Are Returns to Scale with Variable Network Size Adequate for Transport Industry Structure Analysis?

Leonardo J. Basso, Sergio R. Jara-Díaz

Departamento de Ingeniería Civil, Universidad de Chile, Casilla 228-3, Santiago, Chile
{lbasso@ing.uchile.cl, jaradiaz@ing.uchile.cl}

It is customary to analyze transport industry structure using two indices: (1) economies of density and (2) *economies of scale with variable network size*. The latter has been defined to analyze the behavior of costs when output and network size expand simultaneously. After reviewing in detail what is intended with the calculation of RTS under this definition, we show analytically that, when the spatial aspects underlying transport production are taken into account, the seemingly reasonable conditions imposed on the aggregate output descriptions and the network variable conceal implicit output expansions that are not uniquely defined: they happen to depend on the specification of variables and on the evaluation point. Furthermore, most of the multiple output expansions analyzed correspond to cases that are hardly instructive. We conclude that this index is inherently ambiguous, hardly contributes to an adequate analysis of transport industry structure, and should be replaced by the calculation of economies of spatial scope (*Journal of Economic Literature* L91, L11, D40).

Key words: transportation economics; returns to scale; transport production; multioutput cost functions
History: Received: March 2005; revision received: November 2005; accepted: February 2006.

1. Introduction

Industries characterized by network technologies are disproportionately represented in econometric cost studies. There are two related reasons for this. First, network technologies are usually thought to be characterized by economies of scale. This has resulted in most of them being regulated over the years, which, in turn, has meant much better than average data availability for cost function estimation. Furthermore, the opening of such industries to competitive entry has often focused important policy debates on the extent of scale economies that may or may not be present. Unfortunately, their network structure makes the aggregation problem under discussion particularly severe. If point-to-point transportation (or transmission) movements are viewed as the true cost-causitive outputs of the firm, a firm operating even a relatively small network must be viewed as producing an astronomical number of products. (Panzar 1989, p. 44)

Spady and Friedlaender (1978) showed how sensitive transport cost function analysis was to product specification. The emergence of the new multioutput theory (synthesized by Baumol, Panzar, and Willig 1982) further promoted this discussion. Presently, most of the empirical work on transport cost functions includes a variety of product descriptions, attributes, and network indices. After Caves, Christensen, and Tretheway (1984), it became customary to analyze transport industry structure using two indices: (1) economies of density (RTD) and (2) economies

of scale with variable network size (RTS).[1] The first, aimed at analyzing costs as product grows within a fixed-size network, is calculated as the inverse of the sum of the cost elasticities with respect to products. The second, aimed at studying both product and network growth, includes, in addition, the elasticity of cost with respect to the network size. Both RTD and RTS have become *the* textbook concepts to analyze transport industry structure (Small 1992; Berechman 1993; De Rus and Nash 1997; Braeutigam 1999; Pels and Rietveld 2000).

Some authors, however, have expressed reservations concerning these concepts. Regarding RTD, Panzar (1989), in a simple example, showed that "returns to density are precisely equal to (what has been previously defined to be) the degree of multiproduct economies of scale!" (pp. 43–44), something also mentioned by Hurdle et al. (1989) and Filippini and Maggi (1992). Jara-Díaz and Cortés (1996, hereafter JDC) showed theoretically that, in fact, an improved version of economies of density is scale under a strict multioutput definition. Oum and Zhang (1997) concurred with JDC (1996). Regarding RTS, Panzar (1989), whose observations have received surprisingly little attention in the transportation literature, showed that the Caves et al. (1985) measure of *returns to scale* (RTS) is always equal to one in his

---

[1] Antoniou (1991) traces this distinction back to Koontz (1951), however.

simple example, pointing out that this is not particularly relevant for the analysis. Second, some authors have suggested, either literally or implicitly, that RTS and scope are related (Daughety 1985; Hurdle et al. 1989; Borenstein 1992). JDC (1996) pointed out that network expansions require scope—and not scale—analysis, because they usually imply an increase in the number of products (origin-destination pairs). Finally, Antoniou (1991) argued that not considering the so-called network attributes in the calculation of RTS is unjustified. Oum and Zhang (1997) proposed to consider the relations between attributes and the network variables to calculate RTS.

Yet, despite the caveats and proposals for improvements, RTS still is *the* concept used in applied research to study issues such as ownership, regulatory reform, the scope for competition, or postderegulation market structures.[2] In this paper, we provide a rigorous base to examine what has been intuitively and qualitatively suggested, explaining why and where RTS fails as a tool for industry structure analysis. We begin by discussing output, scale, and scope in transport analysis in §2. Building on this, in §3, we offer a discussion of the current interpretation of RTS. In §4, we then move to the main issue, a close examination of the meaning and use of RTS. We conclude that this index is ill-defined and does not contribute to an adequate analysis of transport industry structure, proposing spatial scope as the relevant concept to be used.

2. Product, Scale, and Scope in Transport

A multioutput cost function $C(\mathbf{w}, \mathbf{Y})$ is defined as the minimum expenditure necessary to produce the output vector $\mathbf{Y}$ at input prices $\mathbf{w}$. Economies of scale exist if an expansion by the same proportion of all products in $\mathbf{Y}$ causes a less than proportional increase in cost. Economies of scope exist if it is cheaper to produce $\mathbf{Y}$ with one firm than to split production into two orthogonal subsets. In other words, scale analysis deals with the (proportional) growth of all products, while scope analysis is related with the addition of new products to the line. Analytically, the (multiproduct) degree of economies of scale at $\mathbf{Y}$, $S(\mathbf{Y})$, is calculated as the inverse of the sum of cost elasticities with respect to products at $\mathbf{Y}$. A value of S larger, equal, or less than one shows increasing, constant, or decreasing returns to scale at $\mathbf{Y}$, respectively, indicating the relative cost efficiency of proportional expansions or reductions of output. The
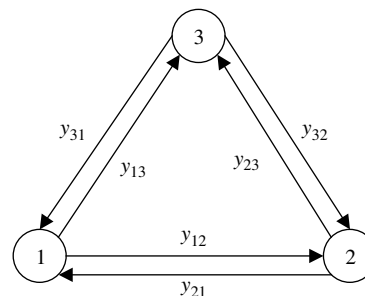

Figure 1   An Illustrative OD Structure

degree of economies of scope at $\mathbf{Y}$ relative to a subset R of products, $SC^R(\mathbf{Y})$, is calculated as $SC^R(\mathbf{Y}) = (C(\mathbf{w}, \mathbf{Y}^R) + C(\mathbf{w}, \mathbf{Y}^{M-R}) - C(\mathbf{w}, \mathbf{Y}))/C(\mathbf{w}, \mathbf{Y})$, where $\mathbf{Y}^R$ is vector $\mathbf{Y}$ with $y_i = 0$, $\forall\, i \notin R \subset M$, and M is the whole product set. A positive SC^R means that it is cheaper to produce $\mathbf{Y}$ with one firm than to split production into two orthogonal subsets R and $M - R$ (Baumol, Panzar, and Willig 1982).

In the case of transport, a firm produces movements of people and goods between many origins and destinations (OD pairs) during different periods. Strictly, then, a transport firm produces a vector $\mathbf{Y} = \{y_{ijkt}\}$, where y_{ijkt} represents flow of type k (goods or people), between origin i and destination j, during period t (Jara-Díaz 1982a, b; Winston 1985; Ying 1992; Braeutigam 1999). Because of our emphasis on the spatial dimension of transport production, let us keep the ij subindices only. Scale economies exist if cost increases less than $\lambda\%$ when all flows increase by $\lambda\%$, while there are economies of scope if it is cheaper to produce all OD flows $\mathbf{Y}$ with a single firm than to specialize production spatially. In the example of Figure 1, regarding the movement of freight on six OD pairs, there are economies of scale at $\mathbf{Y}$ (i.e., $S > 1$) if it is not efficient from a cost viewpoint to have various firms, each one serving a fraction of $\mathbf{Y}$, competing on the six OD pairs. If there are diseconomies of scope for the partition $\{y_{12}, y_{21}, 0, 0, 0, 0\}$, $\{0, 0, y_{23}, y_{32}, y_{13}, y_{31}\}$—i.e., if $SC < 0$—then it would not be cost efficient to have a single firm producing all six flows, but it would be better to have two firms, each one serving one of the subsets (see Jara-Díaz 2000). Conversely, it may well happen that firms exhibit $S = 1$ and $SC > 0$ simultaneously (as illustrated in Jara-Díaz 1982b; Jara-Díaz and Basso 2003), in which case it would be cost efficient to have them competing, each one serving all six OD pairs.[3]

---

[2] Recent papers are Mizutani (2004, rails), Liu and Lynk (1999, airlines), and Filippini and Prioni (2003, buses). For earlier studies, see Oum and Waters (1996), Braeutigam (1999), Pels and Rietveld (2000), and Jara-Díaz (2000).

[3] It is important to note that the OD structure represented in Figure 1 does not represent a physical network nor a route structure but the product vector (Jara-Díaz and Basso 2003). Thus, a firm using a hub-and-spoke route structure (hub in Node 3, say), still produces six outputs, in spite of using only two links.

As the number of OD pairs served is usually huge, output aggregation is necessary for the econometric estimation of cost functions. Many aggregates have been used in the literature, some of which are named "products," such as passenger-kilometers, seat- or vehicle-kilometers, or number of shipments, while others are named attributes or characteristics, such as average length of haul or load factor; since the mid-1980s, network size variables—such as route miles or the number of points served—have also been considered (Jara-Díaz 2000). Both the strict definition of transport output and the need for aggregation have been frequently recognized in the literature.[4] Hereafter, following Jara-Díaz (1982a, b) and Winston (1985), we will call vector **Y** the *true transport product*, as a way to distinguish it from the vector of aggregates, which we will denote by $\widetilde{\mathbf{Y}} = \{\tilde{y}_1, \ldots, \tilde{y}_h, \ldots, \tilde{y}_V\}$. As seen here, the concepts of scale and scope are crystal clear when the vector **Y** is considered; let us move then to the relevant case of aggregates and the definition and use of its associated scale concepts: RTD and RTS.

3. Returns to Density and Returns to Scale from Aggregated Cost Functions

Consider an estimated cost function $\widetilde{C}(\widetilde{\mathbf{Y}}; N)$, where N is the variable representing network size and input prices are suppressed for simplicity. Returns to density (RTD) and returns to scale (RTS) are defined on $\widetilde{C}(\widetilde{\mathbf{Y}}; N)$: "RTD refers to the impact on average cost of expanding all traffic, holding network size constant, whereas RTS refers to the impact on average cost of equiproportionate increases in traffic and network size" (Oum and Waters 1996, p. 429). Analytically,

$$\text{RTS} = \frac{1}{\sum_{h \in H} \tilde{\eta}_h + \eta_N}, \tag{1}$$

where $\tilde{\eta}_h$ is the elasticity of $\widetilde{C}(\widetilde{\mathbf{Y}}; N)$ with respect to aggregate product $\tilde{y}_h$ and η_N is the elasticity with respect to N. As RTD does not include η_N,

---

[4] Ying (1992, p. 231) states, "This multiproduct nature is especially evident for transportation firms, which transport various commodities or passengers from a specific origin to a specific destination over a spatial network.... Empirical studies in transportation have necessarily aggregated output data.... In these situations, some researchers have tried to capture the heterogeneous nature of this single output through a vector of output quality or attribute variables." Braeutigam (1999, p. 68) states that "treating the movement of each commodity from each origin to each destination as a separate product would be desirable. There would be so many outputs however, that estimating a cost function would be impossible." See also Winston (1985), Gillen, Oum, and Tretheway (1990), and Small (1992). The first detailed discussion regarding output definition and aggregation can be traced back to Jara-Díaz (1981).

RTD > RTS because the network size elasticity is positive since, *everything else constant*, serving a larger network implies larger expenses (Pels and Rietveld 2000). Increasing returns to scale (RTS > 1) suggest that, if possible, both network size and products should be increased because serving larger networks would diminish ray average cost; along that ray, the industry would be seen as a natural monopoly. Constant returns to scale together with increasing returns to density (RTD > 1) would indicate that, if possible, traffic should be increased keeping network size constant.

As explicitly shown in Equation (1), the sum of the product elasticities is made over a subset H of aggregates, whose definition is an issue still unresolved in the literature. Most articles do not include the so-called attributes in H, as in Friedlaender et al. (1993, railroads), Kumbhakar (1990, airlines), and Bhattacharyya, Kumbhakar, and Bhattacharyya (1995, buses). Other authors argue that the inclusion of certain elasticities will depend on how the product is expanded, as Caves et al. (1985, railroads) who consider the average length of haul elasticity in some of their RTS calculations, Windle (1988, buses) who include the load factor elasticity in a calculation of RTD, and Caves and Christensen (1988, buses and airlines) who include the load factor elasticity for some RTS calculations.

Gagné (1990), Ying (1992), and Xu et al. (1994) considered the interrelations among aggregates ("products" and "attributes") to calculate a total rather than a partial elasticity. JDC (1996) sustained that the interrelations arose because most elements in vector $\widetilde{\mathbf{Y}}$ are implicit functions of the true output vector **Y**, i.e., $\widetilde{\mathbf{Y}}$ is actually $\widetilde{\mathbf{Y}}(\mathbf{Y})$. For instance, total flow is simply the sum of the y_{ij} over all OD pairs, ton-kilometers are a sum of distance-weighted OD flows and average length of haul is the ratio between the latter and the former. This makes $\widetilde{C}$ an implicit function of **Y** as well, which means that $\widetilde{C}(\widetilde{\mathbf{Y}}; N) \equiv \hat{C}(\mathbf{Y})$ and $\widetilde{C}$ can be used to calculate the elasticities of cost with respect to the true product using the chain rule. Following this, the multiproduct degree of economies of scale S, defined in §2, was shown to be equal to the inverse of the sum of the aggregates' cost elasticities, each one multiplied by a factor α_h that corresponds to the local degree of homogeneity of each aggregate with respect to **Y**. With this method, the decision about which aggregates should be considered in H (be it products or attributes) is no longer arbitrary, although the α_h could be different from zero or one. Oum and Waters (1996) described the JDC (1996) method as "a more rigorous reconsideration" of the problem, although pointing out that it corresponds to the calculation of RTD rather than RTS, because the network is not allowed to vary. This is correct: S in JDC (1996) is calculated at the disaggregated **Y** level, considering

proportional expansions of the true products (flows) and, therefore, keeping the network served constant as no new products (new OD pairs) are considered. Therefore the result advanced by Panzar (1989) using a simple example can be generalized: RTD (as defined in the literature) and S become equivalent provided the α_h defined above are used.[5]

Regarding RTS, where the network size is not fixed, the few and mostly qualitative criticisms summarized in §1 have not generated an agreed measurement procedure. We now identify some problems that show the need to sort this out. First, note that an efficient industry configuration (Baumol, Panzar, and Willig 1982) in terms of production and network size would require firms exhibiting $RTD = RTS = 1$, but this can never be obtained from $\widetilde{C}(\widetilde{\mathbf{Y}}; N)$, even if such efficient configuration existed. Second, as seen in the previous section, when transport product is described in detail (vector $\mathbf{Y}$), S and SC play unambiguously complementary roles, analyzing proportional flow expansions and addition of OD pairs, respectively. With RTD and RTS similar complementary roles are intended, which is why some authors have related RTS to scope. However, the relevant case of $S = 1$ with $SC > 0$, a perfectly well-defined outcome in terms of $\mathbf{Y}$, cannot have a counterpart in $RTD = 1$ and $RTS > 1$. Furthermore, in RTS, *density*—understood as the ratio between (aggregated) product and the network index—is kept constant while in the case of economies of scope, when the network is expanded and new OD flows are produced, nothing is imposed a priori on the average density of the transport system.[6] Hence, while RTS has indeed some relationship with spatial scope, its value relative to *one* (the usual scale reference) will not necessarily lead to the same conclusion as $SC > 0$, as confirmed empirically by Basso and Jara-Díaz (2005). So, if RTD is multioutput scale, *and RTS is not scope*, what is RTS then? We answer this next by moving from the RTS definition to an examination of the implicit constraints it imposes on the true transport output vector; that is, the type of output expansion it analyzes.

---

[5] We should note, however, that this does not close the discussion regarding scale measures within a fixed-size transport network, as RTD seems to be linked to the idea of an invariant route structure, something that is not guaranteed after an expansion of $\mathbf{Y}$; see Jara-Díaz and Basso (2003) and Basso and Jara-Díaz (2006).

[6] The concept of *density* in RTS is somewhat ambiguous as it depends on what aggregates are used to describe output (e.g., total number of passengers, passenger-kilometers, or seat-kilometers). In many cases, authors argue that economies of density are present if the average cost of a direct connection decreases with increments on flows on that connection (see, for example, Hendricks, Piccione, and Tan 1995), implying that density should be measured through flow and not volume-distance or capacity measures.

4. A Disaggregate Examination of RTS

RTS is defined on the aggregates, $\widetilde{\mathbf{Y}}$, and the network variable, N. Let us focus on N first. The two network size variables most frequently used in the literature are the number of points served, PS, used mainly in air transport, and route miles, RM, mostly used in land transport. Because PS is simply the number of nodes in the system (e.g., number of airports connected), when PS grows, the number of products grows. On the other hand, RM has two interpretations; sometimes it is understood as the total length of the physical network available, and sometimes as the total distance *used* within the physical network. Under the first definition, the relation between RM and production is unclear, as the network can grow keeping the number of nodes and the number of products (OD pairs) constant, as opposed to the PS case. Moreover, if RM does not change, density might change in two ways: (1) increasing product in the existing OD pairs *or* (2) *increasing the number of OD pairs for the given network* (Keaton 1990). This last output expansion, described as density (RTD) when RM is used, would be described as scale (RTS) if PS was used.

OBSERVATION 1. There is a source of ambiguity regarding the use of N within the context of RTS, as the concept itself depends on which network descriptor is used.

Because we will need to focus on a specific network size variable for the analysis, we will use PS as it is unambiguous and has a direct interpretation in terms of the true flow vector; we comment on RM at the end of the section.

Next, How does RTS relate to the true output vector? Equation (1) provides no information about this. One of the few explicit graphical interpretations of a relation between flows and RTS is given by Braeutigam (1999; replicated in Pels and Rietveld 2000): in the transport system reproduced in our Figure 2, network size and aggregated flows—represented by ton-kilometers (TK)—vary by the same proportion keeping density constant. As he
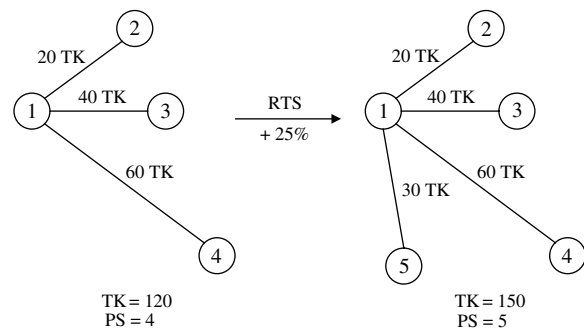

Figure 2 A 25% Network Expansion, Constant Density (Braeutigam 1999)

explains, "the size of the network has been increased by adding another node (5) and a link to serve that node. There are now five nodes (PS = 5), so the number of *nodes* has been increased by 25%, but the volume of traffic over the existing links has not been increased (the density of traffic is unchanged). In other words, the *size* of the network is increased, but the *density* of traffic movements is unchanged. This is the type of output expansions usually envisioned in studies of economies of scale" (Braeutigam 1999, p. 75).

As explicitly stated by the author, the true product is the underlying vector of OD flows y_{ij}, which means up to 12 OD flows on the original network and up to 20 on the final one, whose values are not given. Note that only one link is added and the hub-and-spoke route structure is preserved. How do the TK figures, before and after the network expansion, relate to the flows in **Y**? Let us denote the distance between nodes i and j as d_{ij}. Clearly, the total ton-kilometers on link $1j$, TK_{1j}, is given by the total flow that travels through that link times d_{1j}. Initially, then

$$\text{TK}_{1j}^0 = d_{1j} \cdot \sum_{\substack{i=1 \\ i \neq j}}^{4} (y_{ij}^0 + y_{ji}^0), \quad j \in \{2; 3; 4\}, \qquad (2)$$

where the superscript 0 denotes the original flow values. After the network is expanded, eight new OD pairs are added but only two of these new flows, namely 1–5 and 5–1, will use the new link; the other six will necessarily use some of the other original four links, adding $y_{i5} + y_{5i}$ to the parenthesis of Equation (2). As TK_{1j} remains constant for all $j \in \{2; 3; 4\}$ after Node 5 is added, either the original OD flows diminish (which ones and by how?) or they are kept constant but six out of the eight new OD flows are nil. Besides this ambiguity, unpleasant by itself, neither alternative seems particularly informative for a cost-based industry structure analysis. The hidden constraint, on the OD flows behind the constancy of ton-kilometers on each link reveals a problem with this interpretation of RTS.

Another explicit interpretation of the relation between flows and RTS is given by Oum and Zhang (1997) within the context of their discussion regarding the inclusion of the average length of haul elasticity into the calculation of RTS. They explicitly state that "If an increase in output is accompanied by a change in network size, that is, *if the number of origin-destination pairs also increases with traffic flows in each route*, then theoretical consistency would require investigation on the correlation between length of haul and network size" (p. 310, emphasis added). This implies growth of traffic on all links, which is not compatible with the graphical example and the interpretation of RTS provided in Figure 2.

OBSERVATION 2. The literature on RTS does not provide a clear view regarding its relation with the true output vector **Y**: nothing can be inferred from its analytical definition (Equation (1)); some articles provide contradicting interpretations and most articles do not even touch on the problem.

Then, a rigorous analysis of RTS requires: (1) that we define the problem in terms of the true output vector; (2) to make use of the distinction between the flows on the original OD pairs and those on the added ones. To begin with, it is convenient to recall that the degree of scale economies, when the cost function is differentiable, corresponds to the inverse of the local degree of homogeneity of the cost function (Baumol, Panzar, Willig 1982). Applying this property to RTS as defined in Equation (1), the following identity holds:

$$\widetilde{C}(\lambda \widetilde{\mathbf{Y}}_H, \widetilde{\mathbf{Y}}_K, \lambda N) \equiv \lambda^{1/\text{RTS}} \widetilde{C}(\widetilde{\mathbf{Y}}_H, \widetilde{\mathbf{Y}}_K, N), \qquad (3)$$

where K is the complement set of H; that is, $H \cup K = \{1, 2, \ldots, V\}$, and $\widetilde{\mathbf{Y}}_H$ is a vector containing aggregates $\widetilde{y}_h$ whose elasticities are considered in the calculation of RTS. Identity (3) means that the *calculation of RTS analytically imposes that, after a network expansion, aggregates in H vary in that same proportion while those not in H (subset K) do not vary.* To explore the implications of the definition of RTS under this perspective, let us look at the true production behind the aggregates, as done by JDC (1996). Consider a firm A whose aggregated product vector and network size are described by $\widetilde{\mathbf{Y}}^A$ and N^A, respectively. We assume that a long-run cost function $\widetilde{C}(\widetilde{\mathbf{Y}}; N)$ that fulfils the usual properties[7] and that reproduces industry costs well is available; hence, firm A's production cost is $\widetilde{C}^A = \widetilde{C}(\widetilde{\mathbf{Y}}^A, N^A)$. To analyze *scale with variable network size* as defined in (3), let us consider an expansion of both the network and the aggregates in H by a certain proportion λ; aggregates not in H do not vary. To simplify notation, let us call L the expanded firm, such that $N^L = \lambda N^A$, $\widetilde{\mathbf{Y}}^L = (\lambda \widetilde{\mathbf{Y}}_H^A, \widetilde{\mathbf{Y}}_K^A)$, and $\widetilde{C}^L = \widetilde{C}(\lambda \widetilde{\mathbf{Y}}_H^A, \widetilde{\mathbf{Y}}_K^A, \lambda N^A)$.

To fully understand which true output expansion RTS is analyzing, i.e., how vector **Y** varies, we will try to unveil the characteristics of the expanded firm L in terms of its actual product, the $\mathbf{Y}^L$ vector. Note first that initial aggregates are defined by $\widetilde{\mathbf{Y}}^A(\mathbf{Y}^A)$. Similarly, behind the aggregate description of firm L's product, there is a true flow vector $\mathbf{Y}^L$ whose properties we want to examine. The problem can be formulated as follows: assume $\mathbf{Y}^A$ is known and that both N and part of $\widetilde{\mathbf{Y}}^A(\mathbf{Y}^A)$ expands by λ, forming $\widetilde{\mathbf{Y}}^L = [\lambda \widetilde{\mathbf{Y}}_H^A(\mathbf{Y}^A), \widetilde{\mathbf{Y}}_K^A(\mathbf{Y}^A)]$. Then, the question is: What

---

[7] Mainly, this function must be nondecreasing, concave, and linearly homogenous in input prices and nondecreasing in output.

are the characteristics or properties of $\mathbf{Y}^L$, the true flow vector underlying $\widetilde{\mathbf{Y}}^L$? Schematically,

$$(\mathbf{Y}^A; N^A) \Rightarrow [\widetilde{\mathbf{Y}}_H^A(\mathbf{Y}^A), \widetilde{\mathbf{Y}}_K^A(\mathbf{Y}^A); N^A]$$

$$\xrightarrow[\text{(RTS)}]{N \text{ expansion}} (\lambda \widetilde{\mathbf{Y}}_H^A, \widetilde{\mathbf{Y}}_K^A; \lambda N^A) \equiv (\widetilde{\mathbf{Y}}^L; N^L) \overset{?}{\leftrightarrow} (\mathbf{Y}^L; N^L).$$

Recall that we will use the number of points served, PS, as the network size variable. Because increasing PS, even by one node, implies new OD pairs in the system ($2PS^A$ potential new flows in that case), the expanded firm L will serve two types of flows: the original ones and new ones whose origin or destination is one of the new nodes. How are they assumed to change after the network expansion *within the RTS context*? As stated in Observation 2, nothing can be inferred from Equations (1) and (3) describing RTS, and little has been said in the literature regarding this issue. Recalling the discussion of Braeutigam's (1999) example, the first relevant question is whether the analysis should consider the original OD flows constant after the network expansion or not. We believe that, when one is dealing with new products in addition to old ones, the most reasonable condition for analysis is that the original OD flows (i.e., those served before the network expansion) keep their level.[8] An alternative condition could be that they decrease (as in Braeutigam's 1999 case); we get back to this later.

Let us now discuss what happens with the new flows. In doing so, we consider two of the most popular aggregates in the literature: (1) ton-kilometers (TK), usually an output, and (2) average length of haul (ALH), usually an attribute.[9] First, consider the simple network represented in Figure 3.

PS expands by $\lambda = 3/2$ from two nodes $(1, 2)$ to three $(1, 2, 3)$. The true output of firms A and L are $\mathbf{Y}^A = \{y_{12}, y_{21}, 0, 0, 0, 0\}$ and $\mathbf{Y}^L = \{y_{12}, y_{21}, y_{23}, y_{32}, y_{13}, y_{31}\}$. The values of the aggregates are

$$\text{TK}^A = y_{12} \cdot d_1 + y_{21} \cdot d_1, \tag{4}$$

$$\text{TK}^L = y_{12} \cdot d_1 + y_{21} \cdot d_1 + y_{23} \cdot d_2 + y_{32} \cdot d_2$$
$$+ y_{13} \cdot (d_1 + d_2) + y_{31} \cdot (d_1 + d_2), \tag{5}$$

$$\text{ALH}^A = \frac{y_{12} \cdot d_1 + y_{21} \cdot d_1}{y_{12} + y_{21}} = d_1, \tag{6}$$

Figure 3   Physical Network

$$\text{ALH}^L = (y_{12} \cdot d_1 + y_{21} \cdot d_1 + y_{23} \cdot d_2 + y_{32} \cdot d_2$$
$$+ y_{13} \cdot (d_1 + d_2) + y_{31} \cdot (d_1 + d_2))$$
$$\cdot (y_{12} + y_{21} + y_{23} + y_{32} + y_{13} + y_{31})^{-1}. \tag{7}$$

Let us analyze the constraints imposed by RTS on the new flows for three cases, each defined by the variables in the cost function and the elasticities considered in the calculation of RTS (set H).

Case 1. $\widetilde{C} = \widetilde{C}(\text{TK}, \text{PS})$ and $RTS = [\tilde{\eta}_{\text{TK}} + \eta_{\text{PS}}]^{-1}$.

According to the definition of RTS in Equation (3), TK expands by the same proportion as PS; this is, $\text{TK}^L = (3/2)\text{TK}^A$. From Equations (4) and (5) and after some manipulation we get

$$\frac{d_1}{d_2} = \frac{y_{23} + y_{32} + y_{13} + y_{31}}{0.5 \cdot (y_{12} + y_{21}) - (y_{13} + y_{31})}. \tag{8}$$

As d_1 and d_2 are known, (8) reveals implicit constraints on the new flows magnitudes. For example, the denominator on the right-hand side has to be positive. If $d_1 = d_2$, it should hold that $0.5(y_{12} + y_{21}) = y_{23} + y_{32} + 2(y_{13} + y_{31})$, i.e., the four new flows should add up to less than half the original ones.

Case 2. $\widetilde{C} = \widetilde{C}(\text{TK}, \text{ALH}, \text{PS})$, $H = \{\text{TK}\}$, and $RTS = [\tilde{\eta}_{\text{TK}} + \eta_{\text{PS}}]^{-1}$.

Now, ALH is a variable in the cost function, but is not considered in the calculation of RTS. Hence, TK increases by $3/2$ and ALH remains constant, i.e., $\text{TK}^L = (3/2)\text{TK}^A$ and $\text{ALH}^L = \text{ALH}^A$. Using Equations (4)–(7), we get the constraint

$$0.5(y_{12} + y_{21}) = y_{23} + y_{32} + y_{13} + y_{31}, \tag{9}$$

which reveals that the four new flows have to add up to half the original total flow.

Case 3. $\widetilde{C} = \widetilde{C}(\text{TK}, \text{ALH}, \text{PS})$, $H = \{\text{TK}; \text{ALH}\}$, and $RTS = [\tilde{\eta}_{\text{TK}} + \tilde{\eta}_{\text{ALH}} + \eta_{\text{PS}}]^{-1}$.

Here, both ALH and TK are considered in the RTS calculation.[10] Thus, RTS imposes that $\text{TK}^L = (3/2)\text{TK}^A$ and $\text{ALH}^L = (3/2)\text{ALH}^A$ simultaneously, leading to

$$y_{23} + y_{32} + y_{13} + y_{31} = 0. \tag{10}$$

In this case, RTS imposes that nothing should enter nor depart from the new node.

---

[8] This seems to have been the idea behind Figure 2, on the "wrong" output.

[9] As explained in §2, aggregation of the true output $\mathbf{Y}$ is a necessity; "this poses two interesting problems.... One is a pre-estimation problem, dealing with the search for adequate descriptions of output. The second is an after estimation problem, dealing with the interpretation of results" (JDC 1996, p. 158). It is the second problem that interests us, which is why we chose the two most popular aggregates in the literature. The first problem has been analyzed elsewhere (e.g., Antoniou 1991; Jara-Díaz, Donoso, and Araneda 1991).
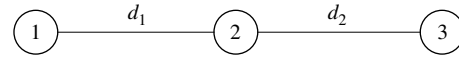
[10] This way of calculating RTS has been justified by some authors, who explain that because the network is allowed to change, it is possible that the length of haul also changes. See, for example, Caves et al. (1985).

Let us now generalize Cases 2 and 3. Equations (9) and (10) reveal constraints on the new flows that do not depend on the link distances, suggesting that they are independent both of the route structure (an endogenous decision) and of the physical network (exogenous information). This happens to be true. For all networks, $T = \sum_{i,j} y_{ij}$ and $ALH = TK/T$, where T is the number of total tons moved. In *Case* 2, $TK^L = \lambda TK^A$ and $ALH^L = ALH^A$. The latter implies $TK^L/T^L = TK^A/T^L$, which combined with the former yields $T^L = \lambda T^A$. Then, if y' and y are the mean of the new and original flows, respectively, their ratio has to fulfill (see the appendix)

$$\frac{y'}{y} = \frac{1}{2} \cdot \frac{PS - 1}{PS}, \qquad (11)$$

which is valid for all networks and depends only on the initial value of PS. Regarding *Case* 3, the conditions are $TK^L = \lambda TK^A$ and $ALH^L = \lambda ALH^A$. The latter implies $TK^L/T^L = \lambda TK^A/T^L$, which combined with the former yields $T^L = T^A$. This shows that RTS imposes that the new flows to be served after the network expansion have to be all zero, irrespective of the physical network, the original number of nodes, or the route structure.

As evident, the implicit constraints on the new flows revealed above differ from case to case, which means that comparability across studies using different output specifications become complex (if not impossible), besides the fact that the unveiled output expansions do not seem to be very informative. Furthermore, a closer look at Equations (8) and (11) exposes yet another important limitation of RTS; namely, that the implicit constraints on **Y** depend on the evaluation point as well, through the topology of the network and the number of nodes. To see this, note that the constraint on flows in Equation (8) (*Case* 1) is network dependent through the values of d_1 and d_2, a property that evidently generalizes to larger networks, and that Equation (11) (generalized *Case* 2) shows that the average of the new flows is between one-fourth and one-half the average of the original flows, depending on the value of PS. We conclude the following.

OBSERVATION 3. Under the condition that the magnitude of old flows remain fixed and PS is the network variable, *the type of output expansion that RTS analyzes is not uniquely defined* in terms of **Y**; it depends on the aggregates included in the cost function, on the elasticities considered in its calculation, and on the evaluation point.

Let us stress that what happens with RTS does not occur with S or with SC^R. Both of them analyze perfectly well-defined and *unique* output variations. S always analyzes the behavior of costs for an equiproportional expansion of current products, i.e., all OD flows grow by the same proportion, and SC^R deals with well-defined partitions of **Y**. Indeed, the actual value of S does depend on the evaluation point **Y**, but the *type* of output expansion does not. Observation 3 implies that RTS comparisons between different evaluation points or across studies are an ineffective exercise, for example, different firms in the same industry.

Now, Observation 3 has been deduced using PS as the network variable, but it does extend to the use of route miles (RM) as well. In this case, the proportion of network growth in Figure 3 would be given by $\lambda = (d_1 + d_2)/d_1$ and some different restrictions on the added flows would appear. These restrictions will be different for different output specifications and set H. They will be dependent on link distances in both the simple and more real networks as well, showing that the output expansion that is being analyzed varies with the network topology.

As for the condition that old flows remain constant, we explained that this is what we think is most reasonable and useful. Other conditions, particularly the one stemming from Figure 2, i.e., that *link* loads are constant, can be analyzed as well. We now show that problems persist. As discussed, in this interpretation, old flows have to diminish after the network expansion to keep the load on the original links unchanged. For the case of our simple network in Figure 3, we would have $\mathbf{Y}^A = \{y_{12}^A, y_{21}^A, 0, 0, 0, 0\}$ and $\mathbf{Y}^L = \{y_{12}^L, y_{21}^L, y_{23}, y_{32}, y_{13}, y_{31}\}$. To keep the load in the original link constant, however, the following condition has to be met:

$$y_{12}^A + y_{12}^A = y_{12}^L + y_{12}^L + y_{13} + y_{31}, \qquad (12)$$

which directly implies that $(y_{12}^A + y_{12}^A) \geq (y_{12}^L + y_{12}^L)$. Assuming, as in Figure 2, that PS is the network variable, we obtain the following conditions on flows, for Cases 1, 2, and 3, respectively:

$$\frac{d_1}{d_2} = \frac{y_{23} + y_{32} + y_{13} + y_{31}}{0.5 \cdot (y_{12}^A + y_{21}^A)}$$

$$= \frac{y_{23} + y_{32} + y_{13} + y_{31}}{0.5 \cdot (y_{12}^L + y_{12}^L + y_{13} + y_{31})}, \qquad (13)$$

$$\tfrac{1}{2}(y_{12}^A + y_{21}^A) = y_{23} + y_{32} = y_{12}^L + y_{21}^L + y_{13} + y_{31}, \qquad (14)$$

$$y_{23} + y_{32} = 0. \qquad (15)$$

Generalization of the simple network cases is a more difficult matter here as the equations that will ensure constancy of load on old links depend on the route structure; it is not obvious how the counterparts of Equation (12) would look for a general case (we just know there would be many equations). However,

if z is the average of firm A's flows and z' is the average of firm L's, it is easy to show that in Case 2 the RTS condition—that is $(\lambda \widetilde{\mathbf{Y}}_H^A, \widetilde{\mathbf{Y}}_K^A; \lambda N^A)$—leads to[11]

$$\frac{z'}{z} = \frac{PS - 1}{PS}. \tag{16}$$

Equations (13)–(16) show that Observation 3 applies to this interpretation regarding old flows as well, and that the different output expansions are still not really informative to the analyst (beginning with the fact that old flows are forced to diminish). Combining the *diminishing old flows* interpretation with the use of RM as the network variable does not improve the situation.

For synthesis, what Observations 1–3 show is that RTS is not uniquely defined in terms of what it analyzes, not even within a particular specification of output aggregates, network size variable, and the H set, which makes it inherently ambiguous. Note that this is not an econometric problem related with either misspecification of the cost function or with biased parameter estimates because of the inclusion of too many or omitted variables; it is an analytically induced ambiguity. And the larger the number of aggregates used to describe transport product (or its "attributes") in the cost function, the more acute the problems just described, because the number of simultaneous constraints increase, irrespective of their consideration or not in the calculation of RTS. All this evidently erodes, if not completely destroys, the usefulness of RTS as an instrument to analyze transport industry structure. Comparisons across studies and within a study are simply meaningless. In addition, each of the implicit constraints that have to be fulfilled by the underlying true product does not seem to be very reasonable. Note that the problems do not depend on the inclusion or absence of attributes elasticities in the calculation of RTS, something that drew some attention in the literature.

So, can RTS be rescued in some way? Antoniou (1991) and Liu and Lynk (1999) argue that in RTS calculations, attributes are held constant but, in fact, they vary. In our terminology, they were suggesting that forcing aggregates in H to vary exactly as the network does, while those not in H do not vary, was not justified. Oum and Zhang (1997, p. 310) take this view, arguing that "theoretical consistency would require investigation of the correlation between length of haul and network size," a relation previously suggested by Caves et al. (1985). Formally (our notation), Oum and

Zhang (1997) explicitly define $ALH = f(N)$ and argue that RTS should be calculated as

$$\text{RTS} = [\tilde{\eta}_{\text{TK}} + \tilde{\eta}_{\text{ALH}} \cdot \eta_N^{\text{ALH}} + \eta_N]^{-1}, \tag{17}$$

where η_N^{ALH} is the elasticity of ALH with respect to the network index. This proposition seems to contribute to solve some of the problems exposed here—particularly regarding the unreasonable constraints on flows—as η_N^{ALH} permits a controlled representation of the true effect of the network expansion. However, what about the other aggregates, which have to vary in the same proportion as the network because of the *constant density* condition? One could think that expanding Oum and Zhang's (1997) procedure to those aggregates could be of assistance. Unfortunately, it is not, as we now discuss. Assuming that functions $\tilde{y}_h(N)$ can be estimated, the total derivative could be taken as the network expands, because if the network grows by λ, aggregates would change to $\tilde{y}_h(\lambda N)$. Then, the presumably improved version of RTS would be

$$\overline{\text{RTS}} = \left[\sum_{h \in H \cup K} \tilde{\eta}_h \eta_N^{\tilde{y}_h} + \eta_N \right]^{-1}. \tag{18}$$

In $\overline{\text{RTS}}$, the value of the aggregates follow what is actually happening with the components of $\mathbf{Y}$ as N varies across firms or in time. $\overline{\text{RTS}}$ examines the behavior of cost as the network expands, accounting for the variations in the level of aggregates (products and attributes). Following this method, industry structure conclusions would be related with the estimated value for $\overline{\text{RTS}}$ as follows: (1) if $\overline{\text{RTS}} > 1$, costs increase less than proportionally with network size, and increasing network size would be cost efficient; (2) if $\overline{\text{RTS}} = 1$, costs increase by the same proportion as the network, and increasing network size would be neutral regarding cost; (3) if $\overline{\text{RTS}} < 1$, costs increase more than proportionally with network size, and a network expansion would not be cost efficient. Are these conclusions reasonable? Note that we are comparing network size with total cost. In many cases, it may happen that costs will increase in a larger proportion than the number of points served ($\overline{\text{RTS}} < 1$, do not increase), but the expansion of the network might be cost efficient because the incremental cost (i.e., the cost of adding the vector of new flows to the line of production) is smaller than the cost of serving those new flows with a different firm. For example, when the number of points served increases from 2 to 3, the network expands by 3/2, but four new flows enter the picture, tripling the number of outputs previously produced. Costs will very likely increase by more than 1.5, but the network expansion might well be

---

[11] If firm A's flows are all identical to z, firm L's flows are all identical to z', and the route structure is hub and spoke before and after the network expansion, then the $PS - 1$ conditions that ensure that the original links maintain their load are all as in Equation (16).

efficiently served by one firm only:[12] the "improved" version fails to properly account for its ultimate goal, namely, to provide useful information for adequate conclusions regarding industry structure.

5. Conclusions

In this article we have examined and interpreted at a flow level the conditions imposed by RTS on the aggregate descriptions of transport product. We showed that what appears as a reasonable approach using aggregates, fails to do the intended job when looked at in detail. In short, because its properties are defined on the aggregates, RTS is ambiguous as a tool to contribute to the analysis of the transport industries structure within the context of varying networks.

We believe that what has prevented the detection of these problems has been the failure to think in terms of the true product **Y**, which is why RTS is still used permanently. The inability to estimate cost functions empirically using a precise description of the product should not be an impediment to make correct economic inferences when using cost functions with aggregates, which presumably reproduce industry costs accurately. Otherwise, the meaning of key multioutput concepts such as economies of scale, specific scale, and economies of scope can be lost. Aggregates are necessary, and we are not challenging their use for econometric purposes. What we are defending is a rigorous view of transport production, which means, among other things, that the *interpretation* of cost functions for the purpose of industry structure analysis should be done in terms of the true product. This is feasible and rewarding once it is recognized that aggregates are synthetic representations of what a transport firm produces.

The best example of the advantages to think in terms of **Y** is the examination of the constant density condition. Apparently, the idea was to perform the analysis of a network expansion considering that the new nodes added to the network are, on average, "like the others" (recall Figure 2). By analyzing things in term of the true product, what we have shown here is, precisely, that the implicit constraints imposed are far from reflecting this.

We believe that the right approach to answer what RTS cannot is economies of spatial scope. The real challenge, however, is how to do this from cost functions that include aggregates to describe production (the only feasible empirical approach), an area where

some work has been done recently (Basso and Jara-Díaz 2005).[13] Because RTS estimates have been used extensively for the analysis of network size and shape in the last 20 years, once the new methods to replace *RTS* are discussed and established, policy conclusions derived from *RTS* should be reexamined.

Appendix
Let $PS^A = PS$. Then, $PS^L = PS + 1$ and $\lambda = (PS + 1)/PS$. Then, T^A and T^L can be written as

$$T^A = \sum_{i=1}^{PS} \sum_{j=1}^{PS} y_{ij} \qquad T^L = \sum_{i=1}^{PS+1} \sum_{j=1}^{PS+1} y_{ij} \quad \text{with } j \neq i.$$

As $T^L = \lambda T^A$ and the original flows do not vary, then

$$\sum_{i=1}^{PS+1} \sum_{j=1}^{PS+1} y_{ij} = \sum_{i=1}^{PS} \sum_{j=1}^{PS} y_{ij} + \sum_{j=1}^{PS+1} y_{(PS+1)j} + \sum_{i=1}^{PS+1} y_{i(PS+1)}$$

$$\equiv \left(\frac{PS+1}{PS} \right) \cdot \sum_{i=1}^{PS} \sum_{j=1}^{PS} y_{ij}$$

$$\Rightarrow \sum_{j=1}^{PS+1} y_{(PS+1)j} + \sum_{i=1}^{PS+1} y_{i(PS+1)}$$

$$\equiv \left(\frac{PS+1}{PS} - 1 \right) \sum_{i=1}^{PS} \sum_{j=1}^{PS} y_{ij} \quad \text{with } j \neq i. \quad \text{(A.1)}$$

The $2 \cdot PS$ new flows are added on the left-hand side, while the $PS(PS-1)$ original flows are added on the right-hand side. If y is the average of the original flows and y' is the average of the new ones, then (A.1) can be rewritten as $2PS \cdot y' = y((PS+1)/PS - 1)PS(PS-1)$, from which the ratio y'/y is obtained as $y'/y = 0.5 \cdot (PS-1)/PS$.

References

Antoniou, A. 1991. Economies of scale in the airline industry: The evidence revisited. *Logist. Transportation Rev.* **27**(2) 159–184.

Basso, L. J., S. R. Jara-Díaz. 2005. Calculation of economies of spatial scope from transport cost functions with aggregate output (with an application to the airline industry). *J. Transport Econom. Policy* **39**(1) 25–52.

---

[12] By constructing a transport cost function analytically from the technology, Jara-Díaz and Basso (2003) are actually able to develop this case: the network expands from two to three points served, and minimum cost increases more than four times ($\overline{RTS} < 1$), but the network expansion is cost efficient ($SC > 0$) because of better fleet utilization.

[13] They proposed and applied a method to calculate economies of spatial scope from this type of cost function, showing that RTS indeed failed to correctly indicate the cost efficiency of network expansions. RTS values for three Canadian airlines in 1980 were 0.88, 0.99, and 1.15, while their respective *SC* values were all positive. The *SC* values were also below one, as they should be in theory, which speaks well of the estimated function.

Basso, L. J., S. R. Jara-Díaz. 2006. Distinguishing multiproduct economies of scale from economies of density on a fixed-size transport network. *Networks Spatial Econom.* **6** 149–162.

Baumol, W. J., J. C. Panzar, R. D. Willig. 1982. *Contestable Markets and the Theory of Industry Structure.* Harcourt, Brace and Jovanovich, New York.

Berechman, J. 1993. *Public Transit Economics and Deregulation Policy.* Elsevier Science Publishers, Amsterdam, The Netherlands.

Bhattacharyya, A., S. Kumbhakar, A. Bhattacharyya. 1995. Ownership structure and cost efficiency: A study of publicly owned passenger-bus transportation companies in India. *J. Productivity Anal.* **8**(1) 35–52.

Borenstein, S. 1992. The evolution of U.S. airline competition. *J. Econom. Perspectives* **6**(2) 45–73.

Braeutigam, R. R. 1999. Learning about transport costs. J. Gomez-Ibañez, W. B. Tye, C. Winston, eds. *Essays in Transportation Economics and Policy.* Brooking Institution Press, Washington D.C., 57–97.

Caves, D. W., L. R. Christensen. 1988. The importance of economies of scale, capacity utilisation and density in explaining interindustry productivity growth. *Logist. Transportation Rev.* **24**(1) 3–32.

Caves, D. W., L. R. Christensen, M. W. Tretheway. 1984. Economies of density versus economies of scale: Why trunk and local airline costs differ. *RAND J. Econom.* **15** 471–489.

Caves, D. W., L. R. Christensen, M. W. Tretheway, R. J. Windle. 1985. Network effects and the measurement of returns to scale and density for U.S. railroads. A. Daughety, ed. *Analytical Studies in Transport Economics.* Cambridge University Press, Cambridge, UK, 97–120.

Daughety, A. F. 1985. Transportation research on pricing and regulation: Overview and suggestions for future research. *Transportation Res. Part A* **19**(5/6) 471–487.

De Rus, G., C. Nash. 1997. *Recent Developments in Transport Economics.* Ashgate, Aldershot, UK.

Filippini, M., R. Maggi. 1992. The cost structure of the Swiss private railways. *Internat. J. Transport Econom.* **19** 307–327.

Filippini, M., P. Prioni. 2003. The influence of ownership on the cost of bus service provision in Switzerland—An empirical illustration. *Appl. Econom.* **35** 683–690.

Friedlaender, A. F., E. Berndt, J. Chiang, M. Showalter, C. Velturro. 1993. Rail costs and capital adjustments in a quasi-regulated environment. *J. Transports Econom. Policy* **27** 131–152.

Gagné, R. 1990. On the relevant elasticity estimates for cost structure analysis of the trucking industry. *Rev. Econom. Statist.* **72** 160–164.

Gillen, D., T. H. Oum, M. W. Tretheway. 1990. Airlines cost structure and policy implications. *J. Transports Econom. Policy* **24** 9–34.

Hendricks, K., M. Piccione, G. Tan. 1995. The economics of hubs: The case of monopoly. *Rev. Econom. Stud.* **62** 83–99.

Hurdle, G., R. Johnson, A. Toskow, G. Warden, M. Williams. 1989. Concentration, potential entry and performance in the airline industry. *J. Indust. Econom.* **38**(2) 119–139.

Jara-Díaz, S. R. 1981. Transportation cost functions: A multiproduct approach. Unpublished doctoral dissertation, Massachusetts Institute of Technology, Cambridge, MA.

Jara-Díaz, S. R. 1982a. The estimation of transport cost functions: A methodological review. *Transportation Rev.* **2** 257–278.

Jara-Díaz, S. R. 1982b. Transportation product, transportation function and cost function. *Transportation Sci.* **16** 522–539.

Jara-Díaz, S. R. 2000. Transport production and the analysis of industry structure. J. Polak, A. Heertje, eds. *Analytical Transport Economics: An International Perspective.* Elgar, Cheltenham, UK, 27–50.

Jara-Díaz, S. R., L. J. Basso. 2003. Transport cost functions, network expansion and economies of scope. *Transportation Res. Part E* **39**(4) 271–288.

Jara-Díaz, S. R., C. Cortés. 1996. On the calculation of scale economies from transport cost functions. *J. Transport Econom. Policy* **30** 157–170.

Jara-Díaz, S. R., P. Donoso, J. Araneda. 1991. Best partial flow aggregation in transportation cost functions. *Transportation Res. Part B* **25** 329–339.

Keaton, M. 1990. Economies of density and service levels on U.S. railroads: An experimental analysis. *Logist. Transportation Rev.* **26** 211–227.

Koontz, H. D. 1951. Economic and managerial factors underlying subsidy needs of domestic trunk line air carriers. *J. Air Law Commerce* **18**(2) 127–156.

Kumbhakar, S. C. 1990. A reexamination of returns to scale, density and technical progress in U.S. airlines. *Southern Econom. J.* **57** 428–442.

Liu, Z., E. L. Lynk. 1999. Evidence on market structure of the deregulated U.S. airline industry. *Appl. Econom.* **31** 1083–1092.

Mizutani, F. 2004. Privately owned railways' cost function, organization size and ownership. *J. Regulatory Econom.* **25**(3) 297–322.

Oum, T. H., W. Waters, II. 1996. A survey of recent developments in transportation cost function research. *Logist. Transportation Rev.* **32**(4) 423–463.

Oum, T. H., Y. Zhang. 1997. A note on scale economies in transport. *J. Transport Econom. Policy* **31** 309–315.

Panzar, J. 1989. Technological determinants of firm and industry structure. R. Schmalensee, R. Willig, eds. *Handbook of Industrial Organization.* North-Holland, New York, 3–59.

Pels, E., P. Rietveld. 2000. Cost functions in transport. K. J. Button, D. Hensher, eds. *Handbook of Transport Modelling.* Elsevier, Amsterdam, The Netherlands, 321–333.

Small, K. 1992. *Urban Transportation Economics.* Harwood Academic Publishers, Philadelphia, PA.

Spady, R., A. Friedlaender. 1978. Hedonic cost functions for the regulated trucking industry. *Bell J. Econom.* **9** 159–179.

Windle, R. J. 1988. Transit policy and the cost structure of urban bus transportation. J. S. Dodgson, N. Topham, eds. *Bus Deregulation and Privatization: An International Perspective.* Avebury, Aldershot, UK, 119–140.

Winston, C. 1985. Conceptual developments in the economics of transportation: An interpretive survey. *J. Econom. Literature* **23** 57–94.

Xu, K., R. Windle, C. Grimm, T. Corsi. 1994. Re-evaluating returns to scale in transport. *J. Transport Econom. Policy* **28** 275–286.

Ying, J. 1992. On calculating cost elasticities. *Logist. Transportation Rev.* **28**(3) 231–235.