

## MODELOS PD Y MIXTOS PARA VIAJES INTERURBANOS EN TREN Y BUS

Andrés Iacobelli y Juan de Dios Ortúzar  
Departamento de Ingeniería de Transporte  
Pontificia Universidad Católica de Chile  
Casilla 306, Cód. 105, Santiago 22  
Tel: 552 2375, Anexo 4822; Fax: 552 4054

### RESUMEN

Se diseñó un experimento de preferencias declaradas para viajes en bus y tren entre Santiago, Chillán y Concepción. El experimento, del tipo elección, se programó en computadores portátiles con ayuda de un software especializado. Además se consultó sobre las características del viaje e información socioeconómica de los individuos. Se consideraron cuatro atributos (y sus respectivos niveles de variación): tiempo de viaje (tres niveles), tarifa (tres niveles), comodidad (dos niveles) y confiabilidad o atraso del servicio (dos niveles).

Para definir el conjunto de opciones a presentar a los entrevistados, se usó un diseño factorial fraccional que dio como resultado nueve opciones. Se agregó a éstas una tarjeta con las características actuales del viaje, de forma de medir consistencia en las respuestas. Se tomaron 112 encuestas, en las cuáles se detectó 14 personas que elegían siempre el tren, independiente de las variaciones en los atributos. En todos estos casos se arguyeron razones de seguridad y comodidad. Dado lo particular de esas respuestas se trataron en forma separada. Con los datos de las restantes 98 encuestas, se dispone de 10 elecciones para cada individuo, lo que se traduce en una muestra de 980 individuos para la etapa de calibración de la muestra.

Conocidas las imperfecciones que pueden resultar de los modelos puros de PD, se procedió a agregar a la muestra datos de PR. Estos provienen de dos fuentes: aquellos recolectados durante el experimento de PD (98 individuos) y otros provenientes de un estudio de viajes interurbanos efectuado por la Universidad de Chile (407 individuos). Se calibraron distintas especificaciones de modelos con datos mixtos, incluyendo estructuras jerárquicas complejas y novedosas. Los resultados se comparan con otros estudios realizados en el país.

## **1. INTRODUCCION**

La estimación de modelos de elección discreta se basa generalmente en una única fuente de datos. Lo más común es que sean datos de "preferencias reveladas" (PR), esto es, observaciones sobre lo que los usuarios efectivamente eligen. Estas elecciones reportadas se relacionan con las restantes alternativas disponibles de cada usuario a través del nivel de servicio de cada una y las características socioeconómicas del viajero. Además, es necesario recurrir a datos provenientes de análisis de redes y algunas veces a información socioeconómica más agregada, para poder estimar los modelos y luego ocuparlos para predecir. Las principales desventajas de esta técnica son:

- i) Relativamente alto costo debido a la gran cantidad de encuestas que se debe efectuar;
- ii) Muchas veces no se dispone de suficiente información en el conjunto de elecciones, debido a problemas de correlación entre los atributos y/o poca variabilidad de los mismos. Esto se traduce en dificultades para identificar cómo el usuario transa entre los distintos atributos al hacer su elección.

Un segundo tipo de datos son los correspondientes a información de "preferencias declaradas" (PD). Esta técnica se basa en respuestas a situaciones hipotéticas de viaje y se está usando cada vez más en ingeniería de transporte, particularmente en aquellos casos en que alguna de las alternativas no existe o en que ciertas características de ellas son muy difíciles de medir. Si bien el enfoque corrige algunas deficiencias presentadas por las PR, existen potenciales sesgos en las respuestas de los entrevistados, por lo que las elecciones eventuales podrían no ser totalmente consistentes con lo declarado.

La modelación con datos mixtos de PR y PD aparece actualmente como una tercera alternativa (ver Ben-Akiva y Morikawa, 1990; Bradley y Daly, 1991), que permite combinar las características más valiosas de ambos tipos de datos. El objetivo de este trabajo es precisamente experimentar con distintas formas de efectuar esta tarea.

El trabajo comienza con un resumen de la base teórica que sustenta la posterior modelación. Le sigue un análisis de los datos ocupados y las modificaciones que debieron efectuarse previo a su utilización; finalmente se presentan los resultados de la modelación y las principales conclusiones que se rescatan de la experiencia.

## **2. ASPECTOS ECONOMETRICOS**

### **2.1 Residuos Estocásticos en PR y PD**

El problema fundamental que se genera al intentar combinar datos de PR y PD, es la distinta naturaleza de los errores que tienen asociados. En general los errores en PR están relacionados con los atributos (variable independiente), en cambio en PD los errores están asociados a la elección (variable dependiente). Así, con datos de PR es razonable asumir que los errores de predicción sean los mismos que los errores de estimación; sin embargo, en PD



se modela la conducta de *pseudo* individuos (generados a partir de las distintas elecciones propuestas a cada entrevistado), y los errores involucrados no tienen porque ser iguales que al intentar predecir el comportamiento individual (ver Ortúzar y Garrido, 1993). Si suponemos que la información de PD puede tener errores asociados a la variable dependiente (elección), los resultados de tales experimentos reflejarían una *pseudo* utilidad  $W_{ik}$ , distinta a la utilidad tradicional  $U_{ik}$ , tal que:

$$U_{ik} = W_{ik} + \eta_{ik} \quad (1)$$

donde  $\eta_{ik}$  representa el error asociado a la variable dependiente. Por otro lado, de la forma tradicional postulada en modelos de utilidad aleatoria se puede obtener lo siguiente:

$$U_{ik} = V_{ik} + \varepsilon_{ik} = W_{ik} + \eta_{ik}$$

luego:

$$W_{ik} = V_{ik} + (\varepsilon_{ik} - \eta_{ik}) \quad (2)$$

A partir de esta última expresión se podría intentar considerar a ambos errores como uno y utilizar directamente en la estimación las técnicas econométricas que normalmente se usan con datos de PR. Sin embargo, dada la existencia del parámetro de dispersión  $\lambda$  en modelos logit (ver por ejemplo, Ortúzar y Willumsen, 1990), que depende de la varianza del error, es incorrecto utilizar los datos de PD de la misma manera que los de PR. Wardman (1991) señala que este tipo de errores produce diferencias de escala en los parámetros que deben ser corregidas explícitamente en la modelación.

Ben-Akiva y Morikawa (1990) desarrollaron un marco teórico adecuado para el caso de datos provenientes de distintas fuentes; en éste los datos se dividen en primarios y secundarios. Los primarios proveen información directa acerca de los principales parámetros a modelar; los secundarios proveen de información adicional (indirecta) sobre estos parámetros. Es importante notar que este marco no es exclusivo para desarrollar modelos conjuntos con datos de PD y PR, sino que se puede aplicar a otros casos que se enmarquen en su definición, como por ejemplo, datos agregados y desagregados.

En el contexto de este trabajo los datos de PR constituyen el conjunto primario, ya que captan el comportamiento real de los individuos, y los datos de PD conforman el segundo grupo. Se postula que la diferencia entre los errores en PR y PD puede ser representada como una función de las varianzas de los errores de cada uno de los datos ( $\varepsilon$  y  $\eta$ ):

$$\sigma_{\eta}^2 = \mu^2 \sigma_{\varepsilon}^2 \quad (3)$$

donde  $\mu$  es un factor de escala. A partir de ello se puede obtener las siguientes funciones de utilidad para una alternativa  $A_i$  cualquiera:

$$U_i^{PR} = \theta X_i^{PR} + \alpha Y_i^{PR} + \varepsilon_i \quad (4)$$

$$\mu U_i^{PD} = \mu (\theta X_i^{PD} + \beta Z_i^{PD} + \eta_i) \quad (5)$$

donde  $\alpha$ ,  $\beta$  y  $\theta$  son parámetros (vectores) a ser estimados;  $X^{ra}$  y  $X^{rd}$  son los vectores de los atributos comunes a los dos tipos de datos;  $Y^{ra}$  y  $Z^{rd}$  son los vectores de atributos que sólo pertenecen a los datos de PR o PD respectivamente.

Si se asume que los errores estocásticos  $\varepsilon$  y  $\eta$  distribuyen Gumbel con media cero y distinta varianza, las probabilidades de elección quedan definidas en base a las funciones de utilidad indirecta en una estructura tipo Logit. La maximización de la función de verosimilitud conjunta asociada es un problema no lineal dado que  $\mu$  multiplica al resto de los parámetros a estimar. Para resolverlo se han empleado variadas técnicas (Bradley y Daly, 1991; Ben-Akiva y Morikawa, 1990) con buenos resultados. A continuación se describirá el método de estimación simultánea, desarrollado por Bradley y Daly (1991), que fue el usado en esta investigación dada la experiencia de Ortúzar y Garrido (1993).

Antes de explicar el método, parece interesante hacer algunos comentarios respecto a las suposiciones que están involucradas en su desarrollo:

- i) Para hacer uso de datos mixtos es necesario, en primer lugar, suponer que la utilidad marginal de cada variable común es igual en ambos contextos.
- ii) Parece razonable suponer que en datos de PR los errores sean independientes, ya que cada dato proviene de una persona distinta. Sin embargo, una de las principales ventajas de PD es que pueden conseguirse varias "observaciones" de un sólo individuo. Luego, la hipótesis de independencia puede ser una mala aproximación en ciertos casos. Dado que no existe una teoría que considere satisfactoriamente este aspecto, y que los estudios previos sugieren que el efecto práctico de esta suposición es limitado, se mantendrá la hipótesis de independencia para los datos de PD (ya sea en modelación mixta como pura).
- iii) La forma funcional de las distribuciones de los errores será en definitiva lo que determine la forma funcional de los modelos de probabilidad utilizados para estimar los coeficientes  $\alpha$ ,  $\beta$  y  $\theta$ . A este respecto, Morikawa *et al.* (1990) asumen que los errores  $\varepsilon$  y  $\eta$  distribuyen Normal, lo que conduce a modelos de probabilidad Probit. Si bien esa suposición facilita el estudio analítico, trae muchas complicaciones para el trabajo empírico. Por otro lado, la suposición de que los errores distribuyen Gumbel conduce a modelos de probabilidad tipo Logit. Ahora bien, desde el punto de vista teórico, ambos tipos de modelos son razonables (Bradley y Daly, 1991), aunque cada uno tiene ventajas e imperfecciones. Debido a que los resultados con ambos no presentan diferencias significativas y a la mayor simplicidad práctica que ofrecen los modelos Logit, se ha optado por estos últimos en el contexto de este trabajo.

## 2.2 El Método de Estimación Simultánea

A partir de las ecuaciones (4) y (5) se puede observar que el factor de escala  $\mu$  es el nexo esencial y único en la estimación de los coeficientes  $\alpha$ ,  $\beta$  y  $\theta$ . Además, es importante señalar que la aparición del producto  $\mu\theta$  es un elemento clave, ya que transforma la estimación con datos mixtos en un problema no lineal. Éste puede resolverse usando ALOGIT 3.2 (Daly, 1992) con una estructura de árbol artificial (jerárquica).

La Figura N°1 ilustra la estructura que consiste básicamente en separar los tipos de datos disponibles, descolgando directamente de la raíz las alternativas de PR, y cada una de las alternativas de PD en un nido particular. Las alternativas de PR no están disponibles para las de PD y similarmente las segundas no están disponibles para las primeras. Es por esta razón (que no está todo el conjunto disponible) que el valor de  $\mu$  no tiene la usual restricción de consistencia de estar entre cero y uno (ver Ortúzar, 1982). Sin embargo,  $\mu$  debe a tomar el mismo valor para cada alternativa de PD (para que sea efectivamente un factor de escala). Si  $\mu$  es mayor que uno esto sólo constituye un reflejo de que los datos de PD tienen menos "ruido" que los de PR; y si es menor que uno, es una indicación de lo contrario.

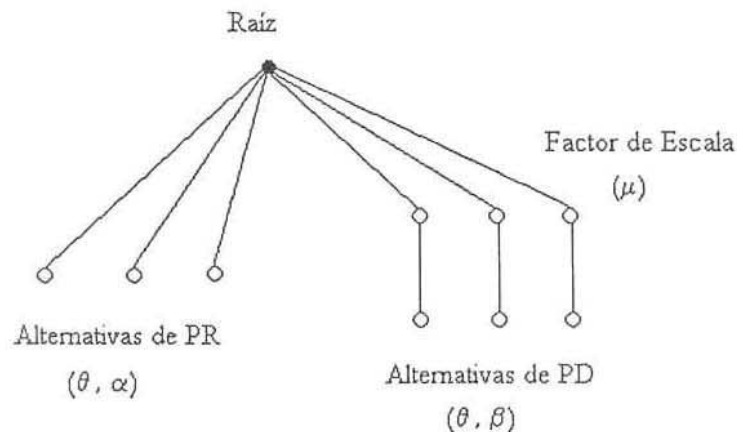


Figura N°1: Estructura Artificial para Estimación con Datos Mixtos

## 3. LOS DATOS UTILIZADOS

Sin duda que un aspecto fundamental para realizar una buena modelación es la calidad de los datos con que se cuenta. Dada la finalidad científica y experimental (no predictiva) de los modelos que se presentan al final de este trabajo, no se procedió en forma absolutamente rigurosa en la toma de datos. Nos referimos específicamente a la cantidad y total aleatoriedad de las encuestas, que no se pudo asegurar debido a restricciones presupuestarias y de tiempo. Sin embargo, esto no invalida la confiabilidad de los datos obtenidos que, como se detalla a continuación, fueron tomados con extremo cuidado y seriedad, destacando la participación directa del primer autor en el caso de las encuestas de PD.



### 3.1. Los Datos de PD

Para conseguir esta información se preparó especialmente un experimento de PD que se implementó en computadores portátiles. La idea era entrevistar a los usuarios en el mismo medio en que viajaban, de forma de asegurar que conocían el tipo de viaje y se habían preguntado alguna vez qué modo elegir para hacer su viaje al sur. Con Santiago como origen, se escogieron los destinos Chillán (a 400 Km.) y Concepción (a 520 Km.), dado que cumplían con el requisito de mediana distancia y tienen un volumen de viajeros, modos y frecuencias que los hacen un destino adecuado.

Luego de realizar un estudio de la oferta disponible y con la experiencia de una encuesta piloto, se definieron los niveles y atributos que se detallan en la Tabla N°1 (ver Iacobelli, 1994).

Tabla N°1: Valores de atributos y niveles

Nivel	Atributo	Concepción		Chillán	
		Bus	Tren	Bus	Tren
Nivel 0	Tiempo de Viaje (horas)	7	9	5	5 1/2
	Tarifa (\$)	P	1,3*P	0,7*P	1,3*P
	Comodidad	Pullman	Económica	Pullman	Económica
	Atraso (min)	20	20	20	20
Nivel 1	Tiempo de Viaje (horas)	8	9	6	5 1/2
	Tarifa (\$)	P	P	0,7*P	P
	Comodidad	Pullman	Salón	Pullman	Salón
	Atraso (min)	40	20	40	20
Nivel 2	Tiempo de Viaje (horas)	8	8	6	4 1/2
	Tarifa (\$)	1,3*P	P	P	P

Nota: P es la tarifa que pagó el individuo por su viaje.

Se debe destacar que gracias a la flexibilidad que permite el computador, la encuesta se adapta a la persona entrevistada, de forma de ponerla en un escenario lo más realista posible. Específicamente al usuario se le pregunta por el destino y la tarifa que pagó ("P"), factores que varían en los escenarios hipotéticos propuestos. Además de la flexibilidad, la implementación en computadores portátiles se tradujo en una fácil y rápida codificación; finalmente, llamó la atención el respeto que este medio despertaba en los encuestados, lo que facilitó la labor y aseguró una mayor seriedad en las respuestas.

A partir de los atributos y niveles de la Tabla N°1 se procedió a determinar las tarjetas (o pantallas) con las alternativas que serían presentadas a los encuestados. Un diseño ortogonal completo se traduce en 36 tarjetas, lo que obviamente atenta contra la paciencia de los

entrevistados. Para reducir el conjunto de opciones posibles, se optó por un diseño factorial Fraccional (Pearmain *et al.*, 1991) que consta de sólo nueve tarjetas. A éstas se agregó una que reproducía las condiciones de mercado que el usuario había enfrentado al tomar la decisión, a fin de medir la posible inconsecuencia o falta de comprensión del experimento por parte del viajero.

En la Tabla N°2 se hace una descripción de la cantidad de datos recolectados por modo y tipo de viajero, de acuerdo a la ocupación y propósito de viaje (obligado o de placer).

Tabla N°2: Número de encuestas por modo y tipo de viajero (PD)

		No Trabajadores		Trabajadores		Total Parcial		Total Parcial		Total
		Oblig.	Placer	Oblig.	Placer	Oblig.	Placer	Notrab.	Trab.	
Bus		7	13	14	17	21	30	20	31	51
Tren	Econ.	2	0	3	3	5	3	2	6	8
	Salón	6	14	9	10	15	24	20	19	39
Total Parcial		15	27	26	30	41	57	42	56	98
Total		42		56		98		98		

Los 98 individuos debieron realizar diez elecciones cada uno; a su vez, cada una de éstas constituye un *pseudo* individuo. La Tabla N°3 detalla estas elecciones según el modo dónde fue tomada la encuesta, lo que permite apreciar que los individuos tienen mayor tendencia a elegir el modo que van ocupando. Este fenómeno se refleja luego en la importancia que tiene la variable Inercia en la modelación.

Notamos que se observó que 14 encuestados en los trenes (12,5 % del total) elegían siempre el modo tren, independiente de las alternativas que se presentaran. Estos, en general, respaldaron su actitud con argumentos de seguridad y comodidad del tren v/s el bus. Por tratarse de un comportamiento particular se dejaron fuera de la modelación, lo que no significa que estos presuntos clientes cautivos del tren deban ser ignorados en estudios futuros de predicción de demanda y valor subjetivo del tiempo.

Tabla N°3: Elecciones de los pseudo individuos v/s modo en que se hizo el experimento

Pseudo Elecciones		Modo en que se tomó la Encuesta					
		Bus	Tren	Bus		Tren	
				Trab	NoTrab	Trab	Notrab
Modo que Elige	Bus	60%	41%	189	118	91	103
	Tren	40%	59%	121	82	159	117
Total		100%	100%	510		470	



### 3.2. Los Datos de PR

Se dispuso de dos fuentes para estos datos:

- a) Los obtenidos a partir de las encuestas de PD (en adelante PR1). Como en este caso no se consultó explícitamente la duración real del viaje y su atraso (y el de su alternativa), no conoce exactamente estas variables. Apoyado en consultas hechas a los operadores de los medios estudiados, se estimó que generalmente el tiempo de viaje tenía una variación positiva o negativa de 15 minutos sobre el tiempo de viaje presupuestado (o de lista). Por esto se supuso que el tiempo total de viaje tenía una distribución Normal con media en el tiempo de lista, varianza de 15 minutos y truncado en más-menos 15 minutos sobre la media. De esta forma se consiguió modelar la variabilidad inherente a este tipo de servicio.
- b) Datos provenientes de un estudio de la Universidad de Chile (ver Gálvez *et al.*, 1992), gentilmente facilitados por estos investigadores (en adelante PR2). Se trata de viajes interurbanos, para usuarios con origen en Santiago y destinos en Concepción y Temuco; como medios de transporte considera las distintas clases de tren, bus y avión, y además se dispone de un conjunto bastante amplio de preguntas de carácter socioeconómico y del tipo de viaje. La Tabla N°4 presenta la cantidad de datos tomados en esta encuesta, de acuerdo a la ocupación, propósito de viaje (obligado o placer) y modo de transporte escogido.

Tabla N°4: Número de encuestas por modo y tipo de viajero (PR - U. de Chile)

		No Trabajadores		Trabajadores		Total Parcial		Total Parcial		Total
		Oblig.	Placer	Oblig.	Placer	Oblig.	Placer	Notrab.	Trab.	
Bus		17	45	58	53	75	98	62	111	173
Tren	Econ.	5	9	16	32	21	41	14	48	62
	Salón	7	10	20	32	27	42	17	52	69
Avión		6	5	59	33	65	38	11	92	103
Total Parcial		35	69	153	150	188	219	104	303	407
Total		104		303		407		407		

## 4. RESULTADOS DE LA MODELACIÓN

### 4.1 Estructura Artificial para los Datos Mixtos

La estructura de este árbol artificial depende ciertamente de las estructuras particulares de cada tipo de datos. Dada la forma en que se recolectó la información y el tipo de datos que se dispone, se estimó que la estructura que reproducía de forma más adecuada estas condiciones era la que aparece en la Figura N°2. Una alternativa interesante es suponer que los datos de PR2 deben descolgarse de un nido particular, dado el distinto error asociado, producto de las distintas condiciones en que se tomaron. Esto se hace únicamente para distinguir los datos de PR, ya que no se pone en duda el factor de escala en los de PD. Sin embargo, al probar esa



estructura arrojó un EMU para el nido en cuestión no significativamente (95%) distinto de uno (ver Ortúzar, 1982), por lo que se colapsó en la estructura mostrada.

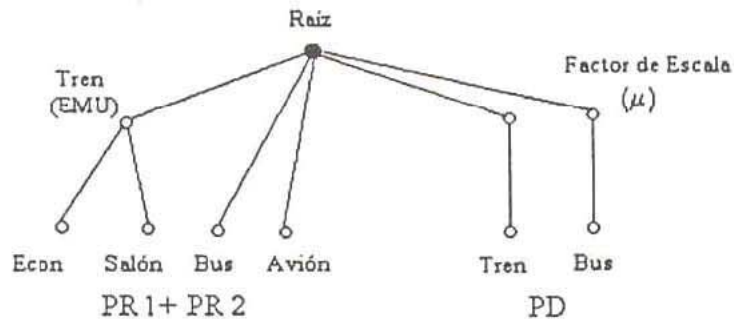


Figura N°2 : Estructura Artificial para DM

Es interesante notar que para los datos de PD y PR1 se ocupa una alternativa compuesta para tren, que incluye las clases económica y salón diferenciadas por la variable comodidad. En el caso de PR2 éstas se presentan explícitamente como alternativas distintas, pero dada la obvia correlación entre dichas opciones se utiliza una estructura de logit jerárquico. Se debe hacer notar, sin embargo, que en el trabajo de Gálvez *et al.* (1992) se usa una estructura logit multinomial (MNL) para estos mismos datos. Ese modelo no sólo entrega resultados de peor calidad que el nuestro, sino que es claramente rechazado por el valor del parámetro estructural correspondiente, como se muestra más adelante.

#### 4.2. Modelo de Utilidad Representativa

Respecto a la forma funcional de la utilidad representativa destacan los siguientes aspectos:

- a) Típicamente la utilidad representativa del modo  $i$  ( $V_i$ ) está definida por una función lineal (en los parámetros) como sigue:

$$V_i = k_i + \theta_c c_i f(I) + \theta_t t_i + \sum \theta_x X_{ix} \quad (6)$$

donde  $k_i$  es la constante modal del modo  $A_i$ ,  $c_i$  es su costo,  $f(I)$  es alguna función del Ingreso ( $I$ ) del individuo,  $t_i$  es el tiempo de viaje del modo  $A_i$ ,  $X_{ix}$  son otros atributos del medio o del individuo, y  $\theta_c$ ,  $\theta_t$  y  $\theta_x$  son los parámetros a estimar.

- b) En base a un enfoque determinístico desarrollado originalmente por Train y McFadden (1978) se han propuesto distintas formas funcionales para introducir el ingreso en función de utilidad representativa total. En primer lugar es razonable suponer que el ingreso deba dividir al costo, de manera de que este último afecte en

distinta proporción a gente con distinto salario; la discusión se ha centrado en qué ingreso ocupar. Tradicionalmente se ha usado la tasa salarial  $w$ , sin embargo, la que ha dado mejores resultados en nuestro país (ver Jara-Díaz y Ortúzar, 1989) y parece más apropiada desde el punto de vista teórico es la especificación con tasa de gasto  $g$  (Jara-Díaz y Farah, 1987). En ella básicamente se establece que para salarios fijos (que no dependen del número de horas trabajadas) lo correcto es dividir el ingreso (en este caso familiar) por el número de horas disponibles para gastarlo. Explícitamente :

$$w = IFAM / W \quad \text{ó} \quad g = IFAM / (T-W) \quad (7)$$

donde  $W$  es el número de horas trabajadas en el período de análisis considerado ( $T$ ), e  $IFAM$  el ingreso familiar de los individuos. La incerteza (o ignorancia) del modelador se introduce por los ya mencionados errores estocásticos  $\varepsilon_i$ , quedando la función de utilidad de la siguiente forma en cada caso:

$$U_i = V_i (c_i/w, t_i, \text{otros}) + \varepsilon_i \quad (8)$$

para la tasa salarial, y

$$U_i = V_i (c_i/g, t_i, \text{otros}) + \varepsilon_i \quad (9)$$

para el caso de los modelos con tasa de gasto.

#### 4.3. Descripción de las Variables

En la etapa de calibración se usaron las siguientes variables:

- a) Constantes modales. Pese que en total hay seis alternativas disponibles, debido a la no disponibilidad mutua entre distintos tipos de datos, bastó con considerar tres constantes modales: una para el tren de PD, otra para la clase económica del tren en PR y una tercera para el avión.
- b) Variables genéricas (comunes). Corresponden al vector  $X$  de las ecuaciones (4) y (5). Están constituidas por el tiempo de viaje y las distintas especificaciones de la tarifa (dividida por el ingreso), ya que son las únicas comunes a todos los datos:
  - Tiempo de Viaje: Tiempo de viaje en el vehículo (horas).
  - Tarifa/g: Como la tasa de gasto ( $g$ ) es en función del Ingreso familiar y para los desempleados las horas trabajadas  $W$  son cero, éstos resultan con tasas de gasto menores que los asalariados (ver Jara-Díaz y Farah, 1987).



- Tarifa/IfamPC: En este caso, la tarifa es dividida por el ingreso familiar *per capita*. Esta especificación se usa sólo en el modelo segmentado para desempleados, ya que desgraciadamente no se dispone de esta información para los trabajadores (en cuyo caso se usa el ingreso líquido reportado).
- c) Variables específicas. Corresponden a los vectores  $\underline{Y}$  y  $\underline{Z}$  de la ecuaciones (4) y (5). A continuación se detallan especificando en qué tipo de datos se incorporan:
- Comodidad (PD): Variable muda en la función de utilidad del tren. Vale uno si el tren es clase salón y cero si es económica.
  - Confiabilidad (PD): Atraso promedio que presenta el medio (horas).
  - Frecuencia (PR1 y PR2): Expresada en salidas semanales, por lo que se espera un coeficiente positivo. Notar que en el contexto interurbano en general los viajes se planifican con anterioridad, por lo que no se puede relacionar esta variable con el tiempo de espera tradicional de los viajes urbanos (ver Gálvez *et al.*, 1992).
  - Inercia (PD): Refleja la tendencia a elegir el modo en que se está viajando en el experimento de preferencias declaradas (ver Tabla N°3). Este fenómeno también fue percibido por Bradley y Daly (1991).
- d) Variables estructurales. Son aquellas propias de la estructura jerárquica de árbol mostrada en la Figura N°2.
- EMU Tren (PR): Es el parámetro de la máxima utilidad esperada para el nido del tren (ver por ejemplo, Ortúzar y Willumsen, 1990). Su valor debe estar entre cero y uno. Un valor cercano a uno sugiere colapsar la estructura a un modelo logit simple.
  - Factor de Escala (PD): Es el estimador  $\mu$  presentado en las ecuaciones (3), (4) y (5).

#### 4.4. Resultados Principales

La Tabla N°5 muestra modelos puros de preferencias reveladas (PR2 y PR1+PR2), de preferencias declaradas (PD) y modelos mixtos (los restantes). En primer lugar es importante señalar que todos se caracterizan por tener signos correctos y buenos índices de bondad de ajuste. La única variable que no resulta significativamente distinta de cero al 95% de confianza es la Confiabilidad (en todos los modelos que la postulan como variable explicativa), pero otra variable novedosa, la Comodidad, resulta altamente significativa. Así, muy probablemente, todos los modelos - aisladamente - serían considerados aceptables por un analista con experiencia.

En el caso de los modelos de PR, es además importante destacar que la variable estructural correspondiente a la simple estructura jerárquica postulada es significativamente distinta de uno, lo que implica que los datos avalan la hipótesis nula de correlación entre ambas alternativas de tren (económico y salón).

Dados todos estos antecedentes, resulta de sumo interés hacer notar el elevado valor subjetivo del tiempo (VST) obtenido en ambos casos de PR (cerca de cuatro veces superior al Ingreso familiar promedio de los viajeros). Cabe destacar que los VST se calcularon de la forma tradicional, esto es, como el producto de la tasa de gastos promedio y el cociente entre los parámetros de las variables Tiempo de Viaje y Tarifa/g (ver Gaudry *et al.*, 1989); también es necesario señalar que todos los VST son estadísticamente significativos al 95% de confianza.

En el caso de los modelos PD y mixtos, se debe tener en cuenta que los valores del estadígrafo  $t$  no fueron corregidos para descontar la potencial influencia positiva de suponer que los 980 pseudo individuos de PD (correspondientes a las 98 personas encuestadas) son independientes. En general los VST son bastante menores, aún cuando todavía altos en relación a la práctica internacional; cabe destacar que el modelo mixto usando los datos PR correspondientes a los 98 encuestados, es el que entrega menor VST (sólo 50% mayor que el Ingreso familiar promedio).

No se puede dejar de mencionar que el modelo más interesante - que combina datos PD con datos PR provenientes de distinta fuente - entrega VST que son 2,2 veces superiores al Ingreso familiar promedio de los encuestados. Así, aun cuando los miembros de la muestra PR2 tienen menores ingresos en promedio, aportan VST más altos que las 98 personas encuestadas en nuestro estudio; pensamos que esto se debe, en parte importante, a que incluyen personas que eligen avión.

Finalmente, es importante notar que el factor de escala  $\mu$  es estadísticamente distinto que 1,0 en el modelo mixto que combina todos los datos, a diferencia del modelo que no incluye los datos PR2. En éste último, si bien los datos de PR parecen aportar mayor ruido que los de PD, la evidencia no es decisiva. Cabe destacar que se ha encontrado valores significativamente menores que 1,0 en experiencias anteriores en el país (Ortuzar y Garrido, 1993); así, esto se puede deber en parte a que los datos PD fueron recolectados con mayor cuidado en esta experiencia, y en parte a que los datos PR tenían escasa variabilidad.

La Tabla N°6 presenta los resultados de modelar separadamente a trabajadores y no trabajadores. Como se puede ver, los modelos también son excelentes en cuanto a signos de los parámetros y estadígrafos de ajuste; y nuevamente la única variable no significativa es la Confiabilidad. También se puede ver, una vez más, que la estructura jerárquica postulada es apoyada por los datos (los parámetros estructurales de EMU son significativamente distintos de uno); cabe destacar que los factores de escala de los modelos para no trabajadores, son menores que en los casos anteriores. No obstante, esto no ocurre en el modelo para trabajadores en que el factor no sólo es alto sino que significativamente distinto de 1,0.

Un hecho sorprendente ocurre con los VST, que nuevamente son significativamente distintos de cero; en efecto, los valores para no trabajadores son bastante más altos (contra lo esperado) que el valor estimado para trabajadores (que es sólo 50% mayor que el Ingreso familiar). Estamos estudiando las posibles razones de este resultado contra la intuición. Finalmente, aparte de los modelos presentados en esta sección se ha estimado una enorme variedad de funciones que no es el caso discutir acá; los interesados pueden consultar a Iacobelli (1994).



Tabla N° 5: Distintas mezclas de datos para la misma especificación

Coeficiente (test t)	PR2	PR1+PR2	PD	PD+PR1	PD+PR1+PR2
Variables Comunes					
Tpo. de Viaje (horas)	-0,7002 (4,0)	-0,7128 (4,7)	-0,7518 (9,3)	-0,8196 (3,4)	-0,4194 (4,5)
Tarifa/g (horas)	-0,0600 (7,6)	-0,0649 (7,9)	-0,1496 (6,1)	-0,1828 (3,0)	-0,0692 (8,2)
Variables Específicas					
Comodidad (PD)	-	-	1,4550 (9,4)	0,8314 (3,0)	0,7790 (4,2)
Confiabilidad (horas)	-	-	-0,5135 (1,1)	-0,5996 (1,0)	-0,2664 (1,0)
Frecuencia (salidas/semana)	0,0051 (4,8)	0,0048 (5,0)	-	-	0,0033 (5,1)
Inercia (PD)	-	-	0,6698 (4,6)	0,7502 (2,6)	0,3700 (3,0)
Constantes Modales					
Cte. Tren (PD)	-	-	0,3992 (2,8)	0,3991 (2,1)	0,2325 (2,4)
Cte. Tren Econ. (PR)	-0,4872 (2,6)	-0,7404 (4,5)	-	-	0,7540 (4,6)
Cte. Tren <sup>1</sup> (PR1)	-	-	-	-1,0320 (2,7)	
Cte. Avión (PR)	-0,2999 (0,3)	-0,3202 (0,4)	-	-	1,2970 (2,7)
Variables Estructurales					
EMU Tren (PR2)	0,6612 (13,8)	0,6718 (20,6)	-	-	0,6938 (17,9)
Factor Escala (PD)	-	-	-	1,7620 (2,7)	1,8510 (4,4)
Valor Subjetivo del Tiempo					
VST(\$/min)	123,81	117,44	56,05	45,38	66,54
(test-t)	(3,88)	(4,27)	(5,90)	(6,20)	(5,26)
IFAM <sup>2</sup> (\$/min)	28,34	28,73	30,37	30,37	29,81
Indices					
L(c)	-497,92	-597,06	-679,03	-746,78	-1276,09
L(0)	-415,32	-505,21	-555,98	-615,77	-1063,87
r <sup>2</sup>	0,17	0,15	0,18	0,18	0,17
Tam. Muestral <sup>3</sup>	407	505	980	1078	1485

<sup>1</sup> En el caso de PD + PR1 se ocupa la misma estructura para ambos datos (opción tren agregada) por lo que se necesitan dos constantes diferentes. <sup>2</sup> Ingreso familiar expresado en (\$/min) en base a considerar un promedio de 48 horas semanales de trabajo. <sup>3</sup> Los datos de PD corresponden a pseudo individuos.

**Tabla N° 6: Modelos segmentados por ocupación del viajero**

<b>Coefficiente (test t)</b>	<b>No Trabajadores IfamPC</b>	<b>No Trabajadores Tasa de Gasto</b>	<b>Trabajadores Tasa de Gasto</b>
<b>Variables Comunes</b>			
Tpo. de Viaje (horas)	-0,8520 (2,2)	-0,7128 (2,5)	-0,2772 (3,6)
Tarifa/IfamPC (horas)	-0,0170 (4,2)	-	-
Tarifa/g (horas)	-	-0,0709 (4,2)	-0,0751 (7,0)
<b>Variables Específicas</b>			
Comodidad (PD)	1,7550 (2,1)	1,4790 (2,5)	0,4124 (3,2)
Confiabilidad (horas)	-0,0141 (0,1)	0,0692 (0,1)	-0,2884 (1,2)
Frecuencia (salidas/semana)	0,0048 (2,4)	0,0045 (2,9)	0,0031 (4,4)
Inercia (PD)	0,2957 (1,0)	0,2420 (1,1)	0,3537 (2,8)
<b>Constantes Modales</b>			
Cte. Tren (PD)	-0,9988 (1,9)	-0,8022 (2,1)	0,0185 (0,3)
Cte. Tren Econ. (PR)	-1,2080 (3,6)	-1,2910 (3,7)	-0,5692 (3,0)
Cte. Avión (PR)	-1,1310 (0,6)	-0,1628 (0,1)	2,462 (5,8)
<b>Variables Estructurales</b>			
EMU Tren (PR2)	0,7213 (16,4)	0,7112 (16,7)	0,6222 (9,1)
Factor Escala (PD)	1,109 (2,0)	1,3460 (2,4)	2,6940 (3,5)
<b>Valor Subjetivo del Tiempo</b>			
VST(\$/min)	87,22	69,58	44,91
(test-t)	(2,41)	(3,11)	(4,28)
IFAM <sup>1</sup> (\$/min)	24,22	24,22	29,82
<b>Estadísticos</b>			
L(c)	-443,01	-443,01	-825,98
L(0)	-359,54	-355,75	-689,59
$\rho^2$	0,19	0,20	0,17
Tam. Muestral <sup>2</sup>	566	566	919

<sup>1</sup> Ingreso familiar expresado en (\$/min) en base a considerar un promedio de 48 horas semanales de trabajo. <sup>2</sup> Los datos de PD corresponden a pseudo individuos.



## 5. CONCLUSIONES

En este trabajo se han estimado modelos mixtos de preferencias reveladas y declaradas, utilizando información correspondiente a viajes interurbanos. Esta labor es pionera por los siguientes motivos:

- nunca se había estimado en el país modelos mixtos con este tipo de datos; la única experiencia de datos mixtos anterior utilizó información correspondiente a alternativas hipotéticas en el caso urbano (Ortúzar y Garrido, 1993);
- se combinó información de preferencias reveladas proveniente de distintas fuentes, con datos de preferencias declaradas, lo que tampoco se había realizado nunca en el país;
- finalmente, dentro de la estimación mixta se utilizó una estructura jerárquica, lo que de acuerdo a nuestra información no se había hecho antes en ninguna parte del mundo.

Todos los modelos estimados son de buena calidad, en el sentido de tener buenos índices de bondad de ajuste, signos correctos y parámetros significativos (con la excepción de una variable difícil de medir en la práctica, la Confiabilidad). Los valores subjetivos del tiempo derivados a partir de las funciones estimadas son extraordinariamente altos (no sólo en relación a la práctica internacional, sino que a las recomendaciones actuales del gobierno); no obstante, todos son significativamente distintos de cero al 95% de confianza. Una posible explicación para este hecho es que los datos de PR2 contienen viajeros de avión. Esto es consistente con el resultado que los valores correspondientes a los modelos mixtos son más bajos que los derivados de información sobre preferencias reveladas.

La conclusión menos feliz respecto de nuestros resultados, es que al estratificar la muestra en trabajadores y no trabajadores, se obtienen VST más altos para los no trabajadores; no obstante, este resultado extraño desaparece al considerar sólo los datos de PD y PR1.

## AGRADECIMIENTOS

Los autores desean agradecer el valioso aporte de Rodrigo Garrido en diversas etapas del estudio. Además se agradece la colaboración del René Alvarez en la fase de recolección de datos y, especialmente, la de Carolina Simonetti, en las fases de diseño y análisis estadístico de la información. También deseamos agradecer a Tristán Gálvez y Jorge Vera, por habernos facilitado los datos de la Universidad de Chile, y a Sergio Jara por ayudarnos a resolver nuestras dudas metodológicas respecto al tratamiento más adecuado de la variable ingreso.

Esta investigación ha contado con el importante financiamiento del Fondo Nacional de Desarrollo Científico y Tecnológico, a través del proyecto FONDECYT 774/91.

## REFERENCIAS

- Bradley, M.A. y Daly A.J. (1991) Estimation of logit choice models using mixed stated preference and revealed preference information. **Preprints 6th International Conference of Travel Behaviour**, Quebec, Canadá, Mayo 1991.
- Ben-Akiva, M. y Morikawa, T. (1990) Estimation of travel demand models from multiple data sources. **Proceedings 11th International Symposium on Transportation and Traffic Theory**, Yokohama, Japón, Julio 1990.
- Daly, A.J. (1992) **Alogit 3.2 User's Guide**. Hague Consulting Group, La Haya.
- Gálvez, T., Jara-Díaz, S.R. y Vera, J. (1992) Modelación de partición modal en viajes interurbanos. **Actas VII Congreso Panamericano de Ingeniería de Tránsito y Transporte**, Caracas, Venezuela, Septiembre de 1992.
- Gaudry, M.J., Jara-Díaz, S.R. y Ortúzar, J. de D. (1989) Value of time sensitivity to model specification. **Transportation Research** 23B, 151-158.
- Iacobelli, A. (1994) **Valor Subjetivo del Tiempo en Viajes Interurbanos para Bus y Tren**. Tesis de Magister, Departamento de Ingeniería de Transporte, Pontificia Universidad Católica de Chile, Santiago.
- Jara-Díaz, S.R. y Farah, M. (1987) Transport demand and users' benefits with fixed income: the goods/leisure trade-off revisited. **Transportation Research** 21B, 165-170.
- Jara-Díaz, S.R. y Ortúzar, J. de D. (1989) Introducing the expenditure rate in the estimation of mode choice models. **Journal of Transport Economics and Policy** XXIII, 293-308.
- Morikawa, T., Ben Akiva, M. y McFadden, D. (1990) Incorporating psychometric data in econometric demand models. **Bauff Invitational Symposium on Consumer Decision Making and Choice Behaviour**, Canadá, Mayo de 1990.
- Ortúzar, J. de D. (1982) Fundamentals of discrete multimodal choice modelling. **Transport Reviews** 2, 47-78.
- Ortúzar, J. de D. y Willumsen, L.G. (1990) **Modelling Transport**. John Wiley & Sons, Chichester.
- Ortúzar, J. de D. y Garrido, R. (1991) Rank, rate or choice? an evaluation of SP methods in Santiago. **Proceedings 19th PTRC Summer Annual Meeting**, University of Sussex, Inglaterra, Septiembre 1991.
- Ortúzar, J. de D. y Garrido, R. (1993) Estimation of discrete choice models with mixed SP and RP data. **XIII World Conference on Operations Research (IFORS '93)**, Lisboa, Portugal, Julio 1993.

Pearmain, D., Swanson, J., Kroes, E. y Bradley, M. (1990) **Stated Preferences Techniques: A Guide to Practice**. Steer Davies Gleave y Hague Consulting Group, Londres.

Train, K. y McFadden (1978) The goods/leisure trade-off disaggregate work trip mode choice models. **Transportation Research** 12, 349-353.

Wardman, M. (1991) Stated preference methods and travel demand forecasting: an examination of the scale factor problem. **Transportation Research** 25A, 79-89.