1. Introduction

The product of the activity of transport firms is a set of flows of persons and goods, moved during a number of periods and among many points in space. This multiproduct nature of transport firms can not be directly captured when estimating a cost function, due to the enormous amount of data that would be required. This has made the use of aggregate descriptions of product a necessity; accordingly, measures like ton-kilometers or revenue passenger-miles are of frequent use in this area of the transport economics literature, accompanied by other variables that intend to capture the complexity of product, like average length of haul, average load or number of shipments. This aggregate treatment poses some difficulties when it comes to the analysis of industry structure using output related properties of the cost function. This is indeed the case of the (multiproduct) degree of economies of scale $S$, which is generically calculated as the inverse of the sum of the output elasticities of cost (Panzar and Willig, 1975). The fact that aggregates make the calculation of $S$ obscure was highlighted by Gagné (1990), Ying (1992) and Xu et al. (1994), who observed that aggregates are usually interrelated; for example ton-kilometres is equal to total flow times average length of haul, a fact that had not been taken into account when making calculations of $S$.

In an article in this same journal (Jara-Díaz and Cortés, 1996) we proposed to look at each aggregate product $\tilde{y}_j$ as a function of the detailed product, the disaggregated origin-destination flow vector $Y$; in this manner, a cost function $\tilde{C}$ specified in terms of aggregates $\tilde{Y}$ is in fact an implicit function of $Y$, i.e. $\tilde{C}[Y(Y)]$. By looking at scale in terms of the elasticities of the elements of $Y$, we were able to show that each elasticity of cost with respect to an aggregate output should be multiplied by a factor in order to calculate an internally consistent estimate for $S$. Each factor depends on the analytical relation between the corresponding aggregate $\tilde{y}_j$ and the flow vector $Y$, and happens to be between zero and one for usual output aggregates; this was regarded as “a more rigorous reconsideration of measuring scale economies” by Oum and Waters (1996, pp. 432). Nevertheless,
Oum and Zhang (1997), showed some reservations regarding the appropriateness of the method in the context of a variation of network size.

In this paper we argue that a variation in the size of the network served is unambiguously related with a variation in the number of products in \( Y \) and, therefore, its impact on costs is, in principle, a matter of scope analysis. This will be illustrated using one of the most popular measures of network size, namely the number of points served, \( PS \), whose elasticity has been widely used to obtain estimates of scale economies from an estimated cost function; we will show here that it can be used to analyse scope. In the following section we present a synthesis of the main multioutput concepts within the context of transport activities, in order to show the relation between \( PS \) and the product vector; special emphasis will be given to the notion of *spatial scope*. Next we show that the elasticity of cost with respect to \( PS \) is related with a particular type of economies of spatial scope, providing an alternative interpretation for the results of empirical cost functions including that variable. The concluding section contains a discussion and directions for research.

2. Scale and Scope in transport activities

The concept of scale is related with the maximum proportional variation of products that is made feasible by a proportional variation of inputs. This, a technical property, translates into an economic property, namely the behavior of cost \( C \) as output \( Y \) expands proportionally. On the other hand, the concept of scope is related with the economic analysis of the addition of new outputs to the line of production. In simple words, scope analysis deals with the enlargement of the set of outputs produced, while scale analysis relates with producing more of each component of the same set of outputs (Panzar and Willig, 1975, 1981).

Analytically, the (multioutput) degree of scale economies \( S \) can be shown to be obtainable from the cost function \( C(Y) \) as the inverse of the sum of cost elasticities with respect to the output components. On the other hand, the degree of economies of scope relative to a subset \( R \), \( SC_R \), can be calculated as

\[
SC_R = \frac{1}{C(Y)} \left[ C(Y^R) + C(Y^{M-R}) - C(Y) \right]
\]  

(1)
where $Y^R$ represents vector $Y$ with $y_i = 0, \forall i \notin R \subseteq M$, with $M$ being the set of all products (we have suppressed input prices for simplicity). Thus, a positive $SC_R$ means that it is cheaper to produce $Y$ with a single firm than to split production into two orthogonal subsets $R$ and $M-R$. It can be easily verified that the degree of economies of scope should lie in the interval (-1,1).

In the case of transport, the firm has to use inputs (vehicles, terminals, rights-of-way, energy, labour, and so on) to produce movements of many things - freight or/and passenger - from many origins to many destinations during many different periods. Thus, the output of a transport firm is (see Jara-Díaz, 1981, 1982 a, b, for a detailed discussion)

$$Y = \{y_{ij}^{kt}\} \in R^{K\times N\times T}$$

(2)

where each component $y_{ij}^{kt}$ represents the flow of type $k$ moved from origin $i$ to destination $j$ (OD pair $ij$), within period $t$, for example passengers from Paris to Frankfurt during a specific weekend ($K, N$ and $T$ are the the number of cargo types, the number of OD pairs, and the number of time periods considered in $Y$, respectively). It should be stressed that the spatial dimension of transport output has been somewhat neglected in the literature, in spite of the discussions by many authors like Spady (1985) or Daughety et al. (1985), who emphasise the OD nature of transport product.

Aggregation of output over any dimension (commodity, time or space) involves loosing information associated with the transport processes generated by the system in reference (Jara-Díaz, 1981). As is evident, spatial aggregation destroys information on the geographical context of the origin-destination system in which a transport system operates. Aggregation of output over time may cause distortions when estimating cost functions if periods of distinctive mean flows are being averaged. Finally, commodity aggregation may affect cost estimation since the (minimum) cost of moving the same aggregate weight or volume will generally depend on the composition of that output.

The loss of information due to aggregation over any dimension may cause serious problems of coefficient interpretation when estimating or analysing a cost function. Most reported transport cost functions use one or more basic output aggregates (for example ton-kilometres or total passenger trips) together with other ‘output’ variables or, as called in the literature, ‘output characteristics’ included to somehow control for the ambiguity of the aggregated output indexes. Thus, seasonal and
‘traffic condition’ dummies are in fact trying to capture the effect of the implicit time aggregation on costs. Similarly, variables like traffic mix or insurance value try to grasp commodity aggregation. The first effort to somehow counterbalance spatial aggregation was the use of mean haul length as part of output description within a ‘hedonic’ treatment (Spady and Friedlaender, 1978).

In the last twenty years, the literature on transport cost functions includes an enormous variety of output descriptions; unfortunately, this has not led yet to a universally accepted form of output treatment. On the other hand, there are many aggregate outputs whose relation with the disaggregated components is clear. This is the case of total tons, which is the summation of flows over every OD pair, or ton-kilometres, which is the summation of distance-weighted flows. Some of the aggregate outputs, however, have very particular relations with the flow vector, as is the case of the number of points served (PS). This aggregate has been frequently used in the airline industry analyses made through cost functions, which have developed significantly during the last decade, following deregulation during the eighties and the subsequent relevant changes in the size and structure of the networks operated by the different airlines. This seems to be the main reason for the introduction of PS as an argument in the estimation of cost functions, as a mean to provide information regarding network size. This variable was first introduced to capture firm size by Caves et al. (1984), followed by other studies as Dionne and Gagné (1988), Gillen et al. (1990), Oum and Zhang (1991) and Windle (1991).

As a variable conceived to describe network size, PS has been used in the literature to make a distinction between scale and density. In short, when the elasticity of cost with respect to PS is not included in the calculation, the result is said to represent economies of density (all products vary but the network is constant). On the other hand, it is argued that economies of scale include the variation of the network, and therefore the elasticity of PS is included. As we show below, the variation of PS is indeed associated with an increase in production, but not in terms of scale (increasing the volume of the same set of products) but of scope (increasing the number of products). The main analytical goal of this paper is precisely to show how a specific type of economies of scope can be calculated from cost functions with aggregate products, which include PS as an argument. We must stress that this should be seen as a highly relevant task, as network change has been the most important operating decision in the airline and trucking industries after deregulation. It should be noted that the relation between the spatial dimension of product and economies of scope has been
mentioned earlier in the literature on cost functions by, for example, Jara-Díaz (1981) and Spady (1985), who pointed out that the consideration of space in transport output allows for the identification of scope effects previously understood as scale, or by Daughety (1985, pp.473), who stated that “size” economies actually depend on scope economies “because changes in the product set are involved”.

In what follows, we will emphasise the spatial dimension of transport product, leaving aside for expository purposes the commodity and time dimensions of product in (2). This means that we will consider as the product of a transport firm in a certain time unit, a vector \( Y = \{y_i\} \), where \( y_i \) represents total flow in the \( i \)-th origin-destination pair served by the firm. In this case, according to the explanation and definitions at the beginning of this section, the multiproduct concept of scale applied to the transport firm admits only one possible interpretation: there are economies of scale if an increase by the same proportion of flows in all origin-destination pairs, provoke an increase in costs by a smaller proportion\(^1\). And there are economies of (spatial) scope if it is less costly for one firm to generate all flows in \( Y \) (i.e. to serve all OD pairs) than to specialize production in a spatial sense, letting one firm serve part of the OD pairs and other serve the remainder\(^2\).

To illustrate these concepts, we show in figure 1 a system with three nodes and six OD pairs (please note that this is not a physical network). If there are economies of scale on this OD system up to the level \( Y \) of production, it means that it is not convenient to produce all six flows with more than one transport firm with each firm \( i \) producing a fraction \( \alpha_i \) of \( Y \) such that \( \sum \alpha_i = 1 \).

\(^1\) Indeed, this is the implicit definition used by Griliches (1972) when he stated that “...all the studies examined ask the question What will happen to average costs if total traffic is expanded on the average in the same proportions and having exactly the same distribution over the various commodities, types, routes, and seasons as the previously handled traffic? There may be very little return to scale from a proportionate increase in all kinds of traffic. Whatever decreasing costs there may be are likely to arise only if one can contemplate disproportionate changes in traffic, changes in some kind of traffic but not in others. But that can not be discovered from such studies as we have examined above. It requires a different and much more ad hoc research program”.

\(^2\) Note that there are many type of economies of scope that one can think of. For example regarding different commodities carried (see Harmatuck, 1991, Kim, 1987, Keeler and Formby, 1994), or between different services (Colburn and Talley, 1992, Tauchen et al., 1983).
The existence of economies of scope for a given orthogonal partition of $Y$ has a different meaning. Let us examine, for example, the partition $(y_1, y_2, y_3, y_4, 0, 0), (0, 0, 0, 0, y_5, y_6)$. In this case, the presence of economies of scope with respect to this partition indicates the convenience of serving all six OD pairs with one firm instead of doing it with two firms as shown in figure 2.

Let us analyse the number of points served $(PS)$. If one firm is serving all OD pairs corresponding to $PS$, the dimension of $Y$ is simply $PS (PS-1)$. This corresponds to six in figure 1. If $PS$ increases by one, the number of OD pairs increases by $2 \cdot PS$, as the new node is a potential origin and destination for all other nodes previously served. For example, following figure 3, adding node $d$ to the OD network increases the number of OD pairs by six as shown. Please note that increasing the number of OD pairs does not mean that flows in every pair will be in fact produced by the firm.
In spite of this clear relation between $PS$ and $Y$, the literature on airline transport shows that $PS$ has been used to distinguish between economies of density (the inverse of the sum of the cost elasticities not including $PS$) and economies of scale (which includes the elasticity of cost with respect to $PS$ in the calculation). We have shown, however, that it is not possible to make an interpretation of the variation of costs after a variation of $PS$ in terms of any proportional change in the components of $Y$. What lies behind a unit variation in $PS$ is an increase in the number of OD pairs, which means adding new products. Therefore, the elasticity of cost with respect to $PS$ should be somehow related with scope and not with scale. Exploring this relation is the focus of the next section.

3. **Economies of spatial scope from cost functions including $PS$.**

We will assume that a cost function $\tilde{C}$ with aggregate outputs $\tilde{Y}$ including $PS$ has been estimated. Our goal is to find an explicit relation between scope, as defined and discussed in the preceding section, and the analytical relation between $\tilde{C}$ and $PS$ obtained econometrically (e.g. “marginal cost” $\partial \tilde{C}/\partial PS$ or “product elasticity” $\eta_{PS}$).

Let us consider a transport firm $A$ serving $N$ points from an $M$ points network. Besides, firm $A$ potentially serves all the possible origin-destination pairs among the $N$ points, that is to say, $N(N-1)$ pairs, as we show in Figure 4. Thus, the product of firm $A$, $Y^A$, can be written as

$$Y^A = (y_1, y_2, ..., y_{N(N-1)}, 0, 0, ..., 0)$$

(3)
where $y_k$ represents the flow in origin destination pair $k$.

Note that $Y^A$ has $(M-N)(M+N-1)$ null components. Let us define other firm $B$ serving the rest of the OD pair system of the $M$ points network. The origin destination pairs potentially served by $B$ are $Q(Q+2N-1)$, obtained from a) the flows generated among the $Q$ points served only by firm $B$, that is $Q(Q-1)$ pairs; b) all the flows with origin at any of the $Q$ points and destination at any of the $N$ original points, that is $Q \cdot N$ pairs; and c) all the flows with origin at any of the $N$ original points and destination at any of the $Q$ points, which adds another $QN$ pairs. Graphically

Therefore, firm $B$ produces a flow vector $Y^B$ given by
Finally, we describe a firm $C$, serving all the OD pairs of the system, such that its product $Y^C$ is

$$Y^C = (y_1, y_2, \ldots, y_{N(N-1)+1}, \ldots, y_{N+Q}(N+Q-1))$$

This way, vectors $Y^A$ and $Y^B$ represent an orthogonal partition of $Y^C$. Moreover, the partition is particularly interesting, as firm $A$ represents a local firm serving a closed subset of the network served by $C$. Firm $B$ connects the terminals served by firm $A$ with the rest of the system. The association among firms, points served and OD pairs is presented in Table 1.

**Table 1: Operation synthesis: Firms A, B and C**

<table>
<thead>
<tr>
<th>VECTOR</th>
<th>PS</th>
<th>OD</th>
<th>PRODUCT DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>$Y^A$</td>
<td>$N$</td>
<td>$N(N-1)$</td>
<td>$Y^A = (y_1, y_2, \ldots, y_{N(N-1)}$, $0$, $0$, $\ldots$, $0$)</td>
</tr>
<tr>
<td>$Y^B$</td>
<td>$N+Q$</td>
<td>$Q(Q+2N-1)$</td>
<td>$Y^B = (0, 0, \ldots, 0, y_{N(N-1)+1}, \ldots, y_{N+Q}(N+Q-1))$</td>
</tr>
<tr>
<td>$Y^C$</td>
<td>$N+Q$</td>
<td>$(N+Q)(N+Q-1)$</td>
<td>$Y^C = (y_1, y_2, \ldots, y_{N(N-1)}, y_{N(N-1)+1}, \ldots, y_{N+Q}(N+Q-1))$</td>
</tr>
</tbody>
</table>

According to our formulation, vectors $Y^C$ and $Y^B$ are associated to the same number of points served. However, the corresponding OD pairs are different. Accordingly, aggregate descriptions of product (e.g. pass-kilometres or total passengers) are also different, i.e. $\tilde{f}(Y^B) \neq \tilde{f}(Y^C)$. Moreover, $\tilde{C}[\tilde{f}(Y^B)]$ will be necessarily smaller than $\tilde{C}[\tilde{f}(Y^C)]$.

Starting from this formulation, we will be able to associate the concept of scope with the cost-PS elasticity, $\eta_{PS}$. This latter can be approximated discretely as

$$\eta_{PS} = \frac{\partial \tilde{C}}{\partial PS} \frac{PS}{\tilde{C}} \approx \frac{\Delta \tilde{C}}{\Delta PS} \frac{PS}{\tilde{C}}$$

(6)
where $PS$ and $\tilde{C}$ are the number of points and cost associated to a reference product, respectively. Let us take $Y^A$ as this reference; therefore, $PS = N$ and $\tilde{C} = \tilde{C}(\tilde{Y}(Y^A))$. Now we can evaluate the approximation in eq. (6), arbitrarily choosing $Y^C$ and $Y^B$ as the end point. Analytically,

$$\eta^C_{PS} \approx \frac{\tilde{C}(\tilde{Y}(Y^C)) - \tilde{C}(\tilde{Y}(Y^A))}{N} \frac{N}{N + Q - N} \tilde{C}(\tilde{Y}(Y^A))$$  \hspace{1cm} (7)$$

$$\eta^B_{PS} \approx \frac{\tilde{C}(\tilde{Y}(Y^B)) - \tilde{C}(\tilde{Y}(Y^A))}{N} \frac{N}{N + Q - N} \tilde{C}(\tilde{Y}(Y^A))$$  \hspace{1cm} (8)$$

Let us obtain analytical expressions for $\tilde{C}(\tilde{Y}(Y^A))$ and $\tilde{C}(\tilde{Y}(Y^B))$ from (7) and (8), that is

$$\tilde{C}(\tilde{Y}(Y^A)) = \frac{N}{Q \eta^C_{PS} + N} \tilde{C}(\tilde{Y}(Y^C))$$  \hspace{1cm} (9)$$

$$\tilde{C}(\tilde{Y}(Y^B)) = \frac{N + Q \eta^B_{PS}}{N + Q \eta^C_{PS}} \tilde{C}(\tilde{Y}(Y^C))$$  \hspace{1cm} (10)$$

Finally, from (9) and (10), we can develop an expression for the degree of economies of scope $SC$, defined in (1), as a function of cost-$PS$ elasticities, considering that the aggregate product $\tilde{Y}$ can be expressed as a function of the real product $Y$. As $\tilde{C}(\tilde{Y}(Y^A))$ is an estimate $\hat{C}(Y)$ of the cost function expressed in terms of $Y$, and noting that basic products $Y^A$ and $Y^B$ are an evident orthogonal partition of $Y^C$, the following expression is consistent with the definition of $SC$

$$SC(Y^A) = SC(Y^B) = \frac{\{\tilde{C}(\tilde{Y}(Y^A)) + \tilde{C}(\tilde{Y}(Y^B)) - \tilde{C}(\tilde{Y}(Y^C))\}}{\tilde{C}(\tilde{Y}(Y^C))}$$  \hspace{1cm} (11)$$

Replacing terms and simplifying,

$$SC(Y^A) = SC(Y^B) = \frac{-N - (M - N)(\eta^C_{PS} - \eta^B_{PS})}{N + (M - N)\eta^C_{PS}}$$  \hspace{1cm} (12)$$

which is evidently unsigned a priori as $\eta^C_{PS} - \eta^B_{PS}$ is positive from equations (7) and (8). Thus, a positive value of (12) means that is cheaper to serve the $M$ points with a single firm than to split production into two specific orthogonal subsets, where one represents a closed system of $N$ points. This is quite an interesting type of economies of scope, as it deals with the convenience of the creation of a spatially specialised transport firm, serving a subset of nodes. One can picture such a system thinking about the interurban and urban bus services, or international and domestic flights.
Note that with this analytically consistent development from a disaggregated point of view, we have shown that the economic convenience of serving a network structure by one or more firms can be analysed using the same elements required in the literature to calculate the degrees of economies of scale and density. Expression (12) to calculate $SC$, obtained from the cost-$PS$ elasticities approximation, shows that this multioutput microeconomic indicator can be estimated, provided that the number of points served, the functional form for $\tilde{C}$ and its specification, and the cost-$PS$ elasticities measured at different points (which depend on the estimated parameters), are known for the firm (or set of firms) studied.

4. Comments and Conclusions

The research described in this paper was motivated by what we considered an ambiguity in the literature of transport cost functions, namely the treatment of network size indices as elements whose cost elasticities contributed to the calculation of scale economies. As shown in section two, this ambiguity is particularly clear in the case of the variable “number of points served” $PS$, included in many articles regarding the airline industry. Increasing $PS$ is a reflection of the addition of new products to the line of production; thus, the variation of cost due to $PS$ should be related with spatial scope. Our objective has been to both identify and explore this relation.

As a result, we have deduced an analytical expression that links the $PS$ elasticity of cost with a specific type of economies of spatial scope, namely the convenience of separating a closed system of $N$ points out of a total of $M$, and making it a specialised local service. It has been obtained by taking into account two fundamental facts: first, that a $PS$ variation implies a variation in the number of OD pairs served; second, that the vector of aggregated products is a function of the real product vector. The analytical link between scope and the $PS$-elasticity of cost was based upon the interpretation and measure of this elasticity from an estimated cost function with aggregate output. As this elasticity is presently included in the calculation of economies of scale, it is clear that what we have called economies of spatial scope and what the literature identifies as economies of scale are indeed interrelated concepts. On the other hand, as shown in section two, what is presently referred to as economies of density is actually economies of scale. The relation between scope and the network size elasticity of cost, found in this article, suggests an alternative interpretation of the results obtained from transport cost functions with aggregate outputs. This in fact might rescue such aggregate specifications as a useful tool for the analysis of economies of scope. This poses a
demanding challenge for the future, which is to reveal the relation between other network related variables and the possible presence of economies of scope, but this is only part of the challenge, which we will now try to formulate.

In essence, what seems to be lacking in this field is a justified specification of output indices in transport cost functions. As transport production means movements in space, the addition of so-called network size variables to aggregate descriptions of product in transport cost functions is a mean to capture this essential feature. From this perspective, such procedure seems to add richness to the analysis, but its precise contribution to the study of industry structure is far from being clear. At this point one wonders whether the addition of variables to the pre-existing set of aggregates is the right thing to do. In order to understand the spatial dimension of the economic behaviour of transport firms through cost functions, we believe that another type of approach is needed. In our opinion, this approach encompasses various dimensions, beginning with the study of the process of transport production itself. A fresh view of what inputs and outputs are, and what the technical process of transformation of the former into the latter is, would greatly help in re-establishing a research agenda for this topic. On the other hand, most cost function studies in the last twenty years claim that although it is well known what transport output is, aggregates are needed for econometric feasibility; however, only isolated efforts have been developed to get a rigorous idea of what is being lost by accepting the aggregates intuitively in the empirical work. Understanding the relations between these aggregates and the true product has proved rewarding for the correct analysis of scale economies (Jara-Díaz and Cortés, 1996) and, as illustrated in this article, to get also a better idea of the links between the aggregates and scope. Moving this direction should help understanding what “economies of scale with variable network” really is, among other things. Finally, the meaning of scale and scope in transport activities is quite clear in reference to the disaggregated product (the flow vector); this should be carefully considered in order to build the best synthetic descriptions of transport output from product \( Y \), which means searching for optimal transport output aggregation procedures\(^3\). Working along these lines should add both strictness and richness to the transport industry structure analysis.

**ACKNOWLEDGEMENTS**

This research was funded by FONDECYT, Chile. We thank Leonardo Basso for assistance.

REFERENCES


