \frac{\beta_B Q_B + \beta_L Q_L}{s},\, c^L(t) = v_L \beta_L \frac{Q_B + Q_L}{s} \quad \text{if} \quad \beta_B > \beta_L;\\ &c^B(t) = v_B \beta_B \frac{Q_B + Q_L}{s},\, c^L(t) = v_L \frac{\beta_B Q_B + \beta_L Q_L}{s} \quad \text{if} \quad \beta_B < \beta_L. \end{aligned} \tag{4}$$

Thus, the equilibrium terminal costs are functions of Q_B and Q_L. From here on, we replace for terminal costs $c^B(t)$ and $c^L(t)$ with $c^B(Q_B, Q_L)$ and $c^L(Q_B, Q_L)$. Denoting $\partial c^h/\partial Q_h$ as c_h^h and $\partial c^h/\partial Q_k$ as c_k^h for $h, k = B, L$ and $h \neq k$, we obtain

$$\begin{aligned} &c_B^B = \frac{v_B\beta_B}{s},\, c_L^B = \frac{v_B\beta_L}{s},\, c_B^L = \frac{v_L\beta_L}{s},\, c_L^L = \frac{v_L\beta_L}{s} \quad \text{if} \quad \beta_B > \beta_L;\\ &c_B^B = \frac{v_B\beta_B}{s},\, c_L^B = \frac{v_B\beta_B}{s},\, c_B^L = \frac{v_L\beta_B}{s},\, c_L^L = \frac{v_L\beta_L}{s} \quad \text{if} \quad \beta_B < \beta_L. \end{aligned} \tag{5}$$

One interesting observation from (5) is that the structure of per-passenger terminal costs depends on the relative magnitude of β_B and β_L. The passenger type with a higher relative schedule-delay cost shows a higher "own effect" on the terminal cost than the "cross effect." More specifically, when $\beta_B > \beta_L$, the increase in business passenger terminal cost due to one additional business passenger is higher than one additional leisure passenger: $c_B^B > c_L^B$. Intuitively, as business passengers arrive after leisure passengers, adding one more business passenger will push business passengers arrival time forward and increase business passengers' queuing time as well as dwell time. On the other hand, adding one more leisure passenger before the arrival of business passengers does not affect business passengers' dwell time but the business passengers have to wait for a longer time in the queue as more leisure passengers will be processed first. Thus, adding more leisure passengers imposes less externality on business passengers than adding more business passengers. However, regarding the leisure passenger terminal cost, the impact of increasing business passengers is the same as increasing leisure passengers: $c_L^L = c_B^L$. This is because either one more business passenger or one more leisure passenger will push the leisure passengers' arrival time forward by the same amount. When $\beta_B < \beta_L$, we have $c_L^L > c_B^L$, while $c_B^B = c_L^B$. This property does not exist in the runway delay cost function (i.e. conventional airport congestion cost function) which always assumes that it is the total traffic volume rather than the traffic volumes of individual passenger types that matters. [12]

When passengers make air travel (ticket purchasing) decisions, they take the above terminal costs into consideration and so Eq. (2) becomes

$$f_h(t) = f_h(Q_B, Q_L) \equiv p_h + v_h D(Q) + c^h(Q_B, Q_L). \tag{6}$$

The demand equilibrium requires $\rho_h(Q_h)$=$f_h(Q_B, Q_L)$, which leads to

$$p_h = \rho_h(Q_h) - v_h D(Q) - c^h(Q_B, Q_L),\, h = B, L. \tag{7}$$

Thus, the (inverse) demand function with respect to airfare is a function of both Q_B and Q_L. Applying the Cramer's rule with respect to p_h, we obtain Proposition 1:

Proposition 1. *(i) When airlines charge discriminative fares, $\partial Q_h/\partial p_h < 0$ and $\partial Q_h/\partial p_k > 0, \forall h \neq k$. (ii) Under uniform airline pricing where $p_B = p_L = p$, $\partial Q_L/\partial p < 0$ and $\partial Q/\partial p < 0$, but the sign of $\partial Q_B/\partial p$ is ambiguous, depending not only on v_h but also on β_h. In particular, given v_h, when $\beta_B > \beta_L$, only β_L will affect the sign of $\partial Q_B/\partial p$; on the other hand, when $\beta_B < \beta_L$, both β_B* and *β_L will affect the sign of $\partial Q_B/\partial p$.*

[11] This paper considers only two types of passengers, but the idea can be extended in the future to accommodate multiple passenger types in which the cumulative-arrivals curve will consist of multiple line segments, each representing one type. In the case of continuous passenger types, the piece-wise linear cumulative-arrivals pattern could become a smoothed curve.

[12] A similar inequality between the own and cross effects due to one additional passenger also occurs in the road bottleneck model (e.g. Lindsey, 2004; van den Berg and Verhoef, 2010, 2011).

Proof. See Appendix A. □

When discriminative fares are charged to the business and leisure passengers (referred to as "laissez fair" situation), an increase in type-h airfare will suppress the volume of type-h passengers while the reduced congestion levels will stimulate the demand of type-k passengers, as the two passenger groups compete for terminal and runway resources. Regarding the case of uniform pricing, Czerny and Zhang (2011) have obtained similar results when considering runway congestion alone.[13] They found that $\partial Q_B/\partial p$ is more likely to be positive if the difference between v_B and $v_L$is larger, because in this case the decreasing number of leisure passengers due to higher ticket price will significantly reduce the runway congestion cost of the business passengers, which can compensate this group for the increased fare. In our model, not only the runway congestion but also the terminal congestion will have an impact. When $\beta_B > \beta_L$, business passengers arrive at the terminal later than the leisure passengers do, so the decreasing number of leisure passengers does not affect business passengers' schedule delay costs but only shortens their terminal queuing time and the amount of queuing time saving is affected solely by the leisure passengers' arrival rate (determined by β_L). Thus, the larger β_L, the higher the terminal cost savings by business passengers to compensate the fare rise, making an increase in business passenger volume more likely to occur. On the other hand, when $\beta_B < \beta_L$, business passengers arrive at the terminal earlier than the leisure passengers do, so the decreasing number of leisure passengers does not affect business passengers' terminal queuing costs but reduces their schedule delays. Thus, only when business passengers valuate schedule delay costs high enough, i.e. $v_B\beta_B$ is substantially larger than $v_L\beta_L$, will the terminal cost savings from fewer leisure passengers be large enough to induce more business passengers. This mechanism complicates the situation and suggests an exception from the results of Czerny and Zhang (2011): even when v_B is significantly larger than v_L, can still be negative as long as $v_B\beta_B$ - $v_L\beta_L$ is sufficiently negative. This observation is important to airlines, because it suggests that passenger behavior at the terminal (arrival pattern in particular) has an implication on their traffic volume and pricing decisions. In particular, we need to be clear not only about the absolute value of time for different passenger groups, but also about their relative values of schedule delay versus queuing time.

4. First-best outcome

According to the traditional bottleneck literature, social optimum is obtained when there are no queuing costs but only schedule delay costs. This is attainable if time-varying tolls are levied.[14] In the case of airports, time-varying tolls are nonexistent in practice and are unlikely to be implemented. Thus in this section, we only derive first-best outcomes given the terminal cost structure derived in Section 3, implying that passenger queuing at the terminal persists. Furthermore, social welfare may consist of two parts, aeronautic surplus and concession surplus. The aeronautic surplus can be expressed in the following way:

$$S_a \equiv \sum_{h=B,L} \int_0^{Q_h} \rho_h(x)dx - v_h D(Q)Q_h - c^h(Q_B, Q_L)Q_h.$$

On the other hand, whether the surplus from terminal concessions should be considered as part of social welfare is still under debate in the literature, as consumption of food and beverage would occur elsewhere even if passengers do not consume at the airport (e.g., Czerny, 2013). Thus, in this section and in Section 5 below, we consider the situation where concession surplus is not a concern. (The role of concessions will be discussed in Section 6.) Consequently, welfare is given by:

$$W \equiv S_a. \tag{8}$$

The first-best outcomes can be derived by taking the first-order conditions of (8) with respect to Q_B and Q_L ($*$ for "first best"),

$$\frac{\partial W}{\partial Q_B} = p_B^* - \Gamma^* - \phi_B^* = 0, \tag{9}$$

$$\frac{\partial W}{\partial Q_L} = p_L^* - \Gamma^* - \phi_L^* = 0. \tag{10}$$

In Eqs. (9) and (10), $\Gamma \equiv (v_B Q_B + v_L Q_L)D'(Q)$ denotes the marginal external congestion cost at runway, whereas $\phi_h \equiv Q_h c_h^h + Q_k c_h^k$ denotes the marginal external terminal cost due to one additional type-h passenger at terminal. It is fairly straightforward to show that the difference between the first-best business and leisure fares is

$$p_B^* - p_L^* = \phi_B^* - \phi_L^* = (c_B^B - c_L^B)Q_B^* - (c_L^L - c_B^L)Q_L^*, \tag{11}$$

which is, by (5), positive (negative, respectively) if the relative schedule-delay cost of business passengers (β_B) is higher (lower, respectively) than that of leisure passengers (β_L) . Moreover, Eq. (11) will not equal zero unless β_B and $\beta_L$are equal. This is because when the relative schedule-delay costs are not equal, one of the two terms on the RHS of Eq. (11) will be zero and the other term will be either positive or negative. This is different from Czerny and Zhang (2011, 2014b) who found, without and with

[13] D'Alfonso et al. (2013) obtained similar results based on linear demand functions (with different model settings). Proposition 1 here is proved with general demand functions, however.

[14] See Arnott et al. (1990, 1993, 1994) for the definition and construction of time-varying tolls.

carrier price discrimination respectively, that the social optimum can be achieved only by the welfare-optimal uniform airfare. This difference is due fundamentally to the inclusion of terminal costs into the present analysis. When runway congestion is the only concern (as in Czerny and Zhang, 2011, 2014b), the impact on runway congestion due to one additional business passenger is the same as the impact of one additional leisure passenger (which is captured by Γ above). As a result, at the social optimum both passenger types should be charged at the same price. However, this is no longer the case when terminal costs enter the picture. For example, as indicated above, when $\beta_B > \beta_L$, adding one more business passenger has a more adverse impact on business passengers' terminal cost than adding one more leisure passenger; consequently, business passengers should be charged higher than leisure passengers. The above analysis leads to the following proposition:

Proposition 2. *When terminal congestion is taken into account and $\beta_B \neq \beta_L$,* i.e. *the relative schedule-delay cost of business passengers differs from that of leisure passengers, a necessary condition for the first-best outcome is that fares are different between business and leisure passengers. The first-best fare charged to business passengers is higher than that to leisure passengers if and only if $\beta_B > \beta_L$.*

Proposition 2 is related to, but different from, the results of the road bottleneck models which generally consider time-dependent tolls (either a completely flexible time-varying "fine toll" which eliminates queues, or a "coarse toll" that differentiates peak travel time from off-peak travel time) as compared to the cases of no toll or uniform toll. In the case of heterogeneous drivers, both the fine toll (see Arnott et al., 1994) and coarse toll (see van den Berg, 2014) vary across time and the difference in the amount of toll paid by passenger types comes from drivers' self-selection and hence is not imposed by the toll scheme itself.[15] In the context of aviation, a discriminative price based on the time of terminal arrival is not really practical. The only tool available to the airlines to correct the terminal congestion lies in its power to differentiate different groups of passengers (business vs. leisure in this case). In other words, with this framework we provide another perspective to the airlines' price differentiation based on passenger groups. That is, it is because of the lack of time-varying toll at the terminal side that airlines need to price discriminate passengers with differentiated fares to correct part of the distortions at the terminal. Section 7.2 provides more discussion on the hypothetical benchmark case where terminal queue is eliminated by a time-varying fine toll.

Following Proposition 2 and after some calculation, we can show that if the airport charges discriminative tolls between the business and leisure passengers, the first-best outcome can be achieved. In practice, however, an airport is often constrained, politically or legally, from charging discriminative tolls. Even if an airport is allowed to do so, the implementation of such a toll scheme can be hard, as the airport usually is not able to distinguish the business and leisure passengers (while airlines may also have an incentive to cheat as one type of passengers will be charged a lower toll than the other).[16] Therefore, the remainder of the paper will focus only on uniform airport charge.

5. Optimal airport charge

5.1. Airline equilibrium behavior

To solve for the optimal (uniform) airport charge, we first examine the second-stage game when airlines simultaneously choose output levels for business and leisure passengers while taking the airport charge as given. Each carrier's objective is to maximize profit given by:

$$\pi^i\left(q_B^1,\ldots,q_B^n;q_L^1,\ldots,q_L^n\right) \equiv \sum_{h=B,L}(p_h-\tau)q_h^i = \sum_{h=B,L}\left(\rho_h(Q_h)-\upsilon_h D(Q)-c^h(Q_B,Q_L)-\tau\right)\cdot q_h^i \tag{12}$$

where τ denotes the uniform airport charge. The first-order conditions of (12) with respect to q_h^i are:

$$\frac{\partial \pi^i}{\partial q_h^i} = (\rho_h' - \upsilon_h D' - c_h^h)q_h^i + p_h - (\upsilon_k D' + c_h^k)q_k^i - \tau = 0 \tag{13}$$

for $i = 1, 2, \ldots n$ and $h, k = B, L$, $k \neq h$. By imposing symmetry and using superscript N to denote the second-stage Nash equilibrium, (13) can be rewritten as:

$$p_h^N + \frac{1}{n}(\rho_h' Q_h^N - \Gamma^N - \phi_h^N) - \tau = 0, \quad \forall h \neq k. \tag{14}$$

[15] Related to the present analysis, some road bottleneck models have also indicated that the non-anonymous second-best flat tolls should differ across types as the marginal external costs vary across passenger types (e.g. Arnott and Kraus, 1998; Small and Verhoef, 2007). In addition, Knockaert et al. (2010) proposed differentiated coarse tolls in the context of road bottleneck. They "exogenously" attribute homogenous road users into two-type groups, facing two different coarse toll levels and tolling periods. One similarity of Knockaert et al.'s setting to ours is that passengers of different types can be identified and hence forced to pay the toll associated to their types. However, unlike the differentiated coarse tolls in Knockaert et al., the pricing scheme presented in Proposition 2 does not have the tolled and untolled periods and hence it maintains the same arrival rates as in the no-toll equilibrium but reduces queue by affecting demand only. In other words, Proposition 2 is more related to the so-called "public time-invariant toll" (TI toll) discussed by van den Berg and Verhoef (2011); but in van den Berg and Verhoef's paper, this TI toll is uniform across user types.

[16] An airport might price discriminate passengers based on the value of air tickets. For example, the airport can charge at a fixed proportion of the air ticket price. However, this pricing scheme may not achieve first-best either and in many cases it may cause more distortions in congestion pricing than a uniform toll, because the first-best toll levied on business passengers – suppose they pay higher ticket price – is not necessarily higher than that levied on leisure passengers. For an interesting recent study on airport price discrimination, see Haskel et al. (2013).

Thus, at equilibrium as the two passenger types are charged independently, each airline will internalize exactly its own share of the "marginal external runway congestion" (MERC) cost. In addition, each airline will internalize its own share of the "marginal external terminal" (MET) costs imposed by type-h passengers, and exercise its market power. The equilibrium outcomes can be written as functions of τ.

To facilitate further discussion, we define the following second-order derivatives:

$$\pi^i_{hihi} \equiv \frac{\partial^2 \pi^i}{\partial {q^i_h}^2}, \; \pi^i_{hiki} \equiv \frac{\partial^2 \pi^i}{\partial q^i_h \partial q^i_k}, \quad \text{and} \quad \pi^i_{hikj} \equiv \frac{\partial^2 \pi^i}{\partial q^i_h \partial q^j_k}.$$

To guarantee that the first-order conditions produce the local maximum, we assume that at equilibrium the Hessian matrix of $\pi^i$is negative definite. That is,

$$\pi^i_{hihi} = \rho''_h \frac{Q_h}{n} + 2(\rho'_h - v_h D' - c^h_h) - (v_h Q_h + v_k Q_k)\frac{D''}{n} < 0, \text{ and } \pi^i_{BiBi}\pi^i_{LiLi} - \pi^i_{BiLi}\pi^i_{LiBi} > 0, \tag{15}$$

where $\pi^i_{hiki} = \pi^i_{kihi} = -(v_h D' + c^h_k + v_k D' + c^k_h) - (v_h Q_h + v_k Q_k)D''/n < 0$. The first condition in (15) automatically holds for any $n \geq 2$, as the assumption$\rho''_h Q_h + 2\rho'_h < 0$ (indicated in Section 2) implies that $(\rho''_h Q_h/n) + \rho'_h \leq 0$.

The existence of a unique (local) Nash equilibrium requires that the stability condition is satisfied as well. To satisfy the stability condition, we further assume that the maximum absolute eigenvalue of the $\partial q^R_i/\partial q_j$ matrix – that is, the matrix representing the impact of changes in airline j's decision on airline i's best-response output levels – is less than $1/(n - 1)$. Following the approach in Zhang and Zhang (1996), this assumption leads to Lemma 1:

Lemma 1. *Given that the Hessian matrix of π^i is negative definite at the Nash equilibrium, then: (i) the sufficient condition of local stability is that the maximum absolute eigenvalue of $\partial q^R_i/\partial q_j$ is less than $1/(n - 1)$; and (ii) this sufficient condition leads to*

$$\Delta \equiv \begin{vmatrix} \pi^i_{BiBi} + (n-1)\pi^i_{BiBj} & \pi^i_{BiLi} + (n-1)\pi^i_{BiLj} \\ \pi^i_{LiBi} + (n-1)\pi^i_{LiBj} & \pi^i_{LiLi} + (n-1)\pi^i_{LiLj} \end{vmatrix} > 0.$$

Proof. See Appendix B. □

The impact of airport charges on the equilibrium traffic volumes can be obtained by differentiating both sides of (13) with respect to τ, imposing symmetry and using Cramer's rule:

$$\frac{\partial q^{iN}_L}{\partial \tau} = \frac{1}{\Delta}\begin{vmatrix} \pi^i_{BiBi} + (n-1)\pi^i_{BiBj} & 1 \\ \pi^i_{LiBi} + (n-1)\pi^i_{LiBj} & 1 \end{vmatrix} = \frac{1}{\Delta}\left(\pi^i_{BiBi} - \pi^i_{LiBi} + (n-1)(\pi^i_{BiBj} - \pi^i_{LiBj})\right),$$

where $\pi^i_{BiBi} - \pi^i_{LiBi} = \rho''_B Q_B/n + 2\rho'_B - (v_B - v_L)D' - (c^B_B - c^B_L) - (c^B_B - c^L_B) < 0$ for any $n \geq 1$. When there is a monopoly airline ($n = 1$) it is straightforward to show, following Lemma 1, that $\partial q^{iN}_L/\partial \tau = (\pi^i_{BiBi} - \pi^i_{LiBi})/\Delta < 0$. When $n \geq 2$, $\partial q^{iN}_L/\partial \tau < 0$ still holds, because

$$\pi^i_{BiBj} - \pi^i_{LiBj} = \rho''_B \frac{Q_B}{n} + \rho'_B - (v_B - v_L)D' - (c^B_B - c^L_B) < 0.$$

Thus, $\partial Q^N_L/\partial \tau = n(\partial q^{iN}_L/\partial \tau) < 0$.

Regarding the impact of airport charge on (equilibrium) business passenger volume, Proposition 1 suggests that the sign of $\partial Q^N_B/\partial \tau = n(\partial q^{iN}_B/\partial \tau)$ is in general ambiguous. The above analysis leads to the following proposition:

Proposition 3. *When both types of passengers are levied a uniform airport charge, an increase in the airport charge τ will reduce the number of leisure passengers and the total number of passengers, but may reduce or increase the number of business passengers, depending not only on time values v_h but also on relative schedule-delay costs β_h (for $h = B, L$).*

Proof. The impact of τ on leisure passenger volume has been proved above, while the rest is proved in Appendix C. □

5.2. Optimal uniform airport charge

The optimal (uniform) airport charge, denoted by $\tau{**}$, can be obtained by taking the first-order condition of the welfare function:

$$\frac{dW}{d\tau} = \frac{\partial W}{\partial Q_B}\frac{\partial Q^N_B}{\partial \tau} + \frac{\partial W}{\partial Q_L}\frac{\partial Q^N_L}{\partial \tau} = \left(p^N_B - \Gamma^N - \phi^N_B\right)\frac{\partial Q^N_B}{\partial \tau} + \left(p^N_L - \Gamma^N - \phi^N_L\right)\frac{\partial Q^N_L}{\partial \tau} = 0. \tag{16}$$

From (14), we know that

$$p^N_h = -\frac{1}{n}(\rho'_h Q^N_h - \Gamma^N - \phi^N_h) + \tau, \forall h \neq k.$$

Table 1
The ranges of Φ^N.

	$\beta_L < \beta_B$	$\beta_L > \beta_B$
$\partial Q_B^N/\partial\tau > 0$	$\Phi^N < \phi_L^N < \phi_B^N$ ($\Phi^N < 0$ is possible)	$\Phi^N > \phi_L^N > \phi_B^N$
$\partial Q_B^N/\partial\tau < 0$	$\phi_L^N < \Phi^N < \phi_B^N$	$\phi_L^N > \Phi^N > \phi_B^N$

By replacing p_B^N and p_L^N in (16) with the above expression and rearranging the equation, the optimal airport charge should satisfy the following pricing rule:

$$\tau^{**} = \left(1-\frac{1}{n}\right)\Gamma^N + \left(1-\frac{1}{n}\right)\Phi^N + \frac{1}{n}M^N, \tag{17}$$

where

$$\Phi^N \equiv \frac{\frac{\partial Q_B^N}{\partial\tau}\phi_B^N + \frac{\partial Q_L^N}{\partial\tau}\phi_L^N}{\frac{\partial Q^N}{\partial\tau}} \text{ and } M^N \equiv \frac{\frac{\partial Q_B^N}{\partial\tau}\rho'_B Q_B^N + \frac{\partial Q_L^N}{\partial\tau}\rho'_L Q_L^N}{\frac{\partial Q^N}{\partial\tau}}.$$

The optimal airport charge has three components. The first term on the RHS of (17) is the part of the MERC cost which is not internalized by airlines. This component was identified by Brueckner (2002) and others. The second term is new: it is the weighted sum of MET costs which are not internalized by airlines. The weighted average structure is common for optimal undifferentiated airport charges (see e.g. D'Alfonso et al., 2013). Note that runway congestion never leads to such a weighted structure, as both passenger types suffer the same amount of delay on the runway but not at the terminal. The last component is the adjustment on market power identified by Pels and Verhoef (2004) and others. The last two components are both weighted terms, with the weights being determined by the marginal impacts of airport charge on traffic volumes (business and leisure passengers). Because $\partial Q_B^N/\partial\tau$, the marginal impact of airport charge on business passenger volume, can be either positive or negative while $\partial Q_L^N/\partial\tau$, is always negative, the signs of these weighted terms are ambiguous in general. For example, the market-power adjustment is negative (a downward correction) when $\partial Q_B^N/\partial\tau$ is negative, but it can be positive when $\partial Q_B^N/\partial\tau$ is positive, leading to a markup on airport charge so as to protect business travelers from the overcrowded runway and terminal by squeezing out leisure passengers.

A closer look at the second term reveals that its magnitude and sign rely on the interaction between $\partial Q_h^N/\partial\tau$ and the relative schedule delay costs. Rewrite Φ^N as,

$$\Phi^N = \phi_L^N + (\phi_B^N - \phi_L^N)\frac{\frac{\partial Q_B^N}{\partial\tau}}{\frac{\partial Q^N}{\partial\tau}} = \phi_B^N - (\phi_B^N - \phi_L^N)\frac{\frac{\partial Q_L^N}{\partial\tau}}{\frac{\partial Q^N}{\partial\tau}}. \tag{18}$$

Notice first that the sign of $\phi_B^N - \phi_L^N$ is the same as the sign of $\beta_B - \beta_L$. The second equality of (18) thus suggests that $\Phi^N > \phi_B^N$ if and only if $\beta_L > \beta_B$. However, according to the first equality of (18), the relationship between Φ^N and ϕ_L^N depends also on the sign of $\partial Q_B^N/\partial\tau$. In particular, if $\partial Q_B^N/\partial\tau > 0$, then $\Phi^N > \phi_L^N$ if and only if $\beta_L > \beta_B$. On the other hand, when $\partial Q_B^N/\partial\tau < 0$, the opposite will hold. We summarize the results of this analysis in Table 1.

Ideally, each passenger should be charged at the uninternalized cost she imposes on the others. This is exactly how passengers are charged for runway congestion, the first component of the optimal airport charge in (17). However, this is not the case for the terminal charge, as passengers of different types cannot be distinguished for charges at the terminal although the impacts they impose on others differ across passenger types. As a result, a particular passenger may pay more, or less, than the amount of externality she imposes on the others. In the former situation, the passenger is "over charged" whilst in the latter situation, she is "under charged." The details are stated in Proposition 4:

Proposition 4. *When $\partial Q_B^N/\partial\tau < 0$, the passenger type with the higher relative schedule delay cost will be over-charged relative to the uninternalized terminal cost imposed on other passengers, while the other type under-charged. When $\partial Q_B^N/\partial\tau\, 0$ and $\beta_B > \beta_L$, all passengers will be under-charged and, if the time values differ dramatically between the two passenger types, they may even be subsidized. When $\partial Q_B^N/\partial\tau > 0$ and $\beta_B < \beta_L$, all passengers will pay more than the uninternalized terminal cost they bring about from travelling.*

When $\beta_B = \beta_L$, it can be shown that $\Phi^N = \phi_B^N = \phi_L^N$. That is, regardless if there is one additional business or leisure passenger, the MET cost of this extra passenger is the same. It then follows that the first and second terms on the RHS of Eq. (17) can be combined into one single uninternalized airport congestion cost (terminal + runway), which has been derived and widely known in literature. The conventional modeling approach which does not separate passenger runway and terminal costs will, therefore, lose important features, unless the relative schedule-delay costs are the same across passenger types.

6. Airport concessions

The welfare-maximizing airport's objective function then becomes:

$$W \equiv S_a + S_c. \tag{19}$$

where S_c is the concession surplus.[17] We first specify S_c. During the dwell time at terminal, passengers may consume concession goods. Following D'Alfonso et al. (2013) we assume that the utility of purchasing concession goods, u, for a type-h passenger who arrives at the airport at time t, follows cumulative distribution, $G_h(u; T_d(t))$,[18] on the domain $[0, \bar{u}]$, which satisfies the "first-order stochastic dominance" (FOSD) property with respect to the amount of dwell time. As the passenger will purchase only when the utility of purchasing concession goods exceeds the price, denoted p_c, the probability of purchasing concessions is equal to

$$\overline{G}_h(p_c; T_d(t)) = 1 - G_h(p_c; T_d(t)).$$

The FOSD property implies that

$$\overline{G}_h(p_c; T_d) \geq \overline{G}_h(p_c; \tilde{T}_d), \quad \forall T_d > \tilde{T}_d$$

which shows that the longer the dwell time, the more likely the passenger purchases concessions.

From the discussion in Section 3, we know that for each unit of time, s passengers will start their dwell time. Thus, the concession demand for each level of dwell time can be written as

$$x_h(p_c, T_d) = s \cdot \overline{G}_h(p_c; T_d).$$

Then, for any $\beta_h > \beta_k$, the surplus from purchasing concession goods can be written as

$$S_c = \int_{T_d(t_m)}^{T_d(\underline{t})} \int_{p_c}^{\bar{u}} s\overline{G}_h(x; z)dxdz + \int_{T_d(\bar{t})}^{T_d(t_m)} \int_{p_c}^{\bar{u}} s\overline{G}_k(x; z)dxdz + (p_c - c_c)\left(\int_{T_d(t_m)}^{T_d(\underline{t})} s\overline{G}_h(p_c; z)dz + \int_{T_d(\bar{t})}^{T_d(t_m)} s\overline{G}_k(p_c; z)dz\right) \quad (20)$$

where c_c is the (constant) cost of one unit of concession goods, $T_d(\underline{t}) = \hat{t} - \underline{t} = Q/s$, $T_d(t_m) = \hat{t} - (\hat{t} - Q_h/s) = Q_h/s$, and $T_d(\bar{t}) = \hat{t} - \hat{t} = 0$. Recall $T_d(t)$ denotes the dwell time of a passenger arriving at the airport at time t. Following Zhang and Zhang (1997) we consider "myopic" passengers (using the terminology of Flores-Fillol et al., 2014). That is, they do not take concession consumption into account when booking a flight and as a result, concession activities do not affect the travel demand.[19]

6.1. First-best outcome

Similar to Section 4, the first-best outcomes with concessions can be derived by taking the first-order conditions of (19) with respect to Q_B, Q_L and p_c. Note that concession price p_c only affects concession surplus; and it is easy to show that

$$\frac{\partial W}{\partial p_c} = \frac{\partial S_c}{\partial p_c} = -s(p_c - c_c)\left(\int_{T_d(t_m)}^{T_d(\underline{t})} g_h(p_c; z)dz + \int_{T_d(\bar{t})}^{T_d(t_m)} g_k(p_c; z)dz\right). \quad (21)$$

The term in the second pair of brackets of Eq. (21) is always positive and hence to maximize concession surplus, the optimal concession price should always be set at the cost, a result that is consistent to Yang and Zhang (2011) and D'Alfonso et al. (2013). Therefore, to simplify the analysis we assume $p_c = c_c$ for the remainder of this section. Consequently, (20) can be rewritten as

$$S_c = \int_{\frac{Q_h}{s}}^{\frac{Q}{s}} \int_{c_c}^{\bar{u}} s\overline{G}_h(x; z)dxdz + \int_0^{\frac{Q_h}{s}} \int_{c_c}^{\bar{u}} s\overline{G}_k(x; z)dxdz.$$

The first-best outcomes can then be derived by taking the first-order conditions of (19) with respect to Q_B and Q_L as,

$$\frac{\partial W}{\partial Q_B} = p_B^* - \Gamma^* - \phi_B^* + \frac{\partial S_c^*}{\partial Q_B} = 0, \quad (22)$$

$$\frac{\partial W}{\partial Q_L} = p_L^* - \Gamma^* - \phi_L^* + \frac{\partial S_c^*}{\partial Q_L} = 0, \quad (23)$$

where

$$\frac{\partial S_c}{\partial Q_h} = \begin{cases} \int_{c_c}^{\bar{u}} \left[\overline{G}_k\left(x; \frac{Q}{s}\right) - \overline{G}_k\left(x; \frac{Q_h}{s}\right) + \overline{G}_h\left(x; \frac{Q_h}{s}\right)\right]dx & \text{if } \beta_h > \beta_k \\ \int_{c_c}^{\bar{u}} \overline{G}_h\left(x; \frac{Q}{s}\right)dx & \text{if } \beta_h < \beta_k \end{cases}.$$

[17] Some papers assume that a fixed δ ($0 \leq \delta \leq 1$) proportion of passengers' concession surplus is contributed by the air travel activity (e.g. Zhang and Czerny, 2012; D'Alfonso et al., 2013). This is because not all the concession consumption is uniquely associated with traveling. Passengers will consume food and beverage regardless if they come to the airport or not, but other concession activities could generate extra surplus. For example, passengers may shop simply to kill time (while waiting for flights) or mitigate the feelings of insecurity and uncomfortableness due to traveling (Dube and Menon, 2000). Duty-free shops and international brands sometimes offer product lines exclusively for travelers at the airport (Rowe, 1998). Given that the level of δ does not qualitatively change our results, we simply set it to unity. That is, all concession surplus is counted as extra contribution to social welfare.

[18] Note that dwell time, $T_d(t)$, is in effect the early schedule delay time and hence it excludes the time spent waiting in the queue. That is, we assume that consumption of concession goods only occurs after queuing for pre-departure procedures.

[19] Since certain concession services, namely car rentals and car parking, may affect travel decisions, some recent studies started to consider non-myopic passengers who take into account certain concession activities when making travel decisions. Please refer to Section 7.3 for a more detailed discussion.

Table 2
The ranges of Σ^N.

	$T_d(t_m) < \hat{T}_{d1}$ or $T_d(t_m) > \hat{T}_{d2}$	$\hat{T}_{d1} < T_d(t_m) < \hat{T}_{d2}$
$\partial Q_B^N/\partial\tau > 0$	$\Sigma^N > -S_{cL}^N > -S_{cB}^N$ $\Sigma^N > 0$ is possible	$\Sigma^N < -S_{cL}^N < -S_{cB}^N$
$\partial Q_B^N/\partial\tau < 0$	$-S_{cL}^N > \Sigma^N > -S_{cB}^N$	$-S_{cL}^N < \Sigma^N < -S_{cB}^N$

The inclusion of concession surplus adds another layer of complexity to the analysis. The FOSD property of $G_h(\cdot\,;T_d)$ suggests that the marginal change in concession surplus should always be positive regardless the difference between β_B and β_L, because $Q/s > Q_h/s$ and hence $\overline{G}_k(x;Q/s) \geq \overline{G}_k(x;Q_h/s)$. Thus, taking concession surplus into consideration will lead to a markdown on the first-best airfares. However, the impact on the difference between p_B^* and p_L^* is less straightforward. Eq. (11) now becomes

$$p_B^* - p_L^* = \begin{cases} \left(c_B^B - c_L^B\right)Q_B^* + \int_c^{\overline{u}} \overline{G}_L\left(x;\frac{Q_B^*}{s}\right) - \overline{G}_B\left(x;\frac{Q_B^*}{s}\right)dx \text{ if } \beta_B > \beta_L \\ -\left(c_L^L - c_B^L\right)Q_L^* + \int_c^{\overline{u}} \overline{G}_L\left(x;\frac{Q_L^*}{s}\right) - \overline{G}_B\left(x;\frac{Q_L^*}{s}\right)dx \text{ if } \beta_B < \beta_L \end{cases}. \tag{24}$$

The difference between p_B^* and p_L^* depends not only on the relative schedule delay costs but also on the distributions of concession goods utility, $G_h(\cdot\,;T_d)$ for $h = B, L$. Torres et al. (2005) empirically compare the average concession goods expenditure of leisure passengers and business passengers. They find that there exist two dwell time thresholds, $\hat{T}_{d1}$ and $\hat{T}_{d2}$ (where $\hat{T}_{d1} < \hat{T}_{d2}$), such that business passengers on average spend more on concession goods than leisure passengers if the dwell time is lower than $\hat{T}_{d1}$ or higher than $\hat{T}_{d2}$, while leisure passengers on average spend more if the dwell time is between $\hat{T}_{d1}$ and $\hat{T}_{d2}$. Accordingly we further assume that for any level of u, $\overline{G}_B(u;T_d) \geq \overline{G}_L(u;T_d)$ if $T_d \leq \hat{T}_{d1}$ or $T_d \geq \hat{T}_{d2}$; and $\overline{G}_B(u;T_d) < \overline{G}_L(u;T_d)$ otherwise. This assumption, together with Eq. (24), lead to Proposition 5:

Proposition 5. *With concessions: (i) When $\beta_B > \beta_L$, $p_B^* > p_L^*$ if $Q_B^*/s \leq \hat{T}_{d2}$, while $p_B^* < p_L^*$ may occur otherwise; (ii) When $\beta_B < \beta_L$, $p_B^* < p_L^*$ if $Q_L^*/s \leq \hat{T}_{d1}$ or $Q_L^*/s \leq \hat{T}_{d2}$, while $p_B^* > p_L^*$ may occur otherwise.*

Note that Q_h^*/s is the dwell time for passengers arriving at t_m at first-best, given that the relative schedule-delay cost of type-h passengers is higher. Thus, the difference of first-best airfares does not depend on the dwell time of every passenger; rather, it depends only on the passengers who arrive at the point of time that sets the two passenger types apart. A comparison with Proposition 2 shows that the inclusion of concession surplus may change the passenger type which should be charged higher at first-best to protect the other type. This is because the inclusion of concession surplus may create a trade-off between the business and leisure passengers. For example, when $\beta_B > \beta_L$, serving one more leisure passenger incurs a lower MET cost than serving one more business passenger. On the other hand, the increase in concession surplus due to one additional business passenger may be higher if $Q_B^*/s < \hat{T}_{d1}$ or $Q_B^*/s > \hat{T}_{d2}$. Given such a trade-off, at first-best business passengers may or may not be charged higher than leisure passengers once concession surplus is in concern.

6.2. Optimal airport charge

The optimal (uniform) airport charge is again derived by taking the first-order condition of (19) with respect to airport charge τ. The inclusion of concession surplus results in an extra term, Σ^N, to the optimal airport pricing rule:

$$\tau^{**} = \left(1-\frac{1}{n}\right)\Gamma^N + \left(1-\frac{1}{n}\right)\Phi^N + \frac{1}{n}M^N + \Sigma^N, \tag{25}$$

where

$$\Sigma^N \equiv -\frac{\frac{\partial Q_B^N}{\partial\tau}S_{cB}^N + \frac{\partial Q_L^N}{\partial\tau}S_{cL}^N}{\frac{\partial Q^N}{\partial\tau}} = -S_{cL}^N - \left(S_{cB}^N - S_{cL}^N\right)\frac{\frac{\partial Q_B^N}{\partial\tau}}{\frac{\partial Q^N}{\partial\tau}} = -S_{cB}^N + \left(S_{cB}^N - S_{cL}^N\right)\frac{\frac{\partial Q_L^N}{\partial\tau}}{\frac{\partial Q^N}{\partial\tau}},$$

and $S_{ch}^N \equiv \partial S_c^N/\partial Q_h > 0 \forall h \in \{B, L\}$.

The sign of Σ^N relies on the interaction between $\partial Q_h^N/\partial\tau$ and the cut-off dwell times, $\hat{T}_{d1}$ and $\hat{T}_{d2}$. The sign of $S_{cB}^N - S_{cL}^N$ depends on the dwell time for passengers arriving at the airport at t_m, i.e. $T_d(t_m)$. In particular, if $T_d(t_m) \leq \hat{T}_{d1}$ or $T_d(t_m) \geq \hat{T}_{d2}$, $S_{cB}^N - S_{cL}^N \geq 0$; otherwise, $S_{cB}^N - S_{cL}^N < 0$. Then, we can show that the inclusion of concession surplus does not always lead to a markdown in airport charge (Table 2).

Given that $\partial Q_B^N/\partial\tau > 0$, when $T_d(t_m)$ is below $\hat{T}_{d1}$ or above $\hat{T}_{d2}$, Σ^N may become positive, leading to an upward correction on the airport charge. In other cases, taking concession surplus into account will impose a downward pressure on the airport charge.

Proposition 6. *When an increase in airport charge raises the volume of business passengers and the dwell time of the passengers arriving at t_m is very short or very long, the presence of concession activities may raise the optimal airport charge.*

Proposition 6 is consistent with the findings in D'Alfonso et al. (2013), but the underlying reasoning differs. Since terminal and runway congestions are not separated in their paper, the direction of the correction due to concession surplus is affected by

Table 3
Impacts of an increase in terminal service rate s on passenger volumes for any $\beta_h \geq \beta_k$.

Cases	Range of β_h/β_k	Impact (i) on Q_h	Impact (i) on Q_k	Impact (ii) on Q_h	Impact (ii) on Q_k	Impact (iii) on Q_h	Impact (iii) on Q_k
First-best	$\beta_h/\beta_k > (v_B+v_L)^2/(4v_Bv_L)$	+	+	+	−	+	+
	$\beta_h/\beta_k = (v_B+v_L)^2/(4v_Bv_L)$	+	+	0	−	0	0
	$(v_B+v_L)^2/(4v_Bv_L) > \beta_h/\beta_k > 1$	+	+	−	−	−	−
	$\beta_h/\beta_k = 1$	+	+	−	−	−	−
Fixed airfares	$\beta_h/\beta_k > 1$	+	+	+	−	+	+
	$\beta_h/\beta_k = 1$	+	+	0	0	+	+

Note: + means positive impact; 0 means no impact; − means negative impact

the level of congestion delay which is determined by total traffic volume. In the present setting, however, as $T_d(t_m) = Q_h/s$ for any $\beta_h > \beta_k$, the direction of this correction is related solely to the traffic volume of the passengers who have a higher relative early schedule-delay cost (that is, it is not related to the volume of the other passenger type).[20]

More specifically, in D'Alfonso et al. (2013) the impact of concession surplus on airport charge consists of two components: (i) a downward correction on congestion toll to internalize the positive externality of congestion on concessions, and (ii) a correction equal to the expected concession surplus weighted by traffic volume changes across passenger types. A closer look at the last term of (25) reveals two components as well. For example, when $\beta_B > \beta_L$, S_{cB}^N equals to the expected concession surplus of one additional business passenger, $\int_{C_c}^{\bar{u}} \overline{G}_B(x; Q_B/s)dx$, plus the increment in concession surplus of leisure passengers due to increased leisure passenger dwell time by adding one more business passenger into the system, $\int_{C_c}^{\bar{u}} [\overline{G}_L(x; Q_B/s) - \overline{G}_L(x; Q_B/s)]dx$. As business passengers arrive later than leisure passengers, one more business passenger contributes to its own expected surplus from concession purchase and at the same time pushes the arrival time of each leisure passenger a little bit earlier, thereby leading to greater dwell time of all the leisure passengers. However, S_{cL}^N is simply equal to the expected concession surplus of one additional leisure passenger, $\int_{C_c}^{\bar{u}} \overline{G}_L(x; Q/s)dx$. Thus, when $\beta_B > \beta_L$, the last term of (25) can be rewritten as:

$$-\frac{\frac{\partial Q_B^N}{\partial \tau}\int_{C_c}^{\bar{u}}\left[\overline{G}_L\left(x;\frac{Q^N}{s}\right) - \overline{G}_L\left(x;\frac{Q_B^N}{s}\right)\right]dx}{\frac{\partial Q^N}{\partial \tau}} - \frac{\frac{\partial Q_B^N}{\partial \tau}\int_{C_c}^{\bar{u}}\overline{G}_B\left(x;\frac{Q_B^N}{s}\right)dx + \frac{\partial Q_L^N}{\partial \tau}\int_{C_c}^{\bar{u}}\overline{G}_L\left(x;\frac{Q^N}{s}\right)dx}{\frac{\partial Q^N}{\partial \tau}}. \tag{26}$$

The first term of (26) corresponds to the first component in D'Alfonso et al. (2013) mentioned above, indicating the correction due to changes in dwell time. Unlike their findings however, this term is not a clear-cut downward adjustment. Instead, when $\partial Q_B^N/\partial \tau$ is positive, this term becomes a markup.[21] That is, to increase the dwell time of leisure passengers, the airport has an incentive to charge more (less, respectively) and attract more business passengers when $\partial Q_B^N/\partial \tau$ is positive (negative, respectively). The second term of (26) corresponds to the second component found by D'Alfonso et al. (2013), indicating the correction due to changes of expected concession surplus as the composition of the two passenger types changes, which can be either positive or negative.

7. Discussions and policy implications

7.1. Impact of terminal investment

The difference in terminal and runway congestions also has policy implications for terminal capacity investment. As the interaction between runway and terminal congestions is complicated, a comprehensive study on the investment-related issues deserves a separate paper. However, as the first step to such a future study, we examine the impact of terminal expansion, i.e. an increase in terminal service rate s, on passenger volumes in two cases: (a) when airfares are fixed; and (b) when airfares are determined at the first best.[22] We abstract away the concession surplus and further assume $D'' = 0$.[23] As shown in Table 3, in both cases, an increase in terminal service rate s will affect passenger volumes via three channels (please refer to Appendix D for details): (i) a direct increase in passenger number which relates to passengers' sensitivity to price; (ii) a positive or negative impact due to the interaction between the terminal and runway congestions; and (iii) a positive or negative impact due to the interaction between passenger types at terminal alone.

[20] The result that the presence of concessions may lead to a markdown in airport (aeronautical) charge is also consistent with Czerny (2006) who showed that concession services raise the passenger demand and hence, the aeronautical charge can be increased with concessions relative to the situation without concessions. Note that Czerny (2006) examines cases where certain concession activities, such as car rental and car parking, may also affect passengers' travel decision, while both D'Alfonso et al. (2013) and the present paper assume a unidirectional relationship between aeronautical services and concession services.

[21] This term will be a markdown only when $\beta_B < \beta_L$.

[22] The effect on equilibrium passenger volumes determined by airlines is far more complicated than the first-best case, but the first-best case is enough to serve the purpose to illustrate the special features of our model setting.

[23] Inclusion of D'' simply adds another higher-order term which imposes exactly opposite impacts on the two passenger types.

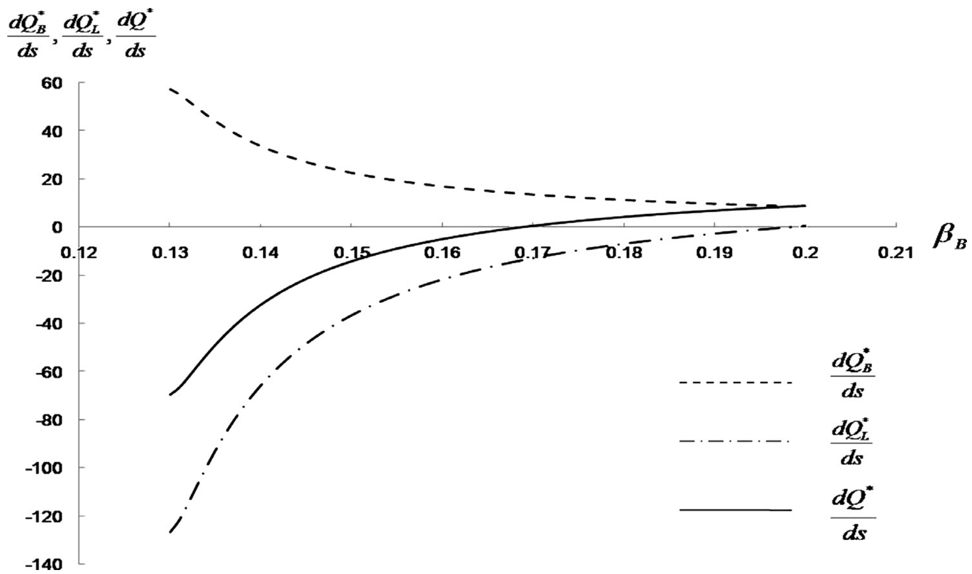

Fig. 3. Numerical example for the impact of s on first-best passenger numbers ($s = 1$, $a_B = 27$, $a_L = 10$, $b_B = 0.4$, $b_L = 0.1$, $v_B = 7$, $v_L = 1$, $\beta_L = 0.1$ and $\beta_B \in [0.13, 0.2]$).

When airfares are fixed, if runway is not congested ($D' = 0$) or if terminal congestion can be modeled the same way as runway congestion ($\beta_B = \beta_L$), impact (ii) will disappear and the decrease in the number of type k passengers would never occur. This indicates the importance of the subtle interaction between terminal and runway congestions. An increase in terminal capacity leads to a direct reduction in terminal costs for both passenger types and hence an increase in the number of both types. The increased passenger numbers interact with runway congestion in two ways. First, an increase in type h passengers raises runway congestion suffered by type k passengers, which causes a downward pressure on the number of type k. Second, an increase in type k passengers also suppresses the number of type h to reduce runway congestion which in turn induces more type k passengers. When β_h is substantially larger than β_k, the negative impact triggered by type h's terminal cost savings dominates the positive impact triggered by type k's terminal cost savings. The above discussion leads to Proposition 7.

Proposition 7. *For given airfares, for any $\beta_h > \beta_k$, as the terminal service rate (s) increases owing to terminal investment, then: (i) the total number of passengers will increase; (ii) the number of type h passengers will increase; but (iii) the number of type k passengers may decrease and the reduction is more likely to occur if the difference in β_h and β_k gets larger and the marginal congestion on runway is more severe.*

Proof. See Appendix D. □

When airfares are set at the first best, if the β_h/β_k ratio is below a certain threshold, impacts (ii) and (iii) can be negative for both passenger types which never occurs in the fixed-fare case. Note that when $D' = 0$ and $\beta_h/\beta_k = 1$, our model is reduced to the case of pure runway congestion and impact (ii) disappears. Then, at the first best, impact (iii) is always negative while it can be positive when $\beta_h/\beta_k$is far above 1. This reveals one difference in terminal and runway congestions: runway expansion is more likely to impose downward pressure on passenger volume while terminal expansion has an opposite effect if passengers differ a lot in relative schedule delay costs.

One counter-intuitive observation from Table 3 is as follows. When airfares are set at the first best and the relative schedule delay costs do not differ a lot between passenger types, an increase in terminal service rate is likely to impose a downward pressure on first-best passenger numbers. Although it is straightforward to prove that at least one of the passenger types' number will increase as s increases, it is still possible that the total passenger number at the first best may decrease. This phenomenon can occur even when the runway congestion is non-existent and gets enhanced when the runway congestion is severe. Fig. 3 illustrates one numerical constellation of such phenomenon with linear demands and no runway congestion, i.e. $\rho_h(Q_h) = a_h - b_h Q_h$ and $D' = 0$, $h \in \{B, L\}$. When β_B is below 0.17 which is very close to β_L, an increase in terminal service rate induces more business passengers but fewer leisure passengers at the first best, and the total number of passengers also decreases. However, as β_B increases above 0.17, the positive impact on business passengers gradually dominates the negative impact on leisure passengers and when β_B approaches to 0.2 both types of passengers increase.

The above phenomenon can be explained in the following way. An increase in terminal capacity will reduce terminal costs for both passenger types and the terminal cost saving for business passengers is larger than leisure passengers. As a result, the social planner has more incentives to increase business passengers than leisure passengers. Since β's measure the relative schedule delay costs (compared to queuing costs), when β_B is close to β_L, the reduction in schedule delay cost for business passengers is not large enough to attract an optimal number of business passengers. Thus, the social planner will have to reduce the amount of leisure passengers to reduce the queuing costs further for business passengers. As a result, at the first best the number of business passengers will increase due to terminal capacity expansion at the cost of the leisure passenger number; and it is possible that the total passenger number will decrease even if the terminal capacity increases. When β_B is far above β_L, the reduction in schedule delay cost for business passengers is already large enough to attract the optimal number of business passengers, leaving some space for the increase in leisure passenger number. Thus, the numbers of both passenger types will increase.

7.2. Benchmark case with time-varying terminal fine toll

In the social optimum, the terminal queue is eliminated by a time-varying terminal fine toll. That is, each passenger pays a toll according to her time of arrival at the terminal and the amount of the toll reflects the corresponding queuing time savings and hence decreases in the schedule delay. Consequently, the total terminal cost (*TC*) borne by the society is the same as total schedule delay costs (*TSDC*). Following Arnott et al.'s (1994) rationale, at social optimum the passenger type with higher schedule delay cost will arrive at the airport later than the other type. Thus, the arrival sequence depends on $v_h\beta_h$. In particular, if $v_B\beta_B > v_L\beta_L$, business passengers arrive closer to the gate closing time, $\hat{t}$; otherwise, leisure passengers arrive after business passengers. We use $\tilde{c}^B$ and $\tilde{c}^L$to denote terminal costs, i.e. schedule delay cost plus time-varying toll, incurred by individual business and leisure passengers respectively.

As shown in Appendix E, after levying the time-varying fine toll, the marginal cost to the society after adding one particular passenger is the same as the cost borne by this additional passenger. As a result, there is no need to make any correction on the first-best airfares as terminal side externality has been fully internalized by the time-varying fine toll. In a word, after imposing time-varying toll at the terminal, the ϕ_h^* term in Eqs. (9) and (10) will vanish and the differentiation in first-best airfares indicated in Proposition 2 will disappear accordingly when concession surplus is not a concern. Nevertheless, different from Czerny and Zhang (2011), in our setting business passengers do not necessarily incur a higher generalized cost (full price) than leisure passengers under this time-varying toll arrangement, because although business passengers incur higher runway cost ($v_BD > v_LD$), they may incur a much lower terminal costs, i.e. $\tilde{c}^B$ can be smaller than $\tilde{c}^L$ when $v_B\beta_B < v_L\beta_L$ holds.

However, when it comes to the behavior of airlines, the equilibrium fare does not take into account the correction made by the terminal fine toll and an airline will continue to charge the part of MET proportional to its market share. That is, $p_h^N = \tau - (\rho_h' Q_h^N - \Gamma^N - \tilde{\phi}_h^N)/n$, where $\tilde{\phi}_h \equiv Q_h\tilde{c}_h^h + Q_k\tilde{c}_h^k = \tilde{c}^h$, $\forall h = B, L$. Thus, airlines over-internalize the terminal externality unless $n \to \infty$ and the second-best airport charge will involve an adjustment to partially correct airlines' over-internalization of terminal external costs (see Appendix E for details). This adjustment tends to be a markdown unless the marginal effect of runway congestion is very strong and business passengers' schedule delay costs are much larger than leisure passengers. In the latter case, a markup is needed to further restrict the number of leisure passengers and induce more business passengers so as to compensate the over-internationalization caused by airlines.

7.3. Travel decisions and concession activities

In Section 6 the model setting for concession activities follows the "conventional" approach which assumes that passengers do not take benefits obtained from concession services into account when making their travel decisions. Several recent papers examine cases where certain concession activities, such as car rental and car parking, may also affect passengers' travel decision (e.g., Czerny, 2006, 2013; Czerny and Lindsey, 2014; Flores-Fillol et al., 2014; Bracaglia et al., 2014; for a related empirical study, see Czerny et al., 2015). In those cases, concession activities as well as concession prices will affect aeronautical surplus. However, it is still unclear to what extent concession activities at the airside after security checks and passport controls can affect travel decisions. Information about products and non-aeronautical services available at the terminals is relatively limited. Passengers may have little knowledge about the prices of these concession activities when purchasing their air tickets, though they can check for car rental and parking prices before booking a flight. In other words, the airport tends to have a better knowledge on an average passenger's concession demand at the terminals than passengers do. Most likely, passengers would not reveal their airside concession demands until they are physically at the terminal. Given that our focus is on the concession activities carried out during the dwell time after security checks, it is, we believe, more reasonable to assume that concession activities do not affect travel decisions in this paper.

Nevertheless, it remains interesting to see how our model can be modified if concession surplus does affect travel decisions. If passengers take their concession surplus into account when booking a flight, concession surplus will, as expected, affect airport arrival decision. In particular, an individual passenger will choose an arrival time, t, which minimizes the terminal-related net cost, i.e. the sum of terminal cost minus concession surplus:

$$g^h(t) \equiv v_h T(t) + \beta_h v_h \cdot (\hat{t} - t - T(t)) - \int_{p_c}^{\bar{u}} \bar{G}_h(x; \hat{t} - t - T(t))dx, \ \forall h \in \{B, L\}. \tag{27}$$

Then, assuming (27) is convex in t, at equilibrium we have:

$$T'(t) = \frac{\beta_h v_h - B_h(\hat{t} - t - T(t))}{v_h(1-\beta_h) + B_h(\hat{t} - t - T(t))} \text{ and } A'(t) = \frac{s}{1 - \left(\beta_h - B_h(\hat{t} - t - T(t))/v_h\right)},$$

where $B_h(T_d) = \int_{p_c}^{\bar{u}} (\partial \bar{G}_h(x; T_d)/\partial T_d)dx > 0$, following the FOSD property. Thus, the functional form of $\bar{G}_h(u; T_d)$ will affect the arrival pattern which will further affect the concession demand via change in T_d. If $B_h(T_d)$ is larger than the schedule delay costs, passengers prefer coming to the airport as early as possible and then there may not exist a pure strategy equilibrium. In a more likely case where $B_h(T_d)$ is less than the schedule delay costs, then some features of the current model will preserve. For example, the last arriving passenger incurs zero dwell time while the first arriving passenger incurs zero queuing time. However, the passengers' arrival pattern can be very complicated. Therefore, a more specific form of $\bar{G}_h(u; T_d)$ would be needed for a comprehensive study on the passengers' terminal arrival behavior, which deserves a separate paper in the future. Note that if there is an equilibrium at the terminal, then each type h passenger should incur the same amount of terminal-related net cost which is also a function of Q_B and Q_L, i.e. $g^h(Q_B, Q_L)$. Then, the first-best airfare will not have the $\partial S_c^*/\partial Q_h$ markdown as suggested in Eqs. (22) and (23), but the adjustment on concession activities won't disappear, since it will be embedded in the terminal-related term (i.e. the previous ϕ_B^*).

8. Concluding remarks

This paper has treated terminal congestion and runway congestion separately, and studied its implication for the design of optimal airport charges and terminal capacity investment. To capture the difference between these two types of congestion, we have adopted a deterministic bottleneck model for the terminal and a conventional congestion model for the runways. We found that the inclusion of terminal congestion leads to the first-best results that differ from the ones in the literature. In particular, the welfare-optimal uniform airfare *does not* yield the first-best outcome. The first-best fare charged to the business passengers is greater than the leisure passengers' fare if and only if the relative schedule-delay cost of business passengers is higher than that of leisure passengers. If concession surplus is taken into account, the comparison between the first-best fares charged to the two passenger types will depend further (in addition to their relative early schedule delay costs) on the dwell time of passengers who arrive at the airport at time t_m.

We have also examined the impact of an airport charge on equilibrium traffic levels and derived the optimal uniform airport charge. We showed that when both types of passengers are levied a uniform airport charge, an increase in airport charge will reduce both the number of leisure passengers and the total number of passengers, but may reduce or increase the number of business passengers. Furthermore, the structure of optimal charge suggests that in terms of the terminal charge, passengers will be under-charged or over-charged and the conditions for either to happen are identified. Finally, when comparing the airport pricing rule with the one found in literature, we found that if the volume of business passengers increases in airport charge, the airport will raise rather than reduce the airport charge in order to lengthen leisure passengers' dwell time and hence increase their chance of purchasing concession goods. Furthermore, we derived the optimal uniform airport charge in the benchmark case where besides this uniform airport charge, a time-varying fine toll is also levied to eliminate terminal queues. Then, in this benchmark case, although both passenger types should be charged a uniform airfare to only internalize the runway congestion externality, airlines have incentives to over-internalize the terminal externality. As a result, the airport in general needs to correct this airline behavior by lowering the airport charge unless the number of business passengers increases as the airport charge increases and the business passengers' schedule delay cost is far above the leisure passengers'.

As clearly shown in this paper, our treatment of terminal congestion and runway congestion in an integrated model gives rise to some new insights about optimal airport pricing. One advantage of this integrated approach is that it explicitly models passengers' behavior in the terminal and relates concession purchases with dwell time rather than just terminal congestion delay. Given the increasing importance of airport concession services, it is important to have a clearer picture of the behavior of different passenger groups in the terminal. The present paper offers a first step of this attempt. There are several potential avenues for future studies. For example, the paper assumes all passengers go through the same line. Some airports may provide dedicated express lines for frequent flyers or first-class and business passengers. To incorporate this, the present model can be modified by having two separate bottlenecks. Another instance would be to empirically test whether the predictions of passenger behavior fit the reality. The idea can also be extended to incorporate continuous passenger types and multiple preferred flight departure times.

Our model can also be extended to study the interaction between berth congestion and yard congestion in the seaport industry, with individual truck arrivals at the port before the gate "cut-off" time and various types of container cargo. Finally, our model setting has important policy implication for future terminal investment. Our analysis suggests that increasing processing speed at the terminal (i.e. increasing terminal capacity) may not always lead to an increase in passenger volume. When there are multiple types of passengers, to achieve first-best, reduction in the volume of one passenger type may have to exceed the increase in the volume of the other passenger type, leading to an overall drop in passenger numbers. This tends to be mitigated as the difference in the relative schedule delay costs increases. Therefore, an increase in terminal capacity should be matched with an increase in total passenger volume only when passengers differ a lot in their relative schedule delay costs. This observation sheds new light on possible future studies that examine the interaction between investment and pricing decisions as well as the interaction between terminal and runway investments.

Acknowledgment

We are very grateful to two anonymous referees whose constructive and thoughtful comments have led to significant improvements of the paper. We also thank Tiziana D'Alfonso, Martin Dresner, Tim Hau, Robin Lindsey, Hai Yang and seminar participants at Hong Kong University of Science and Technology for helpful comments and suggestions on earlier versions of the paper.

Financial support from the Research Grants Council of Hong Kong (RGC) and the Social Science and Humanities Research Council of Canada (SSHRC) is gratefully acknowledged.

Appendix

A. Proof of Proposition 1

Under laissez-fair situation, differentiate both sides of (7) with respect to p_B and use the Cramer's rule. We obtain

$$\begin{aligned}\psi &\equiv \begin{vmatrix}\rho'_B - \upsilon_B D' - c_B^B & -\upsilon_B D' - c_L^B \\ -\upsilon_L D' - c_B^L & \rho'_L - \upsilon_L D' - c_L^L\end{vmatrix} \\ &= \rho'_B\rho'_L - \rho'_B\left(\upsilon_L D' + c_L^L\right) - \rho'_L\left(\upsilon_B D' + c_B^B\right) + \upsilon_L D'\left(c_B^B - c_L^B\right) + \upsilon_B D'\left(c_L^L - c_B^L\right) + c_B^B c_L^L - c_L^B c_B^L > 0.\\ \frac{\partial Q_B}{\partial p_B} &= \frac{1}{\psi}\left(\rho'_L - \upsilon_L D' - c_L^L\right) < 0 \quad \text{and} \quad \frac{\partial Q_B}{\partial p_L} = \frac{1}{\psi}\left(\upsilon_L D' + c_L^B\right) > 0.\end{aligned}$$

$$\text{Similarly,}\ \frac{\partial Q_L}{\partial p_L} = \frac{1}{\psi}\left(\rho'_B - \upsilon_B D' - c_B^B\right) < 0 \quad \text{and} \quad \frac{\partial Q_L}{\partial p_B} = \frac{1}{\psi}\left(\upsilon_B D' + c_B^L\right) > 0. \quad \text{and} \quad \frac{\partial Q_L}{\partial p_B} = \frac{1}{\psi}\left(\upsilon_B D' + c_B^L\right) > 0.$$

Under uniform pricing, $p_B = p_L = p$. Differentiate both sides of (7) with respect to p and use the Cramer's rule. We obtain:

$$\begin{aligned}\frac{\partial Q_L}{\partial p} &= \frac{\partial Q_L}{\partial p_B} + \frac{\partial Q_L}{\partial p_L} = \frac{1}{\psi}\left(\rho'_B - (\upsilon_B - \upsilon_L)D' - (c_B^B - c_B^L)\right),\\ \frac{\partial Q_B}{\partial p} &= \frac{\partial Q_B}{\partial p_B} + \frac{\partial Q_B}{\partial p_L} = \frac{1}{\psi}\left(\rho'_L + (\upsilon_B - \upsilon_L)D' - (c_L^L - c_L^B)\right).\end{aligned}$$

Note that

$$c_B^B - c_B^L = \begin{cases}(\beta_B\upsilon_B - \beta_L\upsilon_L)/s > 0 & \text{if} \quad \beta_B > \beta_L \\ (\upsilon_B - \upsilon_L)\beta_B/s > 0 & \text{if} \quad \beta_B < \beta_L\end{cases} \quad \text{and} \quad c_L^L - c_L^B = \begin{cases}-(\upsilon_B - \upsilon_L)\beta_L/s < 0 & \text{if} \quad \beta_B > \beta_L \\ -(\beta_B\upsilon_B - \beta_L\upsilon_L)/s & \text{if} \quad \beta_B < \beta_L\end{cases}.$$

The sign of $\beta_B\upsilon_B - \beta_L\upsilon_L$ is ambiguous when $\beta_B < \beta_L$. Thus, $\frac{\partial Q_L}{\partial p} < 0$ and the sign of $\frac{\partial Q_B}{\partial p}$ is unclear.

$$\frac{\partial Q}{\partial p} = \frac{\partial Q_B}{\partial p} + \frac{\partial Q_L}{\partial p} = \frac{1}{\psi}\left(\rho'_B + \rho'_L - (c_B^B - c_L^B) - (c_L^L - c_B^L)\right) < 0.$$

B. Proof of Lemma 1

Hessian matrix of $\pi^i$is negative definite is equivalent to $\pi^i_{hihi} < 0$ and $\pi^i_{BiBi}\pi^i_{LiLi} - \pi^i_{BiLi}\pi^i_{LiBi} > 0$.

(i) The sufficient condition of local stability

Following Zhang and Zhang (1996), we derive the sufficient condition for local stability. From the first order condition (14), we have for each airline

$$f_i = \begin{bmatrix}\pi^i_{Bi} \\ \pi^i_{Li}\end{bmatrix} = 0.$$

To derive the best response functions at the Nash equilibrium, let $q_i = [\begin{smallmatrix} q_B^i \\ q_L^i \end{smallmatrix}]$ be the vector of equilibrium quantities of airline *i* = *1,…,n*. Then, we have

$$\frac{\partial f_i}{\partial q_i} = \begin{bmatrix}\pi^i_{BiBi} & \pi^i_{BiLi} \\ \pi^i_{LiBi} & \pi^i_{LiLi}\end{bmatrix},$$

and due to symmetry, for any *j*, *k* ≠ *i*,

$$\frac{\partial f_i}{\partial q_j} = \frac{\partial f_i}{\partial q_k} = \begin{bmatrix}\pi^i_{BiBj} & \pi^i_{BiLj} \\ \pi^i_{LiBj} & \pi^i_{LiLj}\end{bmatrix}.$$

Then,

$$(Df_i)(q_{-i}; q_i) = \begin{bmatrix}\underbrace{\frac{\partial f_i}{\partial q_1} \ \cdots \ \frac{\partial f_i}{\partial q_j}}_{\substack{(n-1)\text{identical2x2 matrix} \\ j\neq i}} & \frac{\partial f_i}{\partial q_i}\end{bmatrix} = \begin{bmatrix}\frac{\partial f_i}{\partial q_{-i}} & \frac{\partial f_i}{\partial q_i}\end{bmatrix}.$$

Denote $q_i^R = [\begin{smallmatrix} q_B^{iR} \\ q_L^{iR} \end{smallmatrix}]$ as the vector of best response functions for airline i. Then,

$$\frac{\partial q_i^R}{\partial q_{-i}} = \begin{bmatrix} \underbrace{\frac{\partial q_i^R}{\partial q_1} \quad \cdots \quad \frac{\partial q_i^R}{\partial q_j}}_{(n-1) identical 2x2 matrix} \\ j \neq i \end{bmatrix} = -\left[\frac{\partial f_i}{\partial q_i}\right]^{-1}\left[\frac{\partial f_i}{\partial q_{-i}}\right].$$

For a certain rival indexed by$j \neq i$, its impact on airline i's best response outputs is

$$\frac{\partial q_i^R}{\partial q_j} = \begin{bmatrix} \frac{\partial q_B^{iR}}{\partial q_B^j} & \frac{\partial q_B^{iR}}{\partial q_L^j} \\ \frac{\partial q_L^{iR}}{\partial q_B^j} & \frac{\partial q_L^{iR}}{\partial q_L^j} \end{bmatrix} = -\begin{bmatrix} \pi_{BiBi}^i & \pi_{BiLi}^i \\ \pi_{LiBi}^i & \pi_{LiLi}^i \end{bmatrix}^{-1} \begin{bmatrix} \pi_{BiBj}^i & \pi_{BiLj}^i \\ \pi_{LiBj}^i & \pi_{LiLj}^i \end{bmatrix}$$

$$= -\frac{1}{A}\begin{bmatrix} \pi_{LiLi}^i & -\pi_{BiLi}^i \\ -\pi_{LiBi}^i & \pi_{BiBi}^i \end{bmatrix} \cdot \begin{bmatrix} \pi_{BiBj}^i & \pi_{BiLj}^i \\ \pi_{LiBj}^i & \pi_{LiLj}^i \end{bmatrix} = \begin{bmatrix} \frac{-a}{A} & \frac{-b}{A} \\ \frac{-c}{A} & \frac{-d}{A} \end{bmatrix},$$

where

$A = \pi_{BiBi}^i \pi_{LiLi}^i - \pi_{BiLi}^i \pi_{LiBi}^i > 0$ (due to negative definite Hessian of π^i)
$a = \pi_{BiBj}^i \pi_{LiLi}^i - \pi_{BiLi}^i \pi_{LiBj}^i$
$b = \pi_{BiLj}^i \pi_{LiLi}^i - \pi_{BiLi}^i \pi_{LiLj}^i$
$c = \pi_{BiBi}^i \pi_{LiBj}^i - \pi_{LiBi}^i \pi_{BiBj}^i$
$d = \pi_{BiBi}^i \pi_{LiLj}^i - \pi_{LiBi}^i \pi_{BiLj}^i$

Similar to Zhang and Zhang (1996), we denote the $2n \times 2n$ best response matrix as

$$T' = \begin{bmatrix} 0 & \frac{\partial q_i^R}{\partial q_j} & \cdots & \frac{\partial q_i^R}{\partial q_j} \\ \frac{\partial q_i^R}{\partial q_j} & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \frac{\partial q_i^R}{\partial q_j} \\ \frac{\partial q_i^R}{\partial q_j} & \cdots & \frac{\partial q_i^R}{\partial q_j} & 0 \end{bmatrix} = \begin{bmatrix} \frac{\partial q_i^R}{\partial q_j} & 0 & \cdots & 0 \\ 0 & \frac{\partial q_i^R}{\partial q_j} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\partial q_i^R}{\partial q_j} \end{bmatrix} \cdot \begin{bmatrix} 0 & I & \cdots & I \\ I & 0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & I \\ I & \cdots & I & 0 \end{bmatrix} = R \cdot E$$

Note that $\|T'\|_p \le \|R\|_p \cdot \|E\|_p = (n-1) \cdot \|R\|_p$. Thus, $\|R\|_p < \frac{1}{n-1}$ implies $\|T'\|_p < 1$, i.e. the local stability condition. Since that R is a diagonal matrix, then $\|R\|_p = \max_{1 \le k \le n} \|R_{k,k}\|_p = \|\frac{\partial q_i^R}{\partial q_j}\|_p$. Therefore, we get the sufficient condition for local stability: $\exists p s.t. \|\frac{\partial q_i^R}{\partial q_j}\|_p < \frac{1}{n-1}$. This sufficient condition is equivalent to $\max|\lambda| < \frac{1}{n-1}$, in which λ is the eigenvalue of $\frac{\partial q_i^R}{\partial q_j}$.

(ii) This sufficient condition leads to $\Delta > 0$

Since $\det(\frac{\partial q_i^R}{\partial q_j} - \lambda I) = \begin{vmatrix} \frac{-a}{A} - \lambda & \frac{-b}{A} \\ \frac{-c}{A} & \frac{-d}{A} - \lambda \end{vmatrix} = 0$, we have $A^2\lambda^2 + A(a+d)\lambda + ad - bc = 0$. Solving this function, we get the two eigenvalues: $\lambda = \frac{-(a+d) \pm \sqrt{(a-d)^2 + 4bc}}{2A}$. Thus, the sufficient condition is equivalent to

$$\max|\lambda| = \frac{|a+d| + \sqrt{(a-d)^2 + 4bc}}{2A} < \frac{1}{n-1}. \tag{B.1}$$

After some tedious algebra, we can show that

$$(a+d)^2 - \left((a-d)^2 + 4bc\right) = 4A(\pi_{BiBj}^i \pi_{LiLj}^i - \pi_{BiLj}^i \pi_{LiBj}^i).$$

Note that

$$\pi_{BiBj}^i - \pi_{BiLj}^i = \rho_B'' \frac{Q_B}{n} + \rho_B' - (c_B^B - c_L^B) < 0$$

$$\pi_{LiLj}^i - \pi_{LiBj}^i = \rho_L'' \frac{Q_L}{n} + \rho_L' - (c_L^L - c_B^L) < 0$$

Therefore, $\pi^i_{BiBj} < \pi^i_{BiLj} < 0$ and $\pi^i_{LiLj} < \pi^i_{LiBj} < 0$, which implies that $\pi^i_{BiBj}\pi^i_{LiLj} - \pi^i_{BiLj}\pi^i_{LiBj} > 0$. Then, we can rewrite

$$\pi^i_{BiBj}\pi^i_{LiLj} - \pi^i_{BiLj}\pi^i_{LiBj} = \frac{(a+d)^2 - ((a-d)^2 + 4bc)}{4A} > 0.$$

We can also show that

$$\Delta = A + (n-1)(a+d) + (n-1)^2(\pi^i_{BiBj}\pi^i_{LiLj} - \pi^i_{BiLj}\pi^i_{LiBj})$$

If $a + d \geq 0$, it is obvious that $\Delta > 0$.
If $a + d < 0$, we have

$$\begin{aligned}
\Delta &= A - (n-1)|a+d| + (n-1)^2(\pi^i_{BiBj}\pi^i_{LiLj} - \pi^i_{BiLj}\pi^i_{LiBj}) \\
&= A - (n-1)|a+d| + (n-1)^2\frac{(a+d)^2 - \left((a-d)^2 + 4bc\right)}{4A} \\
&= \frac{1}{A}\left(A^2 - (n-1)A|a+d| + \frac{(n-1)^2|a+d|^2}{4} - \frac{(n-1)^2\left((a-d)^2+4bc\right)}{4}\right) \\
&= \frac{1}{A}\left(\left(A - \frac{(n-1)|a+d|}{2}\right)^2 - \frac{(n-1)^2\left((a-d)^2+4bc\right)}{4}\right) \\
&= \frac{1}{A}\left(A - \frac{(n-1)|a+d|}{2} + \frac{(n-1)\sqrt{(a-d)^2+4bc}}{2}\right)\left(A - \frac{(n-1)|a+d|}{2} - \frac{(n-1)\sqrt{(a-d)^2+4bc}}{2}\right)
\end{aligned}$$

Inequality (B.1) implies that both brackets in the above expression are positive, which means that $\Delta > 0$.

C. Proof of Proposition 3

The impact of airport charge on the number of business passengers is

$$\begin{aligned}
\frac{\partial q^{iN}_B}{\partial \tau} &= \frac{1}{\Delta}\begin{vmatrix} 1 & \pi^i_{BiLi} + (n-1)\pi^i_{BiLj} \\ 1 & \pi^i_{LiLi} + (n-1)\pi^i_{LiLj} \end{vmatrix} = \frac{1}{\Delta}\left(\pi^i_{LiLi} - \pi^i_{BiLi} + (n-1)(\pi^i_{LiLj} - \pi^i_{BiLj})\right) \\
&= \frac{1}{\Delta}\left(\rho''_L Q_L + n\psi\frac{\partial Q_B}{\partial p} + \psi\frac{\partial Q}{\partial p_B}\right)
\end{aligned}$$

From Proposition 1, we know that the sign of $\frac{\partial Q_B}{\partial p}$ is ambiguous and hence the sign of $\frac{\partial Q^N_B}{\partial \tau} = n\frac{\partial q^{iN}_B}{\partial \tau}$ is ambiguous as well.

$$\frac{\partial Q^N}{\partial \tau} = \frac{\partial Q^N_B}{\partial \tau} + \frac{\partial Q^N_L}{\partial \tau} = \frac{1}{\Delta}\left(\rho''_B Q_B + \rho''_L Q_L + (n+1)\left(\rho'_B + \rho'_L - (c^B_B - c^B_L) - (c^L_L - c^L_B)\right)\right) < 0.$$

D. Impact of terminal capacity increase on passenger numbers

(a) Proof of Proposition 7: impact on passenger numbers given fixed airfares

$$\frac{\partial Q_B}{\partial s} = \frac{1}{\psi}\begin{vmatrix} \partial c^B/\partial s & -\upsilon_B D' - c^B_L \\ \partial c^L/\partial s & \rho'_L - \upsilon_L D' - c^L_L \end{vmatrix} \quad \text{and} \quad \frac{\partial Q_B}{\partial s} = \frac{1}{\psi}\begin{vmatrix} \rho'_B - \upsilon_B D' - c^B_B & \partial c^B/\partial s \\ -\upsilon_L D' - c^L_B & \partial c^L/\partial s \end{vmatrix},$$

where $\psi > 0$ as shown in Appendix A. Therefore, for any $\beta_h > \beta_k$, since $c^B_B c^L_L - c^B_L c^L_B > 0$, it is straightforward to show:

$$\begin{aligned}
\frac{\partial Q_h}{\partial s} &= \frac{1}{\psi s}\left[-c^h\rho'_k + \frac{(\beta_h - \beta_k)}{s}\upsilon_h\upsilon_k Q_h D' + (c^h_h c^k_k - c^h_k c^k_h)Q_h\right] > 0, \\
\frac{\partial Q_k}{\partial s} &= \frac{1}{\psi s}\left[-c^k\rho'_h - \frac{(\beta_h - \beta_k)}{s}\upsilon_h\upsilon_k Q_h D' + (c^h_h c^k_k - c^h_k c^k_h)Q_k\right], \\
\frac{\partial Q}{\partial s} &= \frac{1}{\psi s}\left[-c^B\rho'_L - c^L\rho'_B + (c^B_B c^L_L - c^B_L c^L_B)(Q_B + Q_L)\right] > 0.
\end{aligned}$$

Note that $\partial Q_k/\partial s$, the impact on passengers who have smaller relative schedule delay cost and hence arrive earlier, can be negative if D' is far above zero and $\beta_h - \beta_k$ is very large. That is, if $\beta_B = \beta_L$ or there is no runway congestion($D' = 0$), the terms related to D' in the brackets will disappear and hence $\partial Q_B/\partial s > 0$ and $\partial Q_L/\partial s > 0$ will both hold.

(b) Impact on first-best passenger numbers without concession surplus

Differentiating both sides of Eqs. (9) and (10) with respect to s, dQ_B^*/ds and dQ_L^*/dscan be obtained by solving the following two equations.

$$\frac{\partial S_a^2}{\partial Q_B^2}\frac{dQ_B^*}{ds}+\frac{\partial S_a^2}{\partial Q_B\partial Q_L}\frac{dQ_L^*}{ds}=-\frac{\partial S_a^2}{\partial Q_B\partial s}\quad\text{and}\quad\frac{\partial S_a^2}{\partial Q_B\partial Q_L}\frac{dQ_B^*}{ds}+\frac{\partial S_a^2}{\partial Q_L^2}\frac{dQ_L^*}{ds}=-\frac{\partial S_a^2}{\partial Q_L\partial s}.$$

Since $\partial S_a^2/\partial Q_h^2<0$, $\partial S_a^2/\partial Q_B\partial Q_L<0$and $\partial S_a^2/\partial Q_h\partial s>0$ for any $h\in\{B,L\}$, it is straightforward to show that either dQ_B^*/ds or dQ_L^*/ds(or both) have to be positive.

$$\frac{dQ_B^*}{ds}=\frac{1}{\Omega}[(1)+(2)+(3)],$$

where

$\Omega\equiv\frac{\partial S_a^2}{\partial Q_B^2}\frac{\partial S_a^2}{\partial Q_L^2}-(\frac{\partial S_a^2}{\partial Q_B\partial Q_L})^2>0$, as required by the second-order condition.

$$(1)=-\frac{c^B+\phi_B}{s}\rho_L'>0$$

$$(2)=\begin{cases}\left[4v_Bv_L\beta_B-(v_B+v_L)^2\beta_L\right]\frac{Q_BD'}{s^2}\ge 0 & \text{if}\quad \frac{\beta_B}{\beta_L}\ge\frac{(v_B+v_L)^2}{4v_Bv_L}>1;\ <0\quad\text{if}\quad\frac{(v_B+v_L)^2}{4v_Bv_L}>\frac{\beta_B}{\beta_L}>1\\ \left[-(v_B-v_L)^2\beta_BQ_B-2v_L(v_B+v_L)(\beta_L-\beta_B)Q_L\right]\frac{D'}{s^2}<0 & \text{if}\quad\beta_B<\beta_L\\ \left[4v_Bv_L-(v_B+v_L)^2\right]\frac{\beta Q_BD'}{s^2}<0 & \text{if}\quad\beta_B=\beta_L=\beta\end{cases}$$

$$(3)=\left[4c_B^Bc_L^L-\left(c_L^B+c_B^L\right)^2\right]\frac{Q_B}{s}$$

$$=\begin{cases}\left[4v_Bv_L\beta_B-(v_B+v_L)^2\beta_L\right]\frac{\beta_LQ_B}{s^3}\ge 0 & \text{if}\quad\frac{\beta_B}{\beta_L}\ge\frac{(v_B+v_L)^2}{4v_Bv_L}>1;\ <0\quad\text{if}\quad\frac{(v_B+v_L)^2}{4v_Bv_L}>\frac{\beta_B}{\beta_L}>1\\ \left[4v_Bv_L\beta_L-(v_B+v_L)^2\beta_B\right]\frac{\beta_BQ_B}{s^3}\ge 0 & \text{if}\quad\frac{\beta_L}{\beta_B}\ge\frac{(v_B+v_L)^2}{4v_Bv_L}>1;\ <0\quad\text{if}\quad\frac{(v_B+v_L)^2}{4v_Bv_L}>\frac{\beta_L}{\beta_B}>1\\ \left[4v_Bv_L-(v_B+v_L)^2\right]\frac{\beta^2Q_B}{s^3}<0 & \text{if}\quad\beta_B=\beta_L=\beta\end{cases}$$

Similarly,

$$\frac{dQ_L^*}{ds}=\frac{1}{\Omega}\left[(i)+(ii)+(iii)\right],$$

where

$$(i)=-\frac{c^L+\phi_L}{s}\rho_B'>0.$$

$$(ii)=\begin{cases}\left[-(v_B-v_L)^2\beta_LQ_L-2v_B(v_B+v_L)(\beta_B-\beta_L)Q_B\right]\frac{D'}{s^2}<0 & \text{if}\quad\beta_B>\beta_L\\ \left[4v_Bv_L\beta_L-(v_B+v_L)^2\beta_B\right]\frac{Q_LD'}{s^2}\ge 0 & \text{if}\quad\frac{\beta_L}{\beta_B}\ge\frac{(v_B+v_L)^2}{4v_Bv_L}>1;\ <0\quad\text{if}\quad\frac{(v_B+v_L)^2}{4v_Bv_L}>\frac{\beta_L}{\beta_B}>1\\ \left[4v_Bv_L-(v_B+v_L)^2\right]\frac{\beta Q_LD'}{s^2}<0 & \text{if}\quad\beta_B=\beta_L=\beta\end{cases}$$

$$(iii)=\left[4c_B^Bc_L^L-\left(c_L^B+c_B^L\right)^2\right]\frac{Q_L}{s}=\frac{Q_L}{Q_B}\cdot(3)$$

E. Airfares under time-varying terminal fine toll

When $v_B\beta_B>v_L\beta_L$, the total terminal cost imposed on the society is:

$$TC=TSDC=v_B\beta_B\frac{Q_B^2}{2s}+v_L\beta_L\frac{2Q_BQ_L+Q_L^2}{2s}$$

$$\frac{\partial TC}{\partial Q_B}=v_B\beta_B\frac{Q_B}{s}+v_L\beta_L\frac{Q_L}{s},\quad\text{and}\quad\frac{\partial TC}{\partial Q_L}=v_L\beta_L\frac{Q}{s}$$

Since the first arriving leisure passenger pays no terminal toll[24] and the cost paid by each leisure passenger (terminal toll plus schedule delay cost) is the same at equilibrium, each leisure passenger incurs the following cost at terminal with time-varying toll is:

$$\tilde{c}^L = \nu_L \beta_L \frac{Q}{s} = \frac{\partial TC}{\partial Q_L}.$$

Thus, the time-varying toll incurred by the first-arriving business passenger who arrives the airport at time t_m is:

$$r(t_m) = \tilde{c}^L - SDC_L(t_m) = \nu_L \beta_L \frac{Q}{s} - \nu_L \beta_L \frac{Q_B}{s} = \nu_L \beta_L \frac{Q_L}{s}$$

Therefore, the cost incurred by the first and each subsequent business passenger is:

$$\tilde{c}^B = r(t_m) + SDC_B(t_m) = \nu_L \beta_L \frac{Q_L}{s} + \nu_B \beta_B \frac{Q_B}{s} = \frac{\partial TC}{\partial Q_B}$$

Similarly, when $\nu_B \beta_B < \nu_L \beta_L$, we have the following:

$$\tilde{c}^B = \nu_B \beta_B \frac{Q}{s} = \frac{\partial TC}{\partial Q_B} \text{ and } \tilde{c}^L = \nu_L \beta_L \frac{Q_L}{s} + \nu_B \beta_B \frac{Q_B}{s} = \frac{\partial TC}{\partial Q_L}.$$

(a) First-best fares without concession surplus

The first-best fare is achieved by differentiating the following aeronautic surplus:

$$W = S_a \equiv \sum_{h=B,L} \int_0^{Q_h} \rho_h(x)dx - \nu_h D(Q) Q_h - TC.$$

The first-order conditions of the above function are:

$$\frac{\partial W}{\partial Q_h} = \rho_h^* - \nu_h D - \tilde{c}^h - \Gamma^* = 0, \quad \forall h = B, L.$$

That is, $p_B^* = p_L^* = \Gamma^*$.

(b) Nash equilibrium airfares and second-best airport charges without concession surplus

An airline will maximize the following profit function:

$$\pi^i(q_B^1, \ldots, q_B^n; q_L^1, \ldots, q_L^n) = \sum_{h=B,L} \left(\rho_h(Q_h) - \nu_h D(Q) - \tilde{c}^h(Q_B, Q_L) - \tau\right) \cdot q_h^i.$$

The first order conditions lead to:

$$\frac{\partial \pi^i}{\partial q_h^i} = (\rho_h' - \nu_h D' - \tilde{c}_h^h) q_h^i + p_h - (\nu_k D' + \tilde{c}_h^k) q_k^i - \tau = 0.$$

Thus, the equilibrium airfare will be:

$$p_h^N = \tau - \frac{1}{n}\left(\rho_h' Q_h^N - \Gamma^N - \tilde{\phi}_h^N\right), \quad \forall h = B, L, \text{ where } \tilde{\phi}_h \equiv Q_h \tilde{c}_h^h + Q_k \tilde{c}_h^k = \tilde{c}^h.$$

Second-best airport charge (excluding the time-varying fine toll) is derived from the following equation:

$$\frac{dW}{d\tau} = \frac{\partial S_a}{\partial Q_B}\frac{\partial Q_B^N}{\partial \tau} + \frac{\partial S_a}{\partial Q_L}\frac{\partial Q_L^N}{\partial \tau} = \left(p_B^N - \Gamma^N\right)\frac{\partial Q_B^N}{\partial \tau} + \left(p_L^N - \Gamma^N\right)\frac{\partial Q_L^N}{\partial \tau} = 0$$

That is,

$$\left(\tau - \frac{1}{n}\left(\rho_B' Q_B^N - \Gamma^N - \tilde{\phi}^B\right) - \Gamma^N\right)\frac{\partial Q_B^N}{\partial \tau} + \left(\tau - \frac{1}{n}\left(\rho_L' Q_L^N - \Gamma^N - \tilde{\phi}^L\right) - \Gamma^N\right)\frac{\partial Q_L^N}{\partial \tau} = 0$$

$$\tau^{**} = \left(1 - \frac{1}{n}\right)\Gamma^N - \frac{1}{n}\tilde{\Phi}^N + \frac{1}{n}M^N, \text{ where } \tilde{\Phi}^N \equiv \frac{\frac{\partial Q_B^N}{\partial \tau}\tilde{\phi}_B^N + \frac{\partial Q_L^N}{\partial \tau}\tilde{\phi}_L^N}{\frac{\partial Q^N}{\partial \tau}}.$$

The second term in the above equation is an adjustment to partially correct airlines' over-internalization of terminal external costs. Since the logic applied in Lemma 1 still holds with time-varying terminal toll, the following will hold:

$$\frac{\partial q_L^{iN}}{\partial \tau} = \frac{1}{\tilde{\Delta}}\left(\rho_B'' Q_B + (n+1)\rho_B' - (n+1)(\nu_B - \nu_L)D' - (n+1)(\tilde{c}_B^B - \tilde{c}_B^L)\right) < 0$$

$$\frac{\partial q_B^{iN}}{\partial \tau} = \frac{1}{\tilde{\Delta}}\left(\rho_L'' Q_L + (n+1)\rho_L' + (n+1)(\nu_B - \nu_L)D' - (n+1)(\tilde{c}_L^L - \tilde{c}_L^B)\right),$$

where $\tilde{\Delta} > 0$. As $\tilde{c}_L^L - \tilde{c}_L^B \geq 0$, $\partial q_B^{iN}/\partial \tau$ can be positive only when the marginal runway congestion is very severe. Note again that $\partial Q^N/\partial \tau < 0$ always holds.

[24] To achieve the social optimum, it is not necessary to charge the first-arriving passenger zero time-varying tolls. Rather, the social optimum can still be achieved by adding same amount of toll to all the passengers. However, since this uniform addition of toll is merely an internal transfer and can be rebated to the passengers via reducing the first-best airfares, for simplicity and without loss of generality, we set the terminal toll levied on first-arriving passenger to be zero.

References

Appold, S.J., Kasarda, J.D., 2007. The appropriate scale of US airport retail activities. Journal of Air Transport Management 12, 277–287.
Arnott, R., de Palma, A., Lindsey, R., 1990. Economics of a bottleneck. Journal of Urban Economics 27, 111–130.
Arnott, R., de Palma, A., Lindsey, R., 1993. A structural model of peak-period congestion: a traffic bottleneck with elastic demand. American Economic Review 83, 161–179.
Arnott, R., de Palma, A., Lindsey, R., 1994. The welfare effects of congestion tolls with heterogeneous commuters. Journal of Transport Economics and Policy 28 (2), 139–161.
Arnott, R., Kraus, M., 1998. When are anonymous congestion charges consistent with marginal cost pricing. Journal of Public Economics 67, 45–64.
Arnott, R., Kraus, M., 2003. Principles of transport economics. In: Hall, R. (Ed.), Handbook of Transportation Science. Kluwer Academic Publishers, Dordecht, pp. 700–737.
Ater, I., 2012. Internalization of congestion at US hub airports. Journal of Urban Economics 72, 196–209.
Basso, L.J., 2008. Airport deregulation: effects on pricing and capacity. International Journal of Industrial Organization 26, 1015–1031.
Basso, L.J., Zhang, A., 2007. An interpretative survey of analytical models of airport pricing. In: Darin, Lee (Ed.), Advances in Airline Economics, 2. Elsevier, pp. 89–124.
Basso, L.J., Zhang, A., 2008. Sequential peak-load pricing in a vertical setting: the case of airports and airlines. Canadian Journal of Economics 41, 1087–1119.
Borenstein, S., Netz, J., 1999. Why do all the flights leave at 8am?: Competition and departure-time differentiation in airline markets. International Journal of Industrial Organization 17 (5), 611–640.
Bracaglia, V., D'Alfonso, T., Nastasi, A., 2014. Competition between multiproduct airports. Economics of Transportation 3, 270–281.
Brander, J.A., Zhang, A., 1990. Market conduct in the airline Industry: An empirical investigation. RAND Journal of Economics 21 (4), 567–583.
Brander, J.A., Zhang, A., 1993. Dynamic oligopoly behaviour in the airline industry. International Journal of Industrial Organization 11 (3), 407–435.
Brueckner, J.K., 2002. Airport congestion when carriers have market power. American Economic Review 92 (5), 1357–1375.
Brueckner, J.K., Van Dender, K., 2008. Atomistic congestion tolls at concentrated airports? Seeking a unified view in the internalization debate. Journal of Urban Economics 64 (2), 288–295.
Carlin, A., Park, R.E., 1970. Marginal cost pricing of airport runway capacity. American Economic Review 60, 310–319.
Chu, X., 1995. Endogenous trip scheduling: the Henderson approach reformulated and compared with the Vickrey approach. Journal of Urban Economics 37 (3), 324–343.
Czerny, A.I., 2006. Price-cap regulation of airports: single-till versus dual-till. Journal of Regulatory Economics 30, 85–97.
Czerny, A.I., 2013. Public versus private airport behavior when concession revenues exist. Economics of Transportation 2, 38–46.
Czerny, A.I., Lindsey, R., 2014. Multiproduct Pricing with Core Goods and Side Goods. VU University of Amsterdam Unpublished manuscript.
Czerny, A.I., Shi, Z., Zhang, A., 2015. Can Market Power be Controlled by Regulation of Core Prices Alone?: An Empirical Analysis of Airport Demand and Car Rental Price. Tinbergen Institute Discussion Paper 15-041/III.
Czerny, A.I., Zhang, A., 2011. Airport congestion pricing and passenger types. Transportation Research Part B: Methodological 45, 595–604.
Czerny, A.I., Zhang, A., 2014. Airport peak-load pricing revisited: The case of peak and uniform tolls. Economics of Transportation 3 (1), 90–101.
Czerny, A.I., Zhang, A., 2014. Airport congestion pricing when airlines price discriminate. Transportation Research Part B: Methodological 65, 77–89.
Czerny, A.I., Zhang, A., 2015. Third-degree price discrimination in the presence of congestion externality. Canadian Journal of Economics 48 (4) in press.
D'Alfonso, T., Jiang, C., Wan, Y., 2013. Airport pricing, concession revenues and passenger types. Journal of Transport Economics and PolicyEq. 47 (1), 71–89.
Daniel, J.I., 1995. Congestion pricing and capacity of large hub airports: a bottleneck model with stochastic queues. Econometrica 63, 327–370.
Daniel, J.I., 2009. The Deterministic Bottleneck Model with Non-Atomistic Traffic. University of Delaware, Department of Economics Working Papers 09-08.
Douglas, G.W., Miller, J.C., 1974. Quality competition industry equilibrium, and efficiency in the price constrained airline market. American Economic Review 64, 657–669.
Dube, L., Menon, K., 2000. Ensuring greater satisfaction by engineering salesperson response to customer emotions. Journal of Retailing 76, 285–308.
De Borger, B., Van Dender, K., 2006. Prices, capacities and service levels in a congestible Bertrand duopoly. Journal of Urban Economics 60, 264–283.
Flores-Fillol, R., 2010. Congested hubs. Transportation Research Part B: Methodological 44, 358–370.
Flores-Fillol, R., Iozzi, A., Valletti, T. 2014. Platform Pricing and ConsumerForesight: The Case of Airports. CEIS Working Paper No. 335. Available at SSRN: http://ssrn.com/abstract=2585098 or http://dx.doi.org/10.2139/ssrn.2585098.
Graham, A., 2009. How important are commercial revenues to today's airports? Journal of Air Transport Management 15, 106–111.
Haskel, J., Iozzi, A., Valletti, T., 2013. Market structure, countervailing power and price discrimination: the case of airports. Journal of Urban Economics 74 (1), 12–26.
Ito, H., Lee, D., 2005. Assessing the impact of the September 11th terrorist attacks on U.S. airline demand. Journal of Economics and Business 57 (1), 75–95.
Knockaert, J., Verhoef, E.T., Rouwendal, J., 2010. Bottleneck Congestion: Differentiating the Coarse Charge. Tinbergen Institute Amsterdam/Rotterdam TI Discussion Paper 2010-097/3.
Levine, M.E., 1969. Landing fees and the airport congestion problem. Journal of Law and Economics 12, 79–108.
Lin, M.H., 2013. Airport Congestion Pricing and Airline Network Choices. Osaka University of Economics Unpublished manuscript.
Lin, Y.H., Chen, C.F., 2013. Passengers' shopping motivations and commercial activities at airports – the moderating effects of time pressure and impulse buying tendency. Tourism Management 36, 426–434.
Lindsey, R., 2004. Existence, uniqueness, and trip cost function properties with multiple user classes. Transportation Science 38 (3), 293–314.
Mayer, C., Sinai, T., 2003. Network effects, congestion externalities, and air traffic delays: or why not all delays are evil. American Economic Review 93, 1194–1215.
Mohring, H., 1999. Congestion. In: Gomez-Ibanez, J., TYE, W.B., Winston, C. (Eds.). Essays in Transportation Economics and Policy. A Handbook in Honor of John R. Meyer. Brookings Institution, Washington, DC.
Morrison, S.A., 1983. Estimation of long-run prices and investment levels for airport runways. Research in Transportation Economics 1, 103–130.
Morrison, S.A., 1987. The equity and efficiency of runway pricing. Journal of Public Economics 34, 45–60.
Morrison, S.A., Winston, C., 1989. Enhancing the performance of the deregulated air transportation system. Brookings Papers on Economic Activity: Microeconomics 61–112.
Oum, T.H., Zhang, A., Zhang, Y., 2004. Alternative forms of economic regulation and their efficiency implication for airports. Journal of Transport Economics and PolicyEq. 38 (2), 217–246.
Panzar, J.C., 1979. Equilibrium and welfare in unregulated airline markets. American Economic Review 69, 92–95.
Pels, E., Nijkamp, P., Rietveld, P., 2003. Access to and competition between airports: a case study for the San Francisco Bay area. Transportation Research Part A, 37, 71–83.
Pels, E., Verhoef, E.T., 2004. The economics of airport congestion pricing. Journal of Urban Economics 55 (2), 257–277.
Qian, Z., Zhang, H.M., 2013. The morning commute problem with heterogeneous travellers: the case of continuously distributed parameters. Transportmetrica A: Transportation Science 9 (2), 178–203.
Rowe, I.V., 1998. The Future of Duty-free Retailing: Developments and Trends in Europe and the Middle East. Financial Times Retail and Consumer Publishing, London.
Rupp, N., 2009. Do carriers internalize congestion costs? Empirical evidence on the internalization question. Journal of Urban Economics 65 (1), 24–37.
Saraswati, B., Hanaoka, S., 2012. Space allocation for commercial activities in low-cost airport terminals: Soekarno-Hatta International Airport. Airport Management 6 (4), 397–411.
Silva, H.E., Verhoef, E.T., van den Berg, V., 2014. Airlines' strategic interactions and airport pricing in a dynamic bottleneck model of congestion. Journal of Urban Economics 80, 13–27.

Silva, H.E., Lindsey, R., de Palma, A., van den Berg, V., 2014. On the Existence and Uniqueness of Equilibrium in the Bottleneck Model with Atomic Users. Tinbergen Institute Amsterdam/Rotterdam TI Discussion Paper 14-077.

Small, K.A., Verhoef, E.T., 2007. The Economics of Urban Transportation. Routledge, New York.

Thomas, D., 1997. Retail and leisure developments at London Gatwick. Commercial Airport 38–41.

Tirole, J., 1988. The Theory of Industrial Organization. MIT Press, Cambridge, MA.

Torres, E., Domnguez, J.S., Valdes, L., Aza, R., 2005. Passenger waiting time in an airport and expenditure carried out in the commercial area. Journal of Air Transport Management 11, 363–367.

USDOT, 1997. The value of saving travel time: departmental guidance for conducting economic evaluations.

van den Berg, V, 2014. Coarse tolling with heterogeneous preferences. Transportation Research Part B: Methodological 64, 1–23.

van den Berg, V, Verhoef, E.T., 2010. Why Congestion Tolling could be Good for the Consumer: The Effects of Heterogeneity in the Values of Schedule Delay and Time on the Effects of Tolling. Tinbergen Institute Amsterdam/Rotterdam TI Discussion Paper 2010-016/3.

van den Berg, V, Verhoef, E.T., 2011. Congestion tolling in the bottleneck model with heterogeneous values of time. Transportation Research Part B: Methodological 45, 60–78.

van der Weijde, A.H., 2014. Price Differentiation and Discrimination in Transport Networks. Tinbergen Institute Amsterdam/Rotterdam TI Discussion Paper TI 2014-099/VIII.

Verhoef, E.T., Silva, H.E., 2015. Dynamic Equilibrium at a Congestible Facility under Market Power. Tinbergen Institute Amsterdam/Rotterdam TI Discussion Paper 15-045.

Vickrey, W.S., 1969. Congestion theory and transport investment. American Economic Review 59, 251–261.

Yang, H., Zhang, A., 2011. Price-cap regulation of congested airports. Journal of Regulatory Economics 39 (3), 293–312.

Zhang, A., Czerny, A.I., 2012. Airports and airlines economics and policy: an interpretive review of recent research. Economics of Transportation 1, 15–34.

Zhang, A., Zhang, Y., 1996. Stability of a Cournot–Nash equilibrium: the multiproduct case. Journal of Mathematical Economics 26 (4), 441–462.

Zhang, A., Zhang, Y., 1997. Concession revenue and optimal airport pricing. Transportation Research Part E 4, 287–296.

Zhang, A., Zhang, Y., 2006. Airport capacity and congestion when carriers have market power. Journal of Urban Economics 60 (2), 229–247.