DETERMINACION DE SESGOS POR AGREGACION ESPACIAL DE FLUJOS EN LA ESTIMACION DE FUNCIONES DE COSTO DE TRANSPORTE.

Sergio R. Jara Díaz y Pedro Donoso Sierra

Departamento de Ingeniería Civil, Universidad de Chile

Casilla 228/3, Santiago, Chile

RESUMEN

La caracterización de la oferta de transporte como actividad económica ha sido hecha, históricamente, a partir de descripciones muy sintéticas del producto. El uso de medidas de volumen por distancia, como las toneladas-milla o los pasajeros-kilómetro, fue unánime hasta fines de la década del 70. En sucesivos trabajos a partir de 1981, Jara-Díaz ha justificado una definición del producto generado por una empresa de transporte como un vector de flujos. En ellos se muestra que el uso de descriptores gruesos como los señalados es fuente de errores importantes. En particular, el trabajo empírico evidencia que la medida escalar volumen por distancia usada en funciones de costo, genera estimadores sesgados tanto de los costos marginales m_i como del grado de economías de escala S. En este trabajo se analiza este tipo de sesgos desde el punto de vista econométrico. Para ello, se plantea modelos lineales con especificaciones correcta y agregada del producto. A pesar de usar la misma información básica, se muestra que un modelo agregado es interpretable como uno con variables omitidas con respecto a uno desagregado, a menos que se verifique proporcionalidad entre costos marginales y distancias. Se logra expresar el sesgo en los estimadores de m_i y S en términos de diferencias de sumas ponderadas que involucran solamente información original (gasto, flujos, distancias). Sobre esta base se muestra que, bajo hipótesis estadísticas razonables, el modelo con descripción sintética del producto provoca una sobreestimación sistemática del grado de economías de escala. El mismo fenómeno es demostrable para una estructura más general de costos marginales.

Acta III Congreso Chileno de Ingeniería de Transporte, 1987

1. INTRODUCCION

Durante mucho tiempo el análisis económico de la actividad de las empresas de transporte descansó en una descripción sintética del producto (pasajeros-kilómetro, toneladas-milla, etc.). A finales de la década del 70 se comenzó a hablar de la necesidad de incorporar nuevos descriptores que permitieran representar mejor la actividad económica analizada, como por ejemplo, largo medio de ruta o tipo de carga. Este enfoque fue presentado como la reunión de una descripción cualitativa del producto y la ya tradicional descripción sintética, y fue popularizado particularmente en el trabajo de Spady y Friedlaender (1978). Sólo a comienzos de los 80 se planteó formalmente la necesidad de reconocer que el producto de una empresa de transporte es un vector de flujos de diversos bienes, entre diversos puntos y durante diversos períodos. Esta visión permitió caracterizar la descripción sintética como un agregado, construído implícitamente a partir de la definición vectorial del producto. EL uso de ese agregado en el contexto de la estimación de funciones de costo de transporte fue criticado desde puntos de vista conceptuales, económicos y empíricos por Jara-Díaz (1982,a,b; 1983; 1987) y Jara-Díaz y Winston (1981); se justifica allí el enfoque vectorial, hoy reconocido como la forma correcta de plantear el proceso productivo en transporte (Keeler, 1983; Harker, 1985; Daughety, 1985; Winston ,1985).

En el escaso trabajo empírico hasta aquí realizado con el nuevo enfoque, se ha puesto énfasis en la descripción espacialmente desagregada del producto. Ello ha sido suficiente para mostrar que las características microeconómicas de los sistemas de transporte son erróneamente inferidas a partir de funciones de costo estimadas con producto agregado. En particular, lo son el costo marginal por unidad física transportada en distintos pares origen-destino, m_i, y el grado de economías de escala, S. La evidencia indica que S es notoriamente sobreestimado cuando es calculado a partir de funciones de costo "monoproductivas". Como plantea Winston (1985), "el enfoque agregado monoproductivo sin duda arrojará un estimador incorrecto del grado de economías de escala. Desafortunadamente, el impacto cualitativo de esta imprecisión no es claro". En este trabajo enfrentaremos precisamente este problema, analizando desde un punto de vista econométrico la naturaleza de los sesgos provocados por una errónea descripción del producto. En particular, mostraremos que la sobreestimación de S es prácticamente una propiedad de las funciones de costo con producto agregado.

En la sección siguiente, se plantea modelos lineales con especificaciones correcta y agregada del producto. A pesar de usar la misma información básica, se muestra que un modelo agregado es interpretable como uno con variables omitidas con respecto a uno desagregado. En la sección 3, se derivan expresiones para el sesgo en los estimadores de m_i y S en términos de la información original (gasto, flujos, distancias). Usando hipótesis estadísticas razonables, se muestra la dirección esperada de los sesgos obtenidos. En la sección 4 se propone una generalización motivada en el análisis previo, basada en una forma relativamente general de los costos marginales. En la última sección se discuten los resultados alcanzados y se motiva el trabajo de investigación en agregación parcial de flujos.

2. PLANTEAMIENTO DEL PROBLEMA

Una función de costo $C(\mathbf{w},\mathbf{y})$ representa, como es sabido, el mínimo gasto necesario para producir $\mathbf{y}$ a precios de factores $\mathbf{w}$. En este caso omitiremos el vector $\mathbf{w}$ en lo que sigue, concentrándonos en la descripción de $\mathbf{y}$. Sin perder estrictez, identificaremos las componentes de $\mathbf{y}$ sólo por un subíndice origen-destino, es decir

$$\mathbf{y} = (y_1,\ldots,y_k) \tag{1}$$

donde y_i es el flujo en el par origen-destino i. Si d_i es la distancia asociada al par i, la descripción sintética del producto corresponde a (Jara-Díaz, 1982a)

$$\tilde{y} = \sum_i y_i d_i . \tag{2}$$

Si bien en la estimación de funciones de costo se trata de postular funciones flexibles lo menos restringidas posible, i.e. translog o cuadráticas, plantearemos aquí modelos lineales en el producto, los que siempre son interpretables como una aproximación de primer orden; dado el tipo de fenómeno a estudiar, es suficiente con que el valor de los costos marginales no están restringidos a priori, ya que son los únicos que intervienen en el cálculo de S. Por otra parte, los modelos flexibles son lineales en los parámetros de forma tal que los desarrollos y el enfoque aquí presentados motivan eventuales extensiones posteriores. Definimos el modelo desagregado M_d como aquel que postula que el gasto C es función de $\mathbf{y}$ definido en (1), y el modelo agregado M_a como el que usa $\tilde{y}$.

Formalmente

$$M_d : C = \beta_0 + \beta_1 y_1 + \ldots + \beta_k y_k + \varepsilon \tag{3}$$

$$M_a : C = b_0 + b_1 \tilde{y} + \mu , \tag{4}$$

en que β_0, β_i, b_0 y b_1 son parámetros a estimar, ε y μ son errores. Se puede observar desde ya que, a partir de la ecuación (2), siempre se podrá escribir

$$y_i = d_i^{-1} (\tilde{y} - \sum_{j\neq i} d_j y_j) \tag{5}$$

y, por lo tanto, el modelo M_d puede ser escrito como

$$C = \beta_0 + \frac{\beta_i}{d_i} \tilde{y} + \sum_{j\neq i} y_j \left(\beta_j - \frac{\beta_i}{d_i} d_j \right) + \varepsilon^{1} \tag{6}$$

1 Cabe observar que la independencia lineal entre las observaciones de los flujos, garantiza la independencia lineal de las observaciones de $\tilde{y}$ con cualquier matriz de observaciones de $\mathbf{y}$ sin una columna (flujo).

de donde es evidente que el modelo M_a corresponde a un caso particular del modelo (correcto) M_d en que se ha omitido k-1 variables, a menos que la razón $\frac{\beta_i}{d_i}$ sea constante para todo i, en cuyo caso las formas funcionales propuestas coinciden. Esta observación es básica para el estudio de los sesgos provocados por el modelo M_a, y es el reflejo formal de la ambigüedad que encierra el "producto" $\tilde{y}$, puesta en evidencia una vez que se reconoce su relación con **y** a través de la ecuación (2). En efecto, esa ecuación indica que un mismo valor de $\tilde{y}$ es generable a partir de muchas distintas combinaciones de $\{y_i\}$; como ha mostrado Jara-Díaz (1982b), es esta la raíz de la ambigüedad en costos, ya que un vector $\{y_i\}$ está asociado sólo a una curva de iso-costo en tanto que un valor $\tilde{y}$ "corta" múltiples niveles de costo. La contrapartida en términos de un modelo econométrico es precisamente la interpretación de M_a como un modelo con variables omitidas.

Los modelos M_a y M_d proporcionan los elementos necesarios para realizar un análisis de sesgos en la estimación de m_i y S. La idea central es comparar estimadores a partir del modelo correcto M_d con los estimadores equivalentes obtenidos de M_a asumiendo que en ambos casos el procedimiento mínimo cuadrático es usado. Definamos

β $= \{\beta_i\}_{i=0,k}$

b $= \{b_i\}_{i=0,k}$

g = vector de observaciones de C

Y = matriz de observaciones de $\{y_i\}$

d $= \{d_i\}$

$\hat{\beta}$ y $\hat{b}$ = estimadores mínimo cuadráticos ordinarios de β y b respectivamente.[2]

Entonces se tendrá que

$$\hat{\beta} = \hat{\beta}\ (g\ ,\ Y) \qquad (7)$$

$$\hat{b} = \hat{b}\ (g\ ,d\ ,Y\ ). \qquad (8)$$

Los estimadores del costo marginal $\frac{\partial C}{\partial y_i}$ son

---

2. El caso de mínimos cuadrados generalizados es tratable en forma simi-al que aquí se expondrá.

$$M_d : \hat{m}_i = \hat{\beta}_i(g, Y) \tag{9}$$

$$M_a : \hat{\hat{m}}_i = d_i \hat{b}_1 (g,d,Y) \tag{10}$$

en tanto que los estimadores del grado de economías de escala se obtienen a partir de (Panzar y Willig, 1977) $C(y)/y \nabla C(y)$, i.e.

$$M_d : \hat{S} = \frac{\hat{\beta}_0(g, Y) + \sum_{i=1}^{k} \hat{\beta}_i(g, Y) y_i}{\sum_{i=1}^{k} y_i \hat{\beta}_i(g, Y)} \tag{11}$$

$$M_a : \hat{\hat{S}} = \frac{\hat{b}_0(g, d, Y) + \hat{b}_1(g, d, Y) \sum_{i=1}^{k} y_i d_i}{\hat{b}_1(g, d, Y) \sum_{i=1}^{k} y_i d_i} . \tag{12}$$

De las ecuaciones (9) y (10) se desprende que el sesgo en costos marginales, $\hat{\hat{m}}_i - \hat{m}_i$ es función de los datos **g**, **d** e **Y**. Análogamente, la diferencia $\hat{\hat{S}} - \hat{S}$ será también expresable a partir de **g**, **d** e **Y**. De esta forma, el análisis de sesgos puede ser realizado analíticamente desde los datos del problema.

3. SESGOS EN COSTOS MARGINALES, COSTO FIJO Y GRADO DE ECONOMIAS DE ESCALA.

Considerando el modelo M_d como correctamente especificado, se formulará expresiones para los sesgos en la estimación de variables microeconómicas relevantes calculadas a partir del modelo M_a. Definiremos $Y_c \in R^{nxk}$: matriz de n observaciones de y centradas en la media muestral, $g_c \in R^n$: vector de n observaciones de g centradas en la media muestral, $y_i^c \in R^n$: vector de n observaciones de y_i centradas en la media muestral (columna i-ésima de Y_c).

Comenzaremos por encontrar una relación entre los estimadores mínimo cuadráticos $\hat{\beta}_1 = (\hat{\beta}_1,\ldots,\hat{\beta}_k)$ y $\hat{b}_1$ de ambos modelos, demostrando la siguiente

Proposición 1. **El estimador $\hat{b}_1$ del modelo M_a es una suma ponderada de la razón entre los estimadores $\hat{\beta}_i$ del modelo M_d y las distancias d_i respectivas, donde los ponderadores sólo dependen de la información sobre flujos y distancias.**

Dem. Los estimadores mínimo cuadráticos de los parámetros de los modelos M_d y M_a se calculan a partir de

$$Y_c' \, Y_c \, \hat{\beta}_1 = Y_c' \, g_c \tag{13}$$

$$d' \, Y_c' \, Y_c \, d \, \hat{b}_1 = d' \, Y_c' \, g_c \tag{14}$$

respectivamente. Premultiplicando la primera ecuación por d', igualando los términos resultantes de la mano izquierda y despejando $\hat{b}_1$ se obtiene

$$\hat{b}_1 = (d' Y_c' \, Y_c \, d)^{-1} \, d' Y_c' \, Y_c \, \hat{\beta}_1 \quad \text{o} \tag{15}$$

$$\hat{b}_1 = \frac{\sum_{i=1}^{k} d_i \left[\sum_{j=1}^{k} y_i^{c'} \, y_j^{c} \, \hat{\beta}_j\right]}{\sum_{i=1}^{k} d_i \left[\sum_{j=1}^{k} d_j \, y_i^{c'} \, y_j^{c}\right]} . \tag{16}$$

Manipulando el numerador de la última ecuación y reconociendo que el denominador es una constante con respecto a aquel, se llega finalmente a

$$\hat{b}_1 = \sum_{j=1}^{k} \frac{\sum_{i=1}^{k} d_i \, d_j \, y_i^{c'} \, y_j^{c}}{\sum_{i=1}^{k} d_i \left[\sum_{j=1}^{k} d_j \, y_i^{c'} \, y_j^{c}\right]} \cdot \frac{\hat{\beta}_j}{d_j} = \sum_{j=1}^{k} \theta_j \frac{\hat{\beta}_j}{d_j} \tag{17}$$

con $\theta_j(d, Y)$ y $\sum_{j=1}^{k} \theta_j = 1$, lo que demuestra la proposición. Como corolario, se observa de inmediato que

$$E(\hat{b}_1) = \sum_{j=1}^{k} \theta_j \frac{\beta_j}{d_j} \tag{18}$$

en que β_j es el valor verdadero del costo marginal j-ésimo definido en la ecuación (3). Ahora es posible demostrar la

Proposición 2. Los sesgos en la estimación de costos marginales y costo fijo, y el error relativo en la estimación del grado de economías de escala a partir del modelo M_a evaluado en la media muestral, están dados respectivamente por

$$d_i \, E(\hat{b}_1) - \beta_i = d_i \left(\sum_{j=1}^{k} \theta_j \frac{\beta_j}{d_j} - \frac{\beta_i}{d_i} \right) \tag{19}$$

$$E(\hat{b}_0) - \beta_0 = \sum_{i=1}^{k} \bar{y}_i \left(\beta_i - d_i \sum_{j=1}^{k} \theta_j \frac{\beta_j}{d_j}\right) = \bar{\bar{y}} \left(\sum_{j=1}^{k} \lambda_j r_j - \sum_{j=1}^{k} \theta_j r_j\right) \tag{20}$$

$$\frac{\hat{\hat{S}} - \hat{S}}{\hat{S}} = \frac{1}{\bar{\bar{y}}\, \hat{b}_1} (\hat{b}_0 - \hat{\beta}_0) = \frac{\sum_{j=1}^{k} \lambda_j \hat{r}_j - \sum_{j=1}^{k} \theta_j \hat{r}_j}{\sum_{j=1}^{k} \theta_j \hat{r}_j} \tag{21}$$

donde

$$r_j = \frac{\beta_j}{d_j}, \quad \hat{r}_j = \frac{\hat{\beta}_j}{d_j}, \quad \tilde{y}_i = y_i d_i \quad y$$

$$\lambda_j = \bar{\bar{y}}_j \Big/ \sum_{i=1}^{k} \bar{\bar{y}}_i \qquad \left(\sum_j \lambda_j = 1\right). \tag{22}$$

La propiedad (19) se deduce directamente de las ecuaciones (9), (10) y (18). Para demostrar la propiedad (20), basta recordar que toda estimación mínimo cuadrática pasa por la media de las observaciones, es decir

$$\bar{g} = \hat{b}_0 + \hat{b}_1 \bar{\bar{y}} = \hat{\beta}_0 + \bar{y}' \hat{\beta}_1 \quad ; \tag{23}$$

a partir de esta observación y de la propiedad (19) se demuestra la primera igualdad en (20), en tanto que la segunda igualdad puede comprobarse reemplazando simplemente las expresiones definidas en (22).

$$\hat{\hat{S}} = \bar{g} \,/\, \bar{\bar{y}}\, \hat{b}_1 \quad y \quad \hat{S} = \bar{g} \,/\, \bar{y}' \hat{\beta}_1 \tag{24}$$

a partir de las ecs. (11), (12) y (23) evaluadas en la media muestral. Luego, haciendo la diferencia directa y aplicando la propiedad (20) es fácil demostrar que

$$\frac{\hat{\hat{S}} - \hat{S}}{\hat{S}} = \frac{\bar{y}' \hat{\beta}_1 - \bar{\bar{y}} \hat{b}_1}{\bar{\bar{y}} \hat{b}_1} = \frac{1}{\sum_{j=1}^{k} \theta_j \hat{r}_j} \left(\sum_{j=1}^{k} \lambda_j \hat{r}_j - \sum_{j=1}^{k} \theta_j \hat{r}_j \right) . \tag{25}$$

Es a partir de la proposición 2 que analizaremos el problema de los sesgos, transformado ahora en un análisis de diferencias de promedios pesados. Aunque la utilidad de esta formulación se detectará plenamente en la sección 4, la usaremos también aquí por lo sintético de las expresiones involucradas. En la proposición 1 se muestra que la relación entre el estimador del pseudo costo marginal por volumen-distancia, b_1, y los estimadores de los verdaderos costos marginales, depende de unos θ_j directamente relacionados con los elementos de la matriz de varianzas y covarianzas de los flujos observados. Es posible obtener alguna intuición acerca de los sesgos explorando los efectos que emergen al asumir algunas formas particulares de esta matriz. Una primera idea es motivada por procedimientos sugeridos en la literatura, cual es el de agregar flujos correlacionados (Jara-Díaz, 1983); la ausencia de correlación desfavorece la agregación. Luego, el caso de correlaciones (o covarianzas) nulas entre flujos es atractivo como caso "normal" en que se debe usar flujos espacialmente desagregados. En esta situación, los pesos θ_j y los estimadores β_j toman la forma

$$\theta_j = d_j^2 \, \mathrm{Var}(y_j) \; / \sum_{j=1}^{k} d_j^2 \, \mathrm{Var}(y_j) \tag{26}$$

$$\hat{\beta}_j = \mathrm{Cov}(y_j , g) \; / \; \mathrm{Var}(y_j) = \rho_{jg} \, \sigma_g \, / \, \sigma_j , \tag{27}$$

donde ρ y σ denotan coeficiente de correlación y desviación standard respectivamente.

Conviene introducir aquí el coeficiente de variación v_j de cada flujo j observado, definido como la razón $\sigma_j/\bar{y}_j$, que refleja la dispersión relativa a la media de sus observaciones. Al hacerlo se obtiene:

$$\theta_j = \frac{(\bar{\bar{y}}_j \, v_j)^2}{\sum_{j=1}^{k} (\bar{\bar{y}}_j \, v_j)^2} \tag{28}$$

$$\hat{\beta}_j = \rho_{jg} \frac{\sigma_g}{v_j \, \bar{y}_j} \tag{29}$$

El error relativo en el grado de economías de escala (ERG) se obtiene reemplazando las ecuaciones (28) y (29) en la (21), simplificando hasta obtener:

$$ERG = \frac{\sigma_g}{\bar{\bar{y}} \sum_{j=1}^{k} \theta_j \hat{r}_j} \left[\sum_{j=1}^{k} \rho_{jg} / v_j - \bar{\bar{y}} \sum_{j=1}^{k} \frac{\bar{\bar{y}}_j v_j}{\sum_{i=1}^{k} \bar{\bar{y}}_i^2 v_i^2} \rho_{jg} \right] \quad (30)$$

Hasta ahora sólo se ha asumido covarianzas nulas entre flujos, fenómeno que involucra solamente a las observaciones en y. Es también razonable incorporar un fenómeno de normal ocurrencia en el mismo conjunto, cual es el de observaciones similarmente dispersas en torno a la media de cada flujo; en el extremo ésto equivale a imponer $v_j = v \ \forall j$. Por último, y con el fin de generar una expresión de signo visualizable para ERG en condiciones simplificadas que reflejen propiedades razonables de la información desagregada, asumiremos que no existen flujos que expliquen la variación del gasto más que otros. Cabe señalar que si algún flujo y_i lo hiciese (estadísticamente), el modelo agregado no distorsionaría demasiado los resultados del desagregado, ya que el fenómeno $C = f(y_i)$ quedaría captado en forma similar por M_a o M_d. Extremando este razonable supuesto, asumiremos formalmente $\rho_{jg} = \rho \ \forall_j$. Reemplazando ambos supuestos en la expresión (30) se obtiene

$$ERG = \frac{\rho \ \sigma_g}{\bar{\bar{y}} \sum_{j=1}^{k} \theta_j \hat{r}_j} \left[\frac{k \sum_{i=1}^{k} \bar{\bar{y}}_i^2 - \left(\sum_{i=1}^{k} \bar{\bar{y}}_i \right)^2}{\sum \bar{\bar{y}}_i^2} \right] > 0 \quad (31)$$

en virtud de la desigualdad de Cauchy - Schwarz. Así, bajo razonables hipótesis estadísticas simplificatorias en los datos desagregados **no** agregables, el modelo agregado sobreestima el grado de economías de escala.

También es posible analizar el error en cada costo marginal i-ésimo (Em_i), cuando las correlaciones entre flujos son nulas, $\rho_{jg} = \rho$ y $v_j = v \ \forall j$. Bajo estas condiciones, si se reemplaza las ecuaciones (28) y (29) en la (19) y se simplifica la expresión resultante se obtiene:

$$Em_i = \frac{\rho \ \sigma_g \ d_i \bar{\bar{y}}}{v \ \bar{\bar{y}}_i^2} (\theta_i - \lambda_i) \quad , \quad (32)$$

a partir de lo cual se puede demostrar que

$$Em_i < 0 \iff \sum_{j\neq i} \tilde{\bar{y}}_j (\bar{y}_i - \bar{y}_j) < 0 \ . \tag{33}$$

A pesar de las hipótesis simplificatorias hechas, este resultado sugiere que los costos marginales asociados a pequeños valores de flujo por distancia tenderán a ser subestimados a partir del modelo agregado.

El análisis de sesgos en condiciones más generales no puede limitarse al estudio de propiedades relativas de los pesos θ_i con respecto a los λ_j sino debe también extenderse a las propiedades de la serie ponderada $\hat{r}_j$. Este aspecto queda ilustrado al observar que si los $\hat{r}_i = \hat{\beta}_i/d_i$ son constantes $\forall i$, los sesgos en las expresiones (19), (20) y (21) son nulos[3]. Es decir, el signo del error en la estimación de m_i o S a partir de M_a depende también de la relación entre costos marginales y distancias. Esto motiva una generalización del problema hacia los factores que determinan técnicamente los valores de los costos marginales.

4. GENERALIZACION DEL PROBLEMA: ROL DE LOS COSTOS MARGINALES.

A partir de los desarrollos en la sección anterior, es evidente que el análisis de sesgos está fuertemente influido por las propiedades de los θ_i introducidos en la proposición 1. Por otra parte, se ha mostrado que la relación verdadera entre los costos marginales y los factores del problema es también fundamental. Esto sugiere entrar a analizar casos que comporten una estructura de los costos marginales ligada tanto a los θ_i como al vector de distancias. El teorema que se presenta en esta sección juega precisamente ese rol, postulando una forma de costos marginales que contiene casos particulares con interesantes (y frecuentes) propiedades.

Teorema: Si $\exists f : R \to R_+$ creciente tal que

$$\beta_i = d_i \, f\,(\tilde{\bar{y}}_i/\alpha_i) \quad i=1,\ldots k \tag{34}$$

donde $\alpha_i = d_i \sum_j d_j \, y_i^{c'} \, y_j^c$, entonces la esperanza del signo de ERG es positiva.

Demostración:

En virtud de la ecuación (25), es necesario y suficiente establecer que

$$\sum_j \lambda_j \hat{r}_j > \sum_j \theta_j \hat{r}_j . \tag{35}$$

---

3. Lo que no hace sino confirmar la observación en la sección 2: M_a omite variables sólo si los costos marginales no son proporcionales a la distancia asociada al flujo respectivo.

Para ello es útil aplicar el Lema de ponderadores propuesto y demostrado en el apéndice. Aplicándolo directamente considerando que $\theta_j = \alpha_j / \sum_j \alpha_j$ (ecuación 17) y la definición (22), se obtiene

$$\sum_j \lambda_j \hat{r}_j - \sum_j \theta_j \hat{r}_j = \frac{1}{\bar{\bar{y}} \sum_j \alpha_j} \sum_i \sum_{j<i} \bar{\bar{y}}_i \bar{\bar{y}}_j (\hat{r}_i - \hat{r}_j) \left(\frac{\alpha_j}{\bar{\bar{y}}_j} - \frac{\alpha_i}{\bar{\bar{y}}_i} \right) \qquad (36)$$

Dado que $E(\hat{r}_i) = \beta_i / d_i = f(\bar{\bar{y}}_i / \alpha_i)$ y f es creciente, la esperanza del producto de las expresiones en paréntesis es siempre positivo, lo que prueba el Teorema.

Corresponde analizar el contenido de una estructura de costos marginales como la propuesta. En primer lugar debe notarse que, reemplazando los α_j y manipulando la expresión resultante, se obtiene

$$\beta_i = d_i \ f \ \{ [\ v_i (v_i \ \bar{\bar{y}}_i + \sum_{j \neq i} v_j \ \bar{\bar{y}}_j \ \rho_{ij}) \ ]^{-1} \} \quad , \qquad (37)$$

donde ρ_{ij} es el coeficiente de correlación entre las observaciones de los flujos i y j.

Si se considera la menos restrictiva de las condiciones analizadas en la sección anterior, cual es la de similares coeficientes de variación v_i, β_i parece decrecer con la magnitud del flujo asociado y con la de aquellos positivamente correlacionados con él[4]. En rigor, la influencia de otros flujos sobre un costo marginal no es compatible con un modelo como M_d que, supuesto correcto, no permite analizar complementaridad de costos; sólo el caso $\rho_{ij} = 0$ lo es, ya que ésto reduce la expresión anterior a

$$\beta_i = d_i \ f \ \left(\frac{1}{v_i^2 \ \bar{\bar{y}}_i} \right) \qquad (38)$$

Si, en particular, f es una función afín de la forma $m_0 + mx$ con v_i constante, el costo marginal es de la forma de $m_0 \cdot d_i + K/\bar{y}_i$, con una componente creciente con la distancia y otra decreciente con el flujo. Tal función es compatible con el análisis tecnológico de sistemas elementales de transporte descrito en Jara-Díaz (1982b). Es muy sencillo demostrar que, en este caso, la propiedad (32) relativa al sesgo en costos marginales es igualmente válida.

---

4. El efecto estricto de $\bar{y}_i$ sobre β_i debe ser analizado considerando que ρ_{ij} e $\bar{y}_i$ están relacionados. Puede demostrarse que $(\partial\beta_i / \partial\bar{y}_i) < 0$ para ρ_{ij} pequeños.

Es ilustrativo discutir, por último, el caso $p_{ij} = 0$ descrito a partir de la ecuación (37). Si fuese efectivamente conveniente la producción conjunta de los flujos y_i e y_j, ceteris paribus las observaciones de lo que una una empresa de transporte ha hecho tenderán a mostrar correlación positiva entre ellos. La captura de tal fenómeno en un modelo lineal no puede sino ser a través de costos marginales menores que si esa relación de flujos complementarios no existiese. Lo interesante de ésto radica en que ello contribuye a provocar una sobreestimación del grado de economías de escala en el modelo agregado.

5. SINTESIS Y CONCLUSIONES

En este trabajo se ha presentado y aplicado un enfoque para analizar los sesgos provocados por la agregación de flujos en funciones de costo de transporte. A partir de dos modelos, agregado y desagregado, se estableció la interpretación del primero como uno de variables omitidas con respecto al segundo, a pesar de incorporar el mismo tipo de variables. Los estimadores mínimo cuadráticos ordinarios de los parámetros de ambos modelos sirvieron de base para mostrar los sesgos en costos marginales y grado de economías de escala en forma interpretable como una relación entre sumas ponderadas. Planteando hipótesis estadísticas razonables, se mostró que el modelo con descripción sintética del producto provoca una sobreestimación sistemática del grado de economías de escala. Se demostró el mismo fenómeno para una estructura de costos marginales más general compatible con las hipótesis propuestas; tal estructura contiene casos particulares interpretables económicamente en forma relativamente sencilla.

De esta manera, se ha revelado la raíz econométrica de un error de tratamiento al producto de transporte que ya ha sido mostrado en sus dimensiones económica y técnica. La sobreestimación de S está prácticamente garantizada en el tratamiento agregado del producto. Dado el impacto normativo que ello tiene en términos de organización industrial en el Sector Transporte, y dada la imposibilidad de realizar análisis con desagregación total en la mayoría de los casos, el problema central pasa a una etapa siguiente, cual es la de buscar formas óptimas de agregación parcial en el sentido de distorsionar mínimamente la información económica a ser rescatada de observaciones recogidas. El enfoque aquí presentado será usado precisamente en esta dirección de investigación.

AGRADECIMIENTOS

. Esta investigación ha sido parcialmente financiada a través del proyecto Nº 327 (1987) de FONDECYT.

REFERENCIAS

DAUGHETY, A. (1985). **Analytical Studies in Transport Economics** Cambridge: Cambridge University Press.

HARKER, P. (1985). The State of the Art in the Predictive Analysis of Freight Transport Systems, **Transport Reviews** 5, pp. 143-164.

JARA-DIAZ, S. (1982a) The Estimation of Transport Cost Functions: A Methodological Review, **Transport Reviews** 2, pp. 257-278.

JARA-DIAZ, S. (1982b). Transportation Product, Transportation Function and Cost Functions, **Transportation Science** 16, pp. 522-539.

JARA-DIAZ, S. (1983) Freight Transportation Multioutput Analysis, **Transportation Research** 17A, pp. 429-483.

JARA-DIAZ, S. (1987) Multioutput Analysis of Trucking Operations Using Spatially Desaggregated Flows. Por aparecer en **Transportation Research.**

JARA-DIAZ, S. y WINSTON, C. (1981) Multiproduct Transportation Cost Functions: Scale and Scope In Railroad Operations, Proceedings of the Eighth Conference of the European Association for Research in Industrial Economies, September.

KEELER, T. (1983) **Railroads, Freight and Public Policy**, Washington, D.C. The Brookings Institution.

PANZAR, J. y WILLIG, R. (1977) Economies of Scale in Multi-Output Production, **Quaterly Journal of Economics** 9, pp. 159-179.

SPADY, R. y FRIEDLAENDER, A. (1978) Hedonic Cost Functions for the Regulated Trucking Industry, **Bell Journal of Economics** 9, pp. 159-179.

WINSTON, C. (1985) Conceptual Developments in the Economics of Transportation: An Interpretative Survey, **Journal of Economic Literature** XXIII, pp. 57-94.

APENDICE

Lema

Sean $(t_i)_{i=1,k}$ $(\delta_i)_{i=1,k}$ y $(\gamma_i)_{i=1,k}$ reales tales que: $\sum_i \delta_i \neq 0$ $\sum_i \gamma_i \neq 0$, entonces:

$$\sum_{i=1}^{k} \left[\frac{\delta_i}{\sum_{j=1}^{k} \delta_j}\right] t_i - \sum_{i=1}^{k} \left[\frac{\gamma_i}{\sum_{j=1}^{k} \gamma_j}\right] t_i = \frac{1}{\sum_{j=1}^{k} \delta_j \sum_{j=1}^{k} \gamma_j} \cdot$$

$$\cdot \sum_{i=1}^{k} \sum_{j<i} \delta_i \delta_j (t_i - t_j) \left(\frac{\gamma_j}{\delta_j} - \frac{\gamma_i}{\delta_i}\right)$$

Demostración

$$\sum_{i=1}^{k} \left[\frac{\delta_i}{\sum_{j=1}^{k} \delta_j}\right] t_i - \sum_{i=1}^{k} \left[\frac{\gamma_i}{\sum_{j=1}^{k} \gamma_j}\right] t_i = \frac{1}{\sum_{j=1}^{k} \delta_j \sum_{j=1}^{k} \gamma_j} \sum_{i=1}^{k} \sum_{j=1}^{k} (\delta_i \gamma_j - \delta_j \gamma_i) t_i$$

$$= \frac{1}{\sum_{j=1}^{k} \delta_j \sum_{j=1}^{k} \gamma_j} \left[\sum_{i=1}^{k} \sum_{j<i} (\delta_i \gamma_j - \delta_j \gamma_i) t_i + \right.$$

$$\left. + \sum_{i=1}^{k} \sum_{i<j} (\delta_i \gamma_j - \delta_j \gamma_i) t_i \right] \qquad (1)$$

Pero:

$$\sum_{i=1}^{k} \sum_{i<j} (\delta_i \gamma_j - \delta_j \gamma_i) t_i = \sum_{j=1}^{k} \sum_{i<j} (\delta_i \gamma_j - \delta_j \gamma_i) t_i = \sum_{i=1}^{k} \sum_{j<i} (\delta_j \gamma_i - \delta_i \gamma_j) t_j$$

Al reemplazar esta expresión en (1) se obtiene:

$$\sum_{i=1}^{k}\left[\frac{\delta_i}{\sum_{j=1}^{k}\delta_j}\right]t_i - \sum_{i=1}^{k}\left[\frac{\gamma_i}{\sum_{j=1}^{k}\gamma_j}\right]t_i =$$

$$= \frac{1}{\sum_{j=1}^{k}\delta_j \sum_{j=1}^{k}\gamma_j}\sum_{i=1}^{k}\sum_{j<i}(\delta_i\gamma_j - \delta_j\gamma_i)(t_i - t_j)$$

$$= \frac{1}{\sum_{j=1}^{k}\delta_j \sum_{j=1}^{k}\gamma_j}\sum_{i=1}^{k}\sum_{j<i}\delta_i\delta_j(t_i - t_j)\left(\frac{\gamma_j}{\delta_j} - \frac{\gamma_i}{\delta_i}\right)$$

Q.E.D.