The international platform of Ludwig-Maximilians University's Center for Economic Studies and the Ifo Institute

Working Papers

www.cesifo.org/wp

Convenient Flight Connections vs. Airport Congestion Modeling the 'Rolling Hub'

Jan K. Brueckner
Ming Hsin Lin

CESifo Working Paper No. 5502
Category 11: Industrial Organisation
September 2015

ISSN 2364-1428

CESifo
Center for Economic Studies & Ifo Institute

CESifo *Working Paper No. 5502*

Convenient Flight Connections vs. Airport Congestion Modeling the 'Rolling Hub'

Abstract

This paper provides the first analysis of the trade-off between convenient flight connections and airport congestion. A continuous spatial model illustrates this trade-off in a framework where a small gap between flight operating times raises congestion while also shortening a connecting passenger's layover time. When the passenger's cost per unit of layover time rises, the monopoly airline chooses to narrow the gap between its flights, yielding shorter layovers but more congestion. A discrete spatial model, where flights congest one another only if they operate in the same discrete period, makes this layover-cost effect discontinuous: the monopoly carrier concentrates (deconcentrates) its flights when this cost is high (low) relative to the costs of congestion. When fringe carriers are present, however, the hub carrier always concentrates its flights, either partially or fully. But the presence of a second hub carrier leads to an equilibrium mirroring the monopoly outcome: the carriers concentrate their flights in different periods when the layover cost is high and deconcentrate them otherwise. The paper also presents a welfare analysis, showing that movement from the equilibrium to the social optimum typically requires greater carrier separation.

JEL-Code: L900.

Keywords: airport congestion, layover, rolling hub.

Jan K. Brueckner
Department of Economics
University of California, Irvine
3151 Social Science Plaza
USA – Irvine, CA 92697
jkbrueck@uci.edu

Ming Hsin Lin
Faculty of Economics
Osaka University of Economics
2-2-8 Osumi, Higashiyodogawa-ku
Japan – Osaka 533-8533
linmh@osaka-ue.ac.jp

August 2015

Convenient Flight Connections vs. Airport Congestion Modeling the 'Rolling Hub'

by

Jan K. Brueckner and Ming Hsin Lin*

1. Introduction

After decades of operating hub airports with flight arrivals and departures concentrated in "banks" that facilitate short layover times for connecting passengers, American Airlines in 2002 "depeaked" its hubs, shifting to what it called a "rolling hub" concept. Starting with the Chicago hub, flight operations were spread out, lengthening layover times, with the goal of reducing congestion and improving operational performance, thus saving costs. At the time, American's CEO Donald Carty stated that "[o]ur Chicago experience has improved customer service, reduced costs, improved productivity and allowed us to fly the same schedule with the equivalent of five fewer aircraft and four fewer gates." But Mary Fagan, an American spokesperson, noted that "[if] you're connecting, it may mean an [extra] delay of 10 to 12 minutes," pointing to longer layovers (both quotes are from Carey (2012)).

While Delta and United Airlines followed American's lead in adopting rolling hubs, recent years have seen a reversal of this trend. Hub carriers have mostly "rebanked" their hubs, abandoning the rolling-hub concept. Current American CEO Scott Kirby, quoted by Jean (2015), stated that "although the continuous [rolling] hub lowered operating costs, the lost revenue outweighed the savings," with the revenue losses apparently due to the lower number and timeliness of possible connections. Indeed, Marilyn DeVoe, vice president of American's Miami hub, stated that "[o]ur hubs are all about connecting people, and rebanking allows us to do that more effectively" (Jean, 2015).

Although there is now a large theoretical literature on airport congestion,[1] analytical treatment of trade-off underlying the rolling hub (connection convenience vs. airport congestion) is mostly absent. Mayer and Sinai (2003) argue that hub congestion is the price we pay for convenient connections, and Katz and Garrow (2014) provide evidence on the cost and revenue consequences of depeaking, but no analytical treatment of the rolling hub is available in the literature.[2]

The purpose of the present paper is to provide such a treatment. The goal is to develop formal frameworks where the rolling-hub trade-off is clearly illustrated and to explore the implications of the resulting models. The paper constructs and analyzes two related models. The first is a continuous spatial model, where flight departure times are represented by locations on a circle capturing the daily clock, and where a monopoly carrier serves only two spoke cities out of the hub. Narrowing the gap between in- and outbound flights reduces layover time for connecting passengers, but congestion cost per flight rises as the schedule is compressed. The optimal spacing of flights balances the resulting gain and loss. Although the addition of other carriers may leave the hub airline's flight spacing unchanged, a high overall flight density will cause the carrier to bunch its flights more closely than it would in isolation.

To allow more-interesting interaction effects between carriers to emerge, the analysis then turns to a discrete spatial model, where flights congest one another only if they operate in the same discrete period. The model has two periods and n spoke cities, and a layover cost now arises only if the flights used by a connecting passenger operate in different periods. This cost can be eliminated if the airline concentrates all its flights in one period, but the resulting high level of congestion may make this option unappealing. While an increase in the cost per unit of layover time in the continuous model leads to a marginal reduction in the optimal gap between flights, the effect of a higher layover cost (now a cost for the entire layover rather than a cost per unit of time) is discontinuous in the discrete model. The carrier evenly splits its flights between the periods if this cost is low relative to the costs of congestion, while it concentrates all its flights in one period if the cost is high.

The addition of fringe carriers, each of which operates a single flight, alters this pattern, causing the hub airline to fully or partially concentrate its flights, with the fringe carrier's flights all operating in the other period. However, the equilibrium when two hub carriers share the hub airport mirrors the outcome in the monopoly case. When the layover cost is low, both carriers evenly split their flights between the periods, while a high layover cost leads the carriers to concentrate their flights in different periods.

Although, by eliminating market power, the assumption of perfectly elastic demand makes the monopoly equilibrium efficient, the presence of other carriers leads to inefficient equilibria. The culprit is a congestion externality, which is generated when a carrier shifts a flight into a period where another carrier is operating. As a result, equilibria where several carriers are present in the same period are inefficient, with a movement to the social optimum typically requiring greater separation of the airlines. As usual, congestion tolls can remedy this inefficiency.

Overall, the paper illustrates the trade-off between convenient connections and airport congestion, showing that a higher layover cost leads an airline to reduce layover time by concentrating its flights. The carrier takes self-imposed congestion into account, but since it ignores congestion imposed on other carriers, corrective action may be needed under certain conditions.

The plan of the paper is as follows. Section 2 develops the continuous spatial model. Section 3 develops the discrete spatial model with a monopoly carrier. Section 4 adds fringe carriers to the model, while section 5 analyzes the case with two hub carriers. Section 6 offers conclusions.

2. A Continuous Spatial Model

To illustrate the trade-off between convenient connections and airport congestion in simple fashion, this section of the paper develops a continuous spatial model, where individual flight times are explicitly captured. Let the daily clock be represented by a circle, with the time of a flight indicated by its position on the circle, as shown in Figure 1. This figure is based on the simplest possible hub-and-spoke network, where a monopoly airline serves two endpoint cities, A and B, out of a hub airport H. Passengers travel in three city-pair markets AH, BH and AB, with these trips facilitated by aircraft movements among the cities. One plane flies back and forth between A and H and another plane flies back and forth between B and H, as seen in Figure 2. For simplicity, all passengers are assumed to make one-way trips, allowing the analysis to focus on the A-to-H and H-to-B flights. The first flight carries A-originating passengers whose destinations are H or B, and the second flight carries H-originating passengers flying to B as well as connecting A-to-B passengers. A parallel discussion applies to B-to-H and H-to-A flights (see below).

Suppose the airline operates two A-to-H and two H-to-B flights per day, as shown in Figure 1, with the flights evenly spaced around the circle, which has circumference T. Passengers in each market, whose desired departure times are uniformly distributed around the circle, will choose the most convenient departure times. A-to-H passengers whose desired departure times are located on the upper half the circle will chose the A-to-H flight at the top of the circle, while the remaining A-to-H passengers will choose the A-to-H flight at the bottom of the circle. The average schedule delay (the distance between the preferred and actual departure times) for these passengers is $T/8$, equal to one half the maximum gap between a preferred time and the departure time. A-to-B and H-to-B passengers will self-select across flights in the same fashion. H-to-B passengers with preferred departure times on the left (right) half of the circle will choose the H-to-B flight on the left (right) of the circle, and A-to-B passengers with preferred departure times on the upper (lower) half of the circle will choose the A-to-H flight on the top (bottom) of the circle, with average schedule delay in each case again $T/8$.

The A-to-B trip involves a connection (with a layover) at the hub, and the connecting time is given by the gap between the A-to-H and H-to-B flights, which under the even spacing shown in the figure equals $T/4$. When the airline instead operates f flights on each spoke route, leading to $2f$ flights in total, the schedule-delay and layover-time values are given by $T/4f$ and $T/2f$ respectively.

Instead of spacing its flights evenly, the airline could move its A-to-H and H-to-B flights closer together, which would reduce the connecting time of A-to-B passengers. In Figure 1, the H-to-B flight locations would move counterclockwise, becoming closer to the A-to-H flights and leading to the creation of two flight "banks."[3] A byproduct of this movement, however, would be an increase in airport congestion, as the operating times of the two types of flights would be closer together. To capture this effect in stylized fashion, suppose that the operating cost per flight is a decreasing function of the distance to the closest flight. With even spacing, flights are equidistant, separated by $T/2f$, but if the gap between the A-to-H and H-to-B flights were reduced by d, the distance to the closest flight becomes $T/2f - d \equiv M$, a value that applies to both A-to-H and H-to-B flights. Note that M also equals the connecting time of A-to-B passengers.

Let the airline's congestion cost per flight as a function of this minimum distance be given by $K(M)$, and let congestion cost per passenger be given by $C(M)$. Both the $C(\cdot)$ and $K(\cdot)$ functions are decreasing as well as strictly convex. In addition, let β denote a passenger's cost per unit of layover time, and let γ denote the cost per unit of schedule-delay time, so that layover cost is βM and schedule-delay cost is $\gamma T/4f$. The operating cost per flight, not including congestion cost, is denoted θ.

The demand for flights in each market is assumed to be perfectly elastic, with travel benefit equal to a common value b in all markets. Then, for A-to-H and H-to-B passengers, travel benefit net of the costs of congestion and schedule delay is $b - C(M) - \gamma T/4f$. For an A-to-B passenger, travel benefit net of these same costs as well as the layover cost is $b - 2C(M) - \gamma T/4f - \beta M$. Note that a connecting passenger incurs congestion cost twice since two flights are used.[4]

The monopoly airline will set the fare in each market equal to the relevant net travel benefit, so as to exhaust the consumer's surplus from travel.[5] Fares thus reflect passenger losses from schedule delay, congestion, and layover time, which means that the airline must reduce the amount it charges when any of these costs rises. As a result, the carrier will take into account the effects of schedule delay, congestion, and extended layovers on passenger welfare.

Normalizing the number of passengers per market to unity, and setting fares equal to net travel benefits, the airline's profit is then

$$\pi \;=\; 2[b - C(M) - \gamma T/4f] \;+\; [b - 2C(M) - \gamma T/4f - \beta M] \;-\; 2f[K(M) + \theta]. \qquad (1)$$

The first expression in (1) is revenue from A-to-H and H-to-B passengers (whose total volume equals 2), while the second is revenue from the unitary volume of A-to-B passengers. The last term is airline costs.

Since the choice of the number of flights f is not relevant to the issues at hand, consider the choice of d holding f fixed.[6] Substituting $M = T/2f - d$ in (1), the first-order condition for choice of d is

$$\pi_d \equiv 4C'(T/2f - d) + 2fK'(T/2f - d) + \beta = 0. \qquad (2)$$

This condition, which is assumed to be satisfied for some d between 0 and $T/2f$, says that the increase in passenger congestion costs from closer flight spacing ($-4C' > 0$) plus the increase in airline congestion cost ($-2K' > 0$) should equal the benefit β from lower layover costs. Given convexity of the cost functions, the optimal d rises with β, so that the chosen layover time naturally falls as layover cost rises. In addition, the optimal d falls as f increases, with the lower d serving to hold M constant at the optimal value M^* determined by β as f rises.[7]

While the optimality rule embodied in (2) is unsurprising, its usefulness lies in formalizing the trade-off between convenient flight connections and airport congestion in a model that includes some essential features of hub-and-spoke networks. This model could be extended by introducing additional carriers. "Fringe" carriers, for example, each of which operates a single flight, could locate their flights between the hub carrier's banks. To see the effect, suppose that a single fringe carrier, operating out of H, enters between each of the flight banks of the hub carrier, locating at the midpoint of the gap to minimize congestion. The fringe carriers could serve cities A or B or some other cities, but because of perfectly elastic demand, any competition they provide to the hub carrier has no effect on fares. If, after entry, the distance between a fringe carrier's flight and the nearest flight of the hub carrier is greater than M^*, entry has no effect on the carrier's flight-spacing choice. On the other hand, if this distance is less than M^*, then the hub carrier will move the flights within each bank closer together. Since the gap between the fringe and hub-carrier flights is smaller in this case than the gap between the hub carrier's own flights, moving the latter flights closer together has a double benefit: it increases the distance to the nearest flight (operated by the fringe carrier), thus reducing congestion, and it reduces the layover time of connecting passengers. However, once the flights of all the carriers are evenly spaced, there is no further gain to the hub carrier from narrowing the gap between its flights. The reason is that the resulting M is less than M^*, making the derivative in (2) negative.

Similar discussion applies when multiple fringe-carrier flights locate between the hub carrier's banks. In addition, entry by a second hub carrier has an effect similar to that of fringe-carrier entry. If the gap between the banks of the two hub carriers is greater than M^*, then entry has no effect on the spacing of the original hub carrier's flights, with both carriers behaving as if they were operating in isolation. Otherwise, however, the intra-bank flight distance would fall until all the flights of the two carriers are evenly spaced at a distance smaller than M^*.

While it illuminates a basic tradeoff, the model thus lacks rich implications regarding the interaction between different carriers in a setting where layover time matters. The model developed in the next section leads to more-interesting interaction effects, and it does so by collapsing the continuous spatial setting into a discrete spatial framework in which flights operate in discrete periods and must be present in the same period in order to congest one another. This model also allows a more accurate depiction of airline hub-and-spoke network operations.

3. A Discrete Spatial Model

To construct the discrete spatial model, let the day be broken up into a series of intervals, with each interval containing two discrete periods denoted 0 and 1. A monopoly hub airline serves n spoke cities out of the hub, and it has previously chosen to operate a single flight to each endpoint during each of the day's intervals. Its remaining choice is how to allocate the flights across the 0 and 1 periods within a given interval. This same choice is repeated for each interval over the day.

Let r denote the number of endpoints served during period 0, with $n - r$ endpoints served during period 1. Congestion within a period depends on the number of flights operated during the period. The airline's congestion cost per flight is assumed to equal a constant $k > 0$ times the number of flights operated during the period. Therefore, the airline's congestion cost per flight is $k(n-r)$ in period 1 and kr in period 0. The congestion cost experienced by a passenger takes the same form but with a different constant $c > 0$. While this linear specification lacks generality, it facilitates the analysis.

As usual, passengers traveling between the hub and a spoke city take a single flight, while passengers traveling between two spoke cities make a connection at the hub. The layover cost for this connection depends on whether the connecting flights operate during the same period or different periods. If they operate during the same period, the layover cost is zero. But if they operate during different periods, the layover cost is given by $y > 0$. As a result, the hub airline can eliminate layover costs by concentrating its flights in a single period, but the downside is a high level of congestion during that period. This trade-off is the focus of the analysis.

The number of passengers per city-pair market is again normalized to unity. As a result, the number of hub-to-spoke passengers traveling in period 1 equals $n - r$, while the number of spoke-to-spoke passengers using flights that operate in period 1 is $(n - r)(n - r - 1)/2$. Similarly, the number of hub-to-spoke passengers in period 0 equals r, while the number of spoke-to-spoke passengers using flights that operate in period 0 is $r(r - 1)/2$. The number of spoke-to-spoke passengers whose flights operate in different periods equals the total number of spoke-to-spoke passengers, $n(n - 1)/2$, minus the two previous expressions, a difference that equals $(n - r)r$.

Adapting the fare expressions from section 2 while suppressing the schedule-delay terms, the fares for hub-to-spoke passengers in the two periods are $b - c(n - r)$ and $b - cr$. For spoke-to-spoke passengers, the fares are $b - 2c(n - r)$ for those traveling in period 1 and $b - 2cr$ for those traveling in period 0. For spoke-to-spoke passengers whose flights operate in different periods, requiring a layover, the fare is $b - cr - c(n - r) - y = b - cn - y$. Thus, as in the continuous spatial model, fares reflect passengers losses from congestion and layover time.

Making use of all this information, the profit of the monopoly carrier is given by

$$\begin{aligned} \pi \;=\; & (n-r)(b-c(n-r)) \;+\; r(b-cr) \;+ \\ & \frac{(n-r)(n-r-1)}{2}(b-2c(n-r)) \;+\; \frac{r(r-1)}{2}(b-2cr) \;+ \\ & (n-r)r(b-cn-y) \;-\; (n-r)^2k \;-\; r^2k \;-n\theta. \end{aligned} \tag{3}$$

The first two terms are revenue from hub-to-spoke passengers, the next three terms are revenue from spoke-to-spoke passengers, and last three terms are the airline's costs (congestion cost in period 0, for example, equals r times cost per flight kr, or r^2k). Thus, while passenger and airline congestion costs per flight are linear in the number of flights, total costs are quadratic in r and $n-r$.

The airline's choice problem is how to allocate its flights between the two periods, which involves choosing r to maximize (3). The first and second derivatives of (3) with respect to r are given by

$$\pi_r = (n-2r)(2k+2cn-y) \tag{4}$$

$$\pi_{rr} = -4(k+cn) + 2y. \tag{5}$$

If $y < 2(k+cn)$ holds, so that the layover cost is low relative to the costs of congestion (as represented by $(k+cn)$), then $\pi_{rr} < 0$ holds and profit is strictly concave. The solution is then found by setting π_r equal to zero, which yields $r = n/2$ from (4). However, if $y \geq 2(k+cn)$ holds, with layover costs high and $\pi_{rr} \geq 0$, then π is convex, and the optimal r is given by a corner solution, with $r = 0$ or $r = n$.

Figure 3 provides insight into these outcomes. The figure comes from noting that profit equals travel benefit, given by $[n + n(n-1)/2]b$, minus layover costs, which equal $(n-r)ry$, minus congestion costs, which are given by the remaining elements in (3). By inspection, layover costs are concave in r, with the second derivative equal to $-2y < 0$ (the negative of second term on the RHS of (5)). This concavity, which is shown in the upper panels of Figure 3, means that layover costs are minimized at $r = 0$ or $r = n$. Congestion costs are convex, with second derivative $4(k+cn) > 0$ (the negative of first term on the RHS of (5)), a conclusion can also be shown to hold with nonlinear convex cost functions. This convexity, which is also shown in the upper panels of Figure 3, means that congestion costs are minimized at $r = n/2$. Given these differences in curvature, total costs can be either concave or convex, and the outcome depends on the magnitude of y relative to $2(k+cn)$. In Figure 3a, y is small and the convexity of congestion costs dominates, making total cost convex. In Figure 3b, y is large and the concavity of layover costs dominates, making total costs concave. Finally, since profit equals total travel benefit (which is independent of r) minus total costs, these convexity/concavity conclusions are reversed in Figure 3c, which shows that profit is concave (convex) when y is small (large), with the maximum lying at $r = n/2$ ($r = 0$ or $r = n$). Since the profit maximum thus minimizes congestion plus layover costs, the social planner's goal, it follows that the monopolist's choice is socially optimal. Summarizing these results yields

Proposition 1. *If the layover cost y is small relative to the costs of congestion, then the monopoly carrier divides its flights evenly between the periods. Otherwise, the carrier concentrates all its flights in one period. These outcomes are socially optimal.*

Therefore, a high layover cost encourages concentration of flights, as intuition would suggest, but this effect is discontinous, with flights entirely concentrated when y is above the critical value $2(k + cn)$, and equally divided between the periods otherwise. Proposition 1 implies that, when layover costs are high, the monopoly carrier will concentrate all of its flights in banks, which can be assumed (without loss of generality) to be present in period 1 of each of the day's intervals (so that $r = 0$). The banks will be separated by the empty 0 periods in the various intervals.[8]

4. Adding Fringe Carriers to the Discrete Model

4.1. The equilibrium

Now suppose that $m < n$ fringe carriers are present, each operating a single flight from the hub to one of the hub carrier's spoke cities or some other destination. With perfectly elastic demand, the presence of the fringe carriers has no effect on fares, as noted above, and both the hub and fringe carriers can fill all the flights they choose to operate.

Let u denote the number of fringe carriers providing service in period 0, with $m - u \geq 0$ carriers providing service in period 1. Assuming these carriers have the same costs as the hub carrier,[9] the profits of the fringe carriers operating in periods 0 and 1 are, respectively, given by

$$b - (c+k)(r+u) \ - \ \theta, \qquad b - (c+k)(n-r+m-u) \ - \ \theta. \tag{6}$$

The first expression is comprised of the carrier's revenue from its unitary mass of passengers,

$b - c(r+u)$, minus its costs, $k(r+u)+\theta$, and similarly for the second expression.

The equilibrium division of fringe carriers across periods is determined by equalization of profit per carrier between the periods, which (from (6)) requires equal flight totals in the two periods. Setting $r+u$ and $n-r+m-u$ equal and solving for u yields an equilibrium value of $u = (n+m)/2 - r$, which yields a common flight total of $r+u = n-r+m-u = (m+n)/2 \equiv v$. Note that the equilibrium u is decreasing in r, which means that, as the hub carrier shifts its flights toward period 0 (raising r), the fringe carriers shift their flights toward period 1 (reducing u) to maintain equality of profits. This u solution is relevant as long as it lies between 0 and m, which requires that r lies between $(n+m)/2$ and $(n-m)/2$.

To see how the hub carrier's profit varies with r over this range, the r and $n-r$ terms inside the fare expressions in the first two lines of (3) are replaced by v, the n term inside the fare expression in third line of (3) is replaced by $2v$, and the airline congestion cost expressions are replaced by $(n-r)kv + rkv = nkv$. Once these substitutions are made, the second derivative of the profit function equals $2y > 0$. With the function thus convex over the r interval $[(n-m)/2, (n+m)/2]$, it follows that the endpoints of this interval (which lead to u values of m and 0) yield the highest profit, which is equal at both endpoints. The profit maximum therefore lies at one of the endpoints or at an r value outside the interval. Given symmetry, let $\widehat{r} = (n-m)/2$ and focus on r values in the interval $[0, \widehat{r}]$. Over this interval, $u = m$ holds (generating a kink in the profit function at $\widehat{r}$), with all the fringe carriers operating in period 0.

This convex portion of the profit function is shown in Figure 4. To derive the behavior of profit over the $[0, \widehat{r}]$ interval, flight totals in periods 1 and 0, which equal $n-r$ and $r+m$, are substituted into the hub carrier's profit expression, which then equals

$$\begin{aligned}\widetilde{\pi} \;=\; & (n-r)(b-c(n-r)) \;+\; r(b-c(r+m)) \;+ \\ & \frac{(n-r)(n-r-1)}{2}(b-2c(n-r)) \;+\; \frac{r(r-1)}{2}(b-2c(r+m)) \;+ \\ & (n-r)r(b-c(n+m)-y) \;-\; (n-r)^2k \;-\; r(r+m)k \;-n\theta. \qquad (7)\end{aligned}$$

The first and second derivatives of this function are

$$\widetilde{\pi}_r = -(k+cn)(m-2n+4r)-(n-2r)y \tag{8}$$

$$\widetilde{\pi}_{rr} = -4(k+cn)+2y. \tag{9}$$

To locate the optimal r, suppose first that $y \geq 2(k+cn)$ holds, so that profit is convex over $[0,\widehat{r}]$ from (9). Since $\widetilde{\pi}_r$ evaluated at $\widehat{r}$ equals $m(k+cn-y)$, which is negative under convexity, it follows that profit is downward sloping at $\widehat{r}$. Since profit is convex, it follows that $r=0$ is optimal in this case (which is not shown in Figure 4).

Next suppose that $y < 2(k+cn)$ holds, so that profit is strictly concave. If $y < k+cn$ also holds, then profit is upward sloping at $\widehat{r}$, which means that $\widehat{r}$ is optimal. This case is shown in the lower curve in the $[0,\widehat{r}]$ interval in Figure 4. If, on the other hand, $k+cn \leq y < 2(k+cn)$ holds, profit is concave and downward sloping at $\widehat{r}$, which means that the optimum lies either at $r=0$ or some interior value in $[0,\widehat{r}]$. Setting $\widetilde{\pi}_r$ in (8) equal to zero, the resulting value is negative if $y > (2n-m)(k+cn)/n \equiv \widehat{y}_{eq}$, making the optimal r zero. When the reverse inequality holds, an interior r is optimal. These cases are shown in the upper and middle curves in the $[0,\widehat{r}]$ interval in Figure 4. Letting r_{eq} denote the equilibrium value of r and x_{eq} denote the interior optimum, summarizing yields

$$r_{eq} = \begin{cases} 0 & \text{if } 2(k+cn) \leq y \quad (\textit{convex case}) \\ 0 & \text{if } \widehat{y}_{eq} \leq y < 2(k+cn) \quad (\textit{concave case}) \\ \frac{k(m-2n)+n[c(m-2n)+y]}{2y-4(k+cn)} \equiv x_{eq} < \widehat{r} & \text{if } k+cn < y < \widehat{y}_{eq} \\ \frac{n-m}{2} \equiv \widehat{r} & \text{if } y \leq k+cn, \end{cases}$$

$$\text{where} \quad \widehat{y}_{eq} = (2n-m)(k+cn)/n. \tag{10}$$

This analysis shows that, in the presence of fringe carriers, the hub carrier concentrates its flights in period 1, either partially or totally, never finding an equal split to be optimal. While the carrier, as a monopolist, evenly split its flights when $y < 2(k+cn)$, it now continues to fully concentrate them in period 1 (setting $r=0$) when y is as low as $\widehat{y}_{eq}$. For smaller y values, the carrier sets r somewhere between zero and $(n-m)/2$, so that the period 1 flight volume is never less than $(n+m)/2$.

To understand these conclusions intuitively, observe that, since the fringe carriers allocate themselves between periods so as equalize total flights (at a level v), congestion is unaffected as the hub carrier reduces r past $n/2$ toward $\widehat{r}$, while the greater period-1 concentration of flights lowers layover costs. Therefore, the movement toward $\widehat{r}$ unambiguously raises profit. However, once r reaches $\widehat{r}$ (and u becomes fixed at m), a further reduction affects congestion in the two periods as well as layover costs. The maximization problem then resembles the one faced by the monopolist of section 3, except for the presence of m fringe flights in period 0. Reflecting this similarity, the solution is either an endpoint of the interval $[0, \widehat{r}]$ or an interior point, paralleling the monopoly outcome.[10]

Note that the resulting traffic concentration by the hub carrier is more realistic than the outcome in the monopoly model, where concentration is completely absent below an intermediate value of y. Note also that the critical value $\widehat{y}_{eq}$ above which flights are fully concentrated is an increasing function of the congestion-cost expression $k + cn$. Therefore, it remains true that the hub carrier compares the layover cost to congestion costs in deciding whether to concentrate its flights.

A final observation is that, in constrast to the monopoly case, the effect of y on r is continuous rather than discontinuous in the presence of fringe carriers. As y increases up to $k + cn$, r is constant at $\widehat{r}$, with the solution shifting to x_{eq} in continuous fashion as y passes $k + cn$. As y increases further, r declines smoothly, reaching zero at $\widehat{y}_{eq}$, where it remains as y increases beyong $\widehat{y}_{eq}$.

4.2. The social optimum

As in section 3, maximizing total airline profit, which now includes the profit of the fringe carriers, leads to minimization of total cost. Therefore, total profit is the appropriate objective function for the social planner, and it is given by

$$
\begin{aligned}
W \;=\;& (n-r+m-u)(b-c(n-r+m-u)) \;+\; (r+u)(b-c(r+u)) \;+ \\
& \frac{(n-r)(n-r-1)}{2}(b-2c(n-r+m-u)) \;+\; \frac{r(r-1)}{2}(b-2c(r+u)) \;+ \\
& (n-r)r(b-c(n+m)-y) \;-\; (n-r+m-u)^2k \;-\; (r+u)^2k \;-(n+m)\theta.
\end{aligned}
\tag{11}
$$

The first line of (11) is the combined fringe and hub-carrier hub-to-spoke revenues, while the second line is the intraperiod spoke-to-spoke revenue, adjusted for the extra fringe flights. The last line is interperiod spoke-to-spoke revenue minus the combined costs of the carriers.

The second derivatives of (11) are

$$
W_{rr} \;=\; -4(k+cn)+2y \tag{12}
$$

$$
W_{uu} \;=\; -4(c+k) \;<\; 0. \tag{13}
$$

In addition, $W_{ru} = -2(c+2k+cn)$ and the Hessian determinant equals $-4(c^2(n-1)^2 + 2y(c+k)) < 0$. With this determinant negative, solutions to the first-order conditions cannot represent a maximum.

The maximum can be found using an approach similar to the one underlying Figure 4 in the equilibrium analysis. To begin, observe that since $W_{uu} < 0$ holds, an optimal value of u conditional on r can be found via the first-order condition $W_u = 0$. This condition yields u as a downward sloping function of r over an interval of r values with a lower endpoint $\widetilde{r}$, where $u = m$, and an upper endpoint where $u = 0$. When this function is used to eliminate u in the W expression in (11), total profit becomes a complicated function of r. Just as in the equilibrium analysis, this function is convex, matching the convex curve to the right of $\widehat{r}$ in Figure 4. However, the new curve's lower endpoint, located at $\widetilde{r}$, lies to the right of the endpoint in Figure 4, with $\widetilde{r} > \widehat{r}$.

As in the equilibrium analysis, the socially optimal r lies in the interval $[0, \widetilde{r}]$ below this endpoint. To find the optimum, W is evaluated with $u = m$, with the resulting function denoted $Z(r)$.[11] But since the derivative Z_r is negative at $\widetilde{r}$, the optimum lies to left of $\widetilde{r}$, in contrast to the equilibrium analysis (where $r = \widehat{r}$ could be an equilibrium). Further computation shows that if $y \geq 2(k + cn)$ holds, then Z is convex and the optimal r is zero, matching a similar outcome in the equilibrium analysis. However, if $y < 2(k + cn)$ holds instead, Z is concave. Then, the optimal r, denoted r_{opt}, is either equal to zero, which happens when y lies above a critical value denoted $\widehat{y}_{opt}$, or equal to an interior value given below. These outcomes parallel the equilibrium cases corresponding to the upper and middle curves in Figure 4, and details are provided in the appendix. Summarizing yields

$$r_{opt} \quad = \quad \begin{cases} 0 & \text{if } 2(k+cn) \ \leq \ y \quad (\textit{convex case}) \\ 0 & \text{if } \widehat{y}_{opt} \ \leq y \ < \ 2(k+cn) \quad (\textit{concave case}) \\ \frac{2k(n-m)+c(2n^2-m(1+n))-ny}{4(k+cn)-2y} \equiv x_{opt} & \text{if } y \ < \ \widehat{y}_{opt}, \end{cases}$$

$$\text{where} \quad \widehat{y}_{opt} \quad = \quad \frac{2k(n-m)+c(2n^2-m(1+n))}{n}. \tag{14}$$

From (14), r_{opt} equals $x_{opt} > 0$ when y is small and equals zero otherwise. Note that, while the critical value $\widehat{y}_{opt}$ is not explicitly a function of $k + cn$, it is increasing in both k and c. Therefore, the optimality of concentration still depends on a comparison between layover and congestion costs. The critical value $\widehat{y}_{opt}$ in (14) is easily seen to be less than $\widehat{y}_{eq}$ from (10).

Figure 5 summarizes the information in (14) without presenting a graph. The figure also summarizes the information from (10) and Figure 4 in the same format, and it presents a comparison between the equilibrium and socially optimal r values, allowing an evaluation of the efficiency of the equilibrium. Drawing on (10) and (14), the contents of Figure 5 are displayed formally as follows, where the roman numerals label the cases shown at the bottom of the figure:

$$r_{eq} \quad = \quad \begin{cases} 0 = r_{opt} & \text{if } y \geq \widehat{y}_{eq} & \text{(i)} \\ x_{eq} > 0 = r_{opt} & \text{if } \widehat{y}_{opt} \leq y < \widehat{y}_{eq} & \text{(ii)} \\ x_{eq} > x_{opt} = r_{opt} & \text{if } k+cn \leq y < \widehat{y}_{opt} & \text{(iii)} \\ \widehat{r} \geq x_{opt} = r_{opt} & \text{if } c(n-1) \leq y < k+cn & \text{(iv)} \\ \widehat{r} < x_{opt} = r_{opt} & \text{if } y < c(n-1) & \text{(v)} \end{cases} \tag{15}$$

As seen from Figure 5 and (15), $r_{eq} \geq r_{opt}$ holds, implying the hub carrier's period 0 flights are either excessive or optimal, unless y is small (lying below $c(n-1)$), in which case $r_{eq} < r_{opt}$.

Summarizing these findings along with those of the equilibrium analysis yields

Proposition 2. *In the presence of fringe carriers, the hub carrier fully or partially concentrates its flights, with its period 1 flight volume at least as large as* $(n+m)/2$*. Unless the layover cost* y *is small relative to the costs of congestion, the chosen extent of concentration is either socially optimal or inefficiently low, with* r *too large in the latter case.*

Since the hub-carrier ignores the congestion imposed on the fringe carriers when it shifts another flight to period 0, intuition would suggest that its chosen r should be inefficiently large, reflecting the presence of an externality. While this relationship holds as expected for intermediate values of y, $r_{eq} < r_{opt}$ holds instead when y is small. To understand this apparent anomaly, note that the above intuition applies to a situation in which the equilibrium and social optimum both correspond to interior solutions, where the externality-induced difference in the objection-function slopes can have an effect. The anomaly, however, is associated with the presence of a corner solution for r_{eq}, where this slope difference may not have the anticipated impact.

To understand this point, first recall that the curve used in finding the social optimum has generally the same form as the curve used in finding the equilibrium, shown in Figure 4. However, the kink point at $\widetilde{r}$ on the optimum curve lies to the right of the kink point $\widehat{r}$ on the equilibrium curve, and the optimum curve is downward sloping immediately to the left of $\widetilde{r}$ (recall that the equilibrium curve may slope up to the left of $\widehat{r}$, as seen in Figure 4). Next, it can be shown that, to the left of $\widehat{r}$, the equilibrium curve has a slope that is algebraically larger (less negative or more positive) than the slope of the optimum curve at a common value of r.[12] The reason is that the optimum curve captures the benefits to the fringe carriers of the lower congestion that comes from a decline in r, a gain that is not captured by the equilibrium curve. Given this slope difference, it follows that, at an interior equilibrium (an r value where the hub carrier's profit slope is zero), the slope of the welfare function Z is negative, implying that the socially optimal r lies farther to the left, with $r_{opt} < r_{eq}$. Note that r_{opt} in this case could itself be interior, or it could equal zero.

Suppose, on the other hand, that the equilibrium r is the corner solution $\widehat{r}$, where the hub carrier's profit is upward sloping. In this case, Z could either be upward sloping itself at $\widehat{r}$ (although less steeply), or downward sloping. In the upward-sloping case, concavity of Z means that its derivative is zero somewhere to the right of $\widehat{r}$, with the optimum thus satisfying $\widehat{r} < r_{opt} < \widetilde{r}$ (recall that Z_r is negative at $\widetilde{r}$). In this anomalous case, $r_{eq} < r_{opt}$ holds. Since the slopes of both curves are decreasing in y, and since this case requires a positive slope for each curve at $\widehat{r}$, the case emerges for the smallest values of y. In the case where $r_{eq} = \widehat{r}$ holds but Z is downward sloping at $\widehat{r}$, r_{opt} lies to the left of $\widehat{r}$ and no anomaly arises. This case requires somewhat larger values of y, but if y is larger still, then the hub carrier's profit slope at $\widehat{r}$ switches from positive to negative, and the cases considered in the previous paragraph apply.

The second-to-last case above shows that a corner solution for r_{eq} need not always lead to an anomalous outcome, with $r_{opt} < r_{eq}$ being possible. Observe, however, that this outcome requires the positive profit slope at $\widehat{r}$ to be near zero, so that the Z slope can be negative at $\widehat{r}$. But the corner solution is then "almost" interior, so that the initial intuition regarding the relation between r_{eq} and r_{opt} applies.

The coincidence of the equilibrium and optimum for the highest values of y (where $r_{eq} = r_{opt} = 0$) is also anomalous, but it has an even simpler intuitive explanation. With the hub carrier completely separated from the fringe in equilibrium, no congestion externality arises, making the equilibrium socially optimal.

5. The Discrete Model with Two Hub Carriers

5.1. Equilibrium

Suppose that, instead of facing the group of fringe carriers, the original airline shares the hub with another hub-and-spoke carrier, which serves a collection of spoke cities that may include some of the n original endpoints. While it is uncommon to see separate hubs sharing an airport (the Chicago-O'Hare hubs of United and American provide a rare example), the case is nevertheless interesting.

Let the carriers be denoted 1 and 2, and suppose carrier 1 serves more spoke endpoints than 2, with $n_1 > n_2$ (United is larger than American at Chicago). Letting r_1 and r_2 denote the numbers of cities served by the carriers in period 0, the profit of carrier 1 is given by

$$\pi^1 = (n_1 - r_1)(b - c(n_1 - r_1 + n_2 - r_2)) \; + \; r_1(b - c(r_1 + r_2)) \; + \\ \frac{(n_1 - r_1)(n_1 - r_1 - 1)}{2}(b - 2c(n_1 - r_1 + n_2 - r_2)) \; + \; \frac{r_1(r_1 - 1)}{2}(b - 2c(r_1 + r_2)) \; + \\ (n_1 - r_1)r_1(b - c(n_1 + n_2) - y) \; - \; (n_1 - r_1)(n_1 - r_1 + n_2 - r_2)k \; - \; r_1(r_1 + r_2)k \; - n_1\theta. \tag{16}$$

Note that this profit function parallels that of the fringe case, with $r_2 = u$ and $n_2 = m$.

The r_1 derivatives of the profit function are

$$\pi^1_{r_1} = (k + cn_1)(2n_1 + n_2 - 4r_1 - 2r_2) \; - \; (n_1 - 2r_1)y \tag{17}$$

$$\pi^1_{r_1 r_1} = -4(k + cn_1) \; + \; 2y. \tag{18}$$

The profit function for carrier 2 is gotten by reversing the subscripts in (16), and its derivatives come from reversing the subscripts in (17) and (18).

If $y < 2(k + cn_2) < 2(k + cn_1)$, then both profit functions are strictly concave in their r arguments. The Nash equilibrium is then found by equating the r derivatives to zero and solving simultaneously, yielding $r_1 = n_1/2$ and $r_2 = n_2/2$. Since the reaction functions are easily seen to be downward sloping, this solution, where each carrier evenly splits its flights between the periods, is a stable equilibrium.

If, on the other hand, $2(k + cn_2) < 2(k + cn_1) \leq y$, both profit functions are convex in their r arguments, and corner solutions are optimal in which a carrier operates in the period where the other is not present. The Nash equilibrium thus has $r_2 = n_2$ and $r_1 = 0$ (assuming that carrier 2 operates in period 0).

If $2(k + cn_2) \leq y < 2(k + cn_1)$, then carrier 2's profit function is convex and 1's function is concave. Since carrier 2 will then wish to concentrate its flights in the period where carrier 2 has fewer flights, the Nash equilbrium can be found by setting $r_2 = n_2$ and choosing r_1 to maximize π^1 under this restriction, assuming that the solution has $r_1 < n_1/2$. When this maximization is carried out, the solution has $r_1 = 0$ provided that $n_2 > 2n_1/3$ holds, a condition that will be imposed for simplicity (otherwise an interior r_1 is optimal). See the appendix for details. Summarizing yields

$$(r_{eq1},\ r_{eq2}) \ = \ \begin{cases} (0,\ n_2) & \text{if } 2(k+cn_2)\ \leq\ y \\ (n_1/2,\ n_2/2) & \text{if } y\ <\ 2(k+cn_2). \end{cases} \tag{19}$$

Therefore, the carriers concentrate their flights in different periods when y is large relative to the costs of congestion and evenly split them between the periods when y is small, mirroring the discontinuous outcome in the monopoly case.

The intuition underlying this outcome is that, with r_2 fixed, carrier 1's problem is similar to that faced by a monopolist except that the baseline level of flights (those operated by carrier 2) is not zero in each of the periods. When carrier 2's flights are evenly split, carrier 1's problem is indeed formally identical to the monopoly problem (the flight baseline is equal across periods, although not zero), so that it will evenly split its flights, mimicking carrier 2, when y is small. Since carrier 2's perspective is the same, an even split is the Nash equilibrium when y is small. When y is large, each carrier will prefer a corner solution regardless of the other carrier's choice, with the corner solution that best avoids the other carrier being preferred. Therefore, when one carrier's flights are fully concentrated, the other carrier prefers to fully concentrate its flights in the other period, and this outcome is a Nash equilibrium.

5.2. Social optimum

As before, the social optimum maximizes the combined profit of the two hub carriers, which is given by

$$\begin{aligned} \widetilde{W} \ = \ & (n_1-r_1+n_2-r_2)[b-c(n_1-r_1+n_2-r_2)] \ + \ (r_1+r_2)[b-c(r_1+r_2)] \ + \\ & \left[\frac{(n_1-r_1)(n_1-r_1-1)}{2} \ + \ \frac{(n_2-r_2)(n_2-r_2-1)}{2}\right][b-2c(n_1-r_1+n_2-r_2)] \ + \\ & \left[\frac{r_1(r_1-1)}{2}+\frac{r_2(r_2-1)}{2}\right][b-2c(r_1+r_2)] \ + \\ & [(n_1-r_1)r_1 \ + \ (n_2-r_2)r_2]\,[b-c(n_1+n_2)-y] \ - \ (n_1-r_1+n_2-r_2)^2k \ - \\ & (r_1+r_2)^2k \ -(n_1+n_2)\theta. \end{aligned} \tag{20}$$

The second derivatives of $\widetilde{W}$ are

$$\widetilde{W}_{r_i r_i} \quad = \quad -4(k + cn_i) + 2y, \quad i = 1, 2. \tag{21}$$

The cross partial derivative equals $-2(2k + c(n_1 + n_2))$, and the Hessian determinant equals $-4[c^2(n_1 - n_2)^2 + y((2k + 2cn_1 - y) + 2k + 2cn_2)]$.

If (21) is positive for both carriers, then $\widetilde{W}$ is convex in each of the two r variables, so that a corner solution with $r_1 = 0$ and $r_2 = n_2$ is optimal. If both second derivatives in (21) are instead negative, then it is easily seen that the Hessian expression is negative, which implies that a solution to the first-order conditions cannot be optimal.

In this case, an approach similar to that followed in the fringe analysis can be used to find the solution. Negativity in (21) means an optimal value of r_2 can be found conditional on r_1. When this value is substituted in place of r_2, $\widetilde{W}$ becomes a function r_1. This function is convex, indicating that the endpoints of the r_1 interval over which r_2 lies between 0 and n_2 are preferred to interior values, paralleling the analysis of the fringe case. $\widetilde{W}$ is then evaluated at the lower endpoint of this interval (call it $\widehat{r}_1$) and the range below $\widehat{r}_1$, where $r_2 = n_2$. The resulting function of r_1 (denoted $Q(r_1)$) is concave, and its derivative at $\widehat{r}_1$ is negative, indicating that the optimal r_1 value is smaller. The optimal r_1 equals a value $s_{opt1} > 0$ if $y < y^*$ and equals zero if $y^* \leq y < 2(k + c)n_2$ (see below for the actual $s_{opt,1}$ and y^* values and see the appendix for details).

The case where $2(k + cn_2) \leq y < 2(k + cn_1)$ remains to be considered. In this case, $\widetilde{W}$ is convex in r_2, so that a corner solution is optimal. With r_2 set equal to n_2, $\widetilde{W}$ again reduces to $Q(r_1)$, but since $y > 2(k + cn_2) > y^*$ now holds, the optimal r_1 is zero. Summarizing yields[13]

$$(r_{opt1},\ r_{opt2}) \quad = \quad \begin{cases} (0,\ n_2) & \text{if } y^* \leq\ y \\ (s_{opt1},\ n_2) & \text{if } y\ <\ y^* \end{cases}$$

$$\text{where} \quad y^* \quad = \quad \frac{(n_1 - n_2)(2k + c(2n_1 + n_2))}{n_1} \quad <\ 2(k + cn_2),$$

$$s_{opt1} \quad = \quad \frac{-2k(n_1 - n_2)\ +\ c(-2n_1^2 + n_1 n_2 + n_2^2)\ +\ n_1 y}{-4(k + cn_1) + 2y} \quad < \quad \frac{n_1}{2}. \tag{22}$$

Thus, the smaller hub carrier's flights are fully concentrated at the social optimum while the larger carrier's flights are fully or partially concentrated, with the latter outcome obtaining when y is small.

Figure 6 summarizes the information in (22) along with the results of the equilibrium analysis from (19), and it presents a comparison between the equilibrium and socially optimal r_i values. The information in Figure 6 is displayed formally as follows:

$$r_{eq1} = \begin{cases} 0 = r_{opt1} & \text{if } 2(k + cn_2) \leq y \\ n_1/2 > 0 = r_{opt1} & \text{if } y^* < y < 2(k + cn_2) \\ n_1/2 > s_{opt1} = r_{opt1} & \text{if } y \leq y^* \end{cases}$$

$$r_{eq2} = \begin{cases} n_2 = r_{opt2} & \text{if } 2(k + cn_2) \leq y \\ n_2/2 < n_2 = r_{opt2} & \text{if } y^* < y < 2(k + cn_2) \\ n_2/2 < n_2 = r_{opt2} & \text{if } y \leq y^* \end{cases} \tag{23}$$

Thus, $r_{eq1} > r_{opt1}$ and $r_{eq2} < r_{opt2}$ hold unless y is large, in which case the equilibrium and optimum coincide.

Summarizing this information along with the results of the equilibrium analysis yields

Proposition 3. *When two hub carriers operate together, both carriers concentrate their flights in different periods when the layover cost is large relative to the costs of congestion and divide their flights equally between the periods when it is small. If the equilibrium has concentration of flights, the outcome is socially optimal. But if the equilibrium has flights evenly split, the outcome is inefficient, with the large carrier instead needing to fully or partially concentrate its flights and the small carrier needing to fully concentrate its own flights in a different period.*

The divergence between the equilibrium and the social optimum is again due to the presence of congestion externalities. Since the carriers congest one another but ignore the externality, they are insufficiently separated in situations where they are both present in each period. Thus, moving to optimum from such an equilibrium involves fuller separation of the carriers, with r falling for carrier 1 and rising for carrier 2.

Note that since the equilibrium, when it is inefficient, has interior r solutions for both carriers, a reversal (like that in the fringe case) of the expected relation between the equilibrium and optimal r values (with $r_{opt} > r_{eq}$) does not occur. Moreover, when the equilibrium involves full separation of the carriers, the intercarrier congestion externality is absent and the equilibrium is socially optimal, as in the fringe case. Note also that the undesirability of complete carrier separation when y is small (with $r_{opt1} > 0$) reflects the same force that leads to the even-split equilibrium: avoidance of excessive congestion in period 1.

A final observation is that, if the two hub carriers serve the same number of endpoints, with $n_1 = n_2$, then y^* in (22) is zero and $(r_{opt1},\ r_{opt2}) = (0, n_2)$ holds for all y. Thus, with equal-size carriers, full concentration of flights is always optimal, in contrast to the even split that characterizes the equilibrium for low values of y.

6. Conclusion

This paper has provided the first analysis of the trade-off between convenient flight connections and airport congestion. A continuous spatial model illustrates this trade-off in a framework where a small gap between flight operating times raises congestion while also shortening a connecting passenger's layover time. When the passenger's cost per unit of layover time rises, the monopoly airline chooses to narrow the gap between its flights, yielding shorter layovers but more congestion. The discrete spatial model, where flights congest one another only if they operate in the same discrete period, makes this layover-cost effect discontinuous in the monopoly case: the carrier concentrates (deconcentrates) its flights when this cost is high (low) relative to the costs of congestion. The presence of additional carriers may alter this pattern. When fringe carriers are present, the hub carrier always concentrates its flights, either partially or fully. But the presence of a second hub carrier leads to an equilibrium mirroring the monopoly outcome: the carriers concentrate their flights in different periods when the layover cost is high and deconcentrate them otherwise.

Welfare analysis of the fringe and 2-hub-carrier cases shows that, when the equilibrium involves less than full separation of the carriers, it is inefficient due to congestion externalities. Except for an anomalous outcome in the fringe case due to the presence of a corner solution, movement to the social optimum requires greater separation of the carriers.

The hub-carrier-plus-fringe case appears to best match reality at most of the world's hub airports, where a hub carrier's operations coexist with those of many other airlines, none of which has an appreciable share of the airport's flights.[14] The model predicts that these fringe carriers fully concentrate their flights between the banks of the hub carrier, with that carrier either fully concentrating its own flights or alternately choosing to intermix some of its flights with those of the fringe, operating banks that are not fully concentrated. In the latter case, the intermixing is most likely excessive, with hub carrier needing to more fully concentrate its own flights. This verdict, which applies unless the layover cost is very small, is due to uninternalized congestion, and it overturns any claim that the peaking of flights at hubs is excessive.

Efficient outcomes in both the fringe and 2-hub-carrier cases could be generated in the usual way by appropriate tolls. A hub carrier would be charged a toll equal to the external congestion damage created when it operates an extra flight in a particular period, damage that would depend on the number of other-carrier flights operating in that period. With tolls being charged, the profit function of a carrier would become aligned with the social planner's objective function. Note that, as in past models, the airline internalizes self-imposed congestion, which need not be taken into account in a toll system.

Future research could be devoted to extending the current framework in more-realistic directions. An extension involving two classes of spoke cities (large and small), leading to different market sizes, is unfortunately unworkable due to the added complexity, as are extensions allowing the hub carrier to choose the number of endpoints served or its flight frequency to a fixed number of endpoints. An extension that might be feasible, however, would introduce additional discrete periods beyond two, moving toward a compromise between the two-period and continuous models. Another extension would be to follow an emerging literature (see, for example, Czerny (2013)) by introducing airport concessions, which become more attractive when passengers experience long layovers. If the airline could share concession revenue with the airport in some fashion (via reduced airport charges, for example), then its incentive to deconcentrate its flights would increase.

A final question is whether the model can help explain the introduction of the rolling hub in the early 2000's and its demise a decade later. The introduction appeared to be spurred in part by the general decline in airline traffic after the terrorist attacks in 2001, while the rebanking of hubs occurred as traffic was rebounding following the Great Recession. While a decline in travel benefit b could capture such a demand shift, this change has no effect in the model. Alternatively, the drop in travel demand could be captured by a decline in the number m of fringe carriers. From (10), a drop in m raises the equilibrium level of r when it is positive, causing the hub carrier to deconcentrate its flights, as occurred with adoption of the rolling hub. An eventual rebound in m would explain the later rebanking of hubs. This explanation, however, holds the number of spoke endpoints fixed, even though a demand decline might have reduced n along with m. The model could also generate the introduction and demise of the rolling hub through a cyclical change in the relation between layover cost y and congestion costs, as captured in $k+cn$. However, such a cyclical pattern (with y intially falling in relation to $k+cn$ and then rising), may not be plausible. A final potential explanation for the rolling-hub cycle is pure experimentation: the airlines tried the rolling hub and eventually decided that it was no better than the traditional pattern of concentrated banks. This view is consistent with the findings of Katz and Garrow (2014), which suggest that airline cost savings from depeaking were roughly matched by revenue losses.

Appendix

A.1, Steps leading to (14)

The derivative of W with respect to u is

$$W_u \quad = \quad 2k[m + n - 2(r + u)] \;+\; c[2m + n + n^2 - 2nr - 2(r + 2u)]. \qquad (a1)$$

Setting $W_u = 0$ and solving, the optimal u conditional on r is

$$u(r) \quad \equiv \quad \frac{2m(c + k) \;+\; cn(1 + n - 2r) \;+\; 2k(n - 2r) - 2cr}{4(c + k)}. \qquad (a2)$$

This $u(r)$ solution equals m when $r = \widetilde{r}$, where

$$\widetilde{r} \quad = \quad \frac{2k(n - m) \;+\; c(n^2 + n - 2m)}{2(c + 2k + cn)}. \qquad (a3)$$

Substituting $u(r)$ into W, the resulting function has derivatives

$$W_r|_{u=u(r)} \quad = \quad -\frac{(n - 2r)(c^2(n - 1)^2 + 2y(c + k)}{2(c + k)} \qquad (a4)$$

$$W_{rr}|_{u=u(r)} \quad = \quad \frac{c^2(n - 1)^2 + 2y(c + k)}{c + k} \quad > \quad 0. \qquad (a5)$$

The function $Z(r)$ equals W with u set equal to m, and its derivatives are

$$Z_r \quad = \quad -2k(m - n + 2r) - c(m(1 + n) - 2n^2 + 4nr) - (n - 2r)y \qquad (a6)$$

$$Z_{rr} \quad = \quad -4(k + cn) + 2y < 0 \qquad (a7)$$

Setting $(a6)$ equal to zero and solving for r yields x_{opt}, an expression that is positive when $y < \widehat{y}_{opt}$.

A.2. Steps leading to (19)

The r_1 derivative of π^1 with $r_2 = n_2$ is

$$(k + cn_1)(2n_1 - n_2 - 4r_1) \; - \; (n_1 - 2r_1)y, \tag{a8}$$

and setting ($a8$) equal to zero and solving for r_1 yields

$$r_1 \;\; = \;\; \frac{k(2n_1 - n_2) + n_1(2cn_1 - cn_2 - y)}{4(k + cn_1) - 2y}. \tag{a9}$$

The denominator is positive, and the numerator can be shown to be negative when $2(k+cn_2) < y$ and $n_2 > 2n_1/3$ both hold, implying that the optimal r_1 equals zero.[15]

A.3. Steps leading to (22)

The derivative of $\widetilde{W}$ with respect to r_2 is

$$c[n_1^2 + n_1(n_2 - 2r_1) + 2n_2(n_2 - r_1 - 2r_2)] \; + \; 2k[n_1 + n_2 - 2(r_1 + r_2)] \; - \; (n_2 - 2r_2)y. \tag{a10}$$

Setting ($a10$) equal to zero and solving, the optimal r_2 conditional on r_1 is

$$r_2(r_1) \;\; \equiv \;\; \frac{c[n_1^2 + n_1(n_2 - 2r_1) + 2n_2(n_2 - r_1)] + 2k(n_1 + n_2 - 2r_1) - n_2 y}{4(k + cn_2) - 2y}. \tag{a11}$$

Substituting $r_2(r_1)$ into $\widetilde{W}$, the resulting function has derivatives

$$\widetilde{W}_{r_1}|_{r_2=r_2(r_1)} \;\; = \;\; -\frac{(n_1 - 2r_1)[c^2(n_1 - n_2)^2 + y((2k + 2cn_1 - y) + 2k + 2cn_2)]}{2(k + cn_2) - y} \tag{a12}$$

$$\widetilde{W}_{r_1 r_1}|_{r_2=r_2(r_1)} \;\; = \;\; \frac{2[c^2(n_1 - n_2)^2 + y((2k + 2cn_1 - y) + 2k + 2cn_2)]}{2(k + cn_2) - y}. \tag{a13}$$

Inspection shows that the numerator of (a13) is positive when $-2(k + cn_1) + y < 0, i = 1, 2$ (this conclusion also establishes the negativity of Hessian determinant, which has the opposite sign). Since the denominator is positive, the function $\widetilde{W}|_{r_2=r_2(r_1)}$ is thus convex.

The function $Q(r_1)$ equals $\widetilde{W}$ with r_2 set equal to n_2, and its derivatives are

$$Q_{r_1} = 2k(n_1 - n_2 - 2r_1) + c(2n_1^2 - n_2^2 - n_1(n_2 + 4r_1) - (n_1 - 2r_1)y \qquad (a14)$$

$$Q_{r_1 r_1} = -4(k + cn_1) + 2y < 0. \qquad (a15)$$

Setting $(a14)$ equal to zero yields s_{opt1}, an expression that is positive when $y < y^*$.

Figure 1. A hub carrier's flights in the continuous model

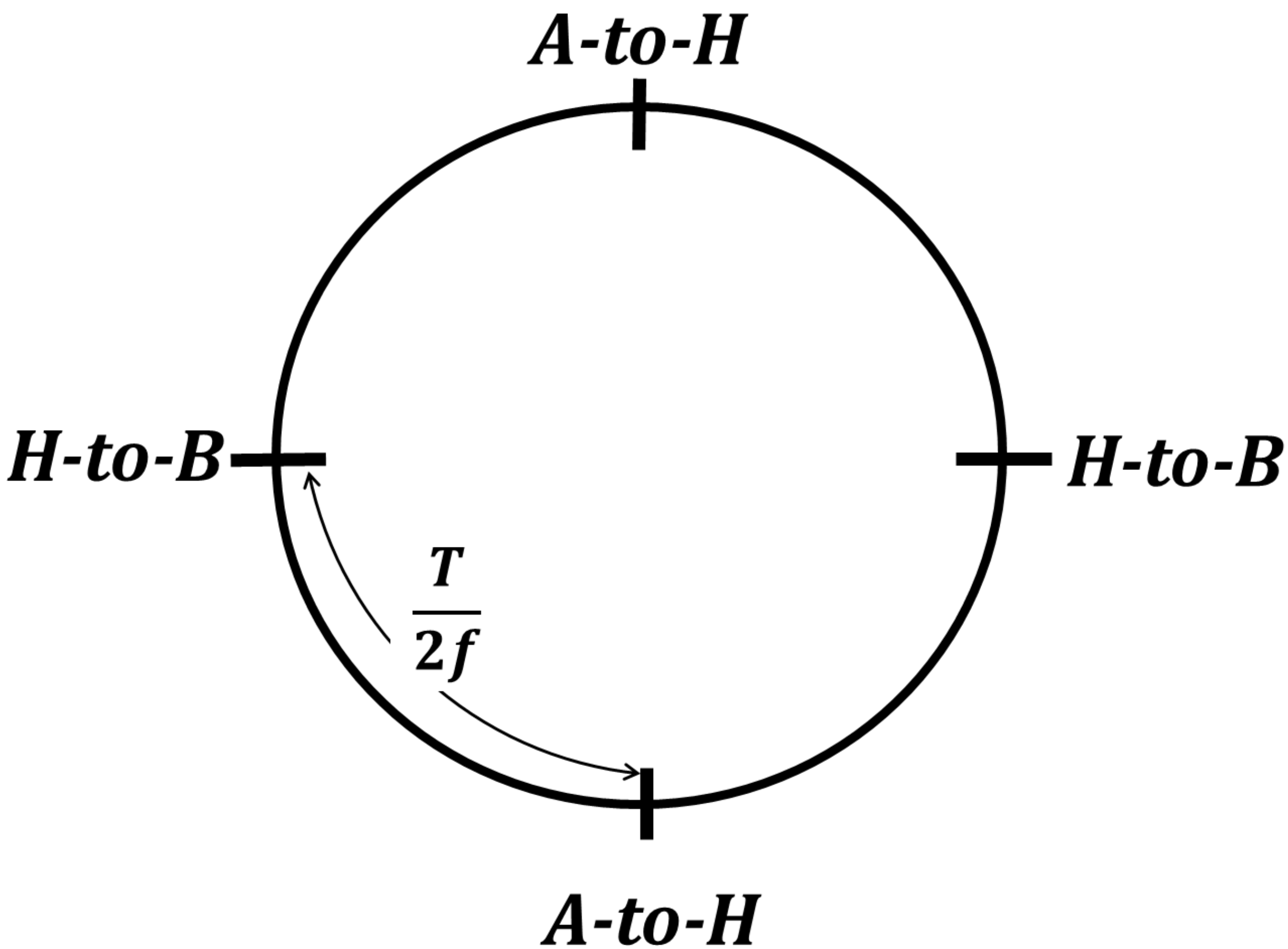


Figure 2. A simple hub-and-spoke network

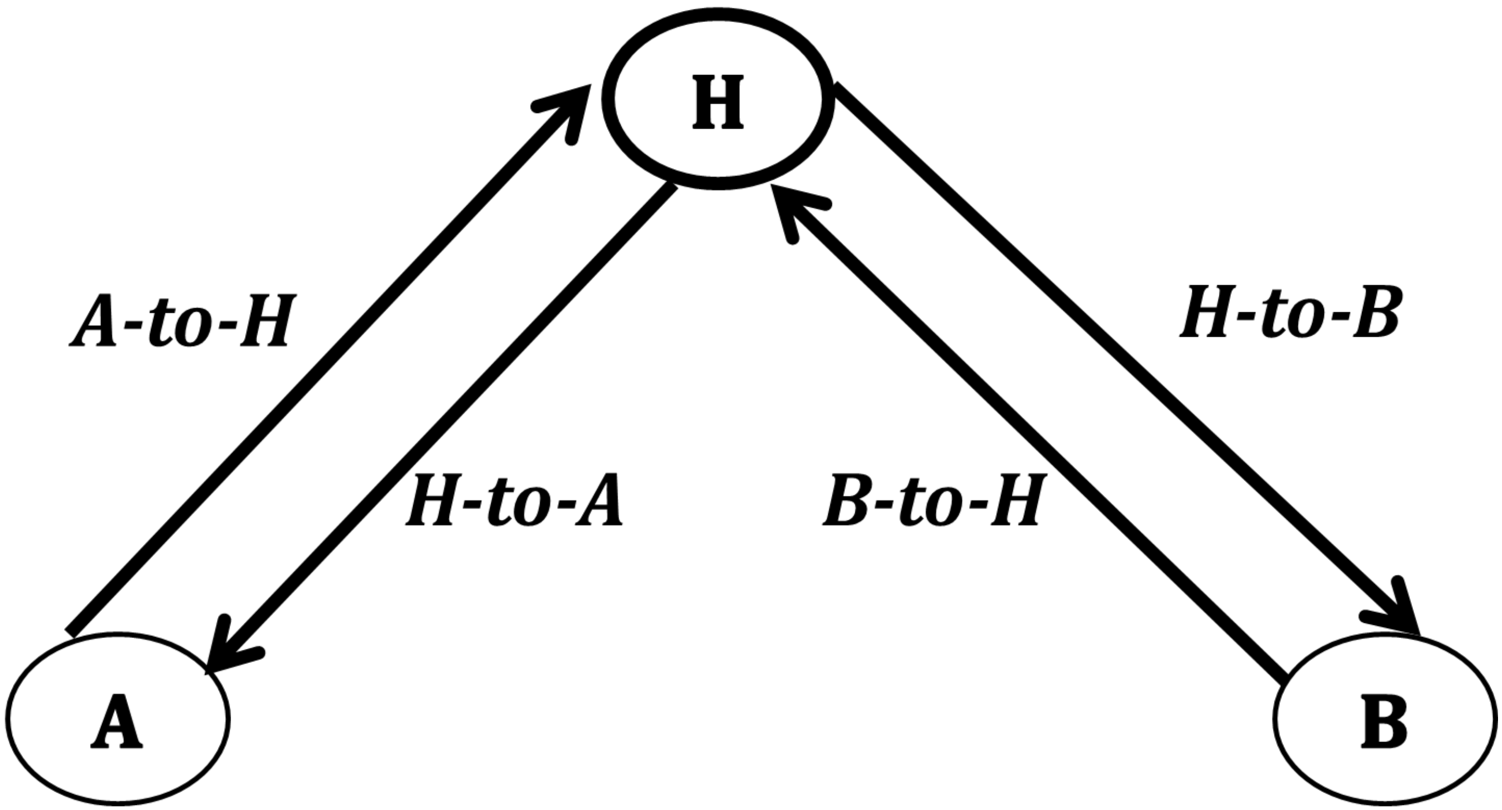


Figure 3a. The monopoly hub carrier's costs

$y < 2(k + cn)$

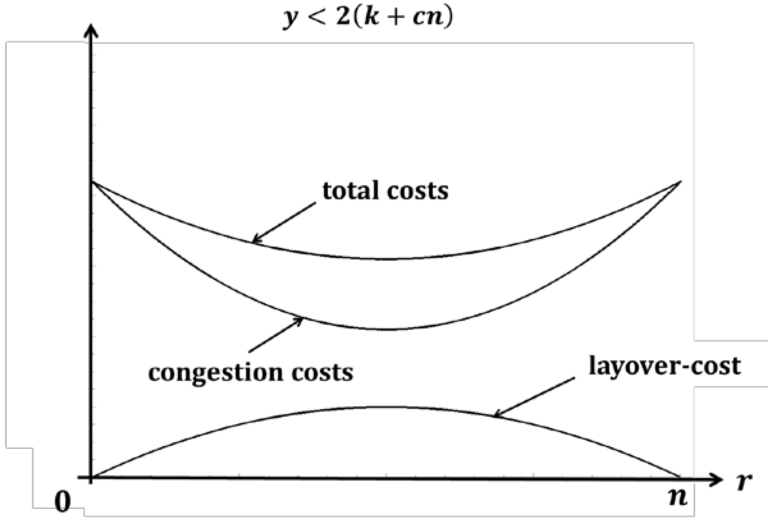


Figure 3b. The monopoly hub carrier's costs

$y > 2(k + cn)$

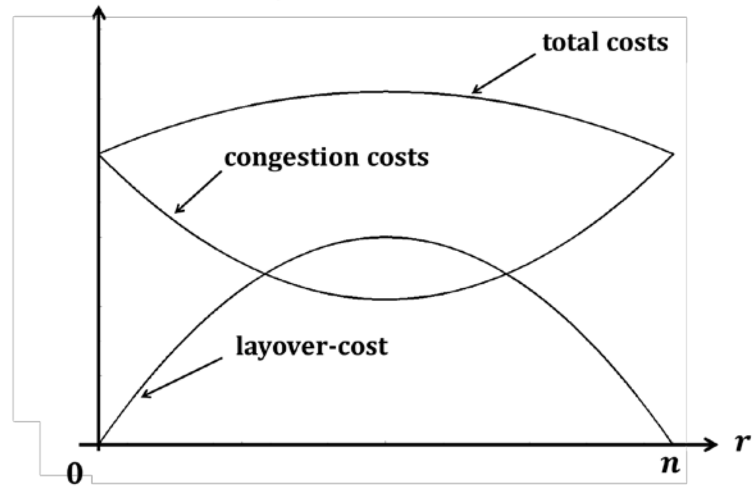


Figure 3c. The monopoly hub carrier's profits

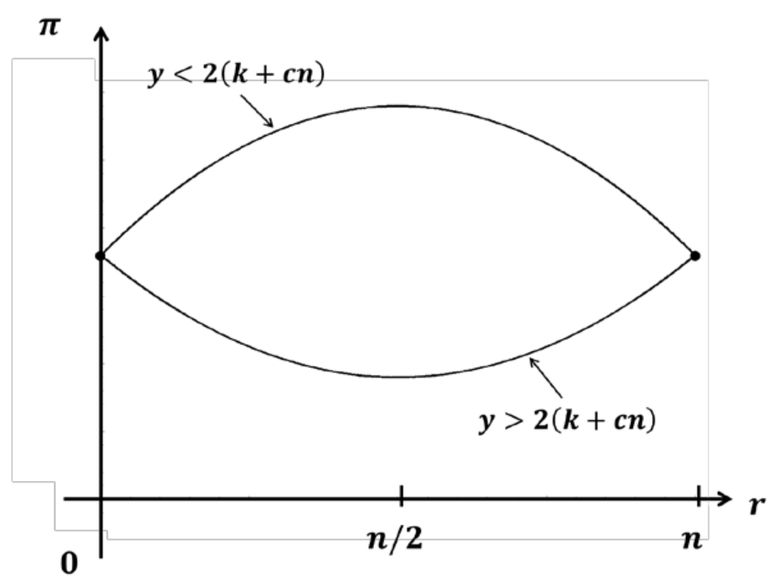


Figure 4. The hub carrier's profit with fringe

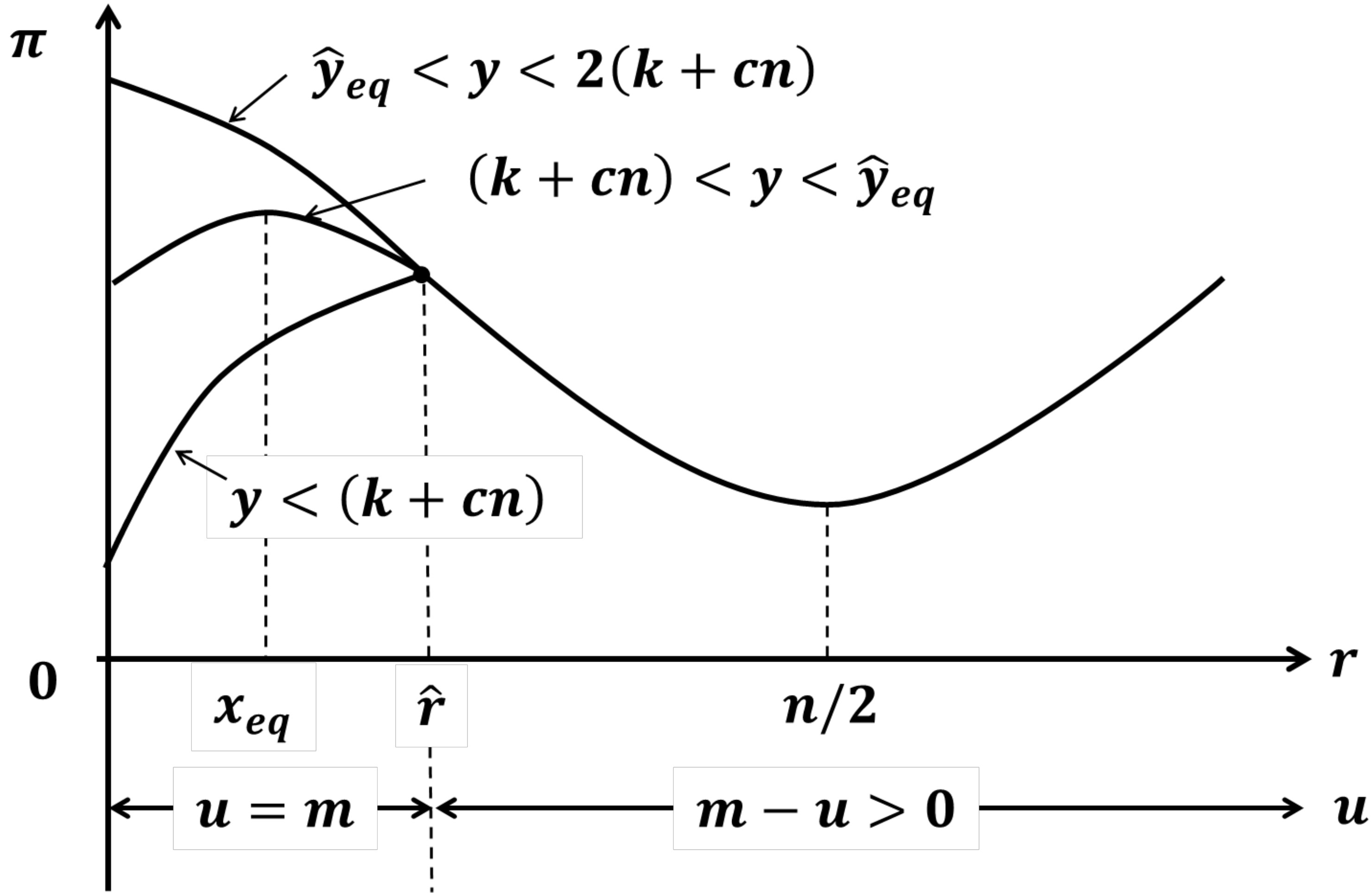


Figure 5. Equilibrium and social optimum (fringe case)

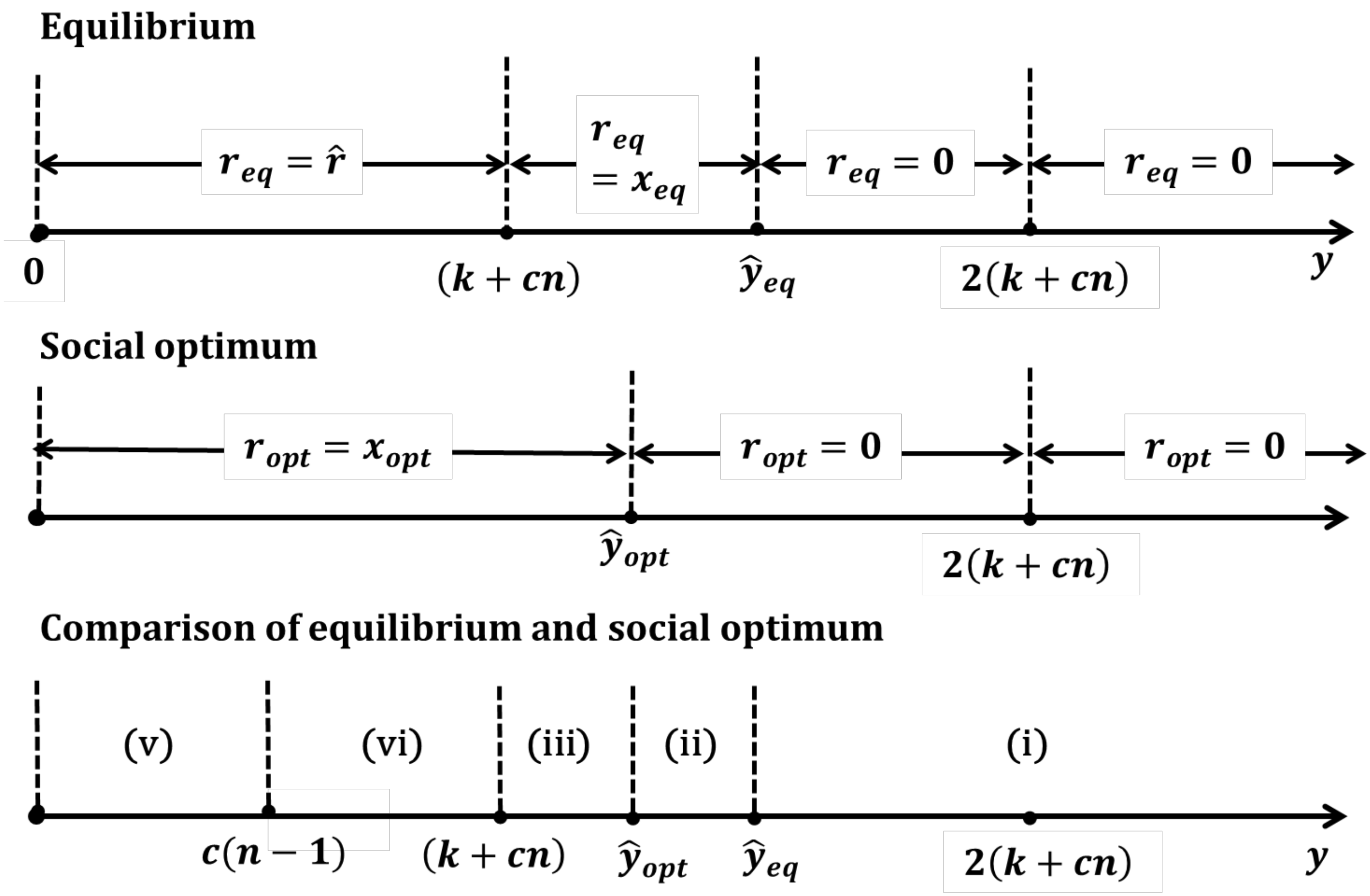
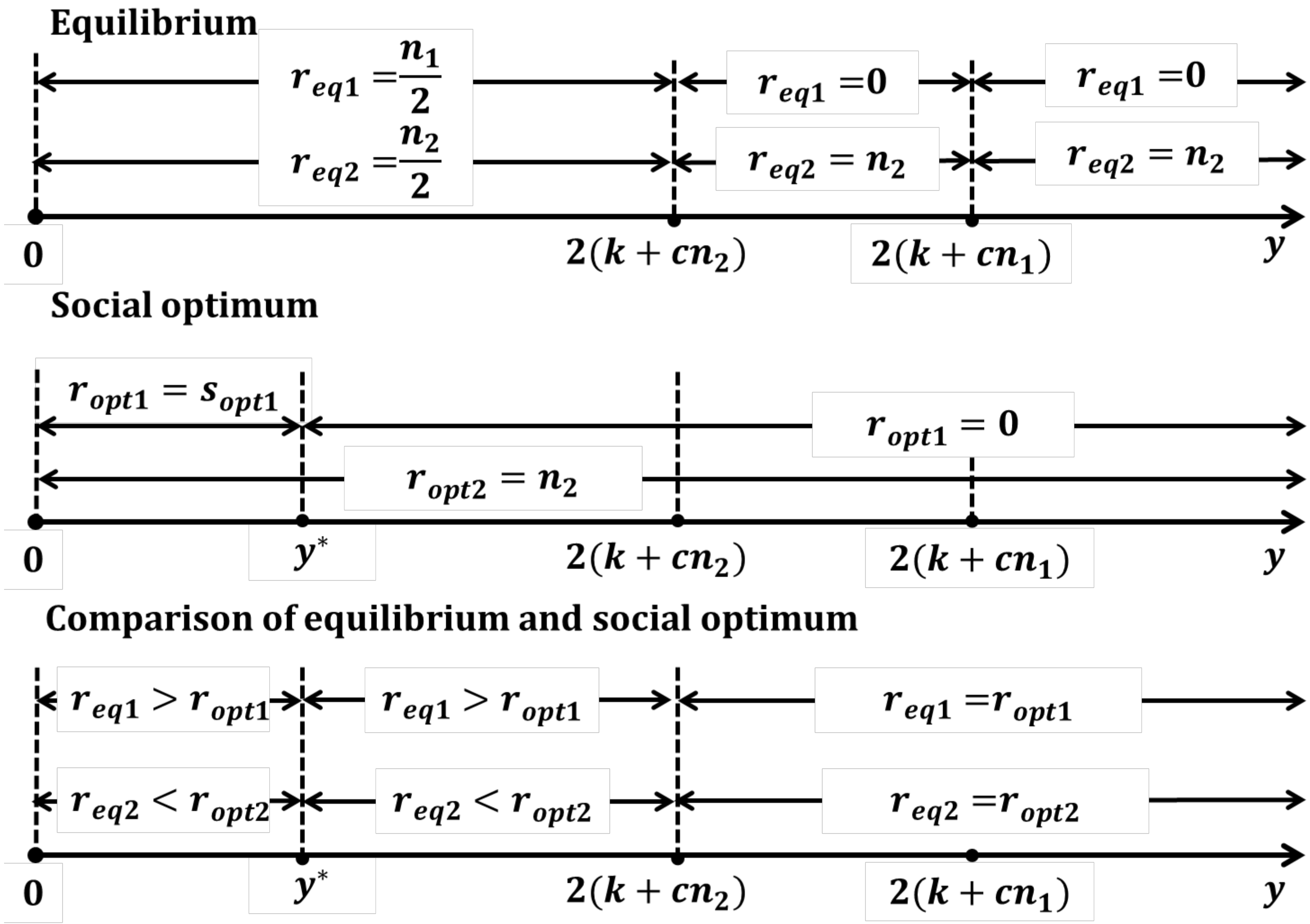


Figure 6. Equilibrium and social optimum (2 hub carriers)



References

BASSO, L., ZHANG, A., 2007. An interpretative survey of analytical models of airport pricing. In: D. Lee (Ed.), *Advances in Airline Economics*, vol. 2, Elsevier, Amsterdam, pp. 89-124.

BRUECKNER, J.K., 2002. Airport congestion when carriers have market power, *American Economic Review* 92, 1357-1375.

BRUECKNER, J.K., 2005. Internalization of airport congestion: A network analysis, *International Journal of Industrial Organization* 23, 599-614.

BRUECKNER, J.K., Schedule competition revisited. *Journal of Transport and Economic Policy* 44, 261-285.

CAREY, C., 2002. American cuts mean longer layovers; but airline says less rushing means less hassle for fliers. *St. Louis Post-Dispatch*, August 16.

CZERNY, A.I., 2013. Public versus private airport behavior when concession revenues exist. *Economics of Transportation* 2, 38-46.

DANIEL, J.I., 1995. Congestion pricing and capacity of large hub airports: A bottleneck model with stochastic queues, *Econometrica* 63, 327-370.

FLORES-FILLOL, R., 2010. Congested hubs. *Transportation Research Part B* 44, 358-370.

JEAN, S., 2015. American plans to 'rebank' its Dallas/Ft. Worth airport hub this Spring. *Dallas Morning News*, August 8.

KATZ, D.S., GARROW, L.A., 2014. Revenue and operational impacts of depeaking at U.S. hub airports. *Journal of Air Transport Management* 34, 57-64.

LIN, M.H., 2013. Airport privatization in congested hub-spoke networks, *Transportation Research Part B* 44, 358-370.

MAYER, C., SINAI, T., 2003. Network effects, congestion externalities, and air traffic delays: Or why all delays are not evil, *American Economic Review* 93, 1194-1215.

Footnotes

*We thank Achim Czerny and Kangoh Lee for helpful comments. Any errors, however, are ours.

[1]Following Daniel (1995) and Brueckner (2002), the literature in the area has expanded substantially. See Basso and Zhang (2007) for a survey.

[2]Brueckner (2005), Flores-Fillol (2010), and Lin (2013) analyze congestion in hub-and-spoke networks without considering the layover-time issue.

[3]Note that schedule delay in the three markets is unaffected by this change.

[4]Doubling of congestion cost is justified by a flight-delay interpretation of the cost. For a connecting passenger, the first flight arrival is delayed due to congestion, and the second flight departure is delayed as well. With the layover time fixed, the total arrival delay is the sum of the two flight delays.

[5]Use of the average schedule delay in deriving the fare expressions is justified by assuming that passengers do not know their preferred departure times until after purchasing the ticket, which means that the expected schedule delay determines their willingness-to-pay for travel. See Brueckner (2010) for one of several related scheduling models built on this assumption.

[6]Were f to be chosen, the first-order condition is $T[C'+K'/2+(2\gamma+\beta)/4]) - 2f^2(K+\theta) = 0$.

[7]Flights serving one-way passengers traveling in the other direction (B to H, B to A, and H to A) can be handled as follows. The plane flying from H to B turns around at B and returns to H carrying B-originating passengers. The original A-to-H aircraft, which has been waiting at the hub, then transports the connecting B-to-A passengers along with H-to-A passengers to city A. The arrival of the B-to-H plane and the departure of the H-to-A plane are close in time, creating congestion in this directional flow (which could be scheduled between the banks operating in the other direction).

[8]Observe that, since the temporal aspect of section 2's model is suppressed in the current framework, the issue of directionality of travel can be ignored.

[9]The fringe carriers are assumed to provide the same service quality as the hub carrier. Otherwise, a quality discount would be subtracted from b in deriving the fringe carriers'

fares, with no effect on the analysis.

[10]The parallel is not exact because the corner solution at $r = 0$ can emerge when the profit function is either convex or concave over the $[0, \widehat{r}]$ interval, while the corner solution at $\widehat{r}$ emerges when the function is concave.

[11]As before, the objective function has a kink at $\widetilde{r}$.

[12]The difference between the slopes of the equilibrium and optimum curves is $m(c + k) > 0$.

[13]The claim regarding the magnitude of y^* follows from the assumption that $n_2 > 2n_1/3$. Observe that, when this inequality holds, $y^* - 2(k + cn_2)$ has the sign of $2cn_1^2 - 2kn_2 - 3cn_1n_2 - cn_2^2 < 2cn_1^2 - 2kn_2 - 3cn_1(2n_1/3) - cn_2^2 = -2kn_2 - cn_2^2 < 0$.

[14]Many of these other flights will be operated by network carriers connecting the given airport to their own hubs. So while the fringe carriers are small at this airport, they need not be small in a global sense.

[15]The numerator of $(a9)$ is negative when $y > (k(2n_1 - n_2) + n_1(2cn_1 - cn_2))/n_1 \equiv y^{\#}$. But since $y^{\#} - 2(k + cn_2) = 2cn_1 - 3cn_2 - kn_2/n_1 < 0$ when $n_2 > 2n_1/3$ holds, it follows that $y > 2(k + cn_2)$ implies $y > y^{\#}$, yielding negativity of $(a9)$'s numerator.