Up in the air?

Price elasticities of air travel demand for direct and indirect travel: theoretical framework and meta-analysis

Thomas de Graaff[1,2] and Eric Pels[1]

[1] Department of Spatial Economics, Vrije Universiteit Amsterdam, the Netherlands
[2] Tinbergen Institute, Amsterdam, the Netherlands

Abstract

Blablaba

Keywords: Price elasticity, airlines, meta-analysis, direct passengers, indirect passengers
JEL codes: R41, R48, R58

1. Introduction

2. Theoretical background

2.1. Demand

Airlines face downward-sloping demand functions for different market segments. The most price conscious segment will be satisfied with just a seat, but this segment is likely to shrink as the duration of the flight increases (Gudmundsson, 2023). In Figure 1 a substantial decrease in price has relatively little effect on demand, compared to the shallow curve. The steep curve represents "business demand", and the shallow curve represents “leisure demand”. But at specific price points, demand is elastic (q_1, p_1) or inelastic (q_2, p_2)

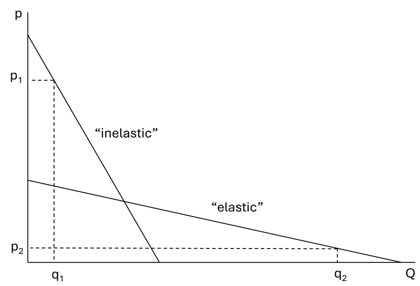


Figure 1: Demand functions

Longer flight durations require higher service levels, in the form of drinks, meals, entertainment, etc., which automatically increase cost levels. However, certain market segments may also require higher levels of service on short flights, such as a relatively high frequency on important business routes to keep the (business) passenger’s schedule delay as low as possible. This may increase the passenger’s willingness-to-pay, an outward shift of the (inverse) demand function, but also means that capacity is added to the market. Conventional airlines solve this problem by accommodating indirect travel, which is usually assumed to be more price sensitive because indirect travel is ’of lesser quality’ compared to a direct connection (i.e., a shallow inverse demand function).

Low-cost airlines typically focus on the core product: a flight (seat) from the origin to the destination (Gudmundsson, 2023). Because there are different market segments, low-cost airlines can add services at an additional cost to the passenger. This “unbundling” implies that low-cost airlines offer more than ’just a seat’ to exploit the willingness to pay for different market segments to generate additional revenue. Conventional airlines are adopting similar strategies, charging passengers separately for a service that used to be included in the fare package.

A key difference between low-cost airlines and conventional airlines remains the importance of frequency of service (see, e.g., Levine, 2002). High frequencies attract passengers sensitive to service levels, but also add indivisible capacity that needs to be filled. Point-to-point airlines know that frequency attracts passengers, but usually do not want the hassle of accommodating transfer traffic. Binggeli and Pompeo (2006) find strong evidence for the existence of S-curves, where an additional flight brings in a disproportionate number of passengers, in markets where there is no competition from low-cost airlines. If low-cost airlines compete, the S-curve becomes less relevant, or unimportant when low-cost airliners compete against each other.

2.2. Market power

The discussion above already mentioned low-cost airlines and unbundling, also by conventional airlines. This means that the competition patterns have changed over time. The (earlier) literature on airline network behavior and pricing often assumes that airlines have market power (see, e.g., Brueckner and Spiller, 1991; Hendricks et al., 1995; Zhang, 1996; Pels and Verhoef, 2004). Brander and Zhang (1993) and Brander and Zhang (1990) and Oum et al. (1993) are common references to justify Cournot competition. Well-known results are the optimality of hub-spoke networks when density economies are present, leading to the creation of fortress hubs (Zhang, 1996), and positive effects for consumers of airline alliances (Brueckner and Whalen, 2000).

When density economies (or, to be more particular: fixed costs) are less important, fortress hubs become less relevant because the hub-spoke network becomes less relevant (see, e.g., Hendricks et al., 1995; Pels et al., 2000; Pels et al., 2001). This conclusion holds for a monopoly airline or airlines competing in a Cournot setting. Hendricks et al. (1999) show in a theoretical model that when airlines compete aggressively in a Bertrand-like setting, the equilibrium result will be that a monopoly airline will operate a hub-spoke network. Duopoly outcomes can exist if no competitor has a path-length advantage in a city pair market, and duopolists do not necessarily opt for a hub-spoke network. However, in general, the idea of airlines competing in a different setting has gained importance in the literature. Daniel (1995) finds that airlines do not internalize congestion, which is what is expected in a competitive market. This finding is supported by Daniel and Harback (2008) and Rupp (2009). When airlines have market power, they are in a position to internalize the congestion they impose on themselves and their own passengers (see, e.g., Brueckner, 2002; Pels and Verhoef, 2004). Contrary to the finding by Daniel (1995), empirical evidence supporting the idea of congestion internalization is provided by Mayer and Sinai (2003) for the U.S. and Santos and Robin (2010) for Europe. The empirical evidence thus is 'mixed', but there is at least some evidence against the Cournot assumption. After all, Cournot competition gives airlines enough market power to internalize the congestion they impose on themselves and their own passengers.

The idea of airline markets moving towards Bertrand behavior may seem intuitive. Conventional airlines offer relatively high frequencies in a hub-and-spoke setting (Brueckner and Zhang, 2001), and schedule flights close together in a competitive setting to attract passengers with a high willingness-to-pay (Givoni and Rietveld, 2009). Kreps and Scheinkman (1983) finds that two-stage capacity-constrained competition leads to Cournot outcomes. But when airlines keep adding capacity to the market in order to compete, the capacity constraint may not be binding for at least some airlines. These airlines must try to fill capacity, and do so by looking at other (indirect) markets or by offering more seats in lower fare classes. On top of that comes competition from low-cost airlines that use distinctively different strategies in order to keep costs of the basic product as low as possible. Perloff et al. (2007) finds that airline behavior in duopoly markets with differentiated goods may be characterized as Bertrand behavior.

2.3. A basic network model

The reason for discussing potential market power has to do with the expected elasticity in a market. Overall demand may be sensitive to price, in the sense that the inverse demand curve is "shallow". But a Cournot oligopolist or monopolist will look at the marginal revenue curve, and at point (q_2, p_2) in Figure 1 marginal revenues are negative because demand at that specific point is inelastic. This idea of course also applies to nonlinear demand functions: marginal revenues are negative when demand at the chosen price point is inelastic. This does not mean that demand needs to be inelastic (at the chosen price point). The market structure will drive the output levels and thus also determines the elasticity at the optimal price point(s). To develop this further, consider the network below, used by a monopoly airline, where A serves as a hub.

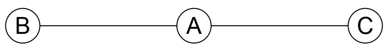


Figure 2: Airline network

In a monopoly setting (see, e.g., Hendricks et al., 1995; Pels et al., 2000; Pels et al., 2001) or in a Cournot oligopoly setting, profits are maximized when

$$(1 + 1/\epsilon^p_{AB}) = \partial C_{AB}(Q_{AB} + Q_{BC})/\partial Q_{AB} \tag{1a}$$

$$(1 + 1/\epsilon^p_{AC}) = \partial C_{AC}(Q_{AC} + Q_{BC})/\partial Q_{AC} \tag{1b}$$

$$\begin{aligned}(1 + 1/\epsilon^p_{BC}) =& \partial C_{AB}(Q_{AB} + Q_{BC})/\partial Q_{BC} + \\ & \partial C_{AC}(Q_{AC} + Q_{BC})/\partial Q_{BC}\end{aligned} \tag{1c}$$

The left-hand sides of Equations (1a)–(1c) are positive when the demand at the optimal output levels is elastic. When capacity is constrained, the marginal cost in, for example, the AB market is the revenue lost in the BC market. In that case, Equation (1a) changes to

$$(1 + 1/\epsilon^p_{AB}) = P^*_{BC} \tag{2}$$

where $P^*(Q_{BC})$ satisfies the condition in Equation (1c). Then Equation (2) implies that $\epsilon^P_{AB} = P_{AB}/(P^*(Q_{BC}) - P_{AB})$. This in turn implies $\epsilon^P_{AB} < -1$, because $P^*(Q_{BC}) > P_{AB}$ would imply $MR_{AB} < MC_{AB}$, which would mean that the airline would not be active on the market between A and B and the flights between A and B are used to accommodate passengers in the BC market. Although technically feasible, this hardly seems a profitable strategy (see, e.g., Hendricks et al., 1995; Pels et al., 2000; Pels et al., 2001). Brons et al. (2002) report an overall mean price elasticity of -1.15. Elastic demand means that airline pricing as described above is feasible, although the standard deviation of 0.62 suggests that in some specific markets demand may be inelastic. Inelastic demand at the optimal output level may be consistent with other forms of competition. For instance, when competition leads the firms to set the price at marginal cost, the optimal output may be found on the inelastic part of the (linear) inverse demand function.

To explore pricing and competition in the above network in more detail, we use linear inverse demand and marginal cost functions $1 - \theta Q$, where Q measures the total output per link (see, e.g., Brueckner and Spiller, 1991; Zhang, 1996). We also add competition in the AB and BC markets. In the BC market, airlines are relatively close substitutes, while in the AB market, airlines are more independent. This reflects the idea that long-haul transfer markets are typically contested by full-service airlines, whereas short-haul markets are often served by full-service airlines competing with low-cost carriers. In case of Cournot competition, the inverse demand functions for airline i are given as:

$$P_{AB,i} = \alpha_{dir} - \beta_{dir} Q_{AB,i} \tag{3a}$$

$$P_{AC,i} = \alpha_{dir} - \beta_{dir} Q_{AC,i} - \gamma_{dir} Q_{AC,j} \tag{3b}$$

$$P_{AC,i} = \alpha_{indir} - \beta_{indir} Q_{BC,i} - \gamma_{dir} Q_{BC,k} \tag{3c}$$

In case of Bertrand competition, demand functions are:

$$Q_{AB,i} = \frac{\alpha_{dir} - P_{AB,i}}{\beta_{dir}} \tag{4a}$$

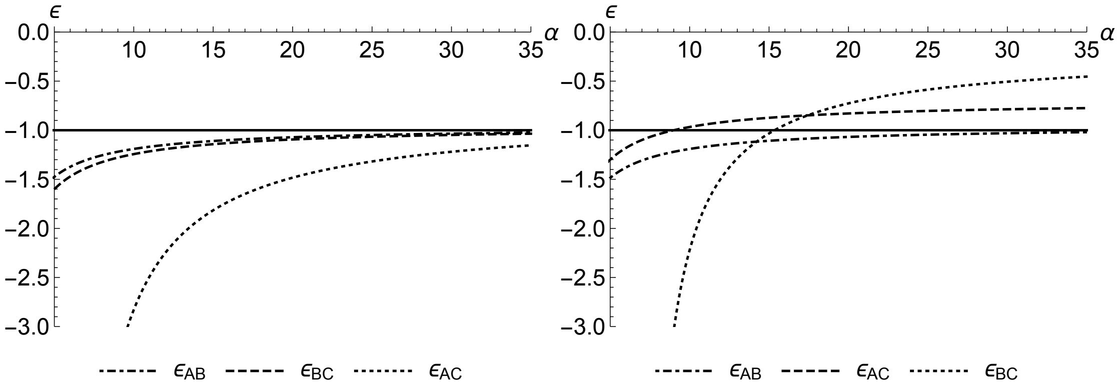
$$Q_{AC,i} = \frac{\alpha_{dir}(\beta_{dir} - \gamma_{dir}) - \beta_{dir} P_{AC,i} + \gamma_{dir} P_{AC,j}}{\beta_{dir}^2 - \gamma_{dir}^2} \tag{4b}$$

$$Q_{BC,i} = \frac{\alpha_{indir}(\beta_{indir} - \gamma_{indir}) - \beta_{indir} P_{BC,i} + \gamma_{indir} P_{BC,k}}{\beta_{indir}^2 - \gamma_{indir}^2} \tag{4c}$$

Finally, in case of mixed Cournot-Bertrand competition, Equations (4a) and (4b) are used to characterize direct markets, and Equation (3c) is used to characterize the indirect market. This reflects the idea that in some direct markets, airline i may have monopoly power or face competition from a low-cost airline, while in the indirect market, airline i faces competition from a similar airline. Appendix A provides the parameterization of the model used to draw Figures 3a and 3b.

Figure 3a shows the elasticities as a function of the market size (measured by α in the direct market) in the direct and indirect markets in a Cournot setting. All elasticitites are below -1, which of course stands to reason. In the direct AB market, the airline i has a monopoly. The AC market is shared with the airline j, the airline i has market power in the part of the market not served by the airline j. The general demand on the market AC may be higher than on the market AB, but the airline i serves fewer passengers than on the market AB, which means that the airline i is on the elastic part of the inverse demand function (that is, closer to the point (q_1, p_1) in Figure 1). In the indirect market BC airline i competes with airline k. The optimality of a hub-spoke system is discussed, e.g., byHendricks et al. (1995), and this is easily extended to the case of Cournot. In such a case, the market is divided between i and k. In a monopoly setting, output already is already relatively low compared to a direct market (assuming demand parameters are the same), and when demand is divided over competitors demand will even be lower. This means that at the optimal price point, the demand is elastic, as seen in Figure 3a.

In case of Bertrand competition in the AC and BC-markets, demand at profit-maximizing levels can be inelastic in these markets (Figure 3b). Competition drives down the price, and in combination with the downward sloping marginal cost curve, this leads to relatively high output levels (i.e. close to the point (q_2, p_2) in Figure 1). In the AB market, the airline i still has a monopoly, so demand remains elastic at the optimal output level.

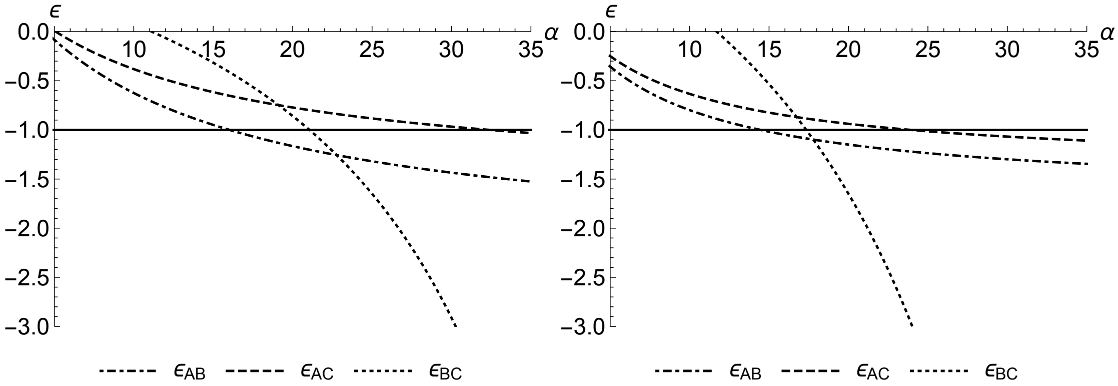


(a) Cournot

(b) Bertrand

Figure 3: airline i's own price elasticities

Finally, when airline i acts as a Bertrand competitor in the direct market (for example, it faces competition from low-cost airline j) and as a Cournot competitor in the market BC (conventional airlines i and k compete with), the result is also mixed (we do not show this in a separate figure). Demand in the AC market is largely inelastic, as expected from Figure 3b. Demand in the indirect market is elastic, as we expect from Figure 3a. The discussion above does not take into account capacity and therefore resembles a long-run perspective. In the short run, the capacity may be fixed. Ideally, the capacity decision would also be modeled in a two-stage model, but then we would be closer to a long-run model again. Figures 4a) and 4b) show the point elasticities in the event that the link capacities are fixed at 35 (see Appendix A for details). In the base case (without capacity restriction), airline i has monopoly power in the market AB and shares the market AC with airline j. As a result, the output in the AB market exceeds the output in the (shared) AC-market. When α $(= \alpha_{dir})$ is relatively large, the output in the AB and BC-markets will be below the unrestricted output because the consumers'willingness-to-pay . This means that in the restricted case, output in the AB and AC markets are closer to the point (q_1, p_1) in Figure 1), so that demand at the optimal (restricted) output level is more elastic. The same holds for the AC-market, but because that market is shared with airline j, the optimal (restricted) output level is lower compared to the market AB for all levels of α_{dir}). On the other hand, when α_{dir}) is relatively low, the optimal unrestricted output levels will also be relatively low. If capacity is too large, the optimal restricted output levels exceed optimal unresticted output levels because capacity need to be filled. After all, airline i incurs the costs of offering an output level equal to the set capacity. In that case, airline i will be closer to a point like (q_2, p_2) in Figure 1), and demand at the optimal (restricted) output levels is inelastic.



(a) Cournot **(b)** Bertrand

Figure 4: airline i's own price elasticities with capacity restrictions

2.4. Expected point elasticities in a network setting

The results in the previous subsection are, of course, highly dependent on the parameterization. Changing the level of substitution between airlines or the slope of the inverse demand function will have an impact on the price elasticities of demand. However, the general patterns will remain similar. For example, if airlines are closer substitutes in the AC market, overall output will be relatively high in this market, but airline i will face relatively strong competition. If the fixed capacity is relatively high, the intersection of the line -1 and the elasticity curves in Figures 4a) and 4b) will occur at a higher level of α_{dir}. Table 1 reports the expected outcome (elastic or inelastic) in the three markets under different market conditions. No numerical values are given, because these depend on the exact parameterization. Table 1 uses α as a measure of market size, because if, for example, α_{dir} increases in Equation 1a, so does α_{dir}/β_{dir} in Equation 4a, while the structure of the marginal cost function remains unchanged. Based on Table 1 the conclusion is that in a Cournot setting demand is elastic at the optimal output levels, because then marginal revenues are non-negative. In a Bertand setting, demand is inelastic (elastic) at the optimal output level when the market size is relatively large (small) and the outputs are large (small). In the most basic case of Bertrand (with homogeneous goods), the price is equal to the marginal cost, and when the market is relatively large, this occurs due to a relatively high output level (so close to the point (q_2, p_2) in Figure 1). In the current setting, there are no homogeneous goods, so a price larger than marginal costs is possible, but the underlying reasoning remains the same.

	market AB no competition	market AC low substitution	market BC high substitution
Cournot	elastic	elastic	elastic
Bertrand	elastic	elastic (low α), inelastic (large α)	elastic (low α), inelastic (large α)
Cournot restricted	inelastic (low α), elastic (large α)	inelastic (low α), elastic (very large α)	inelastic (low α), elastic (large α)
Bertrand restricted	inelastic (low α), elastic (large α)	inelastic (low α), elastic (very large α)	inelastic (low α), elastic (large α)

Table 1: expected level of elasticity, based on airline i's own price elasticities

In a Cournot or Bertrand setting with a capacity restriction, demand is inelastic at the optimal output levels when market size is small and the capacity exceeds the unrestricted optimal output. In this case, the output is relatively large in order to fill the capacity. In the unrestricted case, airlines look at marginal revenues and marginal cost. Increasing output lowers marginal revenues and marginal costs, and airlines look for the optimum. In case of a capacity restriction, the marginal cost becomes a constant, determined by the capacity level. When marginal costs are low due to a relatively large capacity, the output must be increased to achieve the (restricted) optimum again. Note that Figures 4a and 4b show that with a capacity restriction, the point elasticity decreases in α at an increasing rate in the indirect market (i.e., demand becomes more elastic), while the elasticities in direct markets decrease in α at a decreasing rate. Given the link capacity, the marginal and total cost of the link are fixed. Because passengers in the indirect market travel on two links, the airline can increase revenues relative to costs by allowing more direct travelers to travel on the two links, provided that the market size is large enough. If α is relatively large, there are enough passengers in the direct markets to fill a large share of the capacity, so that the output in the indirect market decreases (and hence demand becomes more elastic in that market).

In the next Section, a preferred estimate for the point elasticity for direct and indirect markets is determined. Most, if not all, underlying studies do not report whether Cournot or Bertrand competition is assumed because most papers use an empirical model. But a preferred estimate for the point elasticity will give an indication of what market form may be expected in a specific market, in the sense that an elasticity between -1 and 0 is not consistent with the unrestricted Cournot setting. The next Section also considers the effect of distance, which is not included in the current section because it would add another parameter to the model. While indirect travel also occurs for relatively short distances, typically market BC would be a long haul market in real life.

References

Binggeli, U. and L. Pompeo (2006). 'Analyst Viewpoint Does the S-Curve still exist'. *International Air Transport Association (IATA) Economics Team Report*.

Brander, J. A. and A. Zhang (1990). 'Market Conduct in the Airline Industry: An Empirical Investigation'. *The Rand Journal of Economics* 21.4, pp. 567–583.

— (1993). 'Dynamic oligopoly behaviour in the airline industry'. *International Journal of Industrial Organization* 11.3, pp. 407–435.

Brons, M., E. Pels, P. Nijkamp and P. Rietveld (2002). 'Price elasticities of demand for passenger air travel: a meta-analysis'. *Journal of Air Transport Management* 8.3, pp. 165–175.

Brueckner, J. K. and P. T. Spiller (1991). 'Competition and mergers in airline hub-and-spoke networks'. *International Journal of Industrial Organization* 9, pp. 323–342.

Brueckner, J. K. (2002). 'Airport Congestion When Carriers Have Market Power'. *The American Economic Review* 92.5, pp. 1357–1375.

Brueckner, J. K. and W. Whalen (2000). 'The price effects of international airline alliance'. *Journal of Law and Economics* 43.2, pp. 503–546.

Brueckner, J. K. and Y. Zhang (2001). 'A model of scheduling in airline networks: How a hub- and-spoke system Affects flight frequency, fares and welfare'. *Journal of Transport Economics and Policy* 35.2, pp. 195–222.

Daniel, J. I. (1995). 'Congestion pricing and capacity of large hub airports: a bottleneck model with stochastic queues'. *Econometrica: Journal of the Econometric Society* 63.2, pp. 327–370.

Daniel, J. I. and K. T. Harback (2008). '(When) do hub airlines internalize their self-imposed congestion delays?' *Journal of Urban Economics* 63.2, pp. 583–612.

Givoni, M. and P. Rietveld (2009). 'Airline's choice of aircraft size – Explanations and implications'. *Transportation Research Part A: Policy and Practice* 43.5, pp. 500–510.

Gudmundsson Sveinn, V. (2023). 'Limits to the low-cost niche? Finding sustainable strategies for low-cost long-haul airlines'. *Case Studies on Transport Policy* 12, pp. 1–9.

Hendricks, K., P. Michele and G. Tann (1999). 'Equilibria in Networks'. *Econometrica: Journal of the Econometric Society* 67.6, pp. 1407–1434.

Hendricks, K., M. Piccione and G. Tann (1995). 'The Economics of Hubs: The Case of Monopoly'. *The Review of Economic Studies* 62.1, pp. 83–99.

Kreps, D. and J. Scheinkman (1983). 'Quantity precommitment and Bertrand competition yield Cournot outcomes'. *The Bell Journal of Economics* 14.2, pp. 326–337.

Levine, M. E. (2002). 'Price Discrimination Without Market Power'. *Yale Journal on Regulation* 19.1, pp. 1–36.

Mayer, C. and T. Sinai (2003). 'Network effects, congestion externalities, and air traffic delays: or why not all delays are evil'. *The American Economic Review* 93.4, pp. 1194–1215.

Oum, T. H., A. J. Taylor and A. Zhang (1993). 'Strategic Airline Policy in the Globalizing Airline Networks'. *Transportation Journal* 32.3, pp. 14–30.

Pels, E., P. Nijkamp and P. Rietveld (2000). 'A note on the optimality of airline networks'. *Economics Letters* 69.3, pp. 429–434.

— (2001). 'Erratum to "A note on the optimality of airline networks"'. *Economics Letters* 70.1, p. 145.

Pels, E. and E. T. Verhoef (2004). 'The economics of airport congestion pricing'. *Journal of Urban Economics* 55.2, pp. 257–277.

Perloff, J. M., L. S. Karp and A. Golan (2007). *Estimating market power and strategies*. Cambridge University Press.

Rupp, N. (2009). 'Do carriers internalize congestion costs? Empirical evidence on the internalization question'. *Journal of Urban Economics* 65.1, pp. 24–37.

Santos, G. and M. Robin (2010). 'Determinants of delays at European airports'. *Transportation Research Part B: Methodological* 44.3, pp. 392–403.

Zhang, A. (1996). 'An analysis of fortress hubs in airline networks'. *Journal of Transport Economics and Policy* 30.3, pp. 293–307.

A. Parameterization

This appendix introduces the parameterization of the model used to draw Figures 3a and 3b. First, in line with the common conjecture that indirect travel is of "lesser quality", it is assumed that inverse demand in the BC-market is lower and shallower compared to the AB and AC-markets. Thus, $\alpha_{indir} = \alpha_{dir}/2$, $\beta_{dir} = 1/2$ and $\beta_{indir} = 1/4$. Furthermore, in market BC there is competition from a low-cost airline and in market BC from a conventional airline, as reflected in $\gamma_{dir} = 1/4$ and $\gamma_{indir} = 1/5$. Finally, to ensure that the hub-spoke network is the optimal network (compared to the point-to-point network), $\theta = 0.01$ and the fixed costs per link are $\psi = 5$ (Pels et al., 2000). With these parameters, the airline i sets the following combinations (output, price) in the markets AB, AC and BC at α_{dir}=25 in a Cournot setting: (24.40, 12.80), (19.52, 10.41) and (15.68, 5.17). The total output on link AB (AC) is 40 (35). The corresponding elasticities are: $\epsilon^P_{i,AB} = -1.05$; $\epsilon^P_{i,AC} = -1.07$ and $\epsilon^P_{i,BC} = -1.32$.

In order to draw 4a and 4b, the capacity on both links is fixed at 35. This level is consistent with the required capacity on the link AC at α_{dir}=25. When α_{dir} is lower (higher), the capacity will be too high (low). At α_{dir}=25 airline i sets the following combinations (output, price) in markets AB, AC and BC: (21.56, 14.22), (21.56, 9.52) and (13.44, 5.54). Because capacity is restricted on the AB-link, airline i reduces the output in both the AB and BC-markets. The reduced output on the BC-market requires an increase in output on the AC-market to ensure that the capacity is fully used. The lower output on the AB and BC markets means that the corresponding price elasticities $\epsilon^P_{i,AB}$ and $\epsilon^P_{i,BC}$ decrease to -1.32 and -1.65) respectively. The price elasticity $\epsilon^P_{i,AC}$ increases to -0.88). In markets where the capacity is "too low", output is relatively low (compared to the unrestricted optimum), meaning that we are relatively close to the point (q_1, p_1) in Figure 1). The increase in output on the AC-market means we move closer to the point (q_2, p_2) in Figure 1).