Distinguishing Multiproduct Economies of Scale from Economies of Density on a Fixed-Size Transport Network

Leonardo J. Basso · Sergio R. Jara-Díaz

Abstract In order to produce a certain vector of flows over a fixed-size network, a transport firm has to choose, among many other things, a route structure. In accommodating increasing traffic, transport firms will adjust their route structure to minimize costs. However, transport industry structure analysis considers only, often implicitly, the case where the route structure is fixed. In this paper, economies of density, that represents the latter, are conceptually distinguished from economies of scale on fixed-size networks, *S*, where we allow the route structure to vary. Intuition with a simple example, evidence from the airline industry and the derivation of a formula to calculate *S* from an estimated cost function are provided. Results are both novel and encouraging. Transport firms, while closer to exhaust economies of density, would still have available sizeable economies of scale.

Keywords Scale economies · Returns to density · Air transport networks

1. Introduction

The empirical estimation of a transport cost function has been one of the preferred ways to analyze transport industry structure in the literature in the last decades.

L. J. Basso (✉)
Sauder School of Business, The University of British Columbia,
2053 Main Mall, Vancouver BC, Canada V6T 1Z2
e-mail: basso@sauder.ubc.ca
and
Universidad de Chile, Casilla 228-3, Santiago, Chile
e-mail: lbasso@ing.uchile.cl
Department of Civil Engineering,

S. R. Jara-Díaz
Department of Civil Engineering, Universidad de Chile,
Santiago, Chile
e-mail: jaradiaz@ing.uchile.cl

One of the central points of interest is the calculation of economies of scale.[1] From a theoretical point of view, a transport firm produces movements of different things between many origins and destinations (OD pairs). Strictly then, the product of a transport firm is a vector $\mathbf{Y} = \{y_{ijk}\}$, where y_{ijk} represents flow of type k (goods or persons) between origin i and destination j, as pointed out by Jara-Díaz (1982), Winston (1985) and Braeutigam (1999), among others. These and other researchers have noted that, since vector $\mathbf{Y}$ cannot be used in the empirical work because of its dimension, estimated cost functions are specified in terms of *aggregates* that represent both, products—e.g., passenger-kilometers, total tons, seat–miles—and so-called attributes—e.g., average length of haul, average stage length, average load factor. When a variable describing network size, N (e.g., number of points served, PS), is also included in the estimation of a transport cost function, empirical studies of the transport industry structure make a distinction between returns to density (RTD) and returns to scale (RTS), following Caves et al. (1984). The former assumes constant network size when output increases (increase in density), while the latter assumes that the network grows as well (increase in output through a network expansion but keeping density constant). Accordingly, RTD is calculated as the inverse of the sum of product elasticities of cost, while RTS includes the network elasticity in the summation as well. In short, RTD is aimed at capturing the behavior of costs as the load on each link of a given network increases, while RTS is aimed at capturing the behavior of costs as the network size increases.

Jara-Díaz and Cortés (1996, hereafter JDC) proposed a new approach to transport industry structure analysis, by stressing that behind the aggregates hides vector $\mathbf{Y}$ as defined above, and that this can and must be recognized when doing economic analyses. For example, as economies of scale analyze cost behavior as outputs expand by the same proportion (Panzar and Willig, 1977), a correct calculation of scale economies in transport should be related with the same growth of all flows in vector $\mathbf{Y}$. By looking at scale in terms of the elasticities of cost with respect to the elements of $\mathbf{Y}$, JDC show that an internally consistent calculation of the multiproduct degree of scale economies can be obtained using the elasticities of aggregates with respect to cost—the ones obtained directly from estimation of the cost function—provided each of these elasticities is weighted by the aggregate's local degree of homogeneity with respect to $\mathbf{Y}$. JDC derive these weights for many commonly used aggregates: for instance, they show that the weight for the length of trip or length of haul is zero, as a proportional increase in all flows does not change average distances. On the contrary, the weight for passenger–kilometers is one. Oum and Zhang (1997) point out that since N is not allowed to vary in the JDC setting, their method actually corresponds to an improved version of RTD,[2] which coincides with Panzar's (1989) example suggesting that "returns to density are precisely equal to (what has been previously defined to be) the degree of multiproduct economies of scale!"

[1] Recent examples are Liu and Lynk (1999) and Creel and Farrel (2001) for airlines, Ivaldi and McCullough (2001) and Mizutani (2004) for railways, and Karlaftis and McCarthy (2002) and Filippini and Prioni (2003) for buses. A list of previous studies can be found in Braeutigam (1999).

[2] The theoretical reason for this is that an equiproportional change of the elements of $\mathbf{Y}$ produces no change in N. For example, the number of points served, PS does not change with an increase in all OD flows (see JDC).

Besides the previous observation linking JDC's calculation with *RTD*, Oum and Zhang (1997) investigate the correlation between output attributes and network size, following Ying (1992) and Xu et al. (1994). They argue that, if the network size also varies with output (the *RTS* approach), average distances (length of trip or stage length) may change, leading to weights that could differ from zero. This made the authors suggest that within the context of varying network size (i.e., *RTS*), average distance's elasticity should be taken into account. Here, we show that this could be the case even with a constant network size, because there is an intermediate case—not previously identified—that arises from the difference between *network size* and *route structure*.

As suggested by many, and showed on theoretical grounds by Jara-Díaz and Basso (2003), the route structure is a key endogenous decision for a transport firm. That is, *for any given fixed network size* (fixed PS for example) and output vector **Y**, a transport firm will choose a route structure that minimizes the cost of production.[3] This implies that, as the elements of **Y** increase by the same proportion, a firm may find it optimal to alter its route structure, inducing changes in attributes such as average length of trip. This effect will not be captured by economies of scale with variable network size, *RTS*, because this index is calculated explicitly considering a variable network size, which is not the case. Moreover, recent work by Basso and Jara-Díaz (2005, 2006) suggests that this index is inappropriate altogether for the analysis of network expansions and that what should be calculated instead is economies of spatial scope. Economies of density will not be able to capture the mentioned effect either, as the calculation of *RTD* assumes not only a fixed network size *but also a fixed route structure*, since its objective, as understood from the literature, is to capture what happens with costs when links are increasingly loaded. If the route structure varies, some links may no longer be used and new links may be added. This basic analysis highlights the necessity of an index that captures the economies available to transport firms when the flow vector **Y** increases and the route structure is allowed to vary (which does not require expansions of network size as measured by *PS*). This paper proposes a way to calculate such an index. Note that the proposed method will permit the calculation of what should be called, rigorously, the *multi-product degree of economies of scale*—as opposed to *economies of density* and *economies of spatial scope*—since it captures the effects on costs of equiproportional changes in **Y**, while in RTD the restriction of a fixed route structure is added (often implicitly). The assumptions behind the calculation of each of these indices, is summarized in Fig. 1. In the next section we build on the intuition why these economies of scale are likely to exist and derive the method of calculation by using the JDC approach. We also present evidence from the literature that shows firms' behavior that is consistent with the intuition provided. Section 3 offers, as an illustration, a calculation of the proposed index using a previously published study. Section 4 concludes.

[3] A caveat: in many cases firms will choose a route structure not only to minimize cost but to maximize profit; see for example Oum et al. (1995). In this paper, we stick to the cost minimization framework.

	Route Structure is Fixed	Route Structure is Variable
Network Size is Fixed	Economies of density (*RTD*)	Multiproduct Economies of Scale (*S*, this paper)
Network Size is Variable	Economies of spatial scope	

Fig. 1 Assumptions for the calculation of transport cost structure indices

2. Economies of Scale

2.1. The Intuition

In what follows we will consider an airline-like network for simplicity: network size is directly given by the number of points served, while the absence of a physical network allows us to avoid network assignment issues. To illustrate the subject of interest, let us consider for a moment an airline producing movements on six OD pairs, corresponding to a network size of three nodes (Fig. 2).

The airline will have to choose a route structure to produce these flows. Two out of many possible route structures are hub-and-spoke and direct service in all OD pairs. As a cost function gives the minimum expenditure necessary to produce output vector **Y** for given input prices, the optimal route structure for a given network (size and topology) will change with different values of **Y** (as shown in Jara-Díaz and Basso, 2003). Because in real cases the dimension of **Y** is unmanageable, aggregates are used instead. We will consider two of the most popular aggregates, namely passenger–kilometers (PK), and average length of trip (ALT). The idea of estimating the degree of economies of density is to analyze whether "the marginal cost of carrying an extra passenger on a nonstop route falls as traffic on the route rises" (Brueckner and Spiller, 1994) or more generally, to see if "the average costs of a direct connection decreases with proportionate increases in both flows on that connection" (Hendricks et al., 1995).[4] In terms of **Y**, this translates into the analysis of cost behavior when **Y** is multiplied by a real number and the route structure does not change. This is required to ensure that only the existing links handle the new traffic. How is this to be analyzed from a cost function specified in terms of PK, ALT and PS? Concisely, JDC method is as follows. Both PK and ALT can be written as functions of vector **Y**:

$$PK(Y) = \sum_i y_i \cdot d_i \tag{1}$$

$$ALT(Y) = \sum_i y_i \cdot d_i \bigg/ \sum_i y_i \tag{2}$$

[4] The concepts marginal and average cost should be taken with care here, if one considers the strict description of product **Y**. Roughly, in both cases the authors refer to increases in Q, where Q is the flow on a particular link. Clearly, Q is the sum of some y_{ij}. In particular, average cost should be understood as an incremental average cost on Q.

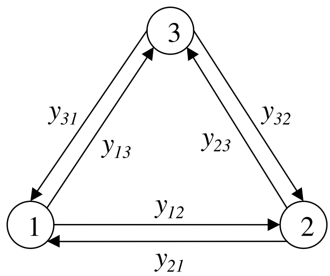

Fig. 2 Network size and OD structure

where i denotes an OD pair (there are six in the case of Fig. 2) and d_i is the distance flow y_i has to travel. It is important to stress that d_i represents the total distance flown by a passenger, which implies that if a hub-and-spoke route structure is chosen—hub on node 3, say—then d_i for flow in OD pair 1–2 will be the sum of two distances. It is easy to see from Eqs. (1) and (2) that the degrees of homogeneity of PK and ALT with respect to $\mathbf{Y}$ are 1 and 0, respectively, provided d_i is constant for all i. This is the argument JDC use to suggest that in a proper calculation of the multiproduct degree of economies of scale, the elasticity of PK should always be considered while that of ALT should never enter the calculation.[5] As explained, Oum and Zhang (1997) argue that, because the elasticity of PS does not enter the calculation, this is actually a computation of economies of density. However, the point can be made clearer if one takes into account all the decisions transport firms can take. Not only does PS remain unchanged—a fact consistent with amplifications of $\mathbf{Y}$—but also the distances d_i remain constant, which implies that the route structure is fixed. Therefore, the JDC method corresponds to economies of density in a stronger sense than that implied by Oum and Zhang: it is the correct way to calculate economies of density because both, the network size *and* the route structure do not vary. The distinction between network size and route structure thus becomes pivotal.[6]

The existence of economies of density and their strong influence on the shaping of hub-and-spoke route structures has been largely documented in empirical analysis of the airline industry (e.g., Caves et al., 1984; Gillen et al., 1990; Kumbhakar, 1990, 1992; Keeler and Formby, 1994; Baltagi et al., 1995). Their existence has been attributed to many different sources but the leading two have been the spreading of fixed cost over larger volumes and economies of aircraft size, which are related to indivisibilities (see e.g., Tretheway and Oum, 1992; Hendricks et al., 1995). But consider the OD system of Fig. 2, assuming a hub and spoke structure is in place (hub in node 3, say) and let us picture what could happen as the flows (in all OD pairs) grow. As flows increase, the airline will be able to exploit the economies of aircraft size by changing to larger airplanes without unduly affecting their frequencies (frequencies might affect the demand side, which is assumed invariant in cost function analysis). If traffic grows enough, some optimal aircraft

[5] JDC analyze many others aggregates. In many cases weights are not 0 or 1 but something in between.

[6] These two concepts are not always distinguished in the literature. For example, Brueckner and Spiller (1994) state that Caves et al. estimate economies of density "by holding the airline's route structure (for example, the number of points served) constant (...)."

size will be achieved (for the distances induced by the route structure), exhausting economies of aircraft size. As traffic increases, the next step would be to increase frequencies only. However, if volumes increase enough, it will start to be costly to keep flows going through the hub. Costs of take-off and landing, extra distances flied and costs provoked by airports' congestion will be enough to overcome any economies still available in the aircraft size side. The airline then might decide to change its route structure, bypassing the hub. This will, among other things, change the average distances flown by passengers. Note that fixed costs related to network size but not to route structure—such as airline counters, station managers, mechanics, ticket offices, advertising—would not be greatly affected by such a decision. Thus, it may well happen that an airline is close to exhausting its economies of density for a given route structure, but might have still economies to exploit by simply changing its route structure and start new direct routes between points that were served indirectly.

It is in this sense that, conceptually, two different notions of multiproduct scale shall be distinguished when a fixed-size transport network is considered: *economies of scale* (which we will denote S to distinguish them from economies of scale with variable network size, RTS) and *economies of density*, RTD. The former analyzes costs when OD flows grow proportionally, the latter analyzes costs when flows grow and the route structure remain unchanged (at the risk of being repetitive, we will stress again that when vector $\mathbf{Y}$ is amplified, the number of OD pairs remained unchanged, so from a theoretical point of view the number of points served remains unchanged: the network size does not vary). Three questions arise then: first, how do we calculate S from an estimated cost function in order to capture the effects we discussed in this section? Second, what do we expect to obtain as a result of a calculation of S in comparison to RTD? Third, is there any evidence that a process such as the one described here has been occurring? We attempt to answer these questions next.

2.2. How to Calculate S

The key is to use the strict description of output, as JDC pointed out. Again, one has to look at changes on the aggregates when $\mathbf{Y}$ changes. The reason why we need to multiply the cost elasticities of aggregates by their local degree of homogeneity with respect to $\mathbf{Y}$ obviously persists when calculating S. The current analysis differs from JDC's however, in that now we want to know what happens to the aggregates when $\mathbf{Y}$ changes but the route structure is not fixed, contrary to their implicit assumption. Let us consider the aggregates PK and ALT in Eqs. (1) and (2), accepting that the d_i are no longer fixed as they were, but are now dependent on the route structure. As argued, the route structure depends on vector $\mathbf{Y}$. Conceptually, we now have:

$$PK(Y) = \sum_i y_i \cdot d_i(Y) \tag{3}$$

$$ALT(Y) = \sum_i y_i \cdot d_i(Y) \Big/ \sum_i y_i \tag{4}$$

As shown by JDC, the weight for PK's cost elasticity is the local degree of homogeneity of PK with respect to $\mathbf{Y}$, α_{PK}, which is given by

$$\alpha_{PK} = \sum_i \frac{\partial PK}{\partial y_i} \cdot \frac{y_i}{PK} \tag{5}$$

Taking into account that the d_i now depend on the route structure, straightforward algebra leads to

$$\frac{\partial PK}{\partial y_h} = \sum_i y_i \cdot \frac{\partial d_i(Y)}{\partial y_h} + d_h(Y) \;\Rightarrow\; \frac{\partial PK}{\partial y_h} \cdot \frac{y_h}{PK} = \left(\sum_i y_i \cdot \frac{\partial d_i(Y)}{\partial y_h} + d_h(Y) \right) \cdot \frac{y_h}{PK}$$

$$\alpha_{PK} = \sum_h \left(\sum_i y_i \cdot \frac{\partial d_i(Y)}{\partial y_h} + d_h(Y) \right) \cdot \frac{Y_h}{PK} \;\Rightarrow$$

$$\alpha_{PK} = 1 + \left\{ \frac{1}{PK} \sum_h Y_h \left(\sum_i y_i \cdot \frac{\partial d_i(Y)}{\partial y_h} \right) \right\}$$

A similar procedure for ALT's degree of homogeneity leads to α_{ALT} equal to the expression in square brackets on the right hand side of the last equation. We therefore get:

$$\alpha_{PK} = 1 + \alpha_{ALT} \tag{6}$$

These complementary roles of PK and ALT are expected: both aggregates are related to each other, not only through $\mathbf{Y}$ but also through another aggregate, namely the total number of passengers.[7] Next, suppose for now that the cost function depends only on PK, ALT and PS. Then, given Eq. (6), S can be calculated provided one can estimate α_{ALT}, the change in the average length of trip with output when the network size is fixed. S is then:

$$S = \left[\eta_{PK}(1 + \alpha_{ALT}) + \alpha_{ALT}\eta_{ALT} \right]^{-1} = \left[\eta_{PK} + \alpha_{ALT}(\eta_{PK} + \eta_{ALT}) \right]^{-1} \tag{7}$$

In Eq. (7), η_{PK} and η_{ALT} are elasticities of cost with respect to PK and ALT, directly obtained from the estimated cost function. Estimation of α_{ALT} is of course an empirical matter. One way to obtain it is to regress ALT on aggregate total traffic (measured in passengers) and PS. The justification for this is that a proportional change in $\mathbf{Y}$ induces a proportional change in total volume, independently of whether the route structure changes or not. One needs to control for PS in order to net out the effect of increased network size in ALT. Indeed, it may be argued that the data may not be consistent with equiproportional changes of $\mathbf{Y}$ but there is no much room for improvement here. The problem of working with aggregates (a necessity in any case) is precisely the loss of information.

[7] Jara-Díaz and Cortes also obtained this sort of dual relations between, for example, the number of seats–kilometers and the average load factor; the sum of the two weights was one in this case.

With α_{ALT} at hand, S can be calculated by multiplying known elasticities by their weights.

If the intuition presented in the previous subsection is correct, one would expect α_{ALT} to be negative, since switching from indirect to direct flights decreases average distances flown by passengers. The difference between S and RTD is then not signed a priori, since PK and ALT cost elasticities usually have opposed signs (η_{ALT} is negative). Since typically η_{PK} is greater than η_{ALT} in absolute value, one would anticipate S to be larger than RTD. These expected results seem reasonable. By accommodating route structures, airlines can exploit economies that are not available otherwise.

Before moving into the example, we would like to make a few remarks. First, other authors have previously suggested the use of econometric regressions as a way to capture indirect effects of aggregate output increases on operating attributes, and the use of these regressions to compute 'total elasticities' instead of partial elasticities. The first attempt in this direction is the railroads study by Caves et al. (1985). They present three measures of economies of scale with variable network size, one of which has the elasticity of average length of haul multiplied by 0.5 while the elasticity of average length of trip is multiplied by one. They argue that data shows that kind of relation between both average distances. Note, however, that this inclusion of the elasticities of length of haul and length of trip was made in the context of *variable network size*. Indeed, in their calculation of RTD these elasticities were not included. Later, Xu et al. (1994) argue that larger firms may not have a cost advantage directly because of size, but that size indirectly influences attributes. They investigate the correlation between attributes and output (ton–kilometers) in the trucking industry, and use the results to calculate economies of scale through total elasticities. These total elasticities were obtained by deriving costs with respect to output so that the elasticity of ton–kilometers was always weighted by one. This was made in the context of variable network size also, since they did not specified a network size variable in their cost function. Oum and Zhang (1997) follow Xu et al. (1994) in the context of the distinction between RTS and RTD. Acknowledging JDC's results, they examine, through econometric regressions, the behavior of attributes only in the case when the network size varies. All of this shows that the concept behind S, as defined here, is a novelty, because the influence of average distances has only been considered in the context of variable network size, and by means of an index (RTS) that may be inadequate to analyze network growth (Basso and Jara-Díaz 2006). In RTD, average distances have not been considered because the route structure was (implicitly) imposed to be fixed. What is obtained when looking at economies of scale more rigorously is an equation that differs from others in two striking ways. First, it incorporates changes in average distances even though the network size does not change. Second, for that incorporation to be consistent with a scale analysis, it has to be coupled with a weight different from one on the elasticity of the output aggregate PK. Lastly, let us highlight that the use of weights in the calculation of S is well-grounded on a multiproduct view of transport firms—in the spirit of JDC approach—and does not rely only on the fact that various descriptions of outputs and attributes are or might be correlated.

Another remark that is important to make is related to the fact that, in many cases, airlines cost functions are not specified in terms of ALT but in terms of average stage length (ASL). ASL represents the average distance of a link in the route structure—where the average takes into consideration the number of passengers on each link—which is different from ALT. Shortly, ASL looks at flows over links rather than at OD levels. Clearly, changes in the route structure will possibly alter ASL. However, while it was argued that ALT should diminish if airlines bypass hubs, it is not clear that ASL will diminish as much. Consideration of an isosceles configuration of three nodes is enough to picture why. However, the calculation of S is not dramatically changed if the cost function is specified in terms of ASL rather than ALT. Simply a second regression, this time of ASL on total passengers P and PS, is needed. The degree of economies of scale can then be calculated as:

$$S = [\eta_{PK}(1 + \alpha_{ALT}) + \alpha_{ASL}\eta_{ASL}]^{-1} \tag{8}$$

Given the above arguments, one would expect α_{ASL} to be smaller, in absolute value, than α_{ALT}. It is also easy to check that calculation of PK using link rather than OD distances and flows yields identical results.

2.3. Evidence from the Literature

So far, it has been explained why a new index is needed, how it can be calculated and the intuition behind it. But, is there any evidence that a process such as the one described in Section 2.1 has been taking place? Let us first look at the evolution of estimates of RTD. Estimates that decrease with newer samples is evidence that airlines have been actively exploiting economies of traffic density and that they will be, therefore, looking for new saving opportunities. To touch on this point let us consider three studies of the US airline industry (Caves et al., 1984; Kumbhakar, 1990 and Liu and Lynk, 1999), summarized in Table 1, which are suited for our purpose since they cover different periods of time for the same industry (the influence of different estimation techniques should not be overlooked though). Both Kumbhakar and Liu and Lynk follow closely the specification of the highly influential article by Caves et al. In every case, RTD is estimated as the inverse elasticity of output, where output is a volume–distance measure. Reported values are from long-run cost functions.

The work by Kumbhakar is particularly informative. Economies of density were important before deregulation, but they are less important in the post-deregulation period. Moreover, when he considers both periods together, his result falls quite close to the result by Caves et al., as the periods are similar. Liu and Lynk's estimate on the other hand, is even smaller than the one of Kumbhakar, which shows that in recent years airlines have continued to exhaust their economies of density and that they are not as important as before.

Table 1 Estimation of economies of density

Study	Sample	RTD (mean)	Cost function specification
Caves et al. (1984)	1970–1981	1.244	Translog
Kumbhakar (1990)	1970–1978	1.367	Symmetric Generalized McFadden
	1978–1984	1.199	
	1970–1984	1.277	
Liu and Lynk (1999)	1984–1991	1.161	Cobb-Douglas

Table 2 Average airplane sizes on scheduled services

Year	Average number of seats	Traffic index*
1985	192	100
1990	195	138
1995	194	174
2000	187	225

Source: Swan (2002); *: Total traffic was measured as passenger kilometers.

Swan's (2002) review of the development of airlines' route structure gives us an even clearer panorama of this. He essentially shows that, while air travel has doubled and even tripled in the last decades, there is a worldwide trend to constant, or even declining, airplane size (see Table 2). He states that this is contrary to the expectation from all sources. Therefore, the new traffic has not been accommodated by changing to bigger airplanes in the last 15 to 20 years. According to Swan, the higher volumes have been accommodated not only with higher frequencies but also through changes in the route structure: "travel is bypassing intermediate stops. Some of this is because bypass non-stops save cost." Swan's findings are supported by Wei and Hansen's (2003) econometric work. They find that, currently, the US domestic airline industry favors more flight rather than larger planes because aircraft operating cost scale economies are not particularly strong nor do they extend very far.

These studies show that the economies of aircraft size—crucial for the existence of economies of density—may be close to be exhausted (if they are not already). This helps to explain why the estimates of *RTD* have been declining. What is important to note however, is that airlines have been reacting to increased traffic not only through higher frequencies but also through changes in the route structure, as Swan mentions. Without economies on the aircraft size side and with increasingly congested airports, hubbing may be more expensive than offering direct flights and, as mentioned, fixed costs are more associated with network size than with route structure. While incentives to offer indirect flights may still arise because of demand reasons, since frequencies in hub-and-spoke structures are higher, with large enough volumes frequencies will still be high in a direct connection, while travel times will be decreased and connecting times removed. The process described in Section 2.1 would then be actually happening according to these studies. But economies of density are not able to capture it, which shows that the methodology to calculate what we have defined as S is in fact needed.

3. Methodological Example

In this section, we present an illustration of the calculation of economies of scale, S. The first step, as required in Eqs. (7) and (8), is to regress ALT and ASL on total passengers, P, and the number of points served, PS, which represents the network size. We use simple log-linear equations on a U.S. airline data set[8] covering the

[8] We consider annual information for five airlines: American Airlines, Continental, Delta, United and Northwest.

period 1980–1989. Estimated coefficients are as follows (standard errors are in parenthesis):

$$\ln ALT = \underset{(0.779)}{9.6512} - \underset{(0.066)}{0.2299} \ln P + \underset{(0.127)}{0.366} \ln PS, \qquad r^2 = 0.205$$

$$\ln ASL = \underset{(0.627)}{8.2105} - \underset{(0.053)}{0.1470} \ln P + \underset{(0.102)}{0.3014} \ln PS, \qquad r^2 = 0.164$$

It follows that $\alpha_{ALT} = -0.2299$ and $\alpha_{ASL} = -0.147$. As predicted, the elasticity of the average length of trip with respect to traffic, after controlling for network size, is negative (see Section 2). Moreover, $|\alpha_{ALT}| > |\alpha_{ASL}|$ as previously discussed. These regressions confirm the intuition that in the presence of small returns to density, airlines will change their route structures bypassing hubs as traffic increases. The relatively low r^2 in the above regressions indicates that these *attributes* are also affected by other exogenous variables besides P and PS, such as input prices (Xu et al., 1994; Oum and Zhang, 1997). As data on input prices were not available for this study we can only note that, provided that the correlation between input prices and P and PS is not 'too large,' the incorporation of input prices should not affect the estimated weights.

To complete the illustration, let us use these estimated coefficients together with the results by Liu and Lynk (1999), as their data set of 11 U.S. airlines includes, as a subset, the five airlines we considered in our regressions, and has the largest overlapping with our data set in terms of the period considered. They estimate the degree of economies of density as the inverse of the elasticity of cost with respect to PK, which yields $RTD = [\eta_{PK}]^{-1} = [0.861]^{-1} = 1.161$. Since their cost function includes ASL rather than ALT as well, Eq. (8) should be used. We get

$$S = [\eta_{PK}(1 + \alpha_{ALT}) + \alpha_{ASL}\eta_{ASL}]^{-1} = [0.861 \cdot (1 - 0.2299) + (-0.147) \cdot (-0.426)]^{-1}$$
$$= 1.378$$

As can be seen, while only mild increasing returns to density are present, strong economies of scale are found. This result reveals the importance of identifying the opportunities open to carriers when accommodating larger traffic. The consequences of recognizing the difference between network size and route structure, and therefore distinguishing economies of density from economies of scale on fixed-size transport network are now apparent.

Before concluding, it should be noted that the JDC approach applies to every aggregate and not only to PK and average distances. In the three articles described in Table 1, the cost function also included as an explanatory variable the average load factor, ALF. JDC argue that if an equiproportional increase in $\mathbf{Y}$ induces changes in ALF, then its cost elasticity should be incorporated to the calculation of RTD, provided it is weighted by its degree of homogeneity with respect to $\mathbf{Y}$. This weight is, again, a matter of econometric analysis: how does ALF varies with (possibly aggregate) output within a fixed network size? Oum and Zhang (1997) regress ALF on aggregate output, PS and input prices. Although they consider a volume–distance descriptor of output, which is not what we have proposed here, their elasticity estimate of 0.044 can be used to see how Liu and Lynk's estimate of RTD changes when the JDC method is applied. The improved version of RTD would be $RTD_2 = [\eta_{PK} + \alpha_{LF}\eta_{LF}]^{-1} = [0.861 + 0.044 \cdot (-0.977)]^{-1} = 1.222$, i.e., larger estimated returns.

But what happens with S? Strictly speaking, it would be necessary to distinguish changes in load factor, for a fixed-size network, when the route structure changes and when it does not. That, however, will lead to a regression of ALF on P, PS and ALT, in order to control for changes in network size and in the route structure. Multicollinearity and causality problems are likely to arise, though. In this paper, only as an illustration, we use the same weight in both cases. We obtain $S_2 = [\eta_{PK}(1+\ \alpha_{ALT}) + \alpha_{ASL}\eta_{ASL} + \alpha_{LF}\eta_{LF}]^{-1} = 1.464$. As with RTD, the (consistent) incorporation of the load factor leads to larger returns to scale. These particular numbers, however, should be taken carefully as the weight was estimated not considering P but a volume–distance output aggregate.

4. Conclusions

A transport firm has to decide many things in order to produce a certain vector of flows at minimum cost. For a fixed-size network, these decisions go beyond fleet size and types of vehicle but include the pivotal decision of route structure. Although this decision is endogenous, considering a fixed route structure when analyzing a proportional output growth is actually a useful imposition as it helps the analyst to assess certain specific characteristics of transport production, such as the existence of decreasing incremental average costs at the link level, that is, economies of density. However, transport firms do adjust their route structure to accommodate increased production. This leads directly to two different notions of 'scale analysis': economies of scale (S) and economies of density (RTD). This distinction is the main conceptual contribution of this paper. In short, for a fixed-size network, density analyzes cost behavior when output (OD flows) increases proportionally and the route structure remains unchanged, while scale analyzes cost behavior when output (OD flows) increases proportionally but the route structure is allowed to change. Note that S and RTD would coincide if proportional output growth does not induce changes in the route structure.[9] It is somewhat surprising that this distinction had not been made before. In our opinion, this has occurred because of two facts: confusion between network size and route structure, and failure to think in terms of the true transport output, the vector of OD flows.

JDC's approach was central to derive the formula for the calculation of S, showing, once again, that even though the use of aggregates is a necessity, economic analysis can and should be made considering a detailed description of product. This, applied to the new concept of multiproduct economies of scale on a fixed-size network, lead to a result with particular features. On one hand, the change in average distances does play a role but, on the other hand, this is balanced by a 'different than one' weight on aggregate output. Some rudimentary empirical work showed that the weights had the expected sign and were ordered in the expected manner. As a result, S was larger than RTD as predicted.

[9] Network size growth, in our view, should be analyzed through economies of spatial scope (Basso and Jara-Díaz, 2005).

Opportunities for further research can be clearly identified. On the empirical side, much more can be done to improve the econometric regressions that relate ASL or ALT to traffic. First, more accurate and complete data (including input prices) should be considered. Second, the fact that ASL, ALT and costs are decisions that are jointly made within the firm should be taken into account. This could be handled through joint estimation or the use of instrumental variables to account for endogeneity. On the theoretical side, this paper provides what should be interpreted as a general approach (and an application) rather than a 'full method' to calculate S. This is so because the formulae we propose work for the particular aggregates considered. While these aggregates have been very popular, many others have been used as well. Analytical derivation of the weights for these other aggregates is one of the urgent things to do. How *attributes* change when the route structure is or is not fixed should be carefully accounted for. Now that things are conceptually clearer, the task should be easier.

Acknowledgments This research was partially funded by Fondecyt-Chile, grant 1050643, and the Millenium Nucleus "Complex Engineering Systems." We would like to thank Anming Zhang for helpful suggestions, and Chunyan Yu and Tae Oum for providing us with the data.

References

Baltagi BH, Griffin JM, Rich DP (1995) Airline deregulation: the cost pieces of the puzzle. Int Econ Rev 36(1):245–258

Basso LJ, Jara-Díaz SR (2005) Calculation of economies of spatial scope from transport cost functions with aggregate output (with an application to the airline industry). J Transp Econ Policy 39:25–52

Basso LJ, Jara-Díaz SR (2006) Is returns to scale with variable network size adequate for transport industry structure analysis? Transp Sience (in press)

Braeutigam RR (1999) Learning about transport costs. In: Gomez-Ibañez J, Tye WB y Winston C (eds) Essays in transportation economics and policy. Brooking, Washington, District of Columbia, pp 57–97

Brueckner JK, Spiller PT (1994) Economies of traffic density in the deregulated airline industry. J Law Econ 37:379–413

Caves DW, Christensen LR, Tretheway MW (1984) Economies of density versus economies of scale: why trunk and local airline costs differ. Rand J Econ 15:471–489

Caves DW, Christensen LR, Tretheway MW, Windle RJ (1985) Network effects and the measurement of returns to scale and density for U.S. railroads. In: Daughety AF (ed) Analytical studies in transport economics. Cambridge University Press, New York

Filippini M, Prioni P (2003) The influence of ownership on the cost of bus service provision in Switzerland—an empirical illustration. Appl Econ 35:683–690

Gillen D, Oum TH, Tretheway M (1990) Airlines cost structure and policy implications. J Transp Econ Policy 24:9–34

Hendricks K, Piccione M, Tan G (1995) The economics of hubs: the case of monopoly. Rev Econ Stud 62:83–99

Ivaldi M, McCullough G (2001) Density and integration effects in class I US Freight railroads. J Regul Econ 19(2):161–182

Jara-Díaz SR (1982) The estimation of transport cost functions: a methodological review. Transp Rev 2:257–278

Jara-Díaz SR, Basso LJ (2003) Transport cost functions, network expansion and economies of scope. Transp Res E 39:271–288

Jara-Díaz SR, Cortés C (1996) On the calculation of scale economies from transport cost functions. J Transp Econ Policy 30:157–170

Karlaftis M, McCarthy P (2002) Cost structures of public transit systems: a panel data analysis. Transp Res E 38:1–18

Keeler JP, Formby JP (1994) Cost economies and consolidation in the U.S. airline industry. International Journal of Transport Economics 21(1):21–45
Kumbhakar SC (1990) A reexamination of returns to scale, density and technical progress in US airlines. South Econ J 57:428–442
Kumbhakar SC (1992) Allocative distortions, technical progress and input demand in U.S. airlines: 1970:1984. Int Econ Rev 33(3):723–737
Liu Z, Lynk EL (1999) Evidence on market structure of the deregulated US airline industry. Appl Econ 31:1083–1092
Mizutani F (2004) Privetly owned railways' cost function, organization size and ownership. J Regul Econ 25(3):297–322
Oum TH, Zhang Y (1997) A note on scale economies in transport. J Transp Econ Policy 31:309–315
Oum TH, Zhang A, Zhang Y (1995) Airline network rivalry. Can J Econ 28:836–857
Panzar J (1989) Technological determinants of firm and industry structure. In: Schmalensee, R, Willig R (eds) Handbook of industrial organization. North-Holland, pp 3–59
Panzar J, Willig R (1977) Economies of scale in multioutput production. Q J Econ 91:481–493
Swan WM (2002) Airline route developments: a review of history. J Air Trans Manag 8:349–353
Tretheway MW, Oum TH (1992) Airline economics: foundations for strategy and policy. Center for Transportation Studies, University of British Columbia
Wei W, Hansen M (2003) Cost economics of aircraft size. J Transp Econ Policy 37:279–296
Winston C (1985) Conceptual developments in the economics of transportation: an interpretive survey. J Econ Lit 23:57–94
Xu K, Windle C, Grimm C, Corsi T (1994) Re-evaluating returns to scale in transport. J Transp Econ Policy 28:275–286
Ying JS (1992) On calculating cost elasticities. Logist Transp Rev 28:231–235