

Análisis de los Enfoques Secuencial y Simultáneo para la Estimación de Modelos Híbridos de Elección Discreta

Sebastián Raveau, Juan de Dios Ortúzar y María Francisca Yáñez
 Departamento de Ingeniería de Transporte y Logística,
 Pontificia Universidad Católica de Chile,
 Casilla 306, Código 105, Santiago 22, Chile
 Tel: 354 4270; E-mail: sraveau@puc.cl, jos@ing.puc.cl, mfyanezc@puc.cl

RESUMEN

La formulación de modelos híbridos de elección discreta, que incluyen atributos tangibles de las alternativas y variables latentes asociadas a las actitudes y percepciones, ha sido un importante tema de discusión en las últimas décadas. Para estimar estos modelos se han desarrollado dos métodos: el secuencial, que aborda la construcción de variables latentes previo a su integración con variables explicativas tradicionales, y el simultáneo, en que ambos procesos se realizan conjuntamente. En este trabajo, se comparan los resultados obtenidos mediante ambos enfoques para casos de elección de modo de transporte a nivel urbano e interurbano.

Palabras claves: variables latentes, modelos de elección discreta, estimación simultánea.

ABSTRACT

The formulation of hybrid discrete choice models, including both tangible alternative attributes and latent variables associated with attitudes and perceptions, has been a major topic of discussion in recent decades. To estimate models with both kinds of variables, two methods have been developed: a sequential one, where the latent variables are built before their integration with the traditional explanatory variables and a simultaneous one, where both process are done together. On this paper, we compare the results obtained using both approaches for urban and intercity modal choice contexts.

Keywords: latent variables, discrete choice models, simultaneous estimation.

1. INTRODUCCIÓN

La inclusión de elementos subjetivos en modelos de elección discreta ha resurgido con fuerza como tema de análisis y discusión, tras perder algo de interés luego de haber sido un tema álgido a principios de los 80 (Ortúzar y Hutt, 1984). Así, en la última década se han planteado modelos híbridos de elección que consideran tanto atributos tangibles de cada alternativa (i.e. las variables explicativas clásicas), como elementos intangibles asociados a percepciones y/o actitudes de las personas, expresadas a través de variables latentes (McFadden, 1986; Ashok *et al.*, 2002). Para estimar modelos con ambos tipos de variables se han desarrollado dos métodos: el secuencial, que aborda la construcción de variables latentes previo a su integración con las variables tradicionales, y el simultáneo, en que ambos procesos se realizan conjuntamente. El segundo debiera producir estimadores más eficientes de los parámetros involucrados (Ben-Akiva *et al.*, 2002), pero ha sido menos utilizado debido a su mayor complejidad.

El principal objetivo de este trabajo es realizar un análisis crítico de los resultados obtenidos a partir de la estimación simultánea de modelos de elección discreta en comparación con los obtenidos en forma secuencial, a fin de identificar posibles discordancias y sus razones de ocurrencia. Para esto se consideran dos bancos de datos, provenientes de encuestas de elección modal para viajes urbanos e interurbanos, en cuya modelación se incorporan variables latentes.

2. ANTECEDENTES GENERALES

2.1 Modelo de Variables Latentes

Las variables latentes son factores que inciden en el comportamiento y percepciones individuales, pero no se pueden cuantificar debido a su intangibilidad (no poseen una escala de medición) o a su subjetividad intrínseca (distintas personas las perciben de diferente forma). La identificación de variables latentes requiere comenzar por complementar una encuesta de preferencias reveladas estándar, con preguntas que intentan captar las percepciones de los individuos sobre ciertos aspectos que de otra manera no sería posible medir, generando indicadores de percepción.

La metodología desarrollada para el tratamiento de variables latentes (Bollen, 1989) consiste en plantear un modelo del tipo MIMIC (*Multiple Indicator Multiple Cause*) en que estas variables (η_{iq}) son explicadas por características de los individuos y las alternativas (s_{iqr}) a través de *ecuaciones estructurales* como (1); a su vez, las variables latentes explican los indicadores de percepción (y_{ipq}) mediante *ecuaciones de medición* como (2):

$$\eta_{ilq} = \sum_r \alpha_{ilr} \cdot s_{iqr} + v_{ilq} \quad (1)$$

$$y_{ipq} = \sum_l \gamma_{ilp} \cdot \eta_{ilq} + \zeta_{ipq} \quad (2)$$

donde i indica una alternativa, l una variable latente, q un individuo, r una variable explicativa y p un indicador; α_{ilr} y γ_{ilp} son parámetros a estimar, mientras que v_{ilq} y ζ_{ipq} son componentes de error que se asumen con media cero y desviación estándar a determinar. Dado que los términos η_{ilq} son desconocidos, ambas ecuaciones deben ser consideradas en forma conjunta en el proceso de obtención de estimaciones de los parámetros.

2.1 Modelo Híbrido de Elección Discreta

En modelación de elecciones discretas se asume tradicionalmente que los individuos actúan de forma racional maximizando su utilidad U_{iq} ; el modelador, que es un observador del sistema, define una utilidad representativa V_{iq} y (como no posee información perfecta) un término de error ε_{iq} asociado a cada alternativa (Ortúzar y Willumsen, 2001), tal que:

$$U_{iq} = V_{iq} + \varepsilon_{iq} \quad (3)$$

La utilidad representativa V_{iq} es función de atributos objetivos de las alternativas X_{ikq} , como tiempo y tarifa; si se incluyen variables latentes se obtiene una función de utilidad como (4), donde θ_{ik} y β_{il} son parámetros a calibrar asociados, respectivamente, a los atributos tangibles y las variables latentes:

$$V_{iq} = \sum_k \theta_{ik} \cdot X_{ikq} + \sum_l \beta_{il} \cdot \eta_{ilq} \quad (4)$$

Como las variables η_{ilq} no son conocidas, el modelo de elección debe abordarse en conjunto con las ecuaciones estructurales (1) y de medición (2) del modelo MIMIC. Finalmente, para caracterizar las decisiones individuales se definen variables binarias d_{iq} , que toman valores según (5), donde $A(q)$ es el conjunto de alternativas disponibles para el individuo q .

$$d_{iq} = \begin{cases} 1 & \text{si } U_{iq} \geq U_{jq}, \quad \forall j \in A(q) \\ 0 & \text{en otro caso} \end{cases} \quad (5)$$

2.3 Metodologías de Estimación

2.3.1 Estimación Secuencial

En este caso el problema se aborda en dos etapas, separando las interacciones entre el modelo de variables latentes y el de elección discreta. En primera instancia se resuelve el modelo MIMIC, para obtener estimadores de los parámetros en las ecuaciones que relacionan las variables latentes con los indicadores de percepción y las variables explicativas. Utilizando estos parámetros en la ecuación (1), se obtienen los valores esperados de las variables latentes para cada individuo y alternativa. Con esto, las variables latentes pueden ser parte del conjunto de variables explicativas del modelo de elección, estimándose junto a las variables tradicionales en una segunda etapa.

Si bien este método posee la desventaja de no utilizar toda la información disponible en conjunto, la aplicación es clara y no presenta dificultades adicionales a las de un modelo tradicional,

razón por la que es el método más utilizado en la práctica (Ashok, *et al.*, 2002; Vredin Johansson *et al.*, 2005). Sin embargo, un problema potencialmente serio de este enfoque es que no garantiza estimadores insesgados de los parámetros involucrados (Bollen, 1989); de igual manera, las desviaciones estándar de los parámetros tienden a subestimarse, produciendo estimadores cuyo nivel de significancia estadística es más elevado que su aporte real al modelo. Este problema puede solucionarse mediante una corrección estadística de las varianzas de los parámetros (Murphy y Topel, 1985), pero este no es un proceso sencillo de realizar. Esta corrección no se realiza en este estudio ya que, a medida que la cantidad de alternativas y parámetros del modelo híbrido aumenta, la dificultad de la corrección de la significancia estadística de las variables latentes crece exponencialmente, complejizando considerablemente el proceso.

2.3.2 Estimación Simultánea

En este enfoque la calibración conjunta se realiza mediante máxima verosimilitud, tomando la probabilidad de replicar la elección individual en base a la utilidad representativa planteada por el modelador; es decir $\Pr(d_{iq} | V_{iq})$. A partir de (4), esta probabilidad condicional puede expresarse en términos de las variables y parámetros del modelo de elección, obteniéndose una probabilidad de elección de la forma $\Pr(d_{iq} | X_{ikq}, \eta_{ilq}, \theta_{ik}, \beta_{il})$.

Como las variables latentes no son observables, es necesario integrar sobre su espacio de variación, condicionándolas en las variables que las explican. Así, la probabilidad de elección está dada por (6), donde $g(\cdot)$ es la función densidad de probabilidad de las variables latentes.

$$\Pr(d_{iq} | X_{ikq}, s_{iqr}, \theta_{ik}, \beta_{il}, \alpha_{ilr}) = \int_{\eta_{ilq}} \Pr(d_{iq} | X_{ikq}, \eta_{ilq}, \theta_{ik}, \beta_{il}) \cdot g(\eta_{ilq} | s_{iqr}, \alpha_{ilr}) \cdot d\eta_{ilq} \quad (6)$$

Para estimar el modelo es necesario introducir la información entregada por los indicadores de percepción pues, de lo contrario, el modelo no sería identificable. Los indicadores no son variables explicativas del modelo sino, al contrario, variables endógenas a las variables latentes tal que la probabilidad de elección a utilizar durante la estimación está dada por (7), donde $f(\cdot)$ es la función densidad de probabilidad de los indicadores.

$$\Pr(d_{iq}, y_{ipq} | X_{ikq}, s_{iqr}, \theta_{ik}, \beta_{il}, \alpha_{ilr}, \gamma_{ilp}) = \int_{\eta_{ilq}} \Pr(d_{iq} | X_{ikq}, \eta_{ilq}, \theta_{ik}, \beta_{il}) \cdot f(y_{ipq} | \eta_{ilq}, \gamma_{ilp}) \cdot g(\eta_{ilq} | s_{iqr}, \alpha_{ilr}) \cdot d\eta_{ilq} \quad (7)$$

Una vez definida la forma funcional del modelo de elección discreta¹, se utiliza máxima verosimilitud simulada para la estimación de los parámetros del modelo híbrido (Bolduc *et al.*, 2008); en nuestro caso se ocupó el Método del Gradiente, realizando en cada iteración 500 simulaciones de las variables latentes a partir de una secuencia pseudo-aleatoria de Halton.

3. APLICACIÓN A UN CASO INTERURBANO

El primer caso de estudio corresponde a una encuesta de preferencias reveladas basada en la elección de viajes interurbanos entre Santiago y cuatro ciudades del sur de Chile. Para el estudio se consideró sólo dos modos de transporte: bus y tren (Egaña, 2009). Se consultó a los encuestados sobre: ingreso, nivel de educación, actividades realizadas durante el viaje, quién

¹ En este estudio sólo se consideró modelos Logit Multinomial, debido a restricciones computacionales experimentadas al estimar el modelo en forma simultánea

pagaba el viaje y si el individuo viajaba en grupo, con bultos, con niños o con *notebook*. Además, se incluyó preguntas acerca de si habían sufrido, o visto personalmente, robos y/o accidentes.

Se recolectó información de tarifas (dos categorías de tren: salón y preferente, y dos para el bus: clásico y ejecutivo), tiempos de viaje y de espera. El tiempo de acceso al terminal no fue considerado pues los terminales de ambos modos son cercanos en las ciudades consideradas.

3.1 Estructura del Modelo de Variables Latentes

Al diseñar el estudio se consideró dos variables latentes: *comodidad* y *seguridad*. Para captar los efectos de estas variables, en el proceso de toma de decisión de los encuestados se solicitó que evaluaran (con notas de 1 a 7) una serie de aspectos asociados a viajar en bus o tren, obteniendo nueve indicadores de percepción: (1) facilidad para viajar con niños y/o bultos, (2) posibilidad de aprovechar el tiempo de viaje para leer, (3) comodidad durante el viaje, (4) puntualidad en el horario de salida, (5) posibilidad de relajarse durante el viaje, (6) seguridad durante el viaje, en el sentido de no sufrir accidentes, (7) seguridad en caso de accidentes, en el sentido de nivel de protección, (8) seguridad ante robos en el vehículo, y (9) seguridad ante robos en el terminal.

Para explicar las variables latentes comodidad y seguridad, se consideró las siguientes variables explicativas: (i) un grupo que caracteriza las condiciones del viaje: individuo viaja solo o en grupo, con o sin computador portátil, con o sin experiencia de accidentes y/o robos anteriores; (ii) dos variables binarias relacionadas con actividades desarrolladas durante el viaje: actividades sociales (jugar con niños y conversar con otros pasajeros) y actividades intelectuales (leer por placer y trabajar/estudiar); (iii) dos variables socioeconómicas: edad e ingreso, esta última considerando dos segmentos, ingreso bajo (inferior a \$400.00 mensuales) e ingreso alto; y (iv) una variable objetiva², que consiste en un índice de fatalidades dado por (8):

$$Índice = \frac{1}{8} \cdot \sum_{2000}^{2007} \text{Fatalidades Año}_i (\text{pax/año}) \cdot \text{Distancia Viaje} (\text{km}) \quad (8)$$

$$\text{Distancia Anual} (\text{km} \cdot \text{veh/año}) \cdot \text{Pasajeros por Viaje} (\text{pax/veh})$$

A partir de las relaciones teóricas entre las variables se plantearon distintas estructuras de causalidad entre las variables explicativas, las variables latentes y los indicadores de percepción, las que luego fueron corroboradas mediante análisis factorial. En la Figura 1 se indican las relaciones que resultaron estadísticamente relevantes entre las variables explicativas y las variables latentes, así como éstas últimas y los indicadores (Egaña, 2009).

3.2 Estructura del Modelo de Elección Discreta

La función de utilidad representativa considerada incluyó variaciones sistemáticas en los gustos para las variables *costo*, según fuente de pago y tipo de pasaje, y *tiempo de viaje*, según sexo. La forma general de la función de utilidad está dada por (9), que corresponde a la mejor especificación obtenida por Egaña (2009) en su estudio, tras considerar diversas formulaciones:

$$V_{iq} = (\theta_{\text{costo}} + \theta_{\text{paga}} \cdot \text{Paga}_q + \theta_{\text{categoría}} \cdot \text{Categoría}_q) \cdot \text{Costo}_{iq} + (\theta_{\text{via}} + \theta_{\text{sexo}} \cdot \text{Hombre}_q) \cdot \text{Tvia}_{iq} + \beta_{\text{com}} \cdot \text{Comodidad}_{iq} + \beta_{\text{seg}} \cdot \text{Seguridad}_{iq} + K_i \quad (9)$$

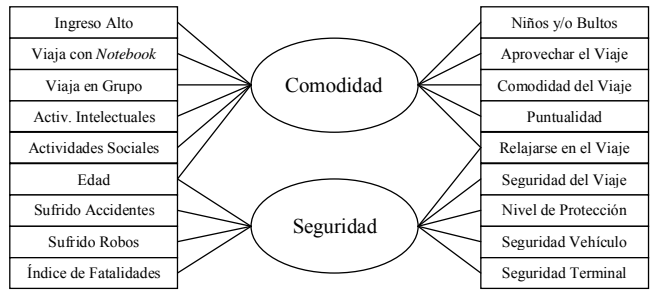


FIGURA 1: Estructura del Modelo de Variables Latentes, Caso Interurbano

No se incluyó como variable explicativa al tiempo de espera pues ambos modos tienen horarios predeterminados que se cumplen con relativa rigurosidad.

3.3 Análisis de Resultados

El modelo híbrido fue calibrado en forma secuencial por Egaña (2009) y en forma simultánea en este trabajo. En la Tabla 1 se presentan los resultados de ambos procesos de estimación, así como el modelo equivalente sin variables latentes. Las estimaciones se acompañan con los respectivos test-t de los parámetros y se incluye también el valor de la log-verosimilitud final de los modelos. Para el modelo híbrido en el método simultáneo, ésta corresponde a la función objetivo del problema; para el método secuencial se consideran las funciones de log-verosimilitud obtenidas por separado para el modelo de ecuaciones estructurales y de elección. Finalmente, para el modelo sin variables latentes se asumió nulos los parámetros de las ecuaciones estructurales y se resolvió sólo el modelo de elección, evaluando luego la función de log-verosimilitud. Es importante recordar que estos resultados se obtienen en forma conjunta con la estimación del modelo de variables latentes para el caso simultáneo (para detalles, ver Raveau, 2009).

TABLA 1: Resultados del Modelo de Elección, Caso Interurbano

Parámetro	Modelo Híbrido Secuencial	Modelo Híbrido Simultáneo	Modelo sin Variables Latentes
Costo	-0,754 (-6,91)	-0,658 (-7,32)	-0,721 (-7,00)
Paga	-0,176 (-2,21)	-0,209 (-2,61)	-0,221 (-2,87)
Categoría	0,925 (9,07)	0,960 (6,82)	1,025 (10,28)
Tiempo de viaje	-0,018 (-3,17)	-0,019 (-3,87)	-0,021 (-3,78)
Hombre	-0,001 (-0,26)	-0,005 (-1,35)	-0,006 (-1,53)
Comodidad	1,528 (5,70)	1,392 (3,84)	-
Seguridad	1,700 (4,79)	0,781 (2,54)	-
ASC Bus	0,358 (0,85)	0,136 (0,29)	-0,533 (1,42)
Numero de observaciones	600	600	600
Log-verosimilitud	-20.870,92	-18.847,10	-34.013,57

En primer lugar, se aprecia que todos los signos obtenidos en los modelos son correctos. Respecto a las magnitudes de los parámetros, se puede ver que los hombres experimentan una mayor des-utilidad que las mujeres frente a un aumento en el tiempo de viaje; por otra parte, viajar en categoría alta implica que los individuos parecen ser indiferentes al costo (éste no es relevante en la elección, producto del compromiso entre precio y calidad). La constante modal, como es esperable, resultó más significativa en el modelo sin variables latentes.

En la Tabla 2 se presentan los valores subjetivos del tiempo de viaje (Gaudry *et al.*, 1989) y las valoraciones de cada unidad de seguridad y comodidad adicional para los individuos que pagan pasaje y no viajan en clase alta (el caso de análisis más interesante).

² Los modelos reportados por la literatura consideran sólo variables socioeconómicas como variables explicativas. No se encontró reporte de la utilización de factores objetivos como en este caso.

Como ambas variables latentes no poseen unidades, no es posible establecer una medida clara de valoración; sin embargo, es posible comparar entre sí los valores estimados secuencial y simultáneamente.

TABLA 2: Valoraciones de Atributos, Caso Interurbano

Valor Subjetivo	Modelo Híbrido Secuencial	Modelo Híbrido Simultáneo	Modelo sin Variables Latentes
Tiempo de Viaje para Hombres	1.234 [\$/hr]	1.661 [\$/hr]	1.668 [\$/hr]
Tiempo de Viaje para Mujeres	1.167 [\$/hr]	1.315 [\$/hr]	1.307 [\$/hr]
Comodidad	1.643 [\$/unidad]	1.606 [\$/unidad]	-
Seguridad	1.828 [\$/unidad]	901 [\$/unidad]	-

Como se aprecia, los hombres tienen una mayor valoración del tiempo que las mujeres, lo que se puede explicar por su mayor tasa de empleo o por la mayor probabilidad de ser jefes de hogar. Un hecho interesante es que el modelo con valoraciones más cercanas a las del modelo híbrido simultáneo no es el secuencial, sino el modelo sin variables latentes (que posee en ambos casos errores menores al 1%), tanto para hombres como mujeres. El método secuencial subestima las valoraciones del tiempo en un 26% para los hombres y un 11% para las mujeres. Respecto de la comodidad, ambos métodos de resolución del modelo híbrido entregan resultados muy similares (2% de error); en cambio, la valoración de la seguridad obtenida con el método secuencial duplica a la del método simultáneo.

La alta significancia estadística de las variables latentes reafirma la importancia de plantear un modelo híbrido por sobre el modelo de elección discreta clásico. Como era de esperar, mediante el método secuencial se subestima la desviación estándar de sus parámetros, obteniéndose test-t mayores para ellos al usar el método simultáneo. Pero, más importante, también se produce una diferencia considerable entre la utilidad marginal de la seguridad en ambos casos, lo que puede deberse a posibles cambios de escala producto de la calibración de las ecuaciones estructurales que las generan. Finalmente, la no inclusión de variables latentes ocasiona que la constante modal deba captar los efectos omitidos, incrementando su significancia estadística.

4. APLICACIÓN A UN CASO URBANO

El transporte urbano representa un caso de análisis atractivo ya que las alternativas disponibles son variadas tanto en cantidad como en características, cubriendo las necesidades y gustos de los usuarios. El banco de datos utilizado en este caso fue la cuarta ola del *Panel de Santiago* (Yáñez et al., 2008), que considera la elección de modo de transporte en viajes al trabajo en punta mañana durante cinco días laborales. Para generar indicadores de percepción y de esta forma mejorar los resultados en modelación de las primeras olas, en la cuarta se incorporó una sección en que los encuestados debían evaluar distintas características de los modos (Rojas, 2009).

El estudio considera diez modos de transporte, puros y combinados: (1) auto-chofer, (2) auto-acompañante, (3) taxi colectivo, (4) metro, (5) bus, (6) auto-chofer/metro, (7) auto-acompañante/metro, (8) taxi colectivo/metro, (9) bus/metro, y (10) taxi colectivo/bus. Los datos utilizados para la modelación provienen de las cinco respuestas proporcionadas por cada uno de los 302 encuestados; sin embargo, se trabajó sin considerar la existencia de correlación entre las observaciones que pertenecen a un mismo individuo debido a limitaciones computacionales en la resolución del modelo híbrido. También se dispone de información acerca del estrato socio-económico, edad y nivel

educacional de los encuestados, así como sus elecciones y alternativas disponibles.

Finalmente, se midió separadamente en terreno los tiempos de viaje, espera y caminata, el costo del viaje y la cantidad de transbordos, para cada alternativa disponible.

4.1 Estructura del Modelo de Variables Latentes

Se consideró tres variables latentes: *accesibilidad/comodidad*, *confiabilidad* y *seguridad*, y se definió siete indicadores de percepción para captar sus efectos. Estos consisten en evaluaciones, con nota de 1 a 7, de los siguientes aspectos de cada modo de transporte puro: (1) seguridad en cuanto a accidentes, (2) seguridad en cuanto a robos, (3) facilidad de acceso, (4) comodidad durante el viaje, (5) disponibilidad de información adecuada, (6) posibilidad de calcular previamente el tiempo de viaje, y (7) posibilidad de calcular previamente el tiempo de espera.

Se consideró cuatro variable explicativas para las ecuaciones estructurales: el nivel educacional, segmentado en las categorías primaria, secundaria, técnica/profesional y universitaria/post-grado, la edad, el sexo y el ingreso segmentado en tres categorías (“bajo” hasta \$350.000, “medio” entre \$350.000 y \$1.000.000 y “alto” desde \$1.000.000). Las interrelaciones del modelo MIMIC fueron estudiadas mediante análisis factorial a fin de garantizar su correcto funcionamiento. En la Figura 2 se señalan los resultados obtenidos de este proceso.

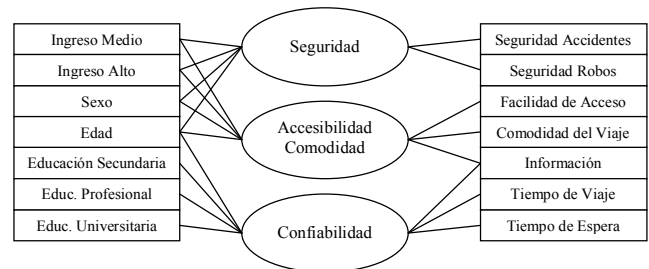


FIGURA 2: Estructura del Modelo de Variables Latentes, Caso Urbano

4.2 Estructura del Modelo de Elección Discreta

La función de utilidad representativa en este caso incluye las distintas variables de tiempo incluidas en la encuesta de preferencias reveladas, así como el número de *transbordos* y considera variaciones sistemáticas según *sexo* para el *tiempo de viaje*. La variable *costo* fue normalizada por la tasa salarial individual w_q . La forma general de la función de utilidad (10)

corresponde a la mejor especificación obtenida entre distintas formulaciones estudiadas; cabe recordar que los modos combinados no poseen variables latentes:

$$V_{iq} = \theta_{costo} \cdot Costo_{iq} / w_q + (\theta_{tvia} + \theta_{sexo} \cdot Sexo_q) \cdot Tvia_{iq} + \theta_{tesp} \cdot Tesp_{iq} + \theta_{tcam} \cdot Tcam_{iq} + \theta_{trans} \cdot Transbordos_{iq} + \beta_{conf} \cdot Confiabilidad_{iq} + \beta_{acc.com} \cdot AccesibilidadComodidad_{iq} + \beta_{seg} \cdot Seguridad_{iq} + K_i \quad (10)$$

4.3 Análisis de Resultados

El modelo híbrido fue estimado tanto en forma secuencial como simultánea en este trabajo. En la Tabla 3 se presentan los resultados de dichas estimaciones, así como el modelo equivalente sin variables latentes. Las estimaciones se acompañan con los respectivos test-t de los parámetros y también se presenta la log-verosimilitud del modelo. El detalle de la calibración del modelo MIMIC asociado se puede ver en Raveau (2009).

TABLA 3: Resultados del Modelo de Elección, Caso Urbano

Parámetro	Modelo Híbrido Secuencial	Modelo Híbrido Simultáneo	Modelo sin Variables Latentes
Costo/tasa salarial	-0,028 (-6,46)	-0,032 (-7,32)	-0,027 (-8,13)
Tiempo de viaje	-0,007 (-4,25)	-0,006 (-4,67)	-0,033 (-4,82)
Sexo	-0,001 (-2,91)	-0,001 (-3,01)	0,030 (2,98)
Tiempo de espera	-0,013 (-0,58)	-0,015 (-1,69)	-0,009 (-0,53)
Tiempo de caminata	-0,019 (-1,69)	-0,022 (-2,89)	-0,016 (-1,80)
Transbordos	-1,060 (-7,85)	-1,102 (-8,21)	-1,110 (-8,20)
Accesibilidad-comodidad	0,590 (4,23)	0,622 (3,79)	-
Confiabilidad	0,339 (2,91)	0,441 (2,70)	-
Seguridad	0,582 (2,01)	0,613 (1,87)	-
ASC Auto Chofer	0,664 (2,81)	0,733 (2,03)	1,220 (5,84)
ASC Auto-Acompañante	-0,903 (-3,83)	-0,889 (-2,12)	-0,800 (-3,64)
ASC Taxi Colectivo	-1,280 (-3,69)	-1,331 (-1,78)	-1,420 (-4,60)
ASC Metro	0,248 (1,13)	0,247 (0,81)	0,241 (1,56)
ASC Auto-Chofer/Metro	0,142 (0,41)	0,223 (0,51)	0,779 (2,65)
ASC Auto-Acomp./Metro	-0,989 (-3,08)	-0,882 (-2,22)	-0,309 (-1,28)
ASC Colectivo/Metro	-0,848 (-2,93)	-0,913 (-1,55)	-0,078 (-0,36)
ASC Bus/Metro	-0,088 (-0,37)	0,342 (1,41)	0,608 (4,59)
ASC Colectivo/Bus	-1,280 (-3,72)	-1,005 (-3,68)	-0,473 (-1,68)
Número de Observaciones	1.107	1.107	1.107
Log-verosimilitud	-55.578,85	-47.883,43	-105.567,06

Se aprecia que todos los signos obtenidos son los esperados de acuerdo a la teoría macroeconómica: las utilidades marginales de las variables latentes son positivas, mientras que los atributos de tiempo, costo y transbordos representan una des-utilidad para los individuos. En todos los casos se obtienen parámetros del tiempo de viaje negativos tanto para hombres como mujeres. Utilizando el método simultáneo, todas las variables son estadísticamente significativas al menos al 90% de confianza, mientras que en los otros modelos esto no ocurre (en especial en relación al tiempo de espera). Las constantes modales del modelo sin variables latentes resultan más significativas que en los modelos híbridos, lo cual es esperable debido a que, al poseer menos variables explicativas, éstas deben explicar (de la mejor forma posible) la información omitida, ajustándose a los patrones de elección de los individuos. Sólo la constante específica del modo bus/metro presenta diferente signo al estimar el modelo en forma secuencial; esto ocurre debido al bajo nivel de significancia de dicho parámetro.

En ambos modelos híbridos es posible concluir que los hombres son más sensibles al tiempo de viaje que las mujeres, sin embargo en el modelo sin variables latentes ocurre lo contrario. Destaca también la diferencia de magnitud de la utilidad marginal del tiempo para las mujeres entre los modelos híbridos y el modelo sin variables latentes (cerca al 500%). En la Tabla 4 se presentan los valores subjetivos del tiempo de viaje, espera y caminata, así como las valoraciones asociadas a los transbordos y a las variables latentes. Como se desconoce la escala de éstas últimas, sólo es posible compararlas entre los distintos métodos de estimación. Para todas las valoraciones presentadas, se considera una tasa salarial promedio de \$3.080 por hora trabajada; este valor se obtuvo a partir de la información recolectada en el *Panel de Santiago*.

TABLA 4: Valoraciones de Atributos, Caso Urbano

Valor Subjetivo	Modelo Híbrido Secuencial	Modelo Híbrido Simultáneo	Modelo sin Variables Latentes
Tiempo de Viaje para Hombres	861 [\$/hr]	693 [\$/hr]	261 [\$/hr]
Tiempo de Viaje para Mujeres	780 [\$/hr]	596 [\$/hr]	3.706 [\$/hr]
Tiempo de Espera	1.363 [\$/hr]	1.467 [\$/hr]	1.015 [\$/hr]
Tiempo de Caminata	2.045 [\$/hr]	2.143 [\$/hr]	1.819 [\$/hr]
Transbordos	1.926 [\$/transb.]	1.778 [\$/transb.]	2.103 [\$/transb.]
Accesibilidad/Comodidad	1.072 [\$/unidad]	1.003 [\$/unidad]	-
Confiabilidad	725 [\$/unidad]	711 [\$/unidad]	-
Valoración de la Seguridad	1.058 [\$/unidad]	989 [\$/unidad]	-

Parece evidente que el valor subjetivo del tiempo derivado a partir del modelo sin variables latentes es errado (para hombres se produce una subestimación, mientras que para mujeres se obtiene un valor extremadamente alto). Esto es una clara desventaja del modelo, que reafirma la importancia de incluir variables latentes en la modelación. Los valores del tiempo de espera y caminata, así como la valoración de transbordos correspondientes al modelo sin variables latentes, se encuentran dentro del rango razonable; sin embargo, los errores relativos con respecto al modelo híbrido simultáneo oscilan entre 15% y 30%. Por otro lado, al utilizar el modelo híbrido secuencial se sobreestima el valor del tiempo (para hombres y mujeres) en casi 30%, esto podría resultar grave al momento de evaluar los beneficios de la implementación de alguna política de transporte. Si bien las diferencias no superan el 10% en el caso de la valoración del tiempo de caminata, espera y transbordos, pueden llegar a ser igualmente graves al evaluar políticas.

Las variables latentes resultan significativas bajo ambos métodos de calibración del modelo híbrido; como es de esperar, utilizando el método secuencial se obtienen test-t mayores debido a la subestimación de las desviaciones estándar propias de la estimación en dos etapas. Un hecho interesante es que, si bien mediante el método simultáneo se obtienen parámetros más elevados, no se aprecian diferencias notorias en las magnitudes. Si bien mediante el método simultáneo se obtienen mayores parámetros, las valoraciones de cada unidad de variable latente son menores que las obtenidas mediante el enfoque secuencial.

5. CONCLUSIONES

En base al análisis realizado en situaciones de elección modal a nivel urbano e interurbano, es posible afirmar que existen diferencias importantes entre la estimación del modelo híbrido mediante los métodos secuencial y simultáneo. En primer lugar, se producen diferencias de escala en la estimación de algunos de los parámetros que acompañan a las variables latentes del modelo híbrido de elección discreta. Esto tiene implicancias no despreciables al momento de utilizar el modelo para predecir o evaluar políticas de transporte. En segundo lugar, al derivar valoraciones subjetivas a partir de las utilidades marginales estimadas, se aprecian diferencias sustanciales entre los modelos que incluyen variables latentes y el modelo tradicional. De igual manera, existen diferencias relevantes entre las valoraciones obtenidas para el modelo híbrido al calibrarlo secuencial o simultáneamente.

Un tema que al parecer no ha sido abordado en la literatura especializada es la utilización de modelos híbridos de elección discreta en modalidad predictiva y/o evaluación de políticas de transporte. El sustancial desarrollo que han tenido estos modelos en el último tiempo se ha limitado a su formulación y calibración, pero no se ha encontrado evidencia acerca de su uso práctico. La problemática surge a partir de la naturaleza intrínseca de las variables latentes; al ser desconocidas, no sólo deben ser predichas para los escenarios de evaluación, sino que varían de individuo en individuo. En este sentido, las diferencias obtenidas entre los enfoques secuencial y simultáneo de estimación pueden resultar en serias discrepancias al momento de utilizar los modelos que incluyen variables latentes.

Ante las diferencias que se obtienen mediante la calibración mediante una u otra técnica, se recomienda utilizar el método simultáneo para la resolución de modelos de elección discreta con variables latentes pues, al ocupar toda la información disponible en la resolución, se obtienen estimadores eficientes e insesgados de los parámetros. El problema es que esto ocurre a costa de una mayor complejidad desde el punto de vista algorítmico. Sin embargo, es esperable que en el corto plazo se desarrollen mejores métodos computacionales para su resolución (en términos de eficiencia y convergencia). Este aspecto representa una dirección de avance incuestionable para los modelos híbridos: desarrollar metodologías para la evaluación y resolución del problema, mediante algoritmos más rápidos, eficientes y precisos.

AGRADECIMIENTOS

Deseamos agradecer la colaboración de Denis Bolduc y Ricardo Álvarez-Daziano, Universidad de Laval, por habernos proporcionado el software experimental para estimar modelos híbridos en forma simultánea y habernos apoyado constantemente durante el proceso. También deseamos agradecer el financiamiento de FONDECYT (Proyecto 1050672) y del Instituto Milenio en Sistemas Complejos de Ingeniería (Proyecto P05-004F).

REFERENCIAS

- Ashok, K., Dillon, W. y Yuan, S. (2002) Extending discrete choice models to incorporate attitudinal and other latent variables. **Journal of Marketing Research** 39, 31-46.
- Ben-Akiva, M.E., Walker, J.L., Bernardino, A.T., Gopinath, D.A., Morikawa, T. y Polydoropoulou, A. (2002) Integration of choice and latent variable models. En H.S. Mahmassani (ed.), **In Perpetual Motion: Travel Behaviour Research Opportunities and Challenges**, 431-470. Pergamon, Amsterdam.
- Bolduc, D., Boucher, N. y Álvarez-Daziano, R. (2008) Hybrid choice modelling of new technologies for car choice in Canada. **Transportation Research Record** 2082, 63-71.
- Bollen, K.A. (1989) **Structural Equations with Latent Variables**. John Wiley and Sons, Chichester.
- Egaña, J.N. (2009) **Inclusión de Variables Latentes en Modelos de Elección Discreta para Usuarios de Buses y Trenes Interurbanos**. Tesis de Magister en Ciencias de la Ingeniería, Departamento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile.
- Gaudry, M.J.I., Jara-Díaz, S.R. y Ortúzar, J. de D. (1989) Value of time sensitivity to model specification. **Transportation Research** 23B, 151-158.
- McFadden, D. (1986) The choice theory approach to market research. **Marketing Science** 5, 275-297.
- Murphy, K.M. y Topel, R.H. (1985) Estimation and inference in two-step econometric models. **Journal of Business and Economic Statistics** 3, 370-379.
- Ortúzar, J. de D. y Hutt G.A. (1984) La influencia de elementos subjetivos en funciones desagregadas de elección discreta. **Ingeniería de Sistemas** IV, 37-54.
- Ortúzar, J. de D. y Willumsen, L.G. (2001) **Modelling Transport**. John Wiley & Sons, Chichester.
- Raveau, S. (2009) **Estimación Simultánea de Modelos de Elección Discreta con Variables Latentes**. Memoria de Título, Departamento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile.
- Rojas, M.R. (2009) **Modelos de Elección Discreta con Variables Latentes para Sistemas de Transporte Urbano**. Memoria de Título, Departamento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile.
- Vredin Johansson, M., Heldt, T. y Johansson, P. (2005) **Latent variables in a travel mode choice model: attitudinal and behavioural indicator variables**. Working Paper, Department of Economics, Uppsala University.
- Yáñez, M.F., Heydecker, B.G. y Ortúzar, J. de D. (2008) A panel data model to forecast the effect of a radical public transport innovation. **4th International Symposium on Travel Demand Management (TDM2008)**, Vienna, Austria.