



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

**PREDICCIÓN CON MODELOS
HÍBRIDOS DE ELECCIÓN DISCRETA:
EL ROL DE ATRIBUTOS TANGIBLES EN
LA DEFINICIÓN DE VARIABLES
LATENTES**

FRANCISCO JOSE BAHAMONDE BIRKE

Tesis para optar al grado de
Magister en Ciencias de la Ingeniería

Profesor Supervisor:
JUAN DE DIOS ORTUZAR SALAS

Santiago de Chile, Enero, 2011

© 2011, Francisco Bahamonde Birke



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

**PREDICCION CON MODELOS
HIBRIDOS DE ELECCION DISCRETA:
EL ROL DE ATRIBUTOS TANGIBLES EN
LA DEFINICION DE VARIABLES
LATENTES**

FRANCISCO JOSE BAHAMONDE BIRKE

Tesis presentada a la Comisión integrada por los profesores:

JUAN DE DIOS ORTUZAR SALAS

LUIS IGNACIO RIZZI CAMPANELLA

MARIA FRANCISCA YAÑEZ CASTILLO

EDUARDO ESTEBAN AGOSIN TRUMPER

Para completar las exigencias del grado de
Magister en Ciencias de la Ingeniería

Santiago de Chile, Enero, 2011

A mi familia

AGRADECIMIENTOS

Ante todo deseo agradecer a mi profesor guía, Juan de Dios Ortúzar, por el enorme apoyo prestado en estos dos años, especialmente en los momentos más difíciles. Aunque aún no lo conocía personalmente cuando decidí solicitarle su supervisión, definitivamente no me equivoqué al dirigirme a su persona; su apoyo ha ido más allá de la simple supervisión de una tesis de postgrado y ha sido un verdadero agrado poder trabajar a su lado.

También quiero agradecer a María Francisca Yáñez y Sebastián Raveau, por motivarme a iniciar este proyecto y por toda la ayuda prestada en su realización. Especialmente, por su paciencia y voluntad para responder a mis consultas, así como su disposición para participar en productivas discusiones, que sin duda elevaron considerablemente la calidad de este trabajo. Agradezco también al profesor Benjamin Heydecker, por su excelente disposición para sentarse junto a nosotros a debatir acerca de este proyecto, y a Julián Arellana por la ayuda prestada y su siempre buena voluntad para responder a mis consultas. También quiero agradecer al Instituto Científico Milenio en Sistemas Complejos de Ingeniería, por el financiamiento parcial de esta investigación, y al profesor Luis Ignacio Rizzi y a Javier Egaña por la facilitación de la base de datos de viajes interurbanos, que permitió completar mi análisis.

Finalmente, deseo agradecer a mi familia por todo el apoyo prestado a lo largo de mi vida universitaria. Si bien a la distancia, siempre estuvieron presentes, apoyándome y respaldándome en cada uno de los proyectos que emprendí. Sin la confianza, que siempre me entregaron, seguramente no estaría acá. En la misma línea, quiero agradecer a todos y cada uno de mis amigos, por la paciencia y por la compañía que me brindaron en estos años, lo que sin duda, me ayudó a compensar la lejanía de mi familia y me permitió disfrutar mi vida universitaria.

A TODOS ELLOS, MUCHAS GRACIAS

INDICE GENERAL

	Pág.
DEDICATORIA.....	ii
AGRADECIMIENTOS	iii
INDICE DE TABLAS	vi
INDICE DE FIGURAS	viii
RESUMEN.....	ix
ABSTRACT.....	x
1. INTRODUCCION	1
1.1 Objetivos.....	2
1.2 Alcances.....	2
1.3 Estructura de la Tesis.....	3
2. MARCO TEORICO	5
2.1 Modelos de Elección Discreta	5
2.1.1 Teoría de la Utilidad Aleatoria.....	5
2.1.2 El Modelo Logit Multinomial	6
2.1.3 El Modelo Logit Jerárquico.....	8
2.1.4 El Modelo Logit Mixto	9
2.2 Modelos Híbridos de Elección Discreta con Variables Latentes.....	10
2.2.1 Estimación de Modelos Híbridos de Elección Discreta con Variables Latentes.....	12
3. GENERACION DEL BANCO DE DATOS	14
3.1 Estructura del Banco de Datos Simulados.....	14
3.2 El Problema del Intercepto	20
4. ESTIMACION Y RESULTADOS	22
4.1 Análisis de Distintas Especificaciones Tendientes a Considerar Facotres Tangibles Propios de las Alternativas.....	22
4.2 Identificabilidad de los Parámetros del Modelo MIMIC.....	26

4.2.1 Efecto de la Normalización en la Estimación del Modelo MIMIC	29
4.3 Resultados y Discusión de la Estimación de los modelos MIMIC.....	33
4.4 El Problema de la Varianza y la Recuperación de Parámetros.....	37
4.4.1 Efecto sobre la Correlación	39
4.4.2 Recuperación de Parámetros	41
4.4.3 Efecto de la Omisión de la Variabilidad inducida por las Variables Latentes.....	44
4.5 Resultados de la Estimación de los Modelos Híbridos de Elección Discreta.....	46
4.6 Inclusión de la Variabilidad de las Variables Latentes.....	50
5. ANALISIS DE RESPUESTA.....	56
5.1 Escenarios Analizados	56
5.2 Indicadores de Respuesta.....	58
5.2.1 Análisis de Respuesta por Nivel de Ajuste	59
5.2.2 Análisis de Respuesta por Sesgo Predictivo	62
6. ANALISIS DE MODELOS DE ELECCION QUE CONSIDERAN INDICADORES DE PERCEPCION EN BASE A DATOS REALES.....	64
7. CONCLUSIONES Y RECOMENDACIONES.....	73
BIBLIOGRAFIA.....	77
ANEXO A – DESARROLLOS MATEMATICOS	82
A.1 De (4.16) a (4.17).....	82
A.2 De (4.21) a (4.22).....	83
ANEXO B – MODELOS ESTIMADOS	84
B.1 Modelo 4	84
B.2 Modelo 5	85

INDICE DE TABLAS

	Pág.
Tabla 3.1 – Variables del Modelo de Elección Discreta	15
Tabla 3.2 – Características Socio-Económicas de los Individuos.....	18
Tabla 3.3 – Variables Explicatorias relacionadas con las Alternativas.....	19
Tabla 4.1 – Efecto de Normalización para la Alternativa Automóvil.....	30
Tabla 4.2 – Efecto de Normalización para la Alternativa Metro	31
Tabla 4.3 – Efecto de Normalización para la Alternativa Bus.....	32
Tabla 4.4 – Estimadores del Modelo MIMIC para la Alternativa Automóvil	34
Tabla 4.5 – Estimadores del Modelo MIMIC para la Alternativa Metro.....	35
Tabla 4.6 – Estimadores del Modelo MIMIC para la Alternativa Bus	36
Tabla 4.7 – Relación entre los Estimadores del Modelo de Elección.....	42
Tabla 4.8 – Análisis Estadístico de la Recuperación de Parámetros.....	43
Tabla 4.9 – Análisis Estadístico de la Recuperación del Parámetro ϕ	44
Tabla 4.10 – Análisis Estadístico del Modelo con Variables Latentes Exactas.....	45
Tabla 4.11 – Resultados de la Estimación de los Modelos de Elección Discreta	47
Tabla 4.12 – Indicadores de Bondad de Ajuste de los Modelos	49
Tabla 4.13 – Comparación entre el Modelo 1 y el Modelo 1C.....	51
Tabla 4.14 – Comparación entre el Modelo 2 y el Modelo 2C.....	52
Tabla 4.15 – Comparación entre el Modelo 3 y el Modelo 3C.....	54
Tabla 4.16 – Indicadores de Bondad de Ajuste de los Modelos C.....	55

Tabla 5.1 – Escenarios Analizados	57
Tabla 5.2 – Log-Verosimilitud Estimada por Escenario.....	59
Tabla 5.3 – Nuevos Escenarios Analizados	61
Tabla 5.4 – Log-Verosimilitud Estimada en los Nuevos Escenarios.....	61
Tabla 5.5 – Indicador χ^2 por Escenario.....	62
Tabla 6.1 – Resultados de la Estimación del Modelo MIMIC, Egaña (2009)	67
Tabla 6.2 – Resultados de la Estimación de los Modelos de Elección Discreta	69
Tabla 6.3 – Bondad de Ajuste de los Modelos Estimados	70
Tabla 6.4 – Partición Modal Predicha e Índice χ^2	71

INDICE DE FIGURAS

	Pág.
Figura 2.1 –Modelo Híbrido de Elección Discreta con Variables Latentes.....	12
Figura 3.1 – Estructura del Modelo MIMIC	16
Figura 4.1 – Estructura del Modelo MIMIC Simplificado.....	23
Figura 6.1 – Estructura del Modelo Estimado por Egaña (2009).....	65

RESUMEN

Es un hecho, que no todas las características tangibles de las alternativas afectan directamente las decisiones a través de cambios en sus utilidades. Algunas características, afectan a las decisiones a través de cambios en la valoración de los atributos subjetivos de las alternativas y éstos son usualmente representados como variables latentes. Este tipo de atributos suelen no ser considerados y las variables latentes son únicamente explicadas por características socio-económicas de los individuos. En este trabajo, utilizando un banco de datos sintéticos, se analiza el impacto de la inclusión u omisión de este tipo de variables, así como el impacto de diferentes especificaciones de los modelos que las consideran.

Utilizando los mismos datos, se analiza también el impacto de las restricciones asociadas al modelo MIMIC, utilizadas para permitir la estimación de los parámetros. Además se analiza el efecto que tiene la inclusión de las variables latentes sobre la variabilidad de la función de utilidad, asociada a los modelos de elección discreta.

Finalmente se utiliza un banco de datos reales, para validar ciertas hipótesis previamente analizadas utilizando los datos sintéticos.

Palabras Clave: Modelos híbridos de elección discreta, Variables latentes, Atributos tangibles, Especificación de modelos, Identificabilidad

ABSTRACT

It is well-known that not all the objective characteristics of alternatives affect decisions directly (through changes in their utility functions). Some attributes can also affect individual decisions through changes in their subjective valuations of hard to measure attributes, which are usually represented as latent variables. However such attributes are not usually taken into consideration and the latent variables are typically explained only by the socioeconomic characteristics of the individuals. In this work, using a synthetic databank, we analyse the impact of including/omitting some of these variables, as well as the impact of different specifications of models that consider them.

Using the synthetic databank, we analyse the impact of constraints required by the MIMIC models to allow the identifiability of their parameters. We also analyze the effect of including the latent variables on the variability of the utility functions associated to the discrete choice model.

Finally, we use a real databank to validate some of the hypotheses previously tested with the synthetic databank.

Keywords: Hybrid discrete choice models, Latent variables, Tangible attributes, Model specification, Identifiability

1. INTRODUCCION

La incorporación de elementos subjetivos en modelos de elección discreta ha resurgido durante las últimas décadas como un tema de discusión y análisis. Es así como los denominados “modelos híbridos de elección discreta”, permiten considerar no sólo atributos tangibles de las alternativas (variables explicativas clásicas como tiempo de viaje, tiempo de espera y costo, entre otros) como en el caso de los modelos tradicionales, sino que también elementos intangibles asociados a las percepciones y actitudes de los usuarios (comodidad, confiabilidad, seguridad, entre otros) incluidos a través de variables latentes (Ashok *et al.*, 2002; Ben Akiva *et al.*, 2002).

Ahora, si bien la literatura muestra un claro avance en la estimación de modelos híbridos y existen publicaciones que prueban empíricamente la superioridad de éstos sobre los modelos tradicionales en términos de ajuste, el uso de estos modelos en modalidad de predicción es aún incipiente (Yáñez *et al.*, 2010). Adicionalmente, existe una carencia de bancos de datos centrados en atributos subjetivos e indicadores de percepción, pues tampoco existe claridad sobre el diseño de una encuesta adecuada que permita la construcción de las variables latentes.

Por otro lado, una de las principales deficiencias de los modelos híbridos reportados hasta ahora en la literatura, es la carencia de factores objetivos (medibles), propios de las alternativas, en el conjunto de variables explicativas del modelo de variables latentes. De hecho, hasta ahora la mayoría de los modelos de variables latentes sólo considera características socioeconómicas del individuo como variables explicativas. Como consecuencia, estos modelos híbridos no son más sensibles que los modelos tradicionales a la hora de evaluar políticas que impacten al sistema de transporte. En particular, no es posible medir el impacto o la valoración de los usuarios acerca de políticas como, por ejemplo, instalación de aire acondicionado en vehículos de transporte público, renovación de asientos, instalación de tableros electrónicos en

paraderos para indicar tiempos de espera, desarrollo de un servicio en línea para planificar viajes, entre otros.

1.1 Objetivos

El objetivo principal de este trabajo consiste en analizar de forma empírica, en base a datos simulados, el potencial predictivo de modelos híbridos de elección discreta que incorporen variables latentes.

En particular, se pretende determinar la relevancia de la inclusión de factores objetivos relativos a las alternativas, así como la importancia que tiene la forma de considerarlos en el modelo. De esta forma se analizan diferentes especificaciones y supuestos, tendientes a permitir la consideración de estos factores, con el fin de cuantificar la importancia de utilizar especificaciones adecuadas.

Dado que tradicionalmente los modelos híbridos de elección discreta no han considerado la inclusión de factores objetivos relativos a las alternativas, no existen en este momento bases de datos confiables con las que se pueda trabajar. Este estudio pretende sentar las bases para el trabajo con este tipo de variables y de esta forma, proveer información que permita, a futuro, desarrollar encuestas y bases de datos orientadas al trabajo con este tipo de modelos.

Finalmente, se intentará, dentro de lo posible (dada la escasez de información relativa a los atributos tangibles de las alternativas), corroborar ciertas conclusiones obtenidas con los datos sintéticos mediante datos reales.

1.2 Alcances

Dado que se trabaja con una base de datos generada en forma sintética, la cual nos permite examinar nuestro objeto de estudio en un ambiente libre de efectos indeseados,

es imperioso suponer que los individuos se comportan de una determinada manera. Por tanto la hipótesis principal de este trabajo es que se cumplen los supuestos de los modelos híbridos de elección discreta con variables latentes. Se considera la existencia de variables explicativas relacionadas con las alternativas que afectan directamente el proceso de elección, así como otras que afectan a las variables latentes.

Los supuestos realizados están enfocados en reproducir de la forma más realista posible el proceso de elección y así poder analizar el efecto de considerar especificaciones incorrectas al estimar modelos econométricos.

1.3 Estructura de la Tesis

Esta tesis está organizada en ocho capítulos. El segundo de éstos corresponde a una revisión teórica y metodológica de los modelos con los que se trabajará. De esta forma, se describen los principales modelos de elección discreta y modelos híbridos de elección discreta, así como sus principales supuestos y las técnicas que permiten su estimación.

El capítulo 3 corresponde a una descripción de la base de datos sintética con la que se trabajará. También en este capítulo, se discute acerca de la estructura de los modelos híbridos y acerca de ciertas restricciones que permiten su estimación, las cuales deben ser tomadas en consideración al generar la base de datos.

En el capítulo 4 se presentan, en primer lugar, las diferentes especificaciones que se analizarán en este trabajo, presentándose posteriormente los resultados de los modelos estimados de acuerdo a estas especificaciones. También en el capítulo 4, se discuten algunos tópicos importantes referidos a la estimación de los modelos.

Un análisis de respuesta de los diferentes modelos estimados, se presenta en el capítulo 5. De esta manera, se analizan los resultados obtenidos por los diferentes modelos al ser utilizados en escenarios distintos a aquél para el que fueron calibrados.

El capítulo 6 presenta un análisis de modelos de elección discreta que incorporan directamente los indicadores de percepción con datos reales. Esto se debe a que los resultados obtenidos en los capítulos 4 y 5 no son concluyentes respecto a este tipo de modelos.

Finalmente en el capítulo 7 se presentan las conclusiones y alcances de este trabajo, así como las recomendaciones orientadas al trabajo futuro con este tipo de modelos.

2. MARCO TEORICO

2.1 Modelos de Elección Discreta

Los modelos de elección discreta buscan reproducir el comportamiento de una población dada al ser enfrentada a una serie de alternativas. Al contrario de los modelos estadísticos tradicionales, que se enfocan en determinar niveles óptimos en un espacio continuo, los modelos de elección discreta ven limitado el universo de posibilidades de elección a un conjunto de alternativas definidas *a priori*. Por tanto, en lugar de recurrir a un proceso de optimización clásico, estos modelos funcionan basados en la comparación entre las diversas alternativas existentes.

Para realizar esta comparación, en concordancia con la teoría microeconómica, se plantea que las diversas alternativas reportan un cierto nivel de utilidad a los individuos, y que éstos eligen aquella alternativa que les reporte un mayor beneficio. En el desarrollo a continuación emplearemos la notación de Ortúzar y Willumsen (2001).

2.1.1 Teoría de la Utilidad Aleatoria

La Teoría de la Utilidad Aleatoria plantea que todos los individuos q , que componen una población Q , actúan en forma determinística y racional maximizando su utilidad neta, sujetos a sus restricciones ambientales. Se supone además, que los individuos cuentan con información completa, por lo que se estaría ante la presencia de un “*homo economicus*”. De esta forma, el individuo optará por la alternativa i , entre un conjunto $A(q)$ de alternativas, si y sólo si $U_{i,q} \geq U_{j,q} \forall j \in A(q)$.

Por otro lado, la teoría asume que el modelador no cuenta con información completa relativa a los individuos o a la valoración por parte de éstos de los distintos atributos de las alternativas. Por esto el modelador sólo puede asignar una utilidad representativa $V_{i,q}$

y se supone que toda la variabilidad es explicada por elementos aleatorios. De esta forma se tiene que:

$$U_{i,q} = V_{i,q} + \varepsilon_{i,q} \quad (2.1)$$

donde $\varepsilon_{i,q}$ representa el componente aleatorio. Dado que, tal como se mencionó anteriormente, un individuo optará por una alternativa i por sobre una alternativa j si y sólo si $U_{i,q} > U_{j,q}$. Luego se tiene que:

$$V_{i,q} + \varepsilon_{i,q} \geq V_{j,q} + \varepsilon_{j,q} \quad (2.2)$$

o reescribiendo lo anterior, agrupando los componentes determinísticos:

$$V_{i,q} - V_{j,q} \geq \varepsilon_{j,q} - \varepsilon_{i,q} \quad (2.3)$$

Dada la naturaleza aleatoria de la expresión $\varepsilon_{j,q} - \varepsilon_{i,q}$ no se puede establecer con certeza si el individuo optará por la alternativa en cuestión y el modelador sólo puede plantear una probabilidad de elección, la cual está dada por:

$$P_{i,q} = \Pr\{V_{i,q} - V_{j,q} \geq \varepsilon_{j,q} - \varepsilon_{i,q}, \forall j \in A(q)\} \quad (2.4)$$

Diferentes distribuciones y supuestos respecto a los errores darán como resultado diversos modelos de elección discreta (Ortúzar y Williams, 2001; Train, 2003).

2.1.2 El Modelo Logit Multinomial (MNL)

Si se asume que los errores $\varepsilon_{i,q}$ son Gumbel independiente e idénticamente distribuidos con media cero y desviación estándar σ , se tiene que la probabilidad de elección de una alternativa i puede ser expresada como (Domencich y Mc Fadden, 1975):

$$P_{i,q} = \frac{e^{\lambda \cdot V_{i,q}}}{\sum_{j \in A(q)} e^{\lambda \cdot V_{j,q}}} \quad (2.5)$$

en que λ corresponde al parámetro de escala de la distribución (y por transposición del modelo MLN), que está asociado a la varianza de los errores, a través de la siguiente ecuación:

$$\lambda = \frac{\pi}{\sigma\sqrt{6}} \quad (2.6)$$

En la inmensa mayoría de los casos se consideran funciones lineales, como (2.7) para describir a la utilidad representativa $V_{i,q}$, de tal forma que los distintos atributos $X_{i,k,q}$, donde k representa al atributo en cuestión, son ponderados por parámetros $\theta_{i,k}$, que deben estimarse.

$$V_{i,q} = \sum_k \theta_{k,i} \cdot X_{k,i,q} \quad (2.7)$$

Esta estructura presenta como grandes ventajas su simplicidad y el hecho de que los parámetros estimados representan la utilidad representativa marginal de los atributos a los cuales se encuentran asociados.

Dado lo anterior, el factor de escala λ es inestimable, ya que en el contexto del MNL multiplica a todos los parámetros $\theta_{i,k}$. Es por tanto que se acostumbra a fijar este factor de escala en uno, con el fin de poder estimar los coeficientes $\theta_{i,k}$; así, estos se obtendrán desfasados por el valor real de λ , lo que en la mayoría de los casos prácticos no provocará inconveniente alguno.

Una de las principales características del MNL es su propiedad de independencia de alternativas irrelevantes; bajo esta, la razón entre las probabilidades de elegir dos alternativas cualesquiera es independiente de la probabilidad de optar por una tercera:

$$\frac{P_{i,q}}{P_{k,q}} = \frac{e^{\lambda \cdot V_{i,q}}}{\sum_{j \in A(q)} e^{\lambda \cdot V_{j,q}}} \bigg/ \frac{e^{\lambda \cdot V_{k,q}}}{\sum_{j \in A(q)} e^{\lambda \cdot V_{j,q}}} = \frac{e^{\lambda \cdot V_{i,q}}}{e^{\lambda \cdot V_{k,q}}} \quad (2.8)$$

Esta característica proviene de las suposiciones realizadas respecto a la estructura de los términos de error (independientes e idénticamente distribuidos) y es una de las principales limitaciones del MNL puesto que no permite trabajar con alternativas correlacionadas. El modelo tampoco permite considerar términos de error heterocedásticos y como se supone que todos los parámetros son fijos, tampoco acepta variaciones en los gustos.

2.1.3 El Modelo Logit Jerárquico (HL)

El modelo HL (Williams, 1977; Daly y Zachary, 1978) permite manejar, en forma simplificada, la existencia de alternativas correlacionadas. Básicamente postula agrupar las alternativas correlacionadas en nidos, de tal forma que los términos de error asociados a las alternativas dentro de un nido sean independientes e idénticamente distribuidos, cumpliendo con los principios del modelo MNL.

$$P_{(i/N),q} = \frac{e^{\Delta \cdot V_{i,q}}}{\sum_{j \in N} e^{\Delta \cdot V_{j,q}}}, \forall i \in N(q) \quad (2.9)$$

De la misma forma, el HL supone que los términos de error asociados a las alternativas no anidadas, así como a la probabilidad de optar por un nido (que se basa en su utilidad representativa) distribuyen igualmente independientes e idénticamente distribuidos. De esta forma, - suponiendo la existencia de un nido N - se tiene que:

$$P_{N,q} = \frac{e^{\lambda \cdot V_{N,q}}}{\sum_{\substack{j \in A(q) \\ j \notin N}} e^{\lambda \cdot V_{j,q}} + e^{\lambda \cdot V_{N,q}}} \quad (2.10)$$

donde la utilidad representativa del nido $V_{N,q}$ está dada por:

$$V_{N,q} = \phi \cdot EMU \quad (2.11)$$

en que ϕ es un parámetro a calibrar y EMU corresponde al valor esperado de la utilidad máxima de todas las alternativas del nido, el cual está definido como:

$$EMU = \ln \sum_{i \in N} e^{\frac{V_{i,q}}{\phi}} \quad (2.12)$$

Finalmente, la probabilidad de elegir una alternativa i , perteneciente al nido N quedará definida de la siguiente manera:

$$P_{i,q} = P_{N,q} \cdot P_{(i/N),q}, \forall i \in N(q) \quad (2.13)$$

El parámetro ϕ es de vital importancia en este modelo, ya que permite captar la correlación existente entre las alternativas de un mismo nido. De hecho, está relacionado con el coeficiente de correlación a través de la siguiente ecuación (Carrasco, 2001):

$$\phi = \sqrt{1 - \rho} \quad (2.14)$$

Si bien, el HL permite solucionar el problema de correlación entre alternativas anidadas, no permite tratar la correlación cruzada entre alternativas en nidos diferentes. Además, al igual que el MNL, no permite tratar estructuras de error heterocedásticas ni introducir variaciones de gustos individuales (para más detalles, ver Carrasco y Ortúzar, 2002).

2.1.4 El Modelo Logit Mixto (ML)

El modelo ML (Cardell y Dunbar, 1980; Ben-Akiva y Bolduc, 1996; McFadden y Train, 2000) permite subsanar las deficiencias que presentan los modelos MNL y HL. Esto ya que permite introducir cualquier estructura de error asociada a las alternativas (incluyendo cualquier tipo de distribución o correlación).

Básicamente el ML consiste en integrar las probabilidades obtenidas por un MNL (e incluso estructuras más generales) sobre la densidad de probabilidad asociada a la

estructura de error introducida. De esta forma la probabilidad de optar por una alternativa i , se estima de la siguiente forma:

$$P_{i,q} = \int L_{i,q}(\beta) \cdot f(\beta|\theta) \cdot d\beta \quad (2.15)$$

donde $L_{i,q}$ corresponde a la probabilidad proveniente del MNL y $f(\beta|\theta)$ es la “función de mezcla” o distribución de los parámetros del modelo, que son variables aleatorias. Notar que esta estructura permite naturalmente manejar variaciones de gustos de los individuos (Train, 2003).

La integral planteada en (2.15) es irresoluble matemáticamente, por lo que se debe recurrir a métodos numéricos, como la máxima verosimilitud simulada para estimar los parámetros asociados al modelo.

2.2 Modelos Híbridos de Elección Discreta con Variables Latentes

No todas las variables que afectan las decisiones de los individuos son cuantificables o medibles. Al contrario, muchas decisiones son motivadas por apreciaciones subjetivas de los individuos sobre atributos absolutamente intangibles. Evidentemente, modelos como los presentados en la sección anterior no son capaces de incorporar adecuadamente el efecto de variables intangibles, por lo que en las últimas décadas se han reportado diferentes modelos tendientes a captar simultáneamente los efectos de las variables tangibles e intangibles (Green, 1984; Keane, 1997).

La literatura reciente reporta un amplio consenso en torno a la utilización de modelos híbridos de elección discreta con variables latentes, como la forma más adecuada de considerar en forma simultánea a variables tangibles e intangibles. Ashok *et al.* (2002), Ben-Akiva *et al.* (2002), Vredin-Johansson *et al.* (2005); Yáñez *et al.* (2010), entre otros, han provisto evidencia empírica acerca de las ventajas de esta especificación.

Estos modelos contemplan la inclusión de un modelo MIMIC (Multiple Indicators, Multiple Causes), que plantea la existencia de variables latentes $\eta_{l,i,q}$ (Bollen, 1989); éstas son explicadas por características tangibles de los usuarios y de las alternativas $s_{i,q,r}$, a través de ecuaciones estructurales como (2.16); de la misma forma, las variables latentes dan lugar a percepciones individuales, cuantificadas por medio de indicadores de percepción $y_{i,p,q}$, a través de ecuaciones de medición como (2.17).

$$\eta_{l,i,q} = \sum_r \alpha_{l,r,i} \cdot s_{r,i,q} + v_{l,i,q} \quad (2.16)$$

$$y_{p,i,q} = \sum_l \gamma_{l,p,i} \cdot \eta_{l,i,q} + \zeta_{p,i,q} \quad (2.17)$$

en donde el subíndice l representa una variable latente, el subíndice r , una variable explicatoria y el subíndice p , un indicador de percepción. Los subíndices q e i , representan al individuo y a la alternativa, respectivamente; $\alpha_{l,r,i}$ y $\gamma_{l,p,i}$ son parámetros a estimar, mientras que $v_{l,i,q}$ y $\zeta_{p,i,q}$ son componentes de error de media cero y una determinada matriz de covarianza. Dado que las variables latentes $\eta_{l,i,q}$ son desconocidas, se debe utilizar ambos conjuntos de ecuaciones simultáneamente para estimar los parámetros en forma insesgada.

Con el fin de poder considerar las variables latentes en el modelo de elección discreta, la utilidad representativa debe ser redefinida. Notar que al considerar las variables latentes a través de la utilidad representativa, no se altera la estructura de los modelos de elección discreta, propiamente tal. Así, la utilidad representativa se re-define como:

$$V_{i,q} = \sum_k \theta_{k,i} \cdot X_{k,i,q} + \sum_l \beta_{l,i} \cdot \eta_{l,i,q} \quad (2.18)$$

La Figura 2.1 presenta una esquematización de los modelos híbridos de elección discreta que facilita su comprensión.

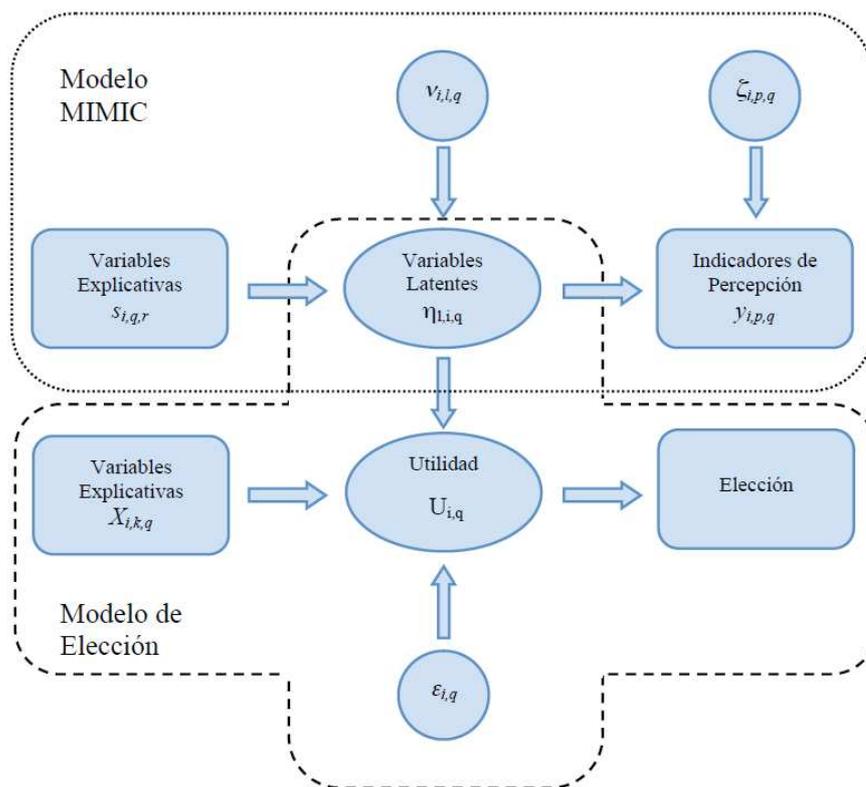


Figura 2.1 –Modelo Híbrido de Elección Discreta con Variables Latentes

2.2.1 Estimación de Modelos Híbridos de Elección Discreta con Variables Latentes

Los modelos híbridos de elección discreta pueden ser estimados tanto en forma simultánea, como en forma secuencial. El primer enfoque consiste básicamente en estimar en conjunto los modelos de elección discreta y de variables latentes (MIMIC). Si bien, este es en principio el enfoque más adecuado para estimar este tipo de modelos, el estado actual de la técnica no permite la flexibilidad requerida para hacerlo

correctamente en el caso general. Por otra parte, se ha demostrado que el enfoque secuencial permite obtener estimadores insesgados (Ben-Akiva *et al.*, 2002; Raveau *et al.*, 2010a). Por esto, sumado a las menores dificultades que presenta en su aplicación, el enfoque secuencial ha prevalecido en la literatura reciente (Ashok *et al.*, 2002; Vredin-Johansson *et al.*, 2005; Yáñez *et al.*, 2010).

El método secuencial consiste en abordar el problema en dos etapas, estimándose en primer lugar el modelo de variables latentes (MIMIC), para luego dar paso a la estimación del modelo de elección discreta, obviándose las interacciones entre éstos.

La estimación del modelo MIMIC se puede realizar mediante mínimos cuadrados en dos etapas (Pagan, 1986) o maximizando la verosimilitud (Goldberger, 1972). Bollen *et al.* (2007), demostraron la validez de ambos enfoques, ya que ambos garantizan estimadores eficientes e insesgados. Para los propósitos de este trabajo, se utilizó el método de maximización de verosimilitud.

Una vez estimado el modelo MIMIC se pueden calcular los valores esperados de las variables latentes, incorporándose éstos directamente en el modelo de elección; este último se resuelve tal como fue planteado en la sección anterior (dependiendo del modelo, se recurre a la maximización de la verosimilitud o a técnicas numéricas, como la máxima verosimilitud simulada, entre otras).

Para los propósitos de este trabajo, además del enfoque secuencial tradicional, se consideró la utilización de un modelo ML, que permite considerar no sólo el valor esperado de las variables latentes, sino que además su variabilidad (que proviene del modelo de variables latentes), lo que representa una reproducción más adecuada de la realidad.

3. GENERACION DEL BANCO DE DATOS

3.1 Estructura del Banco de Datos Simulados

Con el fin de examinar nuestro objeto de estudio en un ambiente libre de efectos indeseados, se trabajó con un banco de datos generados en forma sintética, lo que implicó la obligatoriedad de suponer que los individuos se comportaban de una determinada manera.

La idea era replicar de la forma más realista posible el comportamiento de los individuos. Se consideró, como hipótesis fundamental, que las decisiones de los individuos respondían a un modelo híbrido de elección discreta, cuyo sub-modelo de variables latentes (MIMIC) incluía factores objetivos propios de las alternativas.

De esta forma, se supuso que ciertas características de las alternativas afectaban directamente la decisión y, por tanto, fueron incluidas en el modelo de elección, mientras que otros atributos de éstas formaban parte del modelo de variables latentes, afectando de esta forma la elección de manera indirecta.

Siguiendo la metodología planteada por Williams y Ortúzar (1982), se generó una población de 35.000 individuos, los cuales se comportaban de manera compensatoria de acuerdo a (3.1).

$$U_{i,q} = \sum_k \theta_{k,i} \cdot X_{k,i,q} + \sum_l \beta_{l,i} \cdot \eta_{l,i,q} + \varepsilon_{i,q} \quad (3.1)$$

$$\sum_k \theta_{k,i} \cdot X_{k,i,q} = \theta_i + \theta_{Costo,i} \cdot Costo_{i,q} + \theta_{Tiempo,i} \cdot Tiempo_{i,q}$$

$$\sum_l \beta_{l,i} \cdot \eta_{l,i,q} = \beta_{Sec,i} \cdot Seguridad_{i,q} + \beta_{Com,i} \cdot Comodidad_{i,q} + \beta_{Conf,i} \cdot Confiabilidad_{i,q}$$

Se consideró la existencia de tres alternativas i , las cuales, para facilitar la comprensión, fueron identificadas como automóvil, metro y bus. $Costo_{i,q}$ representa el costo de la alternativa i para el individuo q , mientras que $Tiempo_{i,q}$ está asociado al tiempo de viaje. $Seguridad_{i,q}$, $Comodidad_{i,q}$ y $Confiabilidad_{i,q}$ son las respectivas variables latentes. Los errores $\varepsilon_{i,q}$ distribuyen Gumbel, de moda $-\gamma$ (constante de Euler-Mascheroni) y escala 1 (y por tanto de media 0). Las alternativas 2 y 3 se supusieron correlacionadas con un factor de correlación de 0,3.

Los costos y tiempos de viaje fueron construidos en base a muestras aleatorias tomadas de distribuciones Normal truncadas cuyo dominio fue definido en forma arbitraria (Tabla 3.1). Los términos de error consistentes con la estructura jerárquica planteada fueron generados siguiendo la metodología planteada por Carrasco (2001).

Tabla 3.1 – Variables del Modelo de Elección Discreta

Variable	Media	Desviación Estándar	Rango
Costo ₁	3,0	2,0	[0,0 ; 6,0]
Costo ₂	2,0	0,5	[0,0 ; 4,0]
Costo ₃	1,0	0,5	[0,0 ; 2,0]
Tiempo ₁	1,5	2,0	[0,0 ; 4,0]
Tiempo ₂	2,0	1,0	[0,0 ; 4,0]
Tiempo ₃	3,0	2,0	[0,0 ; 6,0]

Las variables latentes fueron generadas de acuerdo al modelo MIMIC, cuya estructura se presenta en la Figura 3.1. A la izquierda de la figura se encuentran las variables explicativas exógenas del modelo, mientras que a la derecha se presentan los

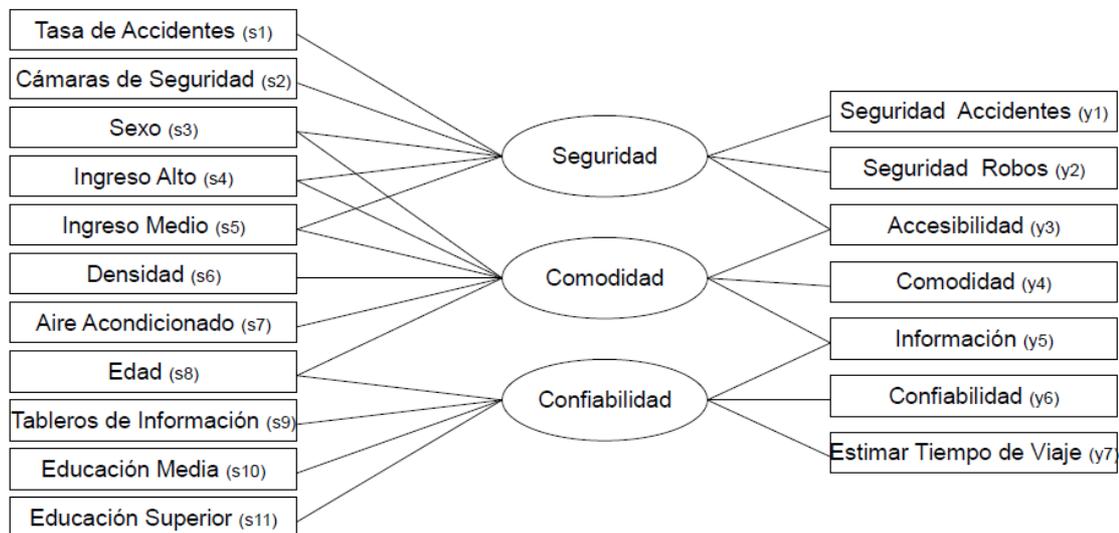


Figura 3.1 – Estructura del Modelo MIMIC

indicadores de percepción, de carácter endógeno. De las primeras, la variable *Sexo* corresponde a un indicador dicotómico, que toma el valor 1 para individuos de sexo femenino, mientras que el ingreso es descrito por medio de dos variables binarias, que toman valores positivos para caracterizar ingreso alto e ingreso medio, respectivamente. La edad de los individuos es representada en forma continua y su nivel educacional es caracterizado por medio de dos variables dicotómicas, las cuales indican un nivel educacional medio o superior.

Además de las variables socioeconómicas anteriores, toman parte del modelo de variables latentes características propias de las alternativas, como la *Tasa de Accidentes*, representada por una variable continua y el *Aire Acondicionado*, que es representado por una variable dicotómica, que toma el valor 1 cuando existe algún sistema de aire acondicionado en el vehículo. Además, se supone que las alternativas de transporte público (metro y bus) son afectadas por variables como *Densidad*, que es una variable continua que indica la densidad de pasajeros a bordo del vehículo, *Cámaras de Seguridad*, que es una variable dicotómica que toma el valor 1 ante la existencia de estos

dispositivos y *Tableros de Información*, que es una variable binaria que toma valores positivos ante la existencia de paneles electrónicos que indican los tiempos de espera en paraderos.

Lo anterior da lugar a las siguientes ecuaciones estructurales (3.2) y de medición (3.3); los términos s_i e y_j están referidos a las variables indicadas en la Figura 3.1.

$$\begin{aligned} \text{Seguridad}_{i,q} = & \alpha_{\text{Sec},s1,i} \cdot s1_{i,q} + \alpha_{\text{Sec},s2,i} \cdot s2_{i,q} + \alpha_{\text{Sec},s3,i} \cdot s3_{i,q} + \alpha_{\text{Sec},s4,i} \cdot s4_{i,q} \\ & + \alpha_{\text{Sec},s5,i} \cdot s5_{i,q} + v_{\text{Sec},i,q} \end{aligned} \quad (3.2)$$

$$\begin{aligned} \text{Comodidad}_{i,q} = & \alpha_{\text{Com},s3,i} \cdot s3_{i,q} + \alpha_{\text{Com},s4,i} \cdot s4_{i,q} + \alpha_{\text{Com},s5,i} \cdot s5_{i,q} + \alpha_{\text{Com},s6,i} \cdot s6_{i,q} \\ & + \alpha_{\text{Com},s7,i} \cdot s7_{i,q} + \alpha_{\text{Com},s8,i} \cdot s8_{i,q} + v_{\text{Com},i,q} \end{aligned}$$

$$\begin{aligned} \text{Confiabilidad}_{i,q} = & \alpha_{\text{Conf},s8,i} \cdot s8_{i,q} + \alpha_{\text{Conf},s9,i} \cdot s9_{i,q} + \alpha_{\text{Conf},s10,i} \cdot s10_{i,q} + \alpha_{\text{Conf},s11,i} \cdot s11_{i,q} \\ & + v_{\text{Conf},i,q} \end{aligned}$$

$$y1_{i,q} = \gamma_{\text{Sec},y1,i} \cdot \text{Seguridad} + \zeta_{y1,i,q} \quad (3.3)$$

$$y2_{i,q} = \gamma_{\text{Sec},y2,i} \cdot \text{Seguridad} + \zeta_{y2,i,q}$$

$$y3_{i,q} = \gamma_{\text{Sec},y3,i} \cdot \text{Seguridad} + \gamma_{\text{Com},y3,i} \cdot \text{Comodidad} + \zeta_{y3,i,q}$$

$$y4_{i,q} = \gamma_{\text{Com},y4,i} \cdot \text{Comodidad} + \zeta_{y4,i,q}$$

$$y5_{i,q} = \gamma_{\text{Com},y5,i} \cdot \text{Comodidad} + \gamma_{\text{Conf},y5,i} \cdot \text{Confiabilidad} + \zeta_{y5,i,q}$$

$$y6_{i,q} = \gamma_{\text{Conf},y6,i} \cdot \text{Confiabilidad} + \zeta_{y6,i,q}$$

$$y7_{i,q} = \gamma_{\text{Conf},y7,i} \cdot \text{Confiabilidad} + \zeta_{y7,i,q}$$

Las variables $v_{i,q}$ y $\zeta_{p,i,q}$ son términos de error de media cero, independientes entre sí y con distribución Normal.

Con el fin de garantizar un balance entre la variabilidad proveniente de los componentes determinísticos y estocásticos de las variables latentes, la desviación estándar de los términos de error de las ecuaciones estructurales $v_{i,q}$ asociados a las variables *Seguridad (Sec)*, *Comodidad (Com)* y *Confiabilidad (Conf)*, tomó los valores 1, 0,7 y 0,5 respectivamente. La desviación estándar de los términos de error $\zeta_{p,i,q}$ asociados a las ecuaciones de medición se consideró igual a uno.

Por otro lado, las características socioeconómicas de los individuos fueron generadas tal como se indica en la Tabla 3.2 y los factores objetivos propios de las alternativas que intervienen en el modelo de variables latentes se generaron de acuerdo a la Tabla 3.3.

Tabla 3.2 – Características Socio-Económicas de los Individuos

Atributo	Distribución	Media	Desviación Estándar	Rango
Sexo	Bernoulli	0,5	0,5	[0,0 ; 1,0]
Ingreso Alto	Bernoulli	0,2	0,4	[0,0 ; 1,0]
Ingreso Medio	Bernoulli	0,45	0,5	[0,0 ; 1,0]
Edad	Normal Truncada	18,0	32,0	[10,0 ; 100,0]
Educación Superior	Bernoulli	0,15	0,4	[0,0 ; 1,0]
Educación Media	Bernoulli	0,75	0,4	[0,0 ; 1,0]

Tabla 3.3 – Variables Explicativas Relacionadas con las Alternativas

Atributo	Distribución	Media	Desviación Estándar	Rango
Tase de Accidentes Auto	Normal Truncada	2,0	2,0	[0,0 ; 4,0]
Tasa de Accidentes Metro	Normal Truncada	0,0	0,2	[0,0 ; 0,5]
Tasa de Accidentes Bus	Normal Truncada	1,0	1,0	[0,0 ; 3,0]
Cámaras de Seguridad Metro	Bernoulli	0,8	0,4	[0,0 ; 1,0]
Cámaras de Seguridad Bus	Bernoulli	0,3	0,5	[0,0 ; 1,0]
Densidad Metro	Normal Truncada	4,0	3,0	[0,0 ; 7,0]
Densidad Bus	Normal Truncada	3,0	2,0	[0,0 ; 5,0]
Aire Acondicionado Auto	Bernoulli	0,6	0,5	[0,0 ; 1,0]
Aire Acondicionado Metro	Bernoulli	0,3	0,5	[0,0 ; 1,0]
Aire Acondicionado Bus	Bernoulli	0,2	0,4	[0,0 ; 1,0]
Tableros de Información Metro	Bernoulli	0,9	0,3	[0,0 ; 1,0]
Tableros de Información Bus	Bernoulli	0,3	0,5	[0,0 ; 1,0]

Para simular el proceso de elección, se construyeron las variables latentes utilizando las ecuaciones estructurales definidas en (3.2). De la misma forma, se determinó la utilidad de las distintas alternativas recurriendo a las funciones de utilidad indicadas en (3.1). En consecuencia se generaron elecciones acordes a una estructura HL.

Se separó una sub-muestra de 5.000 individuos para ser utilizada posteriormente como muestra de validación. De las restantes 30.000 observaciones se obtuvieron, en forma aleatoria, seis sub-muestras de 5.000 individuos cada una. Los modelos estimados y reportados a continuación, corresponden, en todo momento, a la media de estas seis sub-muestras. De la misma forma, la desviación estándar reportada corresponde a la desviación estándar de esta media.

3.2 El Problema del Intercepto

Dada la estructura de los modelos MIMIC, existe colinealidad muy fuerte entre los interceptos de las ecuaciones estructurales y de medición. Por tanto, para permitir la identificación de los parámetros, se debe utilizar interceptos sólo en un conjunto de ecuaciones.

Evidentemente el considerar o no interceptos en las ecuaciones estructurales tendrá un efecto directo en la magnitud de las variables latentes (no así en el nivel de ajuste, el cual será el mismo), las cuales después tomarán parte del modelo de elección discreta. De esta manera, dependiendo del conjunto de ecuaciones en que se consideren los interceptos, las variables latentes tendrán la siguiente forma:

$$\eta_{l,i,q}^1 = Int_{l,i} + \sum_r \alpha_{l,r,i} \cdot s_{r,i,q} + v_{l,i,q} \quad (3.4)$$

$$\eta_{l,i,q}^2 = \sum_r \alpha_{l,r,i} \cdot s_{r,i,q} + v_{l,i,q} \quad (3.5)$$

y, por ende:

$$\eta_{l,i,q}^1 = Int_{l,i} + \eta_{l,i,q}^2 \quad (3.6)$$

en que $\eta_{l,i,q}$ denota a las variables latentes cuando se coloca interceptos en las ecuaciones estructurales y $\eta_{2,l,i,q}$ cuando los interceptos en las ecuaciones de medición; Int_l representa a los interceptos.

Así, al considerar estas variables latentes en el modelo de elección discreta se tendría la siguiente situación:

$$V_{i,q} = \sum_k \theta_{k,i} \cdot X_{k,i,q} + \sum_l \beta_{l,i} \cdot \eta_{1,l,i,q} = \sum_k \theta_{k,i} \cdot X_{k,i,q} + \sum_l \beta_{l,i} \cdot (Int_{l,i} + \eta_{2,l,i,q}) \quad (3.7)$$

$$\sum_k \theta_{k,i} \cdot X_{k,i,q} + \sum_l \beta_{l,i} \cdot \eta_{1,l,i,q} = \sum_k \theta_{k,i} \cdot X_{k,i,q} + \sum_l \beta_{l,i} \cdot \eta_{2,l,i,q} + \sum_l \beta_{l,i} \cdot Int_{l,i} \quad (3.8)$$

sin embargo, aunque $\sum_l \beta_{l,i} \cdot Int_{l,i}$ es diferente para todas las alternativas, es constante para todos los individuos, por lo que es equivalente a sumar un término fijo a cada alternativa. Por tanto, el modelo puede perfectamente ser ajustado a través de las constantes modales. De esta forma, el ajuste obtenido por el modelo híbrido de elección discreta será el mismo (predecirá de igual forma), aunque ambos modelos sean disímiles.

Un corolario interesante de lo anterior, es que las variables latentes no son directamente comparables entre distintas alternativas, ya que no se puede determinar a ciencia cierta su real valor.

Debido a lo anterior, no se consideró la existencia de interceptos en la generación de las variables latentes. No obstante, para propósitos de la estimación se consideró interceptos (ASC) en las ecuaciones de medición.

4. ESTIMACION Y RESULTADOS

4.1 Análisis de Distintas Especificaciones Tendientes a Considerar Factores Tangibles Propios de las Alternativas

El propósito principal de este trabajo es analizar el efecto que podrían tener diferentes especificaciones, tendientes a incorporar los factores propios de las alternativas, sobre la capacidad de predicción de los modelos; para esto se plantean a continuación diversas especificaciones, las cuales serán contrastadas de acuerdo a su capacidad predictiva.

a) Especificación Correcta: Corresponde a la especificación utilizada en la generación del banco de datos. Esta estructura replica el proceso de elección supuesto de los pseudo-individuos simulados. Modelos estimados de acuerdo a esta especificación serán utilizados como información de control.

b) Modelo Híbrido de Elección Discreta Clásico: Corresponde a la estructura más reportada en la literatura (Ben-Akiva *et al.*, 2002; Ashok *et al.*, 2002, Vredin-Johansson *et al.*, 2005; Yáñez *et al.*, 2010). Tradicionalmente los modelos de variables latentes no han considerado las características de las alternativas relacionadas con las variables latentes. Normalmente estos factores son omitidos y sólo se toma en consideración aquellas variables que afectan en forma directa la decisión, tales como el costo y el tiempo de viaje. Por tanto, interesa comparar el nivel de ajuste obtenido por este modelo clásico, con el ajuste obtenido por un modelo estimado en forma adecuada (esto es, replicando la forma en la que los pseudo-individuos realmente generaron sus elecciones). En el modelo clásico se tendría que:

$$U_{i,q} = \sum_k \theta_{k,i} \cdot X_{k,i,q} + \sum_l \beta_{l,i} \cdot \eta_{l,i,q} + \varepsilon_{i,q} \quad (4.1)$$

$$\sum_k \theta_{k,i} \cdot X_{k,i,q} = \theta_i + \theta_{Costo,i} \cdot Costo_{i,q} + \theta_{Tiempo,i} \cdot Tiempo_{i,q}$$

$$\sum_l \beta_{l,i} \cdot \eta_{l,i,q} = \beta_{Sec,i} \cdot Seguridad_{i,q} + \beta_{Com,i} \cdot Comodidad_{i,q} + \beta_{Conf,i} \cdot Confiabilidad_{i,q}$$

que parece ser exactamente la especificación utilizada en la generación de los datos. No obstante, la diferencia radica en el modelo MIMIC, que es una versión simplificada del modelo utilizado en la generación de los datos. Este modelo MIMIC sólo considera variables socioeconómicas como variables explicativas exógenas, según se muestra en la Figura 4.1.

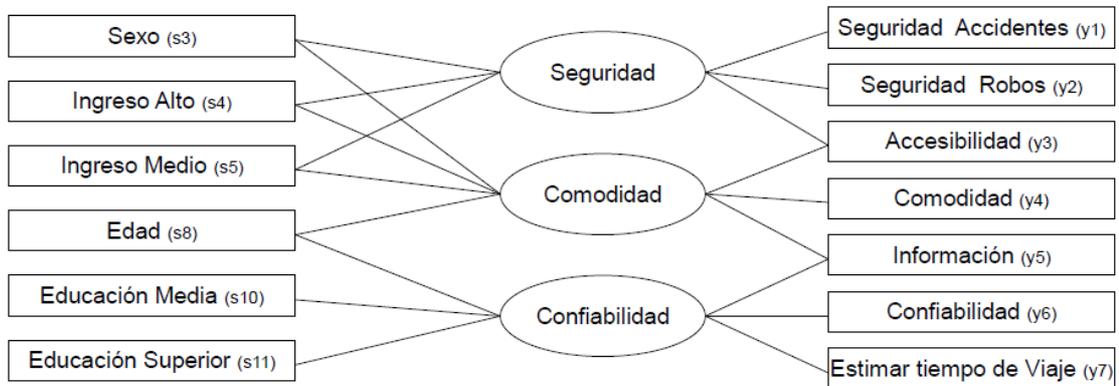


Figura 4.1 – Estructura del Modelo MIMIC Simplificado

Por tanto, el modelo de variables latentes puede ser caracterizado a través de las siguientes ecuaciones estructurales, notar que las ecuaciones de medición son las mismas que se planteó en (3.3).

$$Seguridad_{i,q} = \alpha_{Sec,s3,i} \cdot s3_{i,q} + \alpha_{Sec,s4,i} \cdot s4_{i,q} + \alpha_{Sec,s5,i} \cdot s5_{i,q} + v_{Sec,i,q} \quad (4.2)$$

$$Comodidad_{i,q} = \alpha_{Com,s3,i} \cdot s3_{i,q} + \alpha_{Com,s4,i} \cdot s4_{i,q} + \alpha_{Com,s5,i} \cdot s5_{i,q} + \alpha_{Com,s8,i} \cdot s8_{i,q} + v_{Com,i,q}$$

$$Confiabilidad_{i,q} = \alpha_{Conf,s8,i} \cdot s8_{i,q} + \alpha_{Conf,s10,i} \cdot s10_{i,q} + \alpha_{Conf,s11,i} \cdot s11_{i,q} + v_{Conf,i,q}$$

c) *Modelo Híbrido de Elección Discreta Clásico Mejorado*: Es obvio que un modelo estimado en la forma descrita anteriormente debiera obtener peores resultados que un modelo estimado en forma correcta, ya que considera una menor cantidad de información. En consecuencia, es importante comparar el rendimiento de un modelo que sí considere los factores objetivos propios de las alternativas.

Por tanto, consideraremos un modelo que incluya estas variables directamente en el modelo de elección. De esta forma, las variables latentes serán las mismas que las generadas en el modelo descrito anteriormente, pero el modelo de elección considerará todos los atributos propios de las alternativas. Por tanto, el modelo de elección quedará descrito por las siguientes ecuaciones:

$$U_{i,q} = \sum_k \theta_{k,i} \cdot X_{k,i,q} + \sum_l \beta_{l,i} \cdot \eta_{l,i,q} + \sum_r \delta_{r,i} \cdot s_{r,i,q} + \varepsilon_{i,q} \quad (4.3)$$

$$\sum_k \theta_{k,i} \cdot X_{k,i,q} = \theta_i + \theta_{Costo,i} \cdot Costo_{i,q} + \theta_{Tiempo,i} \cdot Tiempo_{i,q}$$

$$\sum_l \beta_{l,i} \cdot \eta_{l,i,q} = \beta_{Sec,i} \cdot Seguridad_{i,q} + \beta_{Com,i} \cdot Comodidad_{i,q} + \beta_{Conf,i} \cdot Confiabilidad_{i,q}$$

$$\sum_r \delta_{r,i} \cdot s_{r,i,q} = \delta_{s1,i} \cdot s1_{i,q} + \delta_{s2,i} \cdot s2_{i,q} + \delta_{s6,i} \cdot s6_{i,q} + \delta_{s7,i} \cdot s7_{i,q} + \delta_{s9,i} \cdot s9_{i,q}$$

donde $s_{r,i,q}$ son los factores objetivos ligados a las alternativas, que fueron considerados como parte del modelo MIMIC en la generación de la base de datos (Figura 3.1).

d) *Modelo Clásico de Elección Discreta*: A pesar de toda la evidencia empírica a favor de los modelos híbridos de elección discreta (Ashok *et al.*, 2002; Ben-Akiva *et al.*, 2002; Vredin-Johansson *et al.*, 2005; Yáñez *et al.*, 2010), es interesante analizar el comportamiento de un modelo clásico de elección discreta (sin variables latentes) en un experimento controlado.

Esta especificación considera todas las posibles variaciones sistemáticas de gustos, es decir, todas las posibles interacciones entre variables socio-económicas y las variables propias de las alternativas, pero no considera la incorporación de variables latentes. En consecuencia, esta especificación no toma en consideración la información contenida en los indicadores de percepción.

e) Inclusión Directa de los Indicadores de Percepción en el Modelo de Elección Discreta: Aunque la literatura actual recomienda no incluir los indicadores de percepción de forma directa en los modelos de elección discreta, por ser éstos una manifestación de las percepciones y no su causa, se han reportado varios modelos que incorporan estas variables en forma directa (Koppelman y Hauser, 1979; Green, 1984; Harris y Keane, 1999).

Una de las principales limitantes de este enfoque, es el desconocimiento de las percepciones de la población (a diferencia de las características socio-económicas de ésta), lo que obligaría a realizar una nueva encuesta de percepciones cada vez que se deseara utilizar el modelo, lo que a su vez permitiría estimar un nuevo modelo, en base a la nueva información, haciendo obsoleto el modelo anteriormente calibrado.

Debido a estas razones (no causalidad y baja aplicabilidad), este tipo de modelos se ha visto relegado en la literatura reciente. Sin embargo, aprovechando las ventajas que provee el trabajar en un medio ambiente controlado, como un banco de datos generados en forma sintética, se analizará la capacidad predictiva de este tipo de modelos. Para esto, se consideró un modelo de elección discreta clásica (sin variables latentes), en el cual las características propias de las alternativas y las características socio-económicas de los individuos, así como los indicadores de percepción, se incorporaban en forma directa en la función de utilidad. Por razones de economía, no se consideró inicialmente variaciones sistemáticas de gustos o interacciones entre las distintas variables.

f) Inclusión de la Variabilidad de las Variables Latentes: Para los propósitos de este trabajo se consideró inicialmente la estimación de modelos híbridos en forma secuencial,

incorporando el valor esperado de las variables latentes directamente en la función de utilidad del modelo de elección discreta, sin considerar la variabilidad de éstas como se reporta en la literatura reciente (Ashok *et al.*, 2002; Vredin-Johansson *et al.*, 2005; Yáñez *et al.*, 2010).

Sin embargo, y con el fin de poder replicar de mejor forma la naturaleza de las variables latentes, se consideró modelos ML, que permiten tratar su variabilidad al ser incorporadas al modelo de elección discreta. De esta forma se analizaron tres modelos, cuya especificación corresponde a las planteadas en a), b) y c), pero que difieren de los anteriores, al considerar la variabilidad de las variables latentes.

4.2 Identificabilidad de los Parámetros del Modelo MIMIC

Dada la estructura de los modelos de variables latentes, es claro que los estimadores no son únicos. Por tanto es necesario recurrir a ciertas restricciones para permitir la estimación. Los enfoques más populares, para llevar a cabo lo anterior, consisten en fijar tantos parámetros de las ecuaciones de medición como sea requerido (Ben-Akiva *et al.*, 2002; Vredin Johansson *et al.*, 2005) o fijar las varianzas de la ecuaciones estructurales (Bolduc *et al.*, 2008; Raveau *et al.*, 2010b). Stapleton (1978) demostró que ambas normalizaciones se comportan adecuadamente.

Al optarse por restringir las varianzas, se acostumbra a fijar éstas en uno (Bolduc *et al.*, 2008; Raveau *et al.*, 2010b). Sin embargo, dado que en nuestro caso las varianzas son generadas en forma sintética, se conoce *a priori* su valor real. Por tanto interesa analizar qué efecto tendría fijar las varianzas en uno, en lugar de utilizar su valor real, sobre la capacidad de ajuste del modelo. Considerando las ecuaciones estructurales del modelo de variables latentes, se tiene la siguiente situación:

$$\eta_{l,i,q}^1 = \sum_r \alpha_{l,r,i} \cdot s_{r,i,q} + v_{l,i,q} \quad (4.1)$$

en que:

$$Var(\eta_{1,i,q}) = Var\left(\sum_r \alpha_{1,r,i} \cdot s_{r,i,q} + v_{1,i,q}\right) \quad (4.2)$$

pero $\sum_r \alpha_{1,r,i} \cdot s_{r,i,q}$ es de carácter determinístico, por tanto:

$$Var(\eta_{1,i,q}) = Var(v_{1,i,q}) = \sigma_1^2 \quad (4.3)$$

en que σ_1^2 es la varianza de $v_{1,i,q}$. No obstante, al estimar el modelo, por problemas de identificabilidad se debe suponer una varianza σ_2^2 (potencialmente distinta de σ_1^2); luego se tendría la siguiente ecuación estructural:

$$\eta_{2,i,q} = \sum_r \alpha_{2,r,i} \cdot s_{r,i,q} + v_{2,i,q} / Var(\eta_{2,i,q}) = \sigma_2^2 \quad (4.4)$$

Sin embargo, la única fuente de variabilidad estocástica de $\eta_{2,i,q}$ es $v_{2,i,q}$. Por tanto, en este caso, se tiene que:

$$Var(v_{2,i,q}) = \sigma_2^2 \quad (4.5)$$

Ahora como $v_{1,i,q}$ y $v_{2,i,q}$ distribuyen Normal, se puede reescribir el término de error $v_{2,i,q}$ como $\delta v_{1,i,q}$, en que δ es una constante desconocida. Por tanto se tiene que:

$$Var(v_{2,i,q}) = Var(\delta \cdot v_{1,i,q}) = \delta^2 \cdot Var(v_{1,i,q}) = \sigma_2^2 \quad (4.6)$$

Luego reemplazando (4.3) en (4.6) se tiene que:

$$\delta = \frac{\sigma_2}{\sigma_1} \quad (4.7)$$

Finalmente dividiendo (4.4) por δ :

$$\frac{\eta_{l,i,q}^2}{\delta} = \sum_r \frac{\alpha_{l,r,i}}{\delta} \cdot s_{r,i,q} + v_{l,i,q} / \text{Var}\left(\frac{\eta_{l,i,q}^2}{\delta}\right) = \sigma_1^2 \quad (4.8)$$

que es equivalente a las condiciones planteadas en (4.1) y (4.3). Por tanto

$$\frac{\eta_{l,i,q}^2}{\delta} = \eta_{l,i,q}^1 \wedge \frac{\alpha_{l,r,i}}{\delta} = \alpha_{l,r,i} \quad \forall l,r,i \quad (4.9)$$

De esta forma, para las ecuaciones de medición del modelo de variables latentes, se cumple que:

$$y_{p,i,q} = \sum_l \gamma_{l,p,i}^1 \cdot \eta_{l,i,q}^1 + v_{p,i,q} = \sum_l \gamma_{l,p,i}^2 \cdot \eta_{l,i,q}^2 + v_{p,i,q} \quad (4.10)$$

y reemplazando (4.9) en (4.10), se tiene que:

$$\gamma_{l,p,i}^1 = \delta \cdot \gamma_{l,p,i}^2 \quad \forall l,p,i \quad (4.11)$$

Por tanto, es claro que la magnitud de las variables latentes se verá seriamente afectada por la magnitud de las varianzas utilizadas como restricción. Sin embargo, aunque esta variación será distinta para todas las alternativas y variables latentes (suponiendo que la varianza de éstas es disímil), será igual para todos los individuos (sus variables latentes asociadas se verán afectadas por el mismo término δ). Esto reviste gran importancia, ya que una vez que estas variables son incorporadas al modelo de elección discreta, sus estimadores podrán ajustarse sin mayores problemas, obteniéndose el mismo nivel de ajuste que en un modelo en que se hubiesen utilizado las varianzas correctas. Por tanto, independiente de la magnitud de las varianzas utilizadas como restricción, el modelo híbrido se ajustará y predecirá de igual forma.

Un corolario importante de lo anterior, es que ante la menor sospecha de que la variabilidad estocástica asociada a las variables latentes pueda variar de alternativa en alternativa, se hace obligatorio utilizar estimadores específicos asociados a las variables latentes (en desmedro de estimadores genéricas) en la función de utilidad del modelo de elección discreta.

4.2.1 Efecto de la Normalización en la Estimación del Modelo MIMIC

Tal como se mencionó en la sección 4.2, en nuestro caso la variabilidad estocástica asociada a las distintas variables latentes es disímil, ya que la desviación estándar asociada las variables *Seguridad*, *Comodidad* y *Confiabilidad* se fijó en uno, 0,7 y 0,5 respectivamente. Por tanto, para corroborar lo mencionado anteriormente se estimó dos modelos de variables latentes (asumiendo la estructura generadora) utilizándose como restricción, en el primer caso, las varianzas reales (conocidas *a priori*), mientras que en el segundo, se utilizó varianzas iguales a uno, tal como se usa en la práctica.

Una vez estimados los modelos, se determinó la razón entre los parámetros de uno y otro, y ésta fue contrastada con la razón teórica esperada (de acuerdo con las varianzas utilizadas como restricción). Los resultados agregados a nivel de alternativas, se presentan en las Tablas 4.1 a 4.3.

Se debe mencionar que, en virtud de las herramientas computacionales disponibles, la estimación de los modelos se hizo con una precisión de tres decimales, lo que explica la leve inexactitud en el caso de estimadores de menor magnitud. Tal como se observa en las Tablas 4.1 a 4.3, los resultados obtenidos en la práctica están de acuerdo con la teoría. De esta forma, se provee evidencia empírica que respalda lo formulado en la sección 4.2.

Tabla 4.1 – Efecto de Normalización para la Alternativa Automóvil

Automóvil		Modelo Varianzas Correctas	Modelo Varianzas Incorrectas		
Variable Explicada	Variable Explicatoria	Valor Estimado	Valor Estimado	Razón Obs.	Razón Teórica
<i>VL Seguridad</i>	<i>Tasa de Accidentes</i>	-0,802	-0,802	1	1
<i>VL Seguridad</i>	<i>Sexo</i>	1,203	1,203	1	1
<i>VL Seguridad</i>	<i>Ingreso Alto</i>	1,818	1,818	1	1
<i>VL Seguridad</i>	<i>Ingreso Medio</i>	1,01	1,01	1	1
<i>VL Comodidad</i>	<i>Sexo</i>	0,811	1,159	0,7	0,7
<i>VL Comodidad</i>	<i>Ingreso Alto</i>	1,205	1,722	0,7	0,7
<i>VL Comodidad</i>	<i>Ingreso Medio</i>	0,69	0,985	0,701	0,7
<i>VL Comodidad</i>	<i>Aire Acondicionado</i>	0,491	0,701	0,7	0,7
<i>VL Comodidad</i>	<i>Edad</i>	0,01	0,015	0,667	0,7
<i>VL Confiabilidad</i>	<i>Edad</i>	0,015	0,03	0,5	0,5
<i>VL Confiabilidad</i>	<i>Educación Secundaria</i>	0,886	1,772	0,5	0,5
<i>VL Confiabilidad</i>	<i>Educación Superior</i>	1,38	2,759	0,5	0,5
<i>Seguridad Accidentes</i>	<i>VL Seguridad</i>	2,488	2,488	1	1
<i>Seguridad Robos</i>	<i>VL Seguridad</i>	2,189	2,189	1	1
<i>Accesibilidad</i>	<i>VL Seguridad</i>	0,798	0,798	1	1
<i>Accesibilidad</i>	<i>VL Comodidad</i>	2,086	1,46	1,429	1,429
<i>Comodidad</i>	<i>VL Comodidad</i>	2,803	1,962	1,429	1,429
<i>Información</i>	<i>VL Comodidad</i>	0,995	0,696	1,430	1,429
<i>Información</i>	<i>VL Confiabilidad</i>	2,471	1,235	2,001	2
<i>Confiabilidad</i>	<i>VL Confiabilidad</i>	4,183	2,092	2	2
<i>Estimar Tiempo de Viaje</i>	<i>VL Confiabilidad</i>	3,903	1,952	1,999	2

La razón observada (Razón Obs.) corresponde a la relación entre los parámetros estimados por uno y otro modelo, mientras que la razón teórica corresponde a la razón esperada de acuerdo con lo planteado en 4.2.

En algunos casos, especialmente asociados a la variable explicatoria *Edad*, se observan ligeras diferencias entre las razones observadas y esperadas, pero, tal como fue mencionado anteriormente, esto se explica por el nivel de precisión de la herramienta computacional utilizada para estimar los modelos.

Tabla 4.2 – Efecto de Normalización para la Alternativa Metro

Metro		Modelo Varianzas Correctas	Modelo Varianzas Incorrectas		
Variable Explicada	Variable Explicatoria	Valor Estimado	Valor Estimado	Razón Obs.	Razón Teórica
VL Seguridad	Tasa de Accidentes	-1,52	-1,52	1	1
VL Seguridad	Cámaras de Seguridad	0,786	0,786	1	1
VL Seguridad	Sexo	0,797	0,797	1	1
VL Seguridad	Ingreso Alto	1,102	1,102	1	1
VL Seguridad	Ingreso Medio	0,71	0,71	1	1
VL Comodidad	Sexo	0,395	0,564	0,7	0,7
VL Comodidad	Ingreso Alto	0,799	1,142	0,7	0,7
VL Comodidad	Ingreso Medio	0,288	0,411	0,701	0,7
VL Comodidad	Densidad	-0,15	-0,214	0,701	0,7
VL Comodidad	Aire Acondicionado	0,711	1,015	0,7	0,7
VL Comodidad	Edad	0,006	0,008	0,75	0,7
VL Confiabilidad	Edad	0,007	0,014	0,5	0,5
VL Confiabilidad	Educación Secundaria	0,399	0,798	0,5	0,5
VL Confiabilidad	Educación Superior	0,707	1,414	0,5	0,5
VL Confiabilidad	Panel de Información	0,602	1,204	0,5	0,5
Seguridad Accidentes	VL Seguridad	2,504	2,504	1	1
Seguridad Robos	VL Seguridad	2,193	2,193	1	1
Accesibilidad	VL Seguridad	0,796	0,796	1	1
Accesibilidad	VL Comodidad	2,11	1,477	1,429	1,429
Comodidad	VL Comodidad	2,801	1,96	1,429	1,429
Información	VL Comodidad	1,017	0,712	1,428	1,429
Información	VL Confiabilidad	2,471	1,236	1,999	2
Confiabilidad	VL Confiabilidad	4,162	2,081	2	2
Estimar Tiempo de Viaje	VL Confiabilidad	3,867	1,934	1,999	2

La razón observada (Razón Obs.) corresponde a la relación entre los parámetros estimados por uno y otro modelo, mientras que la razón teórica corresponde a la razón esperada de acuerdo con lo planteado en 4.2.

Dada la equivalencia inherente de ambas normalizaciones, y con el objeto de facilitar el análisis respecto a la capacidad de recuperación de parámetros asociada a los modelos, en adelante se trabajará utilizando las varianzas reales (notar que en caso contrario

también es posible realizar el análisis, pero se debe realizar una corrección para compensar la deflexión de las variables latentes).

Tabla 4.3 – Efecto de Normalización para la Alternativa Bus

Bus		Modelo Varianzas Correctas	Modelo Varianzas Incorrectas		
Variable Explicada	Variable Explicatoria	Valor Estimado	Valor Estimado	Razón Obs.	Razón Teórica
<i>VL Seguridad</i>	<i>Tasa de Accidentes</i>	-0,887	-0,887	1	1
<i>VL Seguridad</i>	<i>Cámaras de Seguridad</i>	1,088	1,088	1	1
<i>VL Seguridad</i>	<i>Sexo</i>	-0,286	-0,286	1	1
<i>VL Seguridad</i>	<i>Ingreso Alto</i>	-0,79	-0,79	1	1
<i>VL Seguridad</i>	<i>Ingreso Medio</i>	-0,392	-0,392	1	1
<i>VL Comodidad</i>	<i>Sexo</i>	-0,199	-0,285	0,698	0,7
<i>VL Comodidad</i>	<i>Ingreso Alto</i>	-0,399	-0,57	0,7	0,7
<i>VL Comodidad</i>	<i>Ingreso Medio</i>	-0,202	-0,289	0,699	0,7
<i>VL Comodidad</i>	<i>Densidad</i>	-0,153	-0,218	0,702	0,7
<i>VL Comodidad</i>	<i>Aire Acondicionado</i>	1	1,428	0,7	0,7
<i>VL Comodidad</i>	<i>Edad</i>	-0,004	-0,006	0,667	0,7
<i>VL Confiabilidad</i>	<i>Edad</i>	-0,006	-0,012	0,5	0,5
<i>VL Confiabilidad</i>	<i>Educación Secundaria</i>	-0,2	-0,399	0,501	0,5
<i>VL Confiabilidad</i>	<i>Educación Superior</i>	-0,393	-0,786	0,5	0,5
<i>VL Confiabilidad</i>	<i>Panel de Información</i>	0,812	1,624	0,5	0,5
<i>Seguridad Accidentes</i>	<i>VL Seguridad</i>	2,51	2,51	1	1
<i>Seguridad Robos</i>	<i>VL Seguridad</i>	2,211	2,211	1	1
<i>Accesibilidad</i>	<i>VL Seguridad</i>	0,798	0,798	1	1
<i>Accesibilidad</i>	<i>VL Comodidad</i>	2,097	1,468	1,428	1,429
<i>Comodidad</i>	<i>VL Comodidad</i>	2,79	1,953	1,429	1,429
<i>Información</i>	<i>VL Comodidad</i>	0,993	0,695	1,429	1,429
<i>Información</i>	<i>VL Confiabilidad</i>	2,484	1,242	2	2
<i>Confiabilidad</i>	<i>VL Confiabilidad</i>	4,179	2,089	2	2
<i>Estimar Tiempo de Viaje</i>	<i>VL Confiabilidad</i>	3,871	1,935	2,001	2

La razón observada (Razón Obs.) corresponde a la relación entre los parámetros estimados por uno y otro modelo, mientras que la razón teórica corresponde a la razón esperada de acuerdo con lo planteado en 4.2.

4.3 Resultados y Discusión de la Estimación de los modelos MIMIC

Los modelos de variables latentes se estimaron de acuerdo a dos especificaciones diferentes. Las Tablas 4.4 a 4.6 presentan estos resultados agregados a nivel de alternativas. Los valores “reales” (*target*) de los parámetros de las ecuaciones estructurales y de medición utilizados en la generación de la base de datos se presentan en la tercera columna. A su derecha, se presentan los parámetros estimados de acuerdo al modelo MIMIC que considera tanto variables explicativas socio-económicas como variables tangibles propias de las alternativas (el cual llamaremos Modelo Complejo).

A continuación se presentan los estimadores asociados al modelo de variables latentes que considera exclusivamente variables explicativas socio-económicas (al cual llamaremos Modelo Simplificado). En ambos casos, los parámetros estimados son presentados como porcentaje del valor “real” objetivo (*target*), seguidos por sus respectivos test-t para la hipótesis nula (respecto a cero). Del análisis de los resultados, se desprende que los parámetros estimados de acuerdo a la especificación correcta se acercan bastante a los valores reales objetivos, lo que, sin embargo, no se repite en el caso del modelo simplificado (como era esperable), ya que las variables socio-económicas deben compensar el efecto de la carencia de la información contenida en las variables explicativas propias de las alternativas.

Sin desmedro de lo anterior, los signos asociados a los parámetros estimados de acuerdo al modelo simplificado, son correctos para el 100% de los casos. De esta forma, se garantiza que las variables socio-económicas afectan a las variables latentes en la misma manera supuesta en la generación de la base de datos y por tanto, se debe esperar que las variables latentes continúen afectando el proceso de elección en la misma dirección. De otro modo, las variables socio-económicas estarían afectando (aunque indirectamente) la decisión final de manera inversa.

Tabla 4.4 – Estimadores del Modelo MIMIC para la Alternativa Automóvil

Automóvil		Valor Objetivo	Modelo Complejo		Modelo Simplificado	
			Valor Estimado	Test-t	Valor Estimado	Test-t
Variable Explicada	Variable Explicatoria					
<i>VL Seguridad</i>	<i>Tasa de Accidentes</i>	-0,8	100,25%	-121,051	-	-
<i>VL Seguridad</i>	<i>Sexo</i>	1,2	100,25%	91,304	76,58%	73,704
<i>VL Seguridad</i>	<i>Ingreso Alto</i>	1,8	101,00%	106,644	77,00%	87,159
<i>VL Seguridad</i>	<i>Ingreso Medio</i>	1	101,00%	78,236	76,90%	62,344
<i>VL Comodidad</i>	<i>Sexo</i>	0,8	101,38%	84,943	95,63%	81,523
<i>VL Comodidad</i>	<i>Ingreso Alto</i>	1,2	100,42%	97,809	95,00%	94,425
<i>VL Comodidad</i>	<i>Ingreso Medio</i>	0,7	98,57%	73,674	93,71%	71,04
<i>VL Comodidad</i>	<i>Aire Acondicionado</i>	0,5	98,20%	53,426	-	-
<i>VL Comodidad</i>	<i>Edad</i>	0,01	100,00%	47,931	100,00%	45,695
<i>VL Confiabilidad</i>	<i>Edad</i>	0,015	100,00%	93,902	100,00%	93,902
<i>VL Confiabilidad</i>	<i>Educación Secundaria</i>	0,9	98,44%	109,956	98,44%	109,961
<i>VL Confiabilidad</i>	<i>Educación Superior</i>	1,4	98,57%	130,98	98,57%	130,983
<i>Seguridad Accidentes</i>	<i>VL Seguridad</i>	2,5	99,52%	220,325	131,52%	227,84
<i>Seguridad Robos</i>	<i>VL Seguridad</i>	2,2	99,50%	217,4	131,50%	224,63
<i>Accesibilidad</i>	<i>VL Seguridad</i>	0,8	99,75%	138,675	132,00%	140,401
<i>Accesibilidad</i>	<i>VL Comodidad</i>	2,1	99,33%	185,624	105,00%	187,535
<i>Comodidad</i>	<i>VL Comodidad</i>	2,8	100,11%	203,045	105,82%	205,747
<i>Información</i>	<i>VL Comodidad</i>	1	99,50%	131,121	105,20%	131,815
<i>Información</i>	<i>VL Confiabilidad</i>	2,5	98,84%	191,238	98,84%	191,209
<i>Confiabilidad</i>	<i>VL Confiabilidad</i>	4,2	99,60%	214,328	99,60%	214,326
<i>Estimar Tiempo de Viaje</i>	<i>VL Confiabilidad</i>	3,9	100,08%	212,407	100,08%	212,404

Tabla 4.5 – Estimadores del Modelo MIMIC para la Alternativa Metro

Metro		Valor Objetivo	Modelo Complejo		Modelo Simplificado	
			Valor Estimado	Test-t	Valor Estimado	Test-t
Variable Explicada	Variable Explicatoria					
VL Seguridad	Tasa de Accidentes	-1,5	101,33%	-28,061	-	-
VL Seguridad	Cámaras de Seguridad	0,8	98,25%	50,726	-	-
VL Seguridad	Sexo	0,8	99,63%	63,522	93,63%	60,25
VL Seguridad	Ingreso Alto	1,1	100,18%	69,83	94,36%	66,47
VL Seguridad	Ingreso Medio	0,7	101,43%	56,831	95,71%	54,063
VL Comodidad	Sexo	0,4	98,75%	44,246	85,25%	39,176
VL Comodidad	Ingreso Alto	0,8	99,88%	69,366	86,38%	62,027
VL Comodidad	Ingreso Medio	0,3	96,00%	32,433	83,00%	28,727
VL Comodidad	Densidad	-0,15	100,00%	-53,746	-	-
VL Comodidad	Aire Acondicionado	0,7	101,57%	70,853	-	-
VL Comodidad	Edad	0,006	100,00%	28,542	83,33%	24,515
VL Confiabilidad	Edad	0,007	100,00%	46,662	85,71%	44,349
VL Confiabilidad	Educación Secundaria	0,4	99,75%	55,17	94,00%	52,515
VL Confiabilidad	Educación Superior	0,7	101,00%	77,818	94,29%	73,62
VL Confiabilidad	Panel de Información	0,6	100,33%	57,038	-	-
Seguridad Accidentes	VL Seguridad	2,5	100,16%	216,784	106,36%	218,603
Seguridad Robos	VL Seguridad	2,2	99,68%	211,595	105,86%	213,283
Accesibilidad	VL Seguridad	0,8	99,50%	118,74	105,63%	118,501
Accesibilidad	VL Comodidad	2,1	100,48%	183,223	116,29%	187,968
Comodidad	VL Comodidad	2,8	100,04%	201,513	115,82%	207,988
Información	VL Comodidad	1	101,70%	120,936	117,70%	122,157
Información	VL Confiabilidad	2,5	98,84%	172,096	105,00%	172,992
Confiabilidad	VL Confiabilidad	4,2	99,10%	207,736	105,24%	209,683
Estimar Tiempo de Viaje	VL Confiabilidad	3,9	99,15%	204,175	105,36%	206,019

Tabla 4.6 – Estimadores del Modelo MIMIC para la Alternativa Bus

Bus		Valor Objetivo	Modelo Complejo		Modelo Simplificado	
			Valor Estimado	Test-t	Valor Estimado	Test-t
Variable Explicada	Variable Explicatoria					
VL Seguridad	Tasa de Accidentes	-0,9	98,56%	-96,32	-	-
VL Seguridad	Cámaras de Seguridad	1,1	98,91%	77,608	-	-
VL Seguridad	Sexo	-0,3	95,33%	-23,603	71,67%	-18,131
VL Seguridad	Ingreso Alto	-0,8	98,75%	-51,297	77,75%	-41,493
VL Seguridad	Ingreso Medio	-0,4	98,00%	-32,077	77,50%	-25,892
VL Comodidad	Sexo	-0,2	99,50%	22,776	83,50%	19,52
VL Comodidad	Ingreso Alto	-0,4	99,75%	36,192	83,00%	30,869
VL Comodidad	Ingreso Medio	-0,2	101,00%	22,975	86,00%	19,979
VL Comodidad	Densidad	-0,15	102,00%	44,17	-	-
VL Comodidad	Aire Acondicionado	1	100,00%	-84,192	-	-
VL Comodidad	Edad	-0,004	100,00%	18,853	75,00%	16,372
VL Confiabilidad	Edad	-0,006	100,00%	-41,767	83,33%	-35,352
VL Confiabilidad	Educación Secundaria	-0,2	100,00%	-28,271	79,50%	-23,03
VL Confiabilidad	Educación Superior	-0,4	98,25%	-45,226	81,50%	-38,431
VL Confiabilidad	Panel de Información	0,8	101,50%	107,13	-	-
Seguridad Accidentes	VL Seguridad	2,5	100,40%	217,954	129,36%	224,551
Seguridad Robos	VL Seguridad	2,2	100,50%	213,531	129,45%	219,732
Accesibilidad	VL Seguridad	0,8	99,75%	127,914	128,63%	128,619
Accesibilidad	VL Comodidad	2,1	99,86%	-181,191	118,43%	-186,439
Comodidad	VL Comodidad	2,8	99,64%	-200,362	118,18%	-207,732
Información	VL Comodidad	1	99,30%	-113,718	117,70%	-114,549
Información	VL Confiabilidad	2,5	99,36%	175,695	123,76%	179,465
Confiabilidad	VL Confiabilidad	4,2	99,50%	209,055	123,90%	215,533
Estimar Tiempo de Viaje	VL Confiabilidad	3,9	99,26%	205,602	123,62%	211,77

El punto anterior reviste vital importancia, ya que impone que los signos asociados a las variables latentes deben ser consistentes con los signos supuestos. Por tanto, los modelos híbridos estimados a continuación, utilizando las variables latentes determinadas en función de la estructura simplificada, no debieran por fuerza presentar inconsistencias entre los signos supuestos de las variables latentes (utilizados en la generación de la base de datos) y los signos estimados.

4.4 El Problema de la Varianza y la Recuperación de Parámetros

Tal como se mencionó en el acápite 2.1.2, los parámetros estimados en un modelo de elección discreta se encuentran afectados por el parámetro λ , el cual es inestimable. Dado que este parámetro multiplica a todos los estimadores, se suele asumir igual a uno sin pérdida de generalidad.

Este parámetro está relacionado con la varianza del término de error asociado a la función de utilidad, tal como se indica en la ecuación (2.6). Consideremos dos modelos iguales, cuya única diferencia esté dada por la variabilidad del término de error; en este caso, la relación entre los parámetros estimados sería la siguiente:

$$\theta_v = \frac{\theta_1}{\lambda_1} = \frac{\theta_2}{\lambda_2} \quad (4.12)$$

en que θ_v corresponde al valor real teórico del parámetro (inestimable), θ_1 y θ_2 son los parámetros estimados por los modelos 1 y 2, respectivamente y λ_1 y λ_2 los factores de escala de ambos modelos, asociados a la variabilidad del término de error.

De esta forma se tiene que:

$$\theta_1 = \frac{\lambda_1}{\lambda_2} \cdot \theta_2 \quad (4.13)$$

o bien (tomando el promedio de los coeficientes estimados):

$$\lambda_1 = \left(\frac{\theta_1}{\theta_2} \right) \cdot \lambda_2 \quad (4.14)$$

En el caso de modelos híbridos de elección discreta con variables latentes, la variabilidad no proviene única y exclusivamente de los términos de error asociados a la función de utilidad, sino que también de las variables latentes. Esto, ya que el valor real de estas últimas no es conocido y por tanto son estimadas en base a un modelo que considera términos de error propios. En consecuencia, la variabilidad asociada a la función de utilidad (σ_R^2) será la suma de la varianza debida al término de error propio (σ_P^2) y la varianza inducida por medio de las variables latentes (σ_I^2).

$$\sigma_R^2 = \sigma_P^2 + \sigma_I^2 \quad (4.15)$$

Reemplazando (2.6) en (4.15), se tiene que¹:

$$\frac{\pi^2}{6 \cdot \lambda_R^2} = \frac{\pi^2}{6 \cdot \lambda_P^2} + \frac{\pi^2}{6 \cdot \lambda_I^2} \quad (4.16)$$

y posteriormente despejando λ_R (ver Anexo A):

$$\lambda_R^2 = \lambda_P^2 \cdot \frac{\lambda_I^2}{\lambda_P^2 + \lambda_I^2} \quad (4.17)$$

¹ Para simplificar los cálculos, se asumió que la varianza inducida distribuye como la diferencia entre dos distribuciones Gumbel (Williams, 1977). Una simplificación similar se utilizó con el objeto de expresar esta variabilidad en función del parámetro de escala. Es importante notar que considerar una distribución Normal permitiría llegar a un resultado similar pero el cálculo sería más engorroso.

de donde es claro que $\lambda_R < \lambda_P$. Lo anterior no tendría mayores implicancias al trabajarse con un banco de datos real (ya que no sería posible detectar el problema); sin embargo, el trabajar con datos simulados permite ver que los parámetros no pueden ser recuperados en forma correcta y deben ser corregidos por λ_P/λ_R .

4.4.1 Efecto sobre la Correlación

El resultado anterior debiera tener mayores implicancias cuando existen alternativas correlacionadas. En este caso, la varianza asociada al término de error puede escribirse como:

$$\sigma_P^2 = \sigma_D^2 + \sigma_{DM}^2 \quad (4.18)$$

en que σ_D^2 representa la varianza del término de error común y σ_{DM}^2 representa a la varianza asociada al componente independiente del término de error.

Por tanto, reemplazando (4.18) en (4.15), se tiene que:

$$\sigma_R^2 = \sigma_D^2 + \sigma_{DM}^2 + \sigma_I^2 \quad (4.19)$$

en que σ_I^2 es un término de error independiente para todas las alternativas que proviene del modelo de variables latentes. Debido a lo anterior, se tiene que $\sigma_{DM}^2 + \sigma_I^2$ representan a las componentes independientes del término de error.

Notar que igualmente se pudiese especificar términos de error correlacionados en los modelos de variables latentes, pero no es el caso. En la literatura tampoco se han reportado experiencias en ese sentido.

Introduciendo el coeficiente de correlación, la ecuación (4.19) puede ser descrita de la siguiente forma:

$$\sigma_R^2 = \rho \cdot \sigma_P^2 + \sigma_{DM}^2 + \sigma_I^2 \quad (4.20)$$

en que σ_P^2 corresponde a la varianza indicada en (4.15).

Al estimar un modelo HL, considerando funciones de utilidad cuyos términos de error presentan la estructura de correlación recién señalada, la relación (4.20) puede ser escrita como:

$$\frac{\pi^2}{6 \cdot \lambda_R^2} = \rho \cdot \frac{\pi^2}{6 \cdot \lambda_P^2} + \frac{\pi^2}{6 \cdot \Delta^2} \quad (4.21)$$

en que Δ corresponde al parámetro de escala de la distribución asociada a los términos de error independientes, dentro de un nido dado.

Finalmente operando matemáticamente sobre (4.21) se llega a (ver Anexo A):

$$\sqrt{1 - \rho \cdot \frac{\lambda_R^2}{\lambda_P^2}} = \frac{\lambda_R}{\Delta} = \phi \quad (4.22)$$

Tal como fue mencionado en la sección 4.4, λ_R / λ_P es siempre menor que uno y, por tanto, la expresión anterior tendría que ser necesariamente mayor que la presentada en (2.14). Dado esto, el parámetro ϕ obtenido en modelo híbrido debiera ser, por fuerza, mayor que el parámetro asociado a la correlación real existente entre las alternativas.

Como corolario de lo anterior, se puede concluir que los modelos híbridos de elección discreta no permiten captar de manera adecuada la correlación real existente entre dos alternativas, tendiendo a atenuar el valor real de ésta. Es importante notar que lo anterior

está relacionado con el proceso de elección en sí mismo y que la omisión de las variables latentes no soluciona el problema.

Sin desmedro de lo anterior, los modelos predecirán de forma adecuada ya que el nivel de correlación estimado es consistente con la variabilidad con que trabaja el modelo en cuestión.

4.4.2 Recuperación de Parámetros

Utilizando la base de datos simulada, se estimó un modelo siguiendo la especificación descrita en el acápite a) de la sección 4.1. En nuestro caso simulado, se supuso que el factor de escala de la distribución asociada a los términos de error de la función de utilidad era uno, lo que potencialmente permitiría recuperar los parámetros en forma directa (en caso de no existir el error inducido por medio de las variables latentes).

La Tabla 4.7 presenta los resultados de la estimación y los valores reales objetivos para los distintos parámetros estimados, así como la razón entre éstos. Como se puede observar, los estimadores responden al comportamiento enunciado en la sección 4.2, ya que todos los estimadores se encuentran consistentemente deflactados por una variable menor que uno y no por el factor de escala utilizado en la generación de la base de datos.

El promedio de la razón entre los valores estimados y los valores objetivos es un buen estimador de la relación λ_R / λ_P y será utilizado para corregir los valores objetivos. De esta forma, se realizó un análisis estadístico de la recuperación de parámetros, que se resume en la Tabla 4.8. En ésta se presentan los parámetros estimados, así como su desviación estándar; en la cuarta columna se presenta el valor objetivo corregido,

mientras que en la última columna se presenta el test-t, cuya hipótesis nula corresponde a la igualdad entre los parámetros estimados y observados.

Tabla 4.7 – Relación entre los Estimadores del Modelo de Elección

	Valor Estimado	Valor Objetivo	Razón
<i>ASC_{Auto}</i>	0,429	0,5	0,858
<i>ASC_{Bus}</i>	1,36	1,5	0,907
<i>ASC_{Metro}</i>	-	-	-
<i>Com_{Auto}</i>	0,614	0,75	0,819
<i>Com_{Bus}</i>	0,626	0,75	0,835
<i>Com_{Metro}</i>	0,651	0,75	0,868
<i>Conf_{Auto}</i>	0,593	0,7	0,847
<i>Conf_{Bus}</i>	0,586	0,7	0,837
<i>Conf_{Metro}</i>	0,591	0,7	0,844
<i>Costo_{Auto}</i>	-0,913	-1,1	0,83
<i>Costo_{Bus}</i>	-0,948	-1,1	0,862
<i>Costo_{Metro}</i>	-0,912	-1,1	0,829
<i>Sec_{Auto}</i>	0,461	0,55	0,838
<i>Sec_{Bus}</i>	0,463	0,55	0,842
<i>Sec_{Metro}</i>	0,464	0,55	0,844
<i>Tiempo_{Auto}</i>	-0,606	-0,75	0,808
<i>Tiempo_{Bus}</i>	-0,628	-0,75	0,837
<i>Tiempo_{Metro}</i>	-0,612	-0,75	0,816
			$\bar{X}=0,842$

Como se puede observar en la Tabla 4.8, no existe evidencia estadística que permita, con una confiabilidad del 95%, rechazar la hipótesis nula para ninguno de los parámetros. Esto implica que el modelo es consistente y concordante con la teoría. Es importante destacar que de no aplicarse la corrección, la hipótesis nula sería rechazada en todos los casos, con la excepción de las constantes modales, implicando una potencial inconsistencia del modelo.

Tabla 4.8 – Análisis Estadístico de la Recuperación de Parámetros

	Valor Estimado	Desviación Estándar	Valor Objetivo Corregido	Test-t
<i>ASC_{Auto}</i>	0,429	0,105	0,421	0,074
<i>ASC_{Bus}</i>	1,36	0,121	1,264	0,797
<i>ASC_{Metro}</i>	-	-	-	-
<i>Com_{Auto}</i>	0,614	0,033	0,632	-0,539
<i>Com_{Bus}</i>	0,626	0,0373	0,632	-0,155
<i>Com_{Metro}</i>	0,651	0,0312	0,632	0,616
<i>Conf_{Auto}</i>	0,593	0,0395	0,590	0,084
<i>Conf_{Bus}</i>	0,586	0,0425	0,590	-0,086
<i>Conf_{Metro}</i>	0,591	0,0568	0,590	0,023
<i>Costo_{Auto}</i>	-0,913	0,0124	-0,927	1,098
<i>Costo_{Bus}</i>	-0,948	0,0429	-0,927	-0,498
<i>Costo_{Metro}</i>	-0,912	0,0307	-0,927	0,476
<i>Sec_{Auto}</i>	0,461	0,0162	0,463	-0,143
<i>Sec_{Bus}</i>	0,463	0,0222	0,463	-0,014
<i>Sec_{Metro}</i>	0,464	0,0276	0,463	0,025
<i>Tiempo_{Auto}</i>	-0,606	0,0146	-0,632	1,766
<i>Tiempo_{Bus}</i>	-0,628	0,0185	-0,632	0,205
<i>Tiempo_{Metro}</i>	-0,612	0,0182	-0,632	1,087

En el caso del parámetro ϕ se tiene que, de acuerdo a (2.14):

$$\phi = \sqrt{1 - 0,3} = 0,837$$

sin embargo aplicando la corrección planteada en (4.22), el valor real objetivo corregido con el que se ha de contrastar el valor estimado es:

$$\phi_c = \sqrt{1 - 0,3 \cdot \frac{\lambda_R^2}{\lambda_P^2}} = 0,887$$

En la Tabla 4.9 se presenta el test de hipótesis, cuya hipótesis nula es la igualdad entre el valor estimado y el valor objetivo corregido.

Tabla 4.9 – Análisis Estadístico de la Recuperación del Parámetro ϕ

	Valor Estimado	Desviación Estándar	Valor Objetivo Corregido	Test-t
ϕ	0,893	0,0264	0,887	0, 214

Como se puede ver, no existe evidencia estadística al 95%, que permita rechazar la hipótesis nula, lo que es concordante con lo enunciado en el acápite 4.3.1 y en la Tabla 4.8. Además ratifica que el modelo es consistente.

4.4.3 Efecto de la Omisión de la Variabilidad inducida por las Variables Latentes

En el caso hipotético que las variables latentes fuesen observables directamente por el modelador, la variabilidad inducida por medio de éstas debiese ser nula y, por tanto, no debiese existir ningún factor que afectara de manera artificial los parámetros estimados (en el supuesto que el factor de escala de la distribución asociada a los términos de error fuera conocido).

Aprovechando las ventajas de trabajar con datos simulados, que permiten conocer con exactitud el valor de las variables latentes, se estimó un modelo utilizando el valor exacto de éstas, con el fin de corroborar lo planteado en el párrafo anterior y dar mayor sustento a lo enunciado en la sección 4.4. Los resultados de la estimación se presentan en la Tabla 4.10. También se presenta el test-t cuya hipótesis nula corresponde a la igualdad entre el valor estimado y el valor objetivo no deflactado, así como la razón entre ambos.

Tabla 4.10 – Análisis Estadístico del Modelo con Variables Latentes Exactas

	Valor Estimado	Desviación Estándar	Valor Objetivo	Test-t	Razón
<i>ASC_{Auto}</i>	0,488	0,095	0,5	-0,126	0,976
<i>ASC_{Bus}</i>	1,53	0,0988	1,5	0,304	1,020
<i>ASC_{Metro}</i>	0	-	-	-	-
<i>Com_{Auto}</i>	0,729	0,0194	0,75	-1,082	0,972
<i>Com_{Bus}</i>	0,759	0,0247	0,75	0,364	1,012
<i>Com_{Metro}</i>	0,746	0,0198	0,75	-0,202	0,995
<i>Conf_{Auto}</i>	0,693	0,0254	0,7	-0,276	0,990
<i>Conf_{Bus}</i>	0,67	0,029	0,7	-1,034	0,957
<i>Conf_{Metro}</i>	0,723	0,0274	0,7	0,839	1,033
<i>Costo_{Auto}</i>	-1,08	0,0145	-1,1	1,379	0,982
<i>Costo_{Bus}</i>	-1,09	0,0431	-1,1	0,232	0,991
<i>Costo_{Metro}</i>	-1,11	0,0325	-1,1	-0,308	1,009
<i>Sec_{Auto}</i>	0,546	0,012	0,55	-0,333	0,993
<i>Sec_{Bus}</i>	0,558	0,0165	0,55	0,485	1,015
<i>Sec_{Metro}</i>	0,564	0,0147	0,55	0,952	1,025
<i>Tiempo_{Auto}</i>	-0,728	0,0164	-0,75	1,341	0,971
<i>Tiempo_{Bus}</i>	-0,738	0,0175	-0,75	0,686	0,984
<i>Tiempo_{Metro}</i>	-0,743	0,0191	-0,75	0,366	0,991
ϕ	1,2	0,0289	1,195	0,165	-
					$\bar{X}=0,995$

Como se puede observar, todos los estimadores son recuperados de manera correcta - no existe evidencia estadística, con una confianza del 95%, que permita rechazar la hipótesis de igualdad entre los parámetros estimados y objetivos - sin ser necesario recurrir a la corrección de parámetros. Además el promedio de las razones entre los valores estimados y objetivos es muy cercano a uno (son estadísticamente equivalentes),

lo cual es consistente con el factor de escala de la distribución asociada al término de error utilizado en la generación de los datos.

Lo anterior ratifica lo planteado en forma previa y, en conjunto con lo presentado en los acápites 4.4.1 y 4.4.2 provee evidencia empírica que sustenta el análisis teórico enunciado en la sección 4.4.

4.5 Resultados de la Estimación de los Modelos Híbridos de Elección Discreta

En primer lugar, para las especificaciones que así lo requirieran, se construyeron las variables latentes de acuerdo con los modelos presentados en la sección 4.3. A continuación se estimó los modelos HL correspondientes a las especificaciones a), b), c), d) y e) de la sección 4.1, los cuales serán llamadas Modelo 1, Modelo 2, Modelo 3, Modelo 4 y Modelo 5, respectivamente.

Los parámetros estimados utilizando las diferentes especificaciones, así como sus respectivos test-t, se presentan en la Tabla 4.11. Dado que tanto el Modelo 4, como el Modelo 5 consideran una enorme cantidad de estimadores, por razones de espacio, sólo se presentan las interacciones de primer grado y los estimadores de mayor importancia teórica (para los modelos completos, ver Anexo B). En el caso del Modelo 5, por razones pedagógicas se mantuvieron estimadores estadísticamente no significativos, ya que las variables explicativas del modelo de variables latentes eran dominadas por los indicadores de percepción (para un análisis más completo de este tipo de modelos, ver capítulo 6).

Tabla 4.11 – Resultados de la Estimación de los Modelos de Elección Discreta

Parámetro	Modelo 1		Modelo 2		Modelo 3		Modelo 4		Modelo 5	
	Valor	test-t								
<i>ASC_{Auto}</i>	0,429	4,08	-0,417	-4,51	0,492	4,21			0,449	3,51
<i>ASC_{Bus}</i>	1,36	11,22	0,729	6,61	1,43	10,34	0,976	7,63	1,52	10,53
<i>Com_{Auto}</i>	0,614	18,61	0,567	3,83	0,59	3,92				
<i>Com_{Bus}</i>	0,626	16,79	2,31	2,74	2,59	3,03				
<i>Com_{Metro}</i>	0,651	20,86	0,546	2,96	0,569	3,04				
<i>Conf_{Auto}</i>	0,593	15,01	0,945	4,65	1,08	5,22				
<i>Conf_{Bus}</i>	0,586	13,79	-1,23	-1,93	-1,64	-2,08				
<i>Conf_{Metro}</i>	0,591	10,4	1,54	3,63	1,81	4,21				
<i>Costo_{Auto}</i>	-0,913	-73,87	-0,865	-73,33	-0,913	-73,89	-0,924	-72,17	-1,07	-74,64
<i>Costo_{Bus}</i>	-0,948	-22,09	-0,914	-20,68	-0,948	-21,83	-0,949	-21,97	-1,06	-24,49
<i>Costo_{Metro}</i>	-0,912	-29,7	-0,879	-28,44	-0,912	-29,57	-0,975	-21,46	-1,08	-33,52
<i>Sec_{Auto}</i>	0,461	28,51	0,424	3,82	0,432	3,84				
<i>Sec_{Bus}</i>	0,463	20,87	0,172	0,57	0,099	0,33				
<i>Sec_{Metro}</i>	0,464	16,82	0,336	2,58	0,326	2,49				
<i>Tiempo_{Auto}</i>	-0,606	-41,42	-0,579	-40,7	-0,606	-41,4	-0,606	-41,81	-0,714	-44
<i>Tiempo_{Bus}</i>	-0,628	-33,94	-0,604	-30,51	-0,628	-33,13	-0,621	-32,82	-0,723	-40,47
<i>Tiempo_{Metro}</i>	-0,612	-33,74	-0,591	-32,11	-0,612	-33,55	-0,616	-34,73	-0,731	-38,5
<i>Aire Aco_{Auto}</i>					0,296	9,76	0,414	7,9	0,04	1,15
<i>Aire Aco_{Bus}</i>					0,616	15,1	0,617	15,13	0,032	0,71
<i>Aire Aco_{Metro}</i>					0,478	16,25	0,481	16,33	0,082	2,47
<i>Cam Sec_{Bus}</i>					0,536	14,94	0,583	12,77	0,058	1,46
<i>Cam Sec_{Metro}</i>					0,361	10,6	0,492	7,63	0,022	0,58
<i>Dens_{Bus}</i>					-0,103	-8,09	-0,12	-7,43	-0,001	-0,72
<i>Dens_{Metro}</i>					-0,094	-11,33	-0,133	-8,56	-0,012	-1,32
<i>Ind Acc_{Auto}</i>					-0,378	-26,94	-0,377	-27,28	-0,033	-1,73
<i>Ind Acc_{Bus}</i>					-0,391	-16,11	-0,392	-16,14	-0,015	-0,54
<i>Ind Acc_{Metro}</i>					-0,485	-4,1	-1	-3,99	0,185	1,44
<i>P Info_{Bus}</i>					0,485	13,59	0,535	11,51	0,102	2,3
<i>P Info_{Metro}</i>					0,359	7,84	0,163	1,97	0,011	0,22
<i>Ing Alt_{Auto}</i>							0,65	9,63	0,069	1,08
<i>Ing Alto_{Bus}</i>							-1,45	-13,9	-0,144	-2,31
<i>Ing Med_{Auto}</i>							0,496	8,93	0,081	1,78
<i>Ing Med_{Bus}</i>							-0,813	-16,96	-0,073	-1,77
<i>Edad_{Auto}</i>							2,14	23,33	0,234	2,45
<i>Edad_{Bus}</i>									-0,123	-1,31
<i>Ed Med_{Auto}</i>							0,408	3,83	0,043	0,66
<i>Ed Med_{Bus}</i>							-0,276	-2,36	-0,1	-1,72
<i>Ed Sup_{Auto}</i>									-0,075	-0,9
<i>Ed Sup_{Bus}</i>							-0,761	-9,49	-0,094	-1,25
<i>Sexo_{Auto}</i>									0,051	1,17
<i>Sexo_{Bus}</i>							-1,37	-16,28	-0,065	-1,65
ϕ	0,893	3,56	0,935	1,99	0,863	3,47	0,893	3,48	0,84	6,4

La tabla muestra que en el caso de los modelos 2 y 3, el parámetro asociado a la confiabilidad del modo bus presenta un signo equivocado, con una alta significancia estadística. Esta evidencia debiese ser suficiente para descartar la validez de estas

especificaciones, ya que indica que las características socio-económicas afectarían la elección en forma opuesta a la simulada. En consecuencia, la inclusión directa de variables propias de las alternativas (que son parte del modelo de variables latentes) en el modelo de elección discreta, puede derivar en una inconsistencia interna del modelo estimado.

Es interesante notar que los parámetros asociados al costo y al tiempo de viaje son similares en todos los modelos, lo que ratifica que estas variables fueron consideradas de forma adecuada en todos los modelos.

Finalmente, es importante mencionar que el uso conjunto de los indicadores de percepción y de las variables explicativas del modelo MIMIC (en el Modelo 5) en el modelo de elección discreta, permite atenuar el efecto de la variabilidad inducida por medio de las variables latentes. De esta forma, los parámetros del modelo de elección discreta (léase las constantes modales y los parámetros *Costo*, *Tiempo* y ϕ) son recuperados con una mínima desviación.

Con objeto de comparar la capacidad predictiva de los distintos modelos y dado que éstos obedecen a especificaciones completamente independientes y, por tanto, salvo en el caso del Modelo 2 y el Modelo 3, no pueden ser expresados como una versión restringida de otro, se recurrió a la muestra de validación, generada en conjunto con el banco de datos (ver sección 3.1).

El test χ^2_{FPR} de banda de probabilidad predicha (Ortúzar y Willumsen, 2001) revela que todos los modelos son consistentes con los datos. La Tabla 4.12 presenta indicadores estadísticos del ajuste de los diferentes modelos a la muestra de validación, como *log-verosimilitud*, *test de verosimilitud directa (DLRT)*, ρ^2 y $\bar{\rho}^2$ (ver Ortúzar y Willumsen, 2001). En particular, el estadístico *DLRT* indica que el Modelo 1 se ajusta 1,89 veces mejor a los datos que el Modelo 3 y así sucesivamente.

Tabla 4.12 – Indicadores de Bondad de Ajuste de los Modelos

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
<i>Log- Verosimilitud (θ)</i>	-3696,5	-3913,9	-3697,1	-3701,8	-3170,8
ρ^2	0,327	0,287	0,327	0,326	0,423
$\bar{\rho}^2$	0,304	0,263	0,304	0,303	0,403
<i>DLRT</i>	-	2,58E+94	1,89	214,73	5,3E-229

La Tabla 4.12 muestra que no existe mayor diferencia entre el ajuste obtenido por el Modelo 1 y el Modelo 3. Ambos presentan un mejor ajuste que el Modelo 4, que no considera a los indicadores de percepción. El Modelo 2, que no toma en consideración a los atributos tangibles propios de las alternativas (tal como se estiman habitualmente los modelos híbridos de elección discreta), presenta un ajuste sustancialmente inferior. Esto provee suficiente evidencia para concluir que estos elementos deben ser tomados en consideración, al menos en un caso como el presentado a través de nuestros datos.

Aunque el ajuste obtenido por el Modelo 4 es inferior al obtenido por el Modelo 1, es mejor de lo que se pudiese esperar *a priori*. De todas formas, se debe tomar en cuenta que este modelo considera un número muy elevado de variables explicativas, muchas de las cuales son variables de interacción. Lo anterior implica que tanto su utilización como interpretación puede devenir en un proceso bastante dificultoso. Por tanto, el criterio de parsimonia en conjunto con la peor bondad de ajuste del Modelo 4, proveen suficientes razones para preferir el Modelo 1.

El Modelo 3 obtiene un ajuste similar al Modelo 1; sin embargo, tal como se indicó en forma previa, existen elementos suficientes para descartar su utilización. Un modelo mal especificado puede obtener un ajuste tan bueno como un modelo correctamente especificado, pero puede inducir a una interpretación incorrecta de los resultados. En consecuencia, es importante evaluar en forma adecuada cómo las variables explicativas

de un modelo afectan el proceso de elección de los individuos e intentar replicar éste de la manera más certera posible.

Finalmente el Modelo 5 presenta una bondad de ajuste extremadamente superior al resto de los modelos. Dada la estructura del banco de datos es difícil discernir el origen de esta mejora, ya que el nivel relativo del error asociado a las ecuaciones estructurales del modelo de variables latentes es superior al nivel relativo del error asociado a las ecuaciones de medición (notar que esto no trae problemas al comparar entre sí a los otros modelos). Por tanto, en el capítulo 7 se analizará el ajuste obtenido por un modelo de estas características utilizando un banco de datos real.

4.6 Inclusión de la Variabilidad de las Variables Latentes

Tal como se mencionó en el acápite f) de la sección 4.1, se estimó modelos ML que permiten considerar la variabilidad asociada a las variables latentes en el modelo de elección discreta y no sólo su valor esperado. De esta forma, se estimó modelos según las especificaciones a), b) y c), considerando la variabilidad de las variables latentes. Estos modelos serán llamados Modelo 1A, Modelo 2A y Modelo 3A, respectivamente.

Las Tablas 4.13 a 4.15 presentan los resultados de estos modelos, contrastados con sus contrapartes que no consideran variabilidad. En las primeras tres columnas se presenta el valor estimado para los distintos parámetros, su desviación estándar y su test-t. Posteriormente se presentan los mismos indicadores para el modelo considerando la variabilidad. Finalmente se presenta un test-t, cuya hipótesis nula corresponde a la igualdad entre los parámetros estimados por ambos modelos.

Tabla 4.13 – Comparación entre el Modelo 1 y el Modelo 1A

Parámetro	Modelo 1			Modelo 1A			Diferencia
	Valor	Des. Est.	test-t	Valor	Des. Est.	test-t	test-t
ASC_{Auto}	0,429	0,105	4,08	0,436	0,106	4,11	-0,047
ASC_{Bus}	1,36	0,121	11,22	1,37	0,123	11,16	-0,058
Com_{Auto}	0,614	0,033	18,61	0,616	0,0331	18,62	-0,043
Com_{Bus}	0,626	0,0373	16,79	0,624	0,0372	16,79	0,038
Com_{Metro}	0,651	0,0312	20,86	0,653	0,0314	20,77	-0,045
$Conf_{Auto}$	0,593	0,0395	15,01	0,595	0,0394	15,08	-0,036
$Conf_{Bus}$	0,586	0,0425	13,79	0,584	0,0424	13,76	0,033
$Conf_{Metro}$	0,591	0,0568	10,4	0,598	0,0582	10,27	-0,086
$Costo_{Auto}$	-0,913	0,0124	-73,87	-0,913	0,0124	-73,85	0
$Costo_{Bus}$	-0,948	0,0429	-22,09	-0,949	0,043	-22,06	0,016
$Costo_{Metro}$	-0,912	0,0307	-29,7	-0,912	0,0308	-29,61	0
Sec_{Auto}	0,461	0,0162	28,51	0,461	0,0162	28,55	0
Sec_{Bus}	0,463	0,0222	20,87	0,462	0,0221	20,89	0,032
Sec_{Metro}	0,464	0,0276	16,82	0,466	0,028	16,63	-0,051
$Tiempo_{Auto}$	-0,606	0,0146	-41,42	-0,606	0,0146	-41,41	0
$Tiempo_{Bus}$	-0,628	0,0185	-33,94	-0,628	0,0186	-33,84	0
$Tiempo_{Metro}$	-0,612	0,0182	-33,74	-0,613	0,0182	-33,61	0,039
ϕ	0,893	0,0301	3,56	0,893	0,0304	3,52	0

Como se puede observar, no existen mayores diferencias entre los parámetros estimados siguiendo la especificación utilizada en la generación de la base de datos (Modelos 1 y 1A). De hecho, ninguno de los parámetros estimados por ambos modelos difiere en más de 0,01 unidades en el test-t.

Estos resultados son esperables, ya que ambos modelos obedecen a la especificación correcta y por tanto, el manejo del error es adecuado en ambos casos (el error es

completamente estocástico). Luego, la inclusión de una mayor complejidad, en términos del manejo del error, no debiese afectar a los parámetros estimados.

Tabla 4.14 – Comparación entre el Modelo 2 y el Modelo 2A

Parámetro	Modelo 2			Modelo 2A			Diferencia
	Valor	Des. Est.	test-t	Valor	Des. Est.	test-t	test-t
<i>ASC_{Auto}</i>	-0,417	0,0926	-4,51	-0,428	0,0936	-4,57	0,084
<i>ASC_{Bus}</i>	0,729	0,11	6,61	0,8	0,111	7,22	-0,454
<i>Com_{Auto}</i>	0,567	0,148	3,83	0,629	0,145	4,34	-0,299
<i>Com_{Bus}</i>	2,31	0,845	2,74	2,64	0,824	3,21	-0,280
<i>Com_{Metro}</i>	0,546	0,184	2,96	0,663	0,156	4,25	-0,485
<i>Conf_{Auto}</i>	0,945	0,203	4,65	1,17	0,186	6,29	-0,817
<i>Conf_{Bus}</i>	-1,23	0,638	-1,93	-2,14	0,598	-3,58	1,041
<i>Conf_{Metro}</i>	1,54	0,424	3,63	2,08	0,377	5,53	-0,952
<i>Costo_{Auto}</i>	-0,865	0,0118	-73,33	-0,88	0,013	-67,9	0,854
<i>Costo_{Bus}</i>	-0,914	0,0442	-20,68	-0,95	0,047	-20,21	0,558
<i>Costo_{Metro}</i>	-0,879	0,0309	-28,44	-0,904	0,033	-27,42	0,553
<i>Sec_{Auto}</i>	0,424	0,111	3,82	0,406	0,0976	4,16	0,122
<i>Sec_{Bus}</i>	0,172	0,3	0,57	-0,005	0,267	-0,02	0,44
<i>Sec_{Metro}</i>	0,336	0,13	2,58	0,326	0,12	2,71	0,057
<i>Tiempo_{Auto}</i>	-0,579	0,0142	-40,7	-0,589	0,0148	-39,75	0,488
<i>Tiempo_{Bus}</i>	-0,604	0,0198	-30,51	-0,627	0,0215	-29,2	0,787
<i>Tiempo_{Metro}</i>	-0,591	0,0184	-32,11	-0,607	0,0198	-30,69	0,592
ϕ	0,935	0,033	1,99	0,781	0,071	3,06	-2,161

Más interesante es analizar las diferencias existentes para las especificaciones que consideran el modelo MIMIC simplificado (considerando únicamente variables socio-económicas como variables explicativas). En estos casos, al no estar correctamente especificado el modelo, el término de error estará compuesto por elementos puramente

estocásticos y por elementos provenientes del error inducido por medio de la especificación. Por tanto, una mejor especificación del término de error debiese posibilitar una mejor estimación de los parámetros.

Dado que en la realidad nunca se conocerá con exactitud la especificación correcta, el término de error asociado a estas especificaciones representa de forma más verosímil el error existente en un caso real. Por ello, estas especificaciones (Modelos 2 y 2A y Modelos 3 y 3A) debieran permitir analizar de mejor forma las ventajas de considerar la variabilidad asociada a las variables latentes en la estimación simultánea.

En el caso de los Modelos 2 y 2A, si bien no existe diferencia estadística, al 95%, para rechazar la igualdad de los parámetros estimados por uno y otro modelo, salvo en el caso del parámetro ϕ , sí existen importantes diferencias en términos relativos. De esta manera, en forma agregada, ambos modelos presentan diferencias sustanciales. En el caso de los Modelos 3 y 3A, presentados en la Tabla 4.15, la situación es similar. Si bien, en este caso, los parámetros estimados por uno y otro modelo presentan una mayor similitud que en el caso anterior - en parte debido a la mejor especificación del término de error, producto del trabajo con una mayor cantidad de información relevante - en términos agregados, ambas estimaciones constituyen modelos diferentes.

El problema relativo a los signos incorrectos - asociados a parámetros estimados de acuerdo a especificaciones que consideran el modelo MIMIC simplificado - mencionado en la sección 4.5 se mantiene, e incluso se acentúa, ya que la significancia estadística de estos parámetros es mayor. Más aún, en el Modelo 2A, el estimador asociado a la seguridad del bus también presenta un signo incorrecto (aunque con baja significancia estadística).

Tabla 4.15 – Comparación entre el Modelo 3 y el Modelo 3A

Parámetro	Modelo 3			Modelo 3A			Diferencia
	Valor	Des. Est.	test-t	Valor	Des. Est.	test-t	test-t
<i>ASC_{Auto}</i>	0,492	0,117	4,21	0,509	0,119	4,28	-0,102
<i>ASC_{Bus}</i>	1,43	0,138	10,34	1,48	0,139	10,71	-0,255
<i>Com_{Auto}</i>	0,59	0,151	3,92	0,613	0,147	4,17	-0,109
<i>Com_{Bus}</i>	2,59	0,853	3,03	3,05	0,832	3,66	-0,386
<i>Com_{Metro}</i>	0,569	0,187	3,04	0,588	0,158	3,73	-0,078
<i>Conf_{Auto}</i>	1,08	0,207	5,22	1,24	0,19	6,55	-0,569
<i>Conf_{Bus}</i>	-1,64	0,788	-2,08	-2,36	0,724	-3,26	0,673
<i>Conf_{Metro}</i>	1,81	0,431	4,21	2,23	0,387	5,76	-0,725
<i>Costo_{Auto}</i>	-0,913	0,0124	-73,89	-0,925	0,0134	-69,02	0,657
<i>Costo_{Bus}</i>	-0,948	0,0434	-21,83	-0,969	0,0448	-21,6	0,337
<i>Costo_{Metro}</i>	-0,912	0,0308	-29,57	-0,935	0,0329	-28,39	0,51
<i>Sec_{Auto}</i>	0,432	0,112	3,84	0,352	0,0998	3,53	0,533
<i>Sec_{Bus}</i>	0,0995	0,3	0,33	0,0471	0,265	0,18	0,131
<i>Sec_{Metro}</i>	0,326	0,131	2,49	0,253	0,121	2,09	0,409
<i>Tiempo_{Auto}</i>	-0,606	0,0146	-41,4	-0,614	0,0152	-40,52	0,38
<i>Tiempo_{Bus}</i>	-0,628	0,019	-33,13	-0,642	0,0198	-32,37	0,51
<i>Tiempo_{Metro}</i>	-0,612	0,0183	-33,55	-0,627	0,0196	-31,93	0,559
<i>Aire Aco_{Auto}</i>	0,296	0,0303	9,76	0,3	0,0307	9,76	-0,093
<i>Aire Aco_{Bus}</i>	0,616	0,0408	15,1	0,63	0,0421	14,98	-0,239
<i>Aire Aco_{Metro}</i>	0,478	0,0294	16,25	0,489	0,0305	16	-0,26
<i>Cam Sec_{Bus}</i>	0,536	0,0359	14,94	0,548	0,0369	14,83	-0,233
<i>Cam Sec_{Metro}</i>	0,361	0,0341	10,6	0,369	0,0351	10,53	-0,163
<i>Dens_{Bus}</i>	-0,103	0,0127	-8,09	-0,104	0,013	-8,04	0,055
<i>Dens_{Metro}</i>	-0,094	0,0083	-11,33	-0,095	0,0085	-11,17	0,126
<i>Ind Acc_{Auto}</i>	-0,378	0,014	-26,94	-0,382	0,0143	-26,7	0,2
<i>Ind Acc_{Bus}</i>	-0,391	0,0243	-16,11	-0,399	0,025	-16	0,229
<i>Ind Acc_{Metro}</i>	-0,485	0,118	-4,1	-0,494	0,121	-4,06	0,053
<i>P Info_{Bus}</i>	0,485	0,0357	13,59	0,496	0,0367	13,51	-0,215
<i>P Info_{Metro}</i>	0,359	0,0458	7,84	0,367	0,047	7,8	-0,122
ϕ	0,893	0,031	3,47	0,813	0,058	3,21	-1,4

Con el objetivo de analizar la bondad de ajuste de estos modelos se recurrió nuevamente a la muestra de validación. El test χ^2_{FPR} de banda probabilidad predicha indica que los modelos son consistentes con los datos. La Tabla 4.16 presenta estos resultados:

Tabla 4.16 – Indicadores de Bondad de Ajuste de los Modelos A

	Modelo 1	Modelo 1A	Modelo 2	Modelo 2A	Modelo 3	Modelo 3A
<i>Log- Verosimilitud (θ)</i>	-3696,5	-3696,6	-3913,9	-3918,1	-3697,1	-3696,8
ρ^2	0,327	0,327	0,287	0,287	0,327	0,327
$\bar{\rho}^2$	0,304	0,304	0,263	0,262	0,304	0,304
<i>DLRT</i>	-	1,11	2,60E+94	1,74E+96	1,82	1,35

Como se puede observar, no existe una evidencia clara que respalde las ventajas de considerar la variabilidad asociada a las variables latentes al estimar los modelos de elección discreta. Si bien, en el caso del Modelo 3A se observa un ajuste levemente superior al obtenido por el modelo que no considera la variabilidad, en el caso de los Modelos 1A y 2A la inclusión de la variabilidad en la estimación tiene un efecto negativo sobre el ajuste obtenido.

5. ANALISIS DE RESPUESTA

Dado que en la práctica, uno de los objetivos principales de la calibración de modelos de elección discreta es su utilización para predicción y evaluación de diferentes políticas y medidas, es fundamental saber cómo se comportan cuando son enfrentados a escenarios diferentes a la situación base para la cual fueron calibrados. En los capítulos anteriores nos hemos centrado en problemas relativos a la formulación y estimación de los modelos (tal como es habitual en la literatura), por lo que en este capítulo analizaremos los modelos desde el punto de vista de la flexibilidad que ofrecen para tratar nuevos escenarios.

De esta forma se analizará el comportamiento de los diferentes modelos calibrados en el capítulo anterior frente a cambios producidos en la población o en el nivel de servicio de las distintas alternativas, según la tradición iniciada por Williams y Ortúzar (1982).

5.1 Escenarios Analizados

Dada las características de la base de datos, se pueden distinguir tres tipos de variables. En primer lugar se tienen las variables propias de las alternativas que intervienen directamente en el modelo de elección, las cuales fueron correctamente especificadas en todos los modelos estimados. Luego se tienen las variables propias de las alternativas que intervienen en la generación de las variables latentes, especificadas en forma correcta únicamente en el Modelo 1 y finalmente se tienen las características socio-económicas de los individuos, las cuales forman parte del modelo de elección. Las características socio-económicas fueron especificadas correctamente en todos los

modelos que consideran variables latentes; sin embargo, la omisión de las variables propias de las alternativas en el modelo MIMIC simplificado, tiene un efecto directo sobre la calibración de las variables latentes, tal como se indicó en el capítulo 4.

Dado el elevado número de variables que intervienen en el modelo, se plantearon siete escenarios diferentes, que consideran la variación de una o dos de estas variables. La variación doble se produce al alterar variables cualitativas caracterizadas por dos variables dicotómicas (como el ingreso) o al alterarse características del transporte público y que, por tanto, afectan a las dos alternativas correlacionadas. El resto se considera *ceteris paribus*. Los escenarios estudiados se presentan en la Tabla 5.1.

Tabla 5.1 – Escenarios Analizados

	Costo Auto	Tiempo Metro	Ing. Alto	Ing. Med.	Ed. Sup.	Ed. Med.	Pan. Info. Bus	Dens. Metro	Dens. Bus	Air. Aco. Metro	Air. Aco. Bus
Esc. 1	20%										
Esc. 2		-20%									
Esc. 3			30%	20%							
Esc. 4					50%	-8%					
Esc. 5							100%				
Esc. 6								-30%	-30%		
Esc. 7										100%	100%

Como se puede observar, los dos primeros escenarios consideran variaciones de las características propias de las alternativas que intervienen directamente en el modelo de elección. Los escenarios 3 y 4, consideran variaciones de las características socio-económicas de los individuos y finalmente los escenarios 5, 6 y 7 consideran

alteraciones respecto a los atributos tangibles propios de las alternativas que participan en la generación de las variables latentes.

5.2 Indicadores de Respuesta

En el contexto de un análisis de respuesta existen diversas alternativas e indicadores que permiten analizar el ajuste obtenido por distintos modelos. Una de las principales limitantes, al realizar este tipo de análisis, acostumbra ser que no se conoce el comportamiento de los individuos bajo las nuevas condiciones. Afortunadamente, al trabajar con datos sintéticos se conocen las elecciones en el nuevo escenario, lo que permite utilizar indicadores que contrastan éstas con las predicciones de los modelos.

En este trabajo se utilizaron dos tipos de indicadores. En primer lugar, por un tema de consistencia en relación con el análisis desarrollado en el capítulo 5, se utilizó la log-verosimilitud obtenida por los modelos en los distintos escenarios. En segundo lugar, se utilizó un indicador de sesgo predictivo basado en la diferencia porcentual entre las predicciones hechas por el modelo y los valores reales simulados, formulado como un indicador χ^2 que permite comparar la capacidad de predicción de respuesta en términos globales (Munizaga, 1997). El indicador queda definido como:

$$\chi^2 = \sum_i \frac{(\tilde{N}_i - N_i)^2}{N_i} \quad (5.1)$$

en que \tilde{N}_i es el número de individuos que opta por la alternativa i de acuerdo a la predicción del modelo y N_i es el número real de individuos que opta por la alternativa en cuestión. Si bien el índice es formulado como un indicador χ^2 , el nivel de aceptación

depende de la magnitud o escala de valores, de la varianza de las estimaciones (Munizaga, 1997) y del número de observaciones (es un indicador dimensional).

5.2.1 Análisis de Respuesta por Nivel de Ajuste

La Tabla 5.2 presenta los resultados de la log-verosimilitud calculada en los distintos escenarios.

Tabla 5.2 – Log-Verosimilitud Calculada por Escenario

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Esc. 1	-3561,6	-3733,2	-3562,9	-3569,2	-3056,3
Esc. 2	-3675,9	-3844,6	-3676,4	-3678,4	-3216,4
Esc. 3	-3555,5	-3787,3	-3557,1	-3564,2	-3062,0
Esc. 4	-3703,3	-3933,4	-3706,3	-3709,6	-3254,6
Esc. 5	-3713,3	-3924,5	-3717,9	-3720,7	-3219,7
Esc. 6	-3700,7	-3919,8	-3701,6	-3702,0	-3196,3
Esc. 7	-3650,0	-3831,6	-3651,9	-3659,7	-3183,9

Del análisis de la tabla se desprende que la variación de una sola variable no tiene un efecto significativo sobre la log-verosimilitud obtenida por el modelo. Esto es esperable, ya que en el modelo híbrido intervienen 69 variables diferentes sólo en el modelo de variables latentes y 17 en el modelo de elección discreta, por lo que la perturbación de una de éstas no debiese afectar de manera determinante la decisión.

Sin desmedro de lo anterior, la Tabla 5.2 permite observar ciertas tendencias. En primer lugar, se tiene que al comparar las verosimilitudes promedio, obtenidas por los diversos modelos, en los escenarios 1 y 2, no existen mayores diferencias con respecto a la

verosimilitud obtenida en el caso base (muestra de validación). En cambio, al comparar la verosimilitud promedio, de los distintos modelos en los escenarios 3 y 4 y en los escenarios 5, 6 y 7, se observan diferencias sustanciales con respecto al escenario base. De esta forma, considerando el promedio de los escenarios 3 y 4, el Modelo 1 se ajusta a los datos 12,52 y 3273,74 veces mejor que los Modelos 3 y 4, respectivamente (en el escenario base el ajuste era 1,89 y 214,73 veces mejor). Al considerar el promedio de los escenarios 5, 6 y 7, se observa que el Modelo 1 tiene un ajuste 36,21 y 5985,75 veces mejor que los modelos antes mencionados.

Los resultados presentados son esperables, ya que la perturbación inducida, en el caso de los escenarios 1 y 2, se produce sobre variables que se encuentran correctamente especificadas en todos los modelos (y por ello potencialmente asociadas a parámetros bien estimados). Es por tanto esperable, que cada modelo pueda responder de manera adecuado a cambios producidos sobre estas variables. En los otros escenarios, en cambio, las variables sujetas a variación sólo se encuentran especificadas correctamente en el caso del modelo 1, ya que si bien las variables socio-económicas se encuentra bien especificadas en los Modelos 2 y 3, éstas deben explicar la variabilidad asociada a los factores tangibles propios de las alternativas, provocando que la estimación sea incorrecta. Debido a lo anterior, no es sorprendente que el ajuste de los Modelos 3 y 4 disminuya respecto del ajuste del Modelo 1, al utilizarse estos modelos en escenarios distintos al escenario base para el cual fueron calibrados.

Para corroborar lo anterior, se analizaron dos nuevos escenarios, en los cuales la perturbación de los atributos asociados a las alternativas y a los individuos no sucede en condiciones *ceteris paribus*. Esto con el fin de evitar que la perturbación de una única variable sea absorbida por la estabilidad de las otras. El escenario 8 corresponde a la suma de las alteraciones de los escenarios 3, 4, 5, 6 y 7, mientras que el escenario 9

corresponde a una situación exagerada, que disminuye de manera radical la calidad del servicio ofrecido por el transporte público (especialmente por el bus). Los nuevos escenarios se presentan en la Tabla 5.3.

Tabla 5.3 – Nuevos Escenarios Analizados

	Ing. Alto	Ing. Med	Ed. Sup.	Ed. Med	Pan. Info. Bus	Dens. Metro	Dens. Bus	Air. Aco. Metro	Air. Aco. Bus	Ind. Acc. Bus	Cam Sec. Bus	Edad
Esc. 8	30%	20%	50%	-8%	100%	-30%	-30%	100%	100%			
Esc. 9					0	50%	50%		0	100%	0	-30%

La log-verosimilitud calculada en los nuevos escenarios, para los distintos modelos, es presentada en la Tabla 5.4.

Tabla 5.4 – Log-Verosimilitud Calculada en los Nuevos Escenarios

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Esc. 8	-3638,7	-3845,8	-3643,2	-3650,2	-3136,1
Esc. 9	-2631,2	-4815,1	-2768,5	-2879,2	-2442,4

Esta tabla corrobora la hipótesis planteada previamente. Mientras mayor es la perturbación respecto al escenario base, peor es el ajuste obtenido por los modelos especificados en forma incorrecta, respecto del ajuste obtenido por el modelo correcto. Esto es sumamente importante, ya que en la práctica los modelos consideran un número bastante menor de variables, por lo que la alteración de sólo una de éstas, tendrá un mayor efecto relativo sobre la alteración del escenario base.

Otro hecho importante es la degradación de la verosimilitud, de los Modelos 2 y 5, respecto del Modelo 1, en el escenario 9, que no era observable en los escenarios anteriores. También se debe destacar el enorme descenso de la verosimilitud asociada al Modelo 2, lo que es esperable, ya que es insensible a variaciones de los atributos tangibles de las alternativas que dan lugar a las variables latentes.

Todo lo anterior constituye evidencia irrefutable de la importancia de trabajar con modelos correctamente especificados.

5.2.2 Análisis de Respuesta por Sesgo Predictivo

Tal como se mencionó anteriormente, no sólo se analizó el ajuste obtenido por los modelos, sino que también el sesgo predictivo. Los resultados del indicador χ^2 definido en (5.1) son presentados en la Tabla 5.5.

Tabla 5.5 – Indicador χ^2 por Escenario

	Modelo 1	Modelo 2	Modelo 3	Modelo 4	Modelo 5
Esc. 1	4,81	1,91	4,41	3,85	25,35
Esc. 2	18,58	14,92	13,22	17,90	59,86
Esc. 3	3,84	2,41	3,34	2,97	36,66
Esc. 4	3,68	3,76	3,36	3,42	47,70
Esc. 5	2,53	6,21	2,31	1,95	31,75
Esc. 6	1,63	4,23	1,96	1,72	40,14
Esc. 7	1,66	0,95	1,81	2,02	42,60
Esc. 8	1,07	28,36	0,98	0,43	41,89
Esc. 9	4,28	4845,99	231,89	92,95	251,28

Como se puede observar, salvo en el caso del escenario 9 este indicador no ofrece evidencia que permita discernir entre los Modelos 1, 2, 3 y 4. Sin embargo, no se puede asegurar que esta situación se repita en otros casos, por lo que no se pueden hacer inferencias al respecto. En caso de variaciones mayores (escenario 9), el sesgo predictivo del modelo bien especificado permanece inalterado, degradándose la capacidad predictiva de los otros modelos.

La principal conclusión que nos aporta este indicador, es el enorme sesgo predictivo asociado al Modelo 5. Si bien este modelo presenta claramente el mejor ajuste, su enorme sesgo predictivo nos impide aceptarlo como un modelo válido. Esta consideración reviste suma importancia, ya que al utilizar estos modelos para predicción, el hecho de que predigan en forma acertada es más importante que el hecho de que obtengan un buen ajuste.

6. ANALISIS DE MODELOS DE ELECCION QUE CONSIDERAN INDICADORES DE PERCEPCION EN BASE A DATOS REALES

El análisis realizado en los capítulos 4 y 5 de esta tesis nos ha permitido discernir con certeza entre los Modelos 1, 2, 3 y 4 y ha ratificado la importancia de trabajar con modelos correctamente especificados. Sin embargo, la capacidad predictiva de los modelos que consideran a los indicadores de percepción directamente en el modelo de elección continúa siendo una incógnita.

Estos modelos presentaron un extraordinariamente alto nivel de ajuste, en comparación con los otros modelos – aunque como se mencionó anteriormente no se pudo establecer si eso se debía a la estructura particular del banco de datos generado o a un fenómeno estructural. No obstante, también se observó un importante sesgo predictivo. Todo esto se suma a las consideraciones teóricas presentadas en la sección 4.1, que indican la conveniencia de no utilizar este tipo de modelos.

Por estos motivos, en este capítulo se estudia el comportamiento de modelos que incorporan directamente los indicadores de percepción utilizando un banco de datos real. Este corresponde a una encuesta de preferencias reveladas, en la cual se presentó a los individuos dos alternativas (tren y bus) en el contexto de viajes interurbanos realizados entre la ciudad de Santiago y las ciudades de Curicó, Talca, Linares y Chillán. La encuesta fue llevada a cabo en los terminales interurbanos de Santiago. El tamaño muestral es de 600 individuos, de los cuales 301 optaron por el bus, mientras que 299 optaron por el tren (Egaña, 2009).

Utilizando este banco de datos, Egaña (2009) calibró diferentes modelos de elección discreta. Tras considerar diferentes formulaciones, determinó que la mejor especificación correspondía a un modelo híbrido de elección discreta, cuyo modelo MIMIC es presentado en la Figura 6.1.

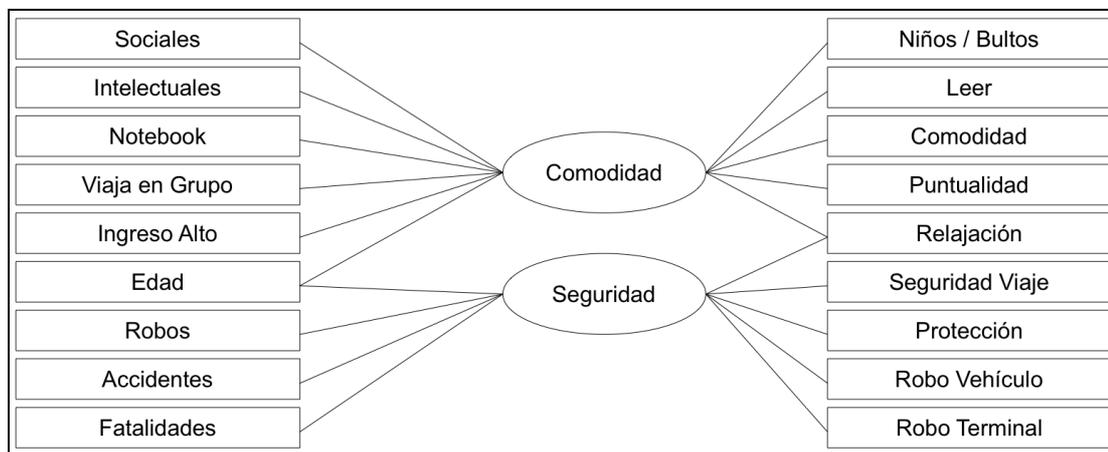


Figura 6.1 – Estructura del Modelo estimado por Egaña (2009)

A la izquierda de la figura se presentan las variables explicativas exógenas del modelo, mientras que a la derecha se presentan los indicadores de percepción, de carácter endógeno; las variables *Comodidad* y *Seguridad*, corresponden a las variables latentes estimadas. Las variables *Sociales* e *Intelectuales* son variables dicotómicas y se relacionan con las actividades desarrolladas por los individuos durante el viaje. La primera indica la presencia de un individuo que interactúa con los demás pasajeros, mientras que la segunda se refiere a individuos que desarrollan actividades tales como leer, trabajar o estudiar. La variable *Notebook* indica que el individuo viaja con algún dispositivo computacional portátil (*netbook* o *laptop*) e *Ingreso Alto* es una variable binaria, indicando un ingreso por sobre los \$400.000; *Viaja en Grupo* es caracterizado en forma dicotómica, mientras que *Edad* es representado de forma continua. Las variables *Robos* y *Accidentes* son binarias y toman el valor 1 cuando el individuo en cuestión ha sufrido o presenciado robos y accidentes, respectivamente. Finalmente *Fatalidades* corresponde a un índice de accidentes fatales asociado a la alternativa bus (no se registran fatalidades para la alternativa tren en el período estudiado), definido de la siguiente manera:

$$Fatalidade \quad s = \frac{\frac{1}{8} \sum_{2000}^{2007} Fatalidade \quad s \text{ Año} \left(\frac{pax}{año} \right) \cdot Distancia \quad Viaje \quad (Km)}{Distancia \quad Anual \left(\frac{Km \cdot veh}{año} \right) \cdot Pasajeros \quad por \quad Viaje \quad \left(\frac{pax}{veh} \right)} \quad (6.1)$$

Respecto a los indicadores de percepción, éstos corresponden a valoraciones asignadas por los individuos a distintas características de las alternativas, y toman un valor mínimo de 1 y un valor máximo de 7. Así, *Niños/Bultos* indica la percepción de la facilidad de viajar con niños o bultos, *Leer* la posibilidad de utilizar el tiempo para leer y *Comodidad* la percepción de comodidad durante el viaje. De la misma manera, *Puntualidad* indica la percepción de la puntualidad en el horario de la salida de la alternativa en cuestión y *Relajación* está asociada a la factibilidad de relajarse durante el viaje. Finalmente *Seguridad Viaje*, *Protección*, *Robo Vehículo* y *Robo Terminal* son indicadores de la percepción de la seguridad en el sentido de no sufrir accidentes durante el viaje, de la protección en caso de sufrir un accidente, de la seguridad ante robos en el vehículo y ante robos en el terminal, respectivamente. Adicionalmente, existe un décimo indicador de percepción, al cual llamaremos *Conveniencia*, que indica la conveniencia de los horarios de salida de las distintas alternativas. Este indicador no resultó relevante en explicar ninguna de las variables latentes utilizadas.

Utilizando el método secuencial, Egaña (2009) estimó los parámetros asociados al modelo de variables latentes, los cuales son presentados en la Tabla 6.1. Además, usando las variables latentes, construidas de acuerdo con la especificación planteada previamente, Egaña (2009) planteó la siguiente función de utilidad representativa:

$$V_{i,q} = \theta_i + (\theta_{Costo} + \theta_{Paga} \cdot Paga_q + \theta_{Categoría} \cdot Categoría_q) \cdot Costo_{i,q} + (\theta_{Tvia} + \theta_{Sexo} \cdot Sexo_q) \cdot Tvia_{i,q} + \beta_{Com} \cdot Com_{i,q} + \beta_{Sec} \cdot Sec_{i,q} \quad (6.2)$$

en que $Costo_{i,q}$ corresponde al costo de la alternativa y $Tvia_{i,q}$ al tiempo de viaje asociada a ésta. $Paga_q$ indica si el individuo paga por el pasaje, mientras que $Categoría_q$ indica si el individuo viaja en la categoría superior de la alternativa en cuestión. La variable dicotómica $Sexo_q$ toma el valor 1 para los individuos de sexo masculino.

Tabla 6.1 – Resultados de la Estimación del Modelo MIMIC, Egaña (2009)

Bus			
Variable Explicada	Variable Explicatoria	Valor Estimado	Test-t
<i>VL Comodidad</i>	<i>Edad</i>	0,007	2,4
<i>VL Comodidad</i>	<i>Ingreso Alto</i>	-0,425	-4,49
<i>VL Comodidad</i>	<i>Notebook</i>	-0,263	-1,17
<i>VL Comodidad</i>	<i>Viaja en Grupo</i>	0,333	2,87
<i>VL Comodidad</i>	<i>Sociales</i>	-0,24	-1,53
<i>VL Comodidad</i>	<i>Intelectuales</i>	-0,206	-2,19
<i>VL Seguridad</i>	<i>Edad</i>	0,01	3,31
<i>VL Seguridad</i>	<i>Robos</i>	-0,2	-1,88
<i>VL Seguridad</i>	<i>Accidentes</i>	-0,494	-4,4
<i>VL Seguridad</i>	<i>Fatalidades</i>	-0,232	-1,37
<i>Niños / Bultos</i>	<i>VL Comodidad</i>	0,717	13,73
<i>Leer</i>	<i>VL Comodidad</i>	0,782	14,46
<i>Comodidad</i>	<i>VL Comodidad</i>	0,965	19,9
<i>Puntualidad</i>	<i>VL Comodidad</i>	0,747	14,64
<i>Relajación</i>	<i>VL Comodidad</i>	0,713	16,02
<i>Relajación</i>	<i>VL Seguridad</i>	0,267	6,19
<i>Seguridad Viaje</i>	<i>VL Seguridad</i>	0,747	14,08
<i>Protección</i>	<i>VL Seguridad</i>	0,897	15,08
<i>Robo Vehículo</i>	<i>VL Seguridad</i>	0,898	14,58
<i>Robo Terminal</i>	<i>VL Seguridad</i>	0,964	13,52
Tren			
Variable Explicada	Variable Explicatoria	Valor Estimado	Test-t
<i>VL Comodidad</i>	<i>Edad</i>	0,005	1,85
<i>VL Comodidad</i>	<i>Ingreso Alto</i>	-0,009	-0,1
<i>VL Comodidad</i>	<i>Notebook</i>	0,379	3,04
<i>VL Comodidad</i>	<i>Viaja en Grupo</i>	0,464	3,89
<i>VL Comodidad</i>	<i>Sociales</i>	-0,166	-1,04
<i>VL Comodidad</i>	<i>Intelectuales</i>	0,286	2,97
<i>VL Seguridad</i>	<i>Edad</i>	0,007	2,28
<i>VL Seguridad</i>	<i>Robos</i>	-0,466	-2,24
<i>VL Seguridad</i>	<i>Accidentes</i>	-0,211	-1,34
<i>VL Seguridad</i>	<i>Fatalidades</i>	0,534	3,12
<i>Niños / Bultos</i>	<i>VL Comodidad</i>	0,636	14,79
<i>Leer</i>	<i>VL Comodidad</i>	0,542	13,2
<i>Comodidad</i>	<i>VL Comodidad</i>	0,644	17,06
<i>Puntualidad</i>	<i>VL Comodidad</i>	0,497	12,25
<i>Relajación</i>	<i>VL Comodidad</i>	0,499	15,16
<i>Relajación</i>	<i>VL Seguridad</i>	0,163	5,07
<i>Seguridad Viaje</i>	<i>VL Seguridad</i>	0,446	9,99
<i>Protección</i>	<i>VL Seguridad</i>	0,698	11,58
<i>Robo Vehículo</i>	<i>VL Seguridad</i>	0,881	15,99
<i>Robo Terminal</i>	<i>VL Seguridad</i>	0,887	13,74

Dado, que no es la intención de esta investigación, obtener el modelo más adecuado, sino que analizar el efecto de la inclusión directa de los indicadores de percepción en la función de utilidad, se asumió como válida la especificación (6.2), en lo que respecta a las variables tangibles (no latentes). A partir de esta estructura, se estimó dos tipos diferentes de modelos: en primer lugar se incorporó únicamente los indicadores de percepción (IP), en reemplazo de las variables latentes, tal como se indica en la ecuación (6.3); en segundo lugar se incorporó directamente tanto los indicadores de percepción, así como las variables explicativas (VE) del modelo MIMIC (ecuación 6.4).

$$V_{i,q} = \theta_i + (\theta_{Costo} + \theta_{Paga} \cdot Paga_q + \theta_{Categoría} \cdot Categoría_q) \cdot Costo_{i,q} + (\theta_{Tvia} + \theta_{Sexo} \cdot Sexo_q) \cdot Tvia_{i,q} + \sum_k \delta_k \cdot IP_{k,i,q} \quad (6.3)$$

$$V_{i,q} = \theta_i + (\theta_{Costo} + \theta_{Paga} \cdot Paga_q + \theta_{Categoría} \cdot Categoría_q) \cdot Costo_{i,q} + (\theta_{Tvia} + \theta_{Sexo} \cdot Sexo_q) \cdot Tvia_{i,q} + \sum_k \delta_k \cdot IP_{k,i,q} + \sum_l \kappa_l \cdot VE_{l,q} \quad (6.4)$$

En el caso de las variables socio-económicas y del intercepto, los estimadores son considerados fijos para la alternativa tren, para permitir las estimación. Los resultados de las estimaciones son presentados en la Tabla 6.2.

Como se puede observar, ciertos parámetros no son estadísticamente significativos; sin embargo éstos tampoco eran estadísticamente significativos en el modelo original. Debido a que las consideraciones que llevaron a mantener estos parámetros en el modelo, no son relevantes en nuestro análisis, se optó por mantener estas variables en todos los modelos, con el fin de que los distintos modelos fueran comparables entre sí.

Tabla 6.2 – Resultados de la Estimación de los Modelos de Elección Discreta

Parámetro	Egaña (2009)		Modelo IP		Modelo IP + VE	
	Valor	test-t	Valor	test-t	Valor	test-t
<i>ASC_{Bus}</i>	0,357	0,85	0,414	0,92	2,12	3,58
<i>Categoría</i>	0,925	9,09	0,917	8,11	0,885	7,15
<i>Comodidad</i>	1,53	5,68				
<i>Costo</i>	-0,754	-6,9	-0,577	-5,33	-0,686	-5,4
<i>Paga</i>	-0,176	-2,21	-0,226	-2,61	-0,255	-2,7
<i>Seguridad</i>	1,7	4,79				
<i>Sexo</i>	-0,001	-0,26	-0,005	-1,13	-0,005	-1,01
<i>Tiempo</i>	-0,018	-3,17	-0,012	-2	-0,006	-0,89
<i>IP Conveniencia</i>			0,306	3,91	0,315	3,35
<i>IP Leer</i>			0,334	2,83	0,375	2,89
<i>IP Relajación</i>			0,416	3,49	0,279	2,06
<i>IP Robo Terminal</i>			0,466	4,5	0,44	3,85
<i>VE Ingreso Alto</i>					-0,814	-2,65
<i>VE Viaja en Grupo</i>					-1,03	-3,4
<i>VE Sociales</i>					-1,07	-2,1
<i>VE Intelectuales</i>					-0,783	-2,61
<i>VE Fatalidades</i>					-3,74	-6,93

Dado que los modelos que consideran los indicadores de percepción directamente en la función de utilidad no incluyen variables latentes, no son directamente comparables con el modelo estimado por Egaña (2009). Ante la ausencia de una muestra de validación, la bondad de ajuste de los distintos modelos, fue comparada de la manera propuesta por Ashok *et al.* (2002). Esta se basa en el Criterio de Información Consistente de Akaike (CAIC, según su sigla en inglés), propuesto por Bozdogan (1987), que se define como:

$$CAIC = -2 \cdot l_M(\theta) + k_M \cdot (\ln n + 1) \quad (6.5)$$

en que $l_M(\theta)$ corresponde a la log-verosimilitud del modelo estimado, k_M es el número de parámetros calibrados y n corresponde al número de observaciones utilizadas en la estimación del modelo.

La bondad de ajuste de los modelos estimados anteriormente, incluyendo el índice *CAIC*, es presentada en la Tabla 6.3. En forma complementaria también se presenta el Criterio de Información de Akaike, (*AIC* según su sigla en inglés), propuesto por Akaike (1973) el cual, más allá de ciertas críticas, tiene mayor difusión en la literatura (Rissanen, 1976; Bozdogan, 2000). El *AIC* está definido como:

$$AIC = -2 \cdot l_M(\theta) + 2 \cdot k_M \quad (6.6)$$

Tabla 6.3 – Bondad de Ajuste de los Modelos Estimados

	Egaña (2009)	Modelo IP	Modelo IP + VE
Log-Verosimilitud	-245,779	-223,48	-171,502
Nº de Parámetros Estimados	8	10	15
<i>CAIC</i>	550,733	520,929	453,958
<i>AIC</i>	507,558	466,96	373,004

El análisis de la Tabla 6.3 revela que los modelos que consideran directamente los indicadores de percepción presentan una evidente superioridad en términos de ajuste,

tanto en términos brutos (log-verosimilitud) como corregidos (*CAIC* y *AIC*), lo que ratifica los resultados obtenidos utilizando la base de datos sintética.

Otro análisis importante respecto a este tipo de modelos, está dado por el sesgo predictivo. La Tabla 6.4 presenta las participaciones modales predichas por los distintos modelos, así como la partición modal observada, y el índice χ^2 definido en (5.1).

Tabla 6.4 – Partición Modal Predicha e Índice χ^2

	Obs.	Egaña (2009)	Modelo IP	Modelo IP + VE
Bus	301	300,96	301,17	301,02
Tren	299	299,04	298,83	298,98
χ^2	-	1,06E-05	1,93E-04	2,66E-06

Tal como se puede observar en la Tabla 6.4, tampoco existe evidencia que permita inferir la existencia de un sesgo predictivo en el caso de los modelos que consideran los indicadores de percepción directamente en la función de utilidad.

En vista de lo anterior, desde el punto de vista estadístico, no existe evidencia empírica para descartar la utilización de este tipo de modelos. Luego el problema recae en las consideraciones teóricas mencionadas con anterioridad, que sí hacen inviable su uso en la práctica.

Básicamente, los modelos de elección discreta tienen como finalidad extrapolar las características de la muestra a la población, lo que es perfectamente realizable al tener como parámetros de entrada ciertas características tangibles de las alternativas o de los individuos, que son en general conocidas. El problema surge al considerar indicadores de percepción como parámetros de entrada; en primer lugar la extrapolación de estos indicadores a la población, a partir de una encuesta, tendrá una menor confiabilidad que

el uso de características tangibles conocidas de la misma, induciéndose así una nueva fuente de error que no puede ser fácilmente dimensionada *a priori* (y que fácilmente podría anular las ventajas obtenidas por el mejor ajuste por estos modelos). En segundo lugar, el uso de indicadores de percepción impide que estos modelos sean utilizados en un instante de tiempo diferente al instante en el cual fue calibrado el modelo. Esto se debe a que cualquier cambio en la población o en la oferta, induciría un cambio en la percepción general. Lo anterior se suma a la imposibilidad de asignar los valores representativos de estos indicadores a las características representativas de la población. Por tanto, la única posibilidad de utilizar estos modelos, es repetir la encuesta utilizada para calibrarlos, lo que obviamente es un sinsentido ya que una nueva encuesta permitiría calibrar un modelo acorde con la nueva información.

También es importante destacar que este tipo de modelos no permite una interpretación clara, ya que como se mencionó anteriormente, los indicadores de percepción no se pueden asignar a ningún grupo poblacional en particular y no se considera un modelo que permita predecir, como variarían estos indicadores de percepción.

Finalmente, se debe mencionar que los indicadores de percepción, no presentan una relación causal directa con las elecciones (aunque son motivados por las mismas actitudes subjetivas), lo que es un argumento más que suficiente para rechazar el uso de estos indicadores como predictores de las decisiones.

7. CONCLUSIONES Y RECOMENDACIONES

Aunque ciertas características tangibles de las alternativas, que pueden participar en la generación de variables latentes son usualmente ignoradas en modelos híbridos de elección discreta, su inclusión debiera afectar la capacidad predictiva de los modelos. Utilizando una base de datos sintética, que incorpora explícitamente el impacto de este tipo de variables, se analizó el efecto de diferentes especificaciones tendientes a considerar estas características en la modelación híbrida.

Los resultados muestran que los estimadores asociados a las variables socio-económicas en el modelo de variables latentes son sensibles a la inclusión/omisión de los atributos tangibles propios de las alternativas, pero aun así continúan afectando a las variables latentes en la forma esperada.

Debido al problema de identificabilidad, se sabe que la estimación de modelos de variables latentes sólo es posible si se fijan los parámetros asociados a las ecuaciones de medición o las varianzas asociadas las ecuaciones estructurales. Con el objeto de determinar el efecto de distintas formas de normalizar, se condujo un análisis teórico y empírico; éste demostró que la normalización no afectaba la capacidad predictiva de los modelos, debido a que, si bien, ésta tiene un efecto directo sobre el valor de las variables latentes estimadas, este efecto es consistente para todos los individuos.

De esta forma, las variables latentes aparecen consistentemente multiplicadas por un factor de escala desconocido. Pero, dado que en el modelo de elección discreta las variables latentes son multiplicadas por parámetros a estimar, éstos pueden ajustarse perfectamente (el efecto del factor de escala es absorbido por los parámetros a estimar).

También se demostró, como corolario de lo anterior, que ante la menor sospecha de que la variabilidad asociada a las variables latentes, varíe de alternativa en alternativa, se

deben utilizar estimadores específicos asociados a éstas en la función de utilidad del modelo de elección discreta.

Respecto a la estimación del modelo de elección discreta, se demostró teórica y empíricamente que la inclusión de variables latentes en su función de utilidad aumenta el componente estocástico de ésta. Debido a lo anterior, el factor de escala de la distribución asociada al término de error se ve afectado provocando que los parámetros no puedan ser recuperados en forma correcta. De esta forma, los estimadores se ven afectados en forma sistemática por una constante que depende de la magnitud de la varianza inducida a través de las variables latentes. En el caso del modelo Logit jerárquico, lo anterior tiene una implicancia mayor, ya que la varianza inducida aumenta el valor de los parámetros estructurales ϕ estimados. Esto tiene como consecuencia que la correlación entre alternativas no pueda ser correctamente estimada en modelos que incluyan variables latentes como variables explicativas.

Es importante notar que el problema mencionado anteriormente no se soluciona excluyendo a las variables latentes del modelo, ya que es un problema inherente al proceso de elección.

La estimación de modelos de elección discreta indica que los resultados obtenidos en el caso de estructuras que incluyan atributos tangibles (propios de las alternativas) afectando las variables latentes, aunque en forma incorrecta (esto es, no consistente con la forma en que fueron generados los datos), pueden obtener una bondad de ajuste similar a la obtenida por modelos bien especificados. En contraste con lo anterior, un modelo que no considere estas características en absoluto, presenta un ajuste a los datos sustancialmente inferior. Es interesante notar, sin embargo, que en el caso del modelo mal especificado (considerando toda la información disponible), sin desmedro de su buen nivel de ajuste, algunos parámetros estimados presentan un signo incorrecto. Esto puede llevar al modelador a interpretar que ciertas variables afectan las decisiones en forma opuesta, a la manera en que realmente actúan.

Utilizando el método de estimación secuencial, no se encontró evidencia empírica que respalde las ventajas teóricas de la inclusión de la variabilidad de las variables latentes, en la función de utilidad del modelo de elección discreta. Este resultado difiere con el encontrado en investigaciones recientes (Yañez, 2010) en que se utilizó datos reales en lugar de simulados.

Una especificación que no considera variables latentes, pero sí considera todas las posibles variaciones sistemáticas de gustos entre los individuos, puede lograr un alto nivel de ajuste, incluso superior al nivel de ajuste esperado *a priori*; sin embargo, éste es inferior al nivel de ajuste obtenido por un modelo bien especificado. De todas maneras, un modelo que considera un número tan elevado de variables puede resultar complicado, tanto desde el punto de vista de su utilización como de su interpretación. Por tanto, de acuerdo al criterio de parsimonia en conjunción con el nivel de ajuste, no parece recomendable optar por este tipo de modelos.

El análisis de especificaciones que consideran la inclusión, en forma directa, de los indicadores de percepción en la función de utilidad asociada al modelo de elección discreta, reveló que estos logran un alto nivel de ajuste a los datos, incluso superior al de los modelos híbridos de elección discreta. Sin embargo, no se recomienda utilizar este tipo de modelos, ya que existen una serie de consideraciones teóricas y prácticas que imponen su descarte como una alternativa válida. Entre estas consideraciones destacan la no causalidad, la imposibilidad de ofrecer una interpretación válida y la obligatoriedad de realizar la encuesta, en forma íntegra, cada vez que se desee utilizar el modelo.

Un análisis de respuesta reveló que modelos mal especificados ven degradado su nivel de ajuste al ser utilizados en escenarios diferentes al escenario para el cual fueron calibrados. Esta degradación depende de la magnitud de la perturbación del escenario y de la partición modal. Un modelo bien especificado no ve afectado su nivel de ajuste en escenarios diferentes.

Obviamente el asumir un determinado comportamiento de los individuos, considerando la influencia de atributos tangibles propios de las alternativas explicando las variables latentes, es un supuesto sumamente fuerte. Es por tanto necesario, validar este supuesto con datos reales, por lo que futuras investigaciones deben estar orientadas en esta dirección. De todas formas, este trabajo provee evidencia empírica de la importancia de trabajar con una especificación correcta del modelo híbrido de elección discreta. En consecuencia, se recomienda fehacientemente a los modeladores, destinar sus esfuerzos a establecer, en la forma más certera posible, el rol desempeñado por las diferentes variables explicativas en el proceso mental de los individuos, conducente a la toma de decisiones. Sólo de esta manera se podrá garantizar que los modelos estimados estén correctamente especificados y que las variables participantes afecten las elecciones en una determinada manera.

BIBLIOGRAFÍA

- Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. *Journal of Marketing Research*, 39, 31-46.
- Arbuckle, J.L. (2005). *Amos 16.0 User's Guide*. SPSS Inc., Chicago.
- 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. y Bolduc, D. (1996). Multinomial probit with a logit kernel and a general parametric specification of the covariance structure. Working Paper, Département d'économique, Université Laval.
- 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.
- Bierlaire, M. (2009). Estimation of discrete choice models with BIOGEME 1.8 (<http://transp-or2.epfl.ch/biogeme/doc/tutorial.pdf>)
- 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.
- Bollen, K.A., Kirby, J.B., Curran, P.J., Paxton, P.M. y Chen, F. (2007). Latent variable models under misspecification: two-stage least squares (2SLS) and maximum likelihood (ML) estimators. *Sociological Methods & Research* 36, 48- 86.
- Bozdogan, H. (1987). Model selection and Akaike's information criterion (AIC): the general theory and its analytical extensions. *Psychometrika* 52, 345-370.
- Bozdogan, H. (2000). Akaike's information criterion and recent developments in information complexity. *Journal of Mathematical Psychology* 44, 62-91.

Carrasco, J.A. (2001). *Elección Discreta con Alternativas Relacionadas Homocedásticas. El Modelo Logit Jerárquico en Profundidad*. Tesis de Magíster en Ciencias de la Ingeniería, Departamento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile.

Carrasco, J.A. y Ortúzar, J. de D. (2002). Review and assessment of the nested logit model. *Transport Reviews* 22, 197-218.

Cardell, N. y Dunbar, F. (1980). Measuring the societal impacts of automobile downsizing, *Transportation Research*, 14A, 423-434.

Daly, A.J. y Zachary, S. (1978). Improved multiple choice models. En D.A. Hensher y M.Q. Dalvi (eds.), *Determinants of Travel Choice*. Saxon House, Westmead.

Domencich, T. y McFadden, D. (1975). *Urban Travel Demand – A Behavioural Analysis*. North Holland, Amsterdam.

Egaña, J.N. (2009). *Inclusión de Variables Latentes en Modelos de Elección Discreta para Usuarios de Buses y Trenes Interurbanos*. Tesis de Magíster en Ciencias de la Ingeniería, Departamento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile.

Goldberger, A.S. (1972). Maximum likelihood estimation of regressions containing unobservable independent variables. *International Economic Review* 13, 1-15.

Green, P. (1984). Hybrid models for conjoint analysis: an expository review. *Journal of Marketing Research* 21, 155-169.

Harris, K.M. y Keane, M.P. (1999). A model of health plan choice: inferring preferences and perceptions from a combination of revealed preference and attitudinal data. *Journal of Econometrics* 89, 131-157.

Keane, M.P. (1997). Modelling heterogeneity and state dependence in consumer choice behaviour. *Journal of Business and Economic Statistics* 15, 310-327.

Koppelman, F. y Hauser, J. (1979). Destination choice for non-grocery-shopping trips. *Transportation Research Record* 673, 157-165.

McFadden, D. y Train, K.E. (2000). Mixed MNL models for discrete response. *Journal of Applied Econometrics* 15, 447-470.

Munizaga, M. (1997). *Implicancias de la Naturaleza de los Datos en la Modelación de Elecciones Discretas*. Tesis de Doctorado en Ciencias de la Ingeniería, Departamento de Ingeniería de Transporte y Logística, Pontificia Universidad Católica de Chile.

- Munizaga, M.A., Heydecker, B.G. y Ortúzar, J. de D. (2000). Representation of heteroskedasticity in discrete choice models. *Transportation Research* 34B, 219-240.
- Ortúzar, J. de D. y Willumsen, L.G. (2001). *Modelling Transport* (3ª Edición). John Wiley and Sons, Chichester.
- Pagan, A. (1986). Two stage and related estimators and their applications. *Review of Economic Studies* 53, 517-538.
- 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.
- Raveau, S., Álvarez-Daziano, R., Yáñez, M.F., Bolduc, D. y Ortúzar, J. de D. (2010a). Sequential and simultaneous estimation of hybrid discrete choice models: some new findings. *Transportation Research Record* 2156, 131-139.
- Raveau, S., Yáñez, M.F. y Ortúzar, J. de D. (2010b). The effect of data variability on the estimation of discrete choice models with latent variables. *5th IMA Conference on Mathematics in Transport*, Londres, 12-14 Abril 2010.
- Raveau, S., Ortúzar, J. de D. y Yáñez, M.F. (2010c). Identification of discrete choice models with latent variables: the role of constraints in sequential and simultaneous estimation. *12th World Conference on Transport Research*, Lisboa, 11-15 Julio 2010.
- Rissanen, J. (1976). Minmax entropy estimation of models for vector processes. En R.K. Mehra y D.G. Lainiotis (eds.), *System Identification*, 97-119. Academic Press, Nueva York.
- Stapleton, D. C. (1978) Analyzing political participation data with a MIMIC model. *Sociological Methodology* 9, 52-74.
- Train, K.E. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press, Cambridge.
- 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.
- Williams, H.C.W.L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning* 9A, 285-344.
- Williams, H.C.W.L. y Ortúzar, J. de D. (1982). Behavioural theories of dispersion and the mis-specification of travel demand models. *Transportation Research* 16B, 167-219.

Yáñez, M.F., Raveau, S. y Ortúzar, J. de D. (2010). Modelling and forecasting with latent variables in discrete choice panel models. *Transportation Research*, 44^a, 744-753.

ANEXOS

ANEXO A – DESARROLLOS MATEMÁTICOS

A.1 De (4.16) a (4.17)

$$\frac{\pi^2}{6 \cdot \lambda_R^2} = \frac{\pi^2}{6 \cdot \lambda_P^2} + \frac{\pi^2}{6 \cdot \lambda_I^2} \quad (4.16)$$

dividiendo por $\pi^2/6$ se obtiene:

$$\frac{1}{\lambda_R^2} = \frac{1}{\lambda_P^2} + \frac{1}{\lambda_I^2}$$

Sumando al lado derecho, se tiene que:

$$\frac{1}{\lambda_R^2} = \frac{\lambda_P^2 + \lambda_I^2}{\lambda_P^2 \cdot \lambda_I^2}$$

Elevando por -1 en ambos lados:

$$\lambda_R^2 = \frac{\lambda_P^2 \cdot \lambda_I^2}{\lambda_P^2 + \lambda_I^2}$$

Y finalmente, aplicando la propiedad conmutativa al lado derecho, se obtiene (4.17)

$$\lambda_R^2 = \lambda_P^2 \cdot \frac{\lambda_I^2}{\lambda_P^2 + \lambda_I^2} \quad (4.17)$$

A.2 De (4.21) a (4.22)

$$\frac{\pi^2}{6 \cdot \lambda_R^2} = \rho \cdot \frac{\pi^2}{6 \cdot \lambda_p^2} + \frac{\pi^2}{6 \cdot \Delta^2} \quad (4.21)$$

dividiendo por $\pi^2/6$ se obtiene:

$$\frac{1}{\lambda_R^2} = \rho \cdot \frac{1}{\lambda_p^2} + \frac{1}{\Delta^2}$$

Reordenando, se tiene:

$$\frac{1}{\lambda_R^2} - \rho \cdot \frac{1}{\lambda_p^2} = \frac{1}{\Delta^2}$$

Multiplicando por λ_R^2 :

$$1 - \rho \cdot \frac{\lambda_R^2}{\lambda_p^2} = \frac{\lambda_R^2}{\Delta^2}$$

Y finalmente, aplicando raíz cuadrada, se obtiene (4.22)

$$\sqrt{1 - \rho \cdot \frac{\lambda_R^2}{\lambda_p^2}} = \frac{\lambda_R}{\Delta} = \phi \quad (4.22)$$

ANEXO B – MODELOS ESTIMADOS

B.1 Modelo 4

Parámetro	Valor Estimado	test-t
ASC_{Bus}	0,976	7,63
$Costo_{Auto}$	-0,924	-72,17
$Costo_{Bus}$	-0,949	-21,97
$Costo_{Metro}$	-0,975	-21,46
$Tiempo_{Auto}$	-0,606	-41,81
$Tiempo_{Bus}$	-0,621	-32,82
$Tiempo_{Metro}$	-0,616	-34,73
$Aire\ Aco_{Auto}$	0,414	7,9
$Aire\ Aco_{Bus}$	0,617	15,13
$Aire\ Aco_{Metro}$	0,481	16,33
$Cam\ Sec_{Bus}$	0,583	12,77
$Cam\ Sec_{Metro}$	0,492	7,63
$Dens_{Bus}$	-0,12	-7,43
$Dens_{Metro}$	-0,133	-8,56
$Ind\ Acc_{Auto}$	-0,377	-27,28
$Ind\ Acc_{Bus}$	-0,392	-16,14
$Ind\ Acc_{Metro}$	-1	-3,99
$P\ Info_{Bus}$	0,535	11,51
$P\ Info_{Metro}$	0,163	1,97
$Ing\ Alt_{Auto}$	0,65	9,63
$Ing\ Alto_{Bus}$	-1,45	-13,9
$Ing\ Med_{Auto}$	0,496	8,93

Parámetro	Valor Estimado	test-t
$Ing\ Medio_{Bus}$	-0,813	-16,96
$Edad_{Auto}$	2,14	23,33
$Ed\ Med_{Auto}$	0,408	3,83
$Ed\ Med_{Bus}$	-0,276	-2,36
$Ed\ Sup_{Bus}$	-0,761	-9,49
$Sexo_{Bus}$	-1,37	-16,28
$Costo_{Auto} \cdot Ed_{Sup}$	0,0634	2,95
$Costo_{Metro} \cdot Sexo$	-0,086	-2,45
$Costo_{Metro} \cdot Ed_{Med}$	0,134	2,62
$Tiempo_{Bus} \cdot Ing_{Alt}$	-0,0852	-2,19
$Aire\ Aco_{Auto} \cdot Ing_{Alt}$	-0,159	-1,89
$Aire\ Aco_{Auto} \cdot Ing_{Med}$	-0,18	-2,63
$Cam\ Sec_{Bus} \cdot Sexo$	-0,12	-1,74
$Cam\ Sec_{Metro} \cdot Ing_{Med}$	-0,176	-2,36
$Dens_{Bus} \cdot Sexo$	0,0453	1,78
$Dens_{Metro} \cdot Edad$	0,096	2,9
$Ind\ Acc_{Metro} \cdot Edad$	1,48	2,75
$Ind\ Acc_{Metro} \cdot Ed_{Sup}$	-0,568	-1,87
$P\ Info_{Bus} \cdot Ing_{Med}$	-0,117	-1,7
$P\ Info_{Metro} \cdot Edad$	0,774	5,05
$P\ Info_{Metro} \cdot Sexo$	-0,248	-3,35
ϕ	0,893	3,48

B.2 Modelo 5

Parámetro	Valor Estimado	test-t
<i>ASC</i> <i>Auto</i>	0,449	3,51
<i>ASC</i> <i>Bus</i>	1,52	10,53
<i>Costo</i> <i>Auto</i>	-1,07	-74,64
<i>Costo</i> <i>Bus</i>	-1,06	-24,49
<i>Costo</i> <i>Metro</i>	-1,08	-33,52
<i>Tiempo</i> <i>Auto</i>	-0,714	-44
<i>Tiempo</i> <i>Bus</i>	-0,723	-40,47
<i>Tiempo</i> <i>Metro</i>	-0,731	-38,5
<i>Aire Aco</i> <i>Auto</i>	0,0396	1,15
<i>Aire Aco</i> <i>Bus</i>	0,0322	0,71
<i>Aire Aco</i> <i>Metro</i>	0,0817	2,47
<i>Cam Sec</i> <i>Bus</i>	0,058	1,46
<i>Cam Sec</i> <i>Metro</i>	0,0216	0,58
<i>Dens</i> <i>Bus</i>	-0,00973	-0,72
<i>Dens</i> <i>Metro</i>	-0,0121	-1,32
<i>Ind Acc</i> <i>Auto</i>	-0,0333	-1,73
<i>Ind Acc</i> <i>Bus</i>	-0,0148	-0,54
<i>Ind Acc</i> <i>Metro</i>	0,185	1,44
<i>P Info</i> <i>Bus</i>	0,102	2,3
<i>P Info</i> <i>Metro</i>	0,0114	0,22
<i>Ing Alt</i> <i>Auto</i>	0,0691	1,08
<i>Ing Alto</i> <i>Bus</i>	-0,144	-2,31
<i>Ing Med</i> <i>Auto</i>	0,0814	1,78
<i>Ing Medio</i> <i>Bus</i>	-0,0725	-1,77
<i>Edad</i> <i>Auto</i>	0,234	2,45
<i>Edad</i> <i>Bus</i>	-0,123	-1,31
<i>Ed Med</i> <i>Auto</i>	0,0426	0,66

Parámetro	Valor Estimado	test-t
<i>Ed Med</i> <i>Bus</i>	-0,1	-1,72
<i>Ed Sup</i> <i>Auto</i>	-0,0747	-0,9
<i>Ed Sup</i> <i>Bus</i>	-0,0938	-1,25
<i>Sexo</i> <i>Auto</i>	0,0513	1,17
<i>Sexo</i> <i>Bus</i>	-0,0653	-1,65
<i>IP Acc</i> <i>Auto</i>	0,12	8,95
<i>IP Acc</i> <i>Bus</i>	0,118	8,33
<i>IP Acc</i> <i>Metro</i>	0,114	9,7
<i>IP Com</i> <i>Auto</i>	0,111	9,96
<i>IP Com</i> <i>Bus</i>	0,124	10,51
<i>IP Com</i> <i>Metro</i>	0,117	11,86
<i>IP Conf</i> <i>Auto</i>	0,0546	4,43
<i>IP Conf</i> <i>Bus</i>	0,0647	4,93
<i>IP Conf</i> <i>Metro</i>	0,0491	4,54
<i>IP ETV</i> <i>Auto</i>	0,0553	4,29
<i>IP ETV</i> <i>Bus</i>	0,0292	2,15
<i>IP ETV</i> <i>Metro</i>	0,0669	5,88
<i>IP Info</i> <i>Auto</i>	0,0709	4,91
<i>IP Info</i> <i>Bus</i>	0,0777	5,11
<i>IP Info</i> <i>Metro</i>	0,0754	5,96
<i>IP Sec Acc</i> <i>Auto</i>	0,0767	6,69
<i>IP Sec Acc</i> <i>Bus</i>	0,0929	7,65
<i>IP Sec Acc</i> <i>Metro</i>	0,094	9,42
<i>IP Sec Ro</i> <i>Auto</i>	0,0956	7,58
<i>IP Sec Ro</i> <i>Bus</i>	0,0832	6,25
<i>IP Sec Ro</i> <i>Metro</i>	0,0899	8,11
ϕ	0,84	6,4