PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

INCLUSIÓN DE VARIABLES LATENTES EN MODELOS DE ELECCIÓN DISCRETA PARA USUARIOS DE BUSES Y TRENES INTERURBANOS

JAVIER NICOLÁS EGAÑA UGRINOVIC

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

Profesor Supervisor:
LUIS IGNACIO RIZZI CAMPANELLA

Santiago de Chile, julio de 2009

© 2009, Javier Nicolás Egaña Ugrinovic

PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

INCLUSIÓN DE VARIABLES LATENTES EN MODELOS DE ELECCIÓN DISCRETA PARA USUARIOS DE BUSES Y TRENES INTERURBANOS

JAVIER NICOLÁS EGAÑA UGRINOVIC

Tesis presentada a la Comisión integrada por los profesores:

LUIS RIZZI

JUAN DE DIOS ORTÚZAR

MARCELA MUNIZAGA

IGNACIO CASAS

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

Santiago de Chile, julio de 2009

A Javier, Lenka, Gabriela y Daniel.

AGRADECIMIENTOS

No resulta fácil comenzar una sección de agradecimientos, sobre todo si se considera que no es posible desvincular el fruto de este trabajo con todo lo que posibilitó que yo estuviese aquí y ahora. Familia, amigos, profesores, compañeros, desconocidos y la suerte han permitido el desarrollo de este trabajo.

No obstante lo anterior, es necesario destacar la participación de algunas personas e instituciones que fueron pilares fundamentales de mi proceso universitario.

Por esto, en primer lugar quisiera agradecer a Luis Rizzi, quien fue mi profesor guía y apoyo incondicional. Gracias profesor por su ayuda, paciencia, cariño, generosidad, tiempo, dedicación, rigurosidad, motivación, amor por la investigación, los conocimientos y oportunidades. Muchas gracias Maestro.

Esta gratitud la extiendo al profesor Juan de Dios Ortúzar, cuyo apoyo fue fundamental durante toda la tesis y en especial durante la última etapa. Le agradezco su tiempo, apoyo, cariño, dedicación, retos, paciencia, impaciencia y ayuda.

Agradezco también a la profesora Marcela Munizaga de la Universidad de Chile, por aceptar ser parte de mi comisión y por los aportes que realizó a este trabajo.

De igual forma, me gustaría destacar la ayuda que me brindaron Paula Iglesias y Francisco Fresard en distintas etapas de mi trabajo.

No puedo dejar de nombrar al profesor Juan Enrique Coeynmans, con quien tuve el placer de compartir innumerables tardes, cafés y conversaciones. Gracias profesor por todo el tiempo que perdimos y las recetas que imaginamos.

Por otra parte, agradezco al Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT), que a través del Proyecto 1050672 entregó fondos para la realización de esta tesis.

Agradezco también a José Espinoza y a mis compañeros de postgrado, por los almuerzos, apoyo, conversaciones, discusiones, partidos de fútbol, buen humor y excelente compañía, en particular, durante los meses de verano.

A todos los amigos que hice en la Universidad, en especial a Cristóbal Pétrovich y Enrique Aldunate, con quienes compartí contables horas de estudio e incontables horas de ocio, en las que fuimos descubriendo los libros y esta hermosa ciudad. Gracias por estar siempre dispuestos a no avanzar con tal de entender.

Por último, agradezco al Departamento de Ingeniería de Transporte y Logística, a todos sus profesores, administrativos y estudiantes por ser parte de una etapa muy feliz de mi vida.

ÍNDICE

Agradecimientos ii

Índice Anexos vi

Índice de Tablas vii

Índice Figuras viii

Resumen ix

Abstract x

1 Introducción 1

1.1 Aspectos generales 1

1.2 Objetivos y alcances del estudio 1

1.3 Estructura del trabajo 1

2 Antecedentes Bibliográficos 2

2.1 Antecedentes teóricos 2

2.2 Antecedentes prácticos 4

3 Marco Teórico 7

3.1 Ecuaciones estructurales 7

3.1.1 Presentación 7

3.1.2 Identificabilidad 9

3.1.3 Estimación 11

3.2 Modelos Logit de elección discreta 14

3.3 Modelos Logit-MIMIC integrados 15

3.3.1 Estimación simultánea de modelos Logit-MIMIC integrados 16

3.3.2 Estimación secuencial de modelos Logit-MIMIC integrados 17

4 Descripción de la Encuesta 18
4.1 Metodología 18
4.2 Análisis de la muestra 20
5 Estimación de Modelos 26
5.1 Análisis factorial exploratorio 26
5.2 Modelo MIMIC 28
5.3 Modelo de elección discreta 37
5.3.1 Modelos base 37
5.3.2 Modelo final 39
5.3.3 Justificación de las técnicas utilizadas 43
5.4 Análisis y Discusiones 45
5.4.1 Elección del mejor modelo 45
5.4.2 Cálculo de valores subjetivos 48
5.4.3 Cálculo de las principales elasticidades Modelo 1 (modelo final) 51
6 Conclusiones 53
Bibliografía 56
Anexos 59

ÍNDICE ANEXOS

Anexo 1: Descripción corrección errores estándar 60
Anexo 2: Notas Variables Indicadoras por Modo.....62
Anexo 3: Promedio respuestas Variables Indicadoras para usuarios de bus 63
Anexo 4: Promedio respuestas Variables Indicadoras para usuarios de tren.....64
Anexo 5: Promedio respuestas Variables Indicadoras Bus por género.....65
Anexo 6: Promedio respuestas Variables Indicadoras Tren por género 66
Anexo 7: Promedio respuestas Variables Indicadoras Bus por tramo de edad.....67
Anexo 8: Promedio respuestas Variables Indicadoras Tren por tramo de edad 68
Anexo 9: Resultados análisis componentes principales.....69
Anexo 10: Indicadores normalizados Modelos 3 y 4.....72
Anexo 11: Inclusión directa de variables del modelo MIMIC en el modelo de elección discreta 74
Anexo 12: Inclusión directa de variables indicadoras en el modelo de elección discreta76

ÍNDICE DE TABLAS

Tabla 4-1: Nivel de Educación por tipo usuario. 22
Tabla 4-2: Ocupación por tipo de usuario. 22
Tabla 4-3: Actividad desarrollada durante el viaje por tipo de usuario. 23
Tabla 4-4: Variables vivenciales por tipo de usuario. 24
Tabla 5-1: Relación entre Indicadores y variables latentes 30
Tabla 5-2: Relación entre Variables objetivas y variables latentes 32
Tabla 5-3: Resultados variables objetivas (ecuación 3.1) 34
Tabla 5-4: Resultados variables indicadoras (ecuación 3.2) 36
Tabla 5-5: Coeficientes modelo Logit Base 38
Tabla 5-6: Coeficientes de modelos con variables latentes 42
Tabla 5-7: Test LR 45
Tabla 5-8: Tabla Éxito Predictivo 46
Tabla 5-9: Tabla Éxito Predictivo Modelo Base 46
Tabla 5-10: Tablas Éxito Predictivo Modelos con Variables Latentes 47
Tabla 5-11: Disposiciones al Pago 49
Tabla 5-12: Sensibilidad ante cambio en magnitud de normalización Modelo 1 50
Tabla 5-13: Elasticidades puntuales Modelo Base y Modelo 1 51

ÍNDICE FIGURAS

Figura 4-1: Distribución de destinos por modo.....20
Figura 4-2: Distribución de motivos de viaje por modo 21
Figura 4-3: Distribución del ingreso por modo.....21

RESUMEN

En este trabajo se estiman modelos de elección discreta de transporte incluyendo variables latentes (comodidad y seguridad), considerando viajes en bus y tren desde Santiago hacia las ciudades de Curicó, Talca, Chillán y Linares. Las variables latentes son conceptos abstractos no medibles directamente (se construyen utilizando ecuaciones estructurales) que representan las percepciones subjetivas de las personas. Para estimar los modelos, se realizaron 299 encuestas a usuarios de tren y 301 a usuarios de bus, durante noviembre del 2008 en el Terminal Alameda y la Estación Central. Los usuarios respondieron preguntas relativas al viaje y sus características socioeconómicas, además de evaluar al bus y al tren en una escala de 1 a 7 en distintos aspectos relacionados con la calidad del servicio. Posteriormente, se estimaron modelos MIMIC y de elección discreta en forma secuencial. En todos los casos los modelos de elección que incluyen variables latentes resultaron superiores al mejor modelo que no las considera, tanto a nivel de ajuste como de predicción, obteniéndose mejoras de alrededor de un 2% en la cantidad de viajes correctamente predichos. Por otra parte, las elasticidades de las probabilidades de elección con respecto a las variables latentes son bastante bajas (alrededor de un 1%), en comparación con las elasticidades del costo y tiempo de viaje (alrededor del 10%).

Sin embargo, el resultado más importante que no ha sido reportado adecuadamente en la literatura, guarda relación con la alta dependencia que tienen los coeficientes de los modelos estimados a las normalizaciones utilizadas en la construcción de las variables latentes. Estas normalizaciones no tienen grandes efectos a nivel de predicción, pero sí hacen muy difícil dar una interpretación cuantitativa a los coeficientes asociados a las variables latentes, tanto en el modelo estructural como de elección, dificultando la obtención de valoraciones económicas por mejoras en la comodidad y seguridad.

ABSTRACT

In this work we incorporate latent variables (comfort and safety) to discrete mode choice models for bus and train trips undertaken from Santiago to the cities of Curicó, Talca, Chillán and Linares. Latent variables are abstract concepts of people´s subjective perceptions that are not directly measurable, but built through structural equations. To estimate the models, 299 interviews of train users and 301 of bus users were undertaken during November 2008 at the Terminal Alameda and the Estación Central. Users responded to questions regarding their travel and socioeconomic characteristics In addition, they had to evaluate bus and train on a scale from 1 to 7 on several aspects related to their quality of service. Subsequently, structural and discrete choice models were estimated following a two step approach. In all cases the models with latent variables were superior to the best model without latent variables, both in fit and prediction, obtaining an improvement of nearly 2% in the number of correctly predicted trips. The elasticities of the choice probabilities with respect to the latent variables were found to be quite low (around 1%) compared with those of the usual variables cost and travel time (about 10%).

However, the most important result which has not been adequately reported in the literature relates to the high dependence of the estimated parameters to the normalization process used for the construction of the latent variables. The normalizations did not have large impacts on prediction, but it is not easy to give a quantitative interpretation to model parameters associated with the latent variables, both in the structural and discrete choice models. This makes it difficult to obtain economic values for improved comfort and safety.

1 INTRODUCCIÓN

1.1 Aspectos generales

Tradicionalmente variables como el tiempo de viaje, el costo, el ingreso, el género, los tiempos de caminata y acceso, entre otras, han sido utilizadas masivamente en la calibración de modelos de elección en transporte. Sin embargo, poco se conoce del proceso mental y de las percepciones subjetivas que llevan a los usuarios a preferir una opción antes que otra. Este trabajo pretende responder parte de dichas cuestiones, para lo que se realizaron encuestas a usuarios de transporte público (buses y trenes interurbanos hacia el sur de Chile) siguiendo la experiencia desarrollada por Rizzi et al. (2005).

1.2 Objetivos y alcances del estudio

El principal objetivo de este trabajo es lograr relacionar las percepciones subjetivas sobre la seguridad y comodidad de los usuarios de buses y trenes, con su elección modal.

Con este fin, se construyen modelos de ecuaciones estructurales para cuantificar las apreciaciones individuales sobre comodidad y seguridad. Posteriormente estos resultados son utilizados como entrada para los modelos de elección discreta. Este enfoque de estimación se denomina “enfoque secuencial” (Murphy y Topel 1985).

1.3 Estructura del trabajo

El trabajo consta de cinco capítulos adicionales a esta introducción. En el capítulo 2, se hace una breve revisión bibliográfica sobre el uso de variables latentes en modelos de elección modal. El capítulo 3 contiene el marco teórico además de la descripción de la encuesta. En el capítulo 4, se detalla la estimación de modelos, que posteriormente son analizados y discutidos en el capítulo 5. Por último, en el capítulo 6 se presentan las principales conclusiones del estudio.

2 ANTECEDENTES BIBLIOGRÁFICOS

En este capítulo se realiza una discusión bibliográfica que revisa la experiencia nacional e internacional con respecto a la inclusión de variables latentes[1] en modelos de elección discreta aplicados al transporte.

2.1 Antecedentes teóricos

Las técnicas de ecuaciones estructurales han sido ampliamente utilizadas en los campos de las ciencias sociales para la construcción de variables latentes, principalmente en la sicología y la sociología. Según Bollen (1989, pp. 4) su desarrollo se debe a las contribuciones de múltiples académicos, sin embargo, destacan los aportes de Jöreskog (1973) asociados a una metodología de estimación de ecuaciones estructurales, Jöreskog y Sörbom (1986) en el desarrollo de LISREL (*Linear Structural Relationships*), programa especializado en la estimación de ecuaciones estructurales y de Bollen (1989) en la condensación y unificación de los conocimientos relativos a ecuaciones estructurales, variables latentes y análisis factorial. La "participación democrática" y la "identificación sindical", entre otros, son ejemplos de variables latentes que han sido construidas con ecuaciones estructurales en el ámbito de la sociología (Bollen, 1989).

Uno de los principales problemas asociados a las ecuaciones estructurales guarda relación con su identificabilidad, es decir, la capacidad de estimar los parámetros de los modelos a partir de las relaciones funcionales planteadas y de los datos disponibles. Bollen (1989) presenta una serie de reglas generales a tener en cuenta para lograr la identificabilidad de estos modelos. Por otra parte, Stapleton (1978) desarrolla condiciones de identificabilidad específicas para el tipo de ecuaciones estructurales que se utilizan en esta tesis. Una excelente revisión bibliográfica sobre ecuaciones

[1] Qué es una variable latente y cómo se construye, será discutido con profundidad en el capítulo 3. Por el momento, basta considerar que por variables latentes se entiende a conceptos abstractos y unidimensionales, no observables directamente.

estructurales puede encontrarse en Golob (2003), donde además se introducen algunos tópicos como identificabilidad y criterios de bondad de ajuste.

Aunque el uso de variables latentes y ecuaciones estructurales en las ciencias sociales ha sido amplio, existen muy pocos trabajos que vinculen estas técnicas con modelos de elección discreta. Estos trabajos se han dado principalmente en las áreas de marketing (elección de bienes de consumo) y transporte (elección de modo de transporte).

La inclusión de variables latentes en modelos de elección discreta se debe primordialmente al impulso dado por el trabajo de Mc Fadden (1986) en donde presenta prototipos de modelos que incluyen indicadores psicométricos en las funciones de utilidad. Posteriormente Ben-Akiva et al. (1999) generan el marco conceptual donde se formaliza la inclusión de variables latentes en modelos de elección discreta. El desarrollo de Ben-Akiva et al. (1999) es importante dado que las variables latentes se construyen a partir de una forma particular de ecuaciones estructurales: los modelos MIMIC (*multiple indicator and multiple cause*). Este tipo de modelos (explicados en la sección 3.1) serán utilizados en esta tesis para construir las variables latentes.

Siguiendo la misma línea de investigación, Walker (2001) presenta un detallado marco teórico y guía metodológica sobre la inclusión de variables latentes (considerando modelos MIMIC) en modelos de elección discreta, donde discute temas de identificabilidad y estimación. Además (Walker, 2001) incluye aplicaciones prácticas considerando distintas funciones de elección (Logit y Probit).

Ashok et al. (2002) presentan modelos de elección discreta en que se incluyen variables latentes a través de análisis factorial[2] y se discuten algunos problemas asociados a la estimación de estos modelos, principalmente debido a la aparición de las integrales multidimensionales que se deben resolver al momento de intentar la calibración.

[2] El análisis factorial es otro caso particular de ecuaciones estructurales.

2.2 Antecedentes prácticos

En general, existen muy pocos trabajos en la literatura donde se incluyan variables latentes en modelos de elección discreta. En nuestro país se han realizado estudios en que variables asociadas al riesgo se han incluido directamente en la función de utilidad. Aunque este tipo de variables no caen dentro de la categoría de "variables latentes", sí dan cuenta del esfuerzo por enriquecer los modelos de elección discreta con atributos menos tradicionales. Dentro de este esfuerzo, podemos destacar la experiencia obtenida a través de la estimación de modelos de elección para rutas interurbanas, que incluían una variable asociada al nivel de riesgo o seguridad, mediante encuestas de preferencias declaradas (Rizzi y Ortúzar, 2003). También se han realizado trabajos similares para el caso de elección de rutas urbanas (Iragüen y Ortúzar, 2004).

Un importante salto cualitativo y metodológico lo dan Rizzi et al. (2005) cuando realizan un estudio de elección discreta para usuarios de buses y trenes interurbanos hacia el sur de Chile, en que incluyen indicadores de calidad de servicio como *proxy* de variables latentes subyacentes. En ese trabajo se consideraron 10 indicadores[3]:

- La *conveniencia* de los horarios de viaje (mayor disponibilidad de viajes diarios).
- La *facilidad* para viajar con niños o bultos pesados.
- El *aprovechamiento del tiempo* en el viaje para leer, escribir, trabajar o realizar otras actividades.
- La *comodidad* durante el viaje (calidad de los asientos, espacios disponible, entre otras).
- La *puntualidad* en el cumplimiento del horario del viaje.
- La *relajación* durante el viaje (capacidad para disfrutar el viaje).
- La *seguridad* al viajar (sensación de que no ocurra un accidente en el transcurso del viaje).
- La *seguridad* en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente).
- La *seguridad* en cuanto a robo/hurto al interior del vehículo/coche tren.
- La *seguridad* en cuanto a robo/hurto en estaciones terminales.

[3] Estos 10 indicadores surgieron de la discusión y experiencia recabada en un grupo focal.

Mediante un análisis de componentes principales, se resumió la información de estos 10 indicadores en tres factores o dimensiones: comodidad, seguridad y conveniencia. Estos factores se cargaron directamente al modelo de elección discreta. Desafortunadamente, los resultados obtenidos a nivel de ajuste general y significancia de los coeficientes asociados a los factores, no fueron satisfactorios. A pesar de lo anterior, se decidió utilizar dicha experiencia como punto de partida para el desarrollo de este estudio, principalmente en términos de la reutilización de todos los indicadores y algunas de las preguntas que se incluyeron en la encuesta.

No obstante, estos trabajos no aplican las técnicas especialmente desarrolladas por Mc Fadden (1986) y Ben-Akiva et al. (1999) para la inclusión de variables latentes en modelos de elección discreta., por lo que no se logra develar en forma explícita el proceso mental de formación de percepciones.

Walker (2001) y Ashok et al. (2002) presentan interesantes aplicaciones prácticas asociadas a elección de productos y decisiones de viaje.

En encuestas de preferencias reveladas vía correo electrónico efectuadas en Suecia (Johansson et al., 2005), los encuestados disponían de tres modos alternativos (bus, auto y tren) para viajar entre las ciudades de Upssala y Estocolmo. Además de las variables tradicionales, como costo y tiempo de viaje, se incluyeron cinco variables latentes al modelo de elección discreta: comodidad, seguridad, flexibilidad, conveniencia y conciencia ambiental. Los resultados a nivel de ajuste general fueron satisfactorios y se concluyó que la inclusión de variables latentes enriquecía los modelos de elección discreta.

Temme et al. (2007) efectuaron una encuesta telefónica en Alemania para usuarios de bus, tren y auto, en que estimaron modelos de elección de modo de transporte que incluían variables latentes, como seguridad, comodidad y flexibilidad. Los resultados obtenidos fueron positivos, resaltando la importancia que los usuarios atribuían a la flexibilidad. En este estudio, a diferencia del desarrollado por Johansson et al. (2005), se estimó los modelos en forma simultánea. Este tema será tratado con mayor detalle en sección 3.3.

Aunque los estudios anteriores[4] utilizan las técnicas descritas por Mc Fadden (1986) y Ben-Akiva et al. (1999), éstos incluyen pocas variables causales que determinen las dimensiones latentes. Esta tesis pretende presentar mayor cantidad de atributos causales, entre los cuales destacan la tasa de accidentabilidad de los modos y vivencias personales de los usuarios. Además, se discutirá tanto la posibilidad de obtener disposiciones al pago por mejoras en los atributos latentes considerados, como las limitaciones prácticas de estas técnicas.

[4] Me refiero a: Walker (2001), Ahok et al. (2002), Johansson et al. (2005) y Temme et al. (2007).

3 MARCO TEÓRICO

En esta tesis se utilizan técnicas de ecuaciones estructurales y modelos de elección discreta. Para una mejor comprensión, este capítulo trata los temas por separado en las secciones 3.1 y 3.2, para luego analizarlos en forma conjunta en la sección 3.3.

3.1 Ecuaciones estructurales

3.1.1 Presentación

Las ecuaciones estructurales son técnicas estadísticas desarrolladas para determinar relaciones causales[5] entre grupos de variables observables y grupos de variables o conceptos no observables directamente.

La principal diferencia entre los modelos basados en ecuaciones estructurales y otro tipo de metodologías, como el análisis de regresión, es que los primeros buscan minimizar la distancia entre la matriz de varianza-covarianza muestral (S) y aquélla resultante del modelo propuesto ($\sum(\theta)$), en vez de minimizar las discrepancias entre observaciones caso a caso.

Según Bollen (1989) existen principalmente tres tipos de modelos generales de ecuaciones estructurales:

- Análisis de trayectoria (*Path Analysis*).
- Métodos de estimación.
- Conceptualización de variables latentes (*latent variables*) y modelos de medición (*measurement models*).

En este trabajo nos centraremos principalmente en el último tipo.

[5] No se pretende hacer un análisis de causalidad en este trabajo. Por "causar" se debe entender "cargar" o "afectar".

Por variables latentes se entiende a conceptos abstractos, unidimensionales, hipotéticos y no observables, que se construyen a través de indicadores y variables objetivas medibles. Tradicionalmente el poder, la inteligencia y la satisfacción, entre otros, han sido modelados como variables latentes. En nuestro estudio, la comodidad y la seguridad corresponden a variables latentes.

Los modelos de medición son aquellas formas funcionales que vinculan estos conceptos no medibles, con variables e indicadores observables y/o medibles. Los modelos MIMIC (*multiple indicator and multiple cause*) son un caso particular de modelos de medición, en que las variables objetivas "causan" a las variables latentes, las que a su vez "causan" a las variables indicadoras. La estructura de un modelo MIMIC típico es la siguiente:

$$\eta = \Gamma x + \zeta \qquad (3.1)^{6}$$

$$y = \Lambda_y \eta + \varepsilon \qquad (3.2)^{7}$$

en que:

η : corresponde al vector de variables latentes[8].

Γ: corresponde al vector de coeficientes que cargan a las variables latentes.

x: corresponde al vector de dimensión *q* de variables objetivas medibles, que "causan" las variables latentes (variables socioeconómicas, características del modo, etc.).

ζ: corresponde al vector de errores estocásticos asociados a (3.1).

y: corresponde al vector de dimensión *p* de variables indicadoras.

Λ_y: corresponde al vector de coeficientes que cargan a los indicadores.

ε: corresponde al vector de errores estocásticos asociados a (3.2).

---

[6] Esta ecuación se denomina "modelo estructural".
[7] Esta ecuación se denomina "modelo de medición" o "modelo de medida".
[8] En la sección 5.1 se explica que las variables latentes utilizadas en este trabajo son dos: $\eta_{seguridad}$ y $\eta_{comodidad.}$

La ecuación 3.1 se conoce como "modelo estructural" y la ecuación 3.2 se denomina "modelo de medición" o "modelo de medida".

Los principales supuestos de este modelo son:

$$E(\zeta) = E(\varepsilon) = 0 \tag{3.3}$$

$$Cov(\zeta, \varepsilon) = 0 \tag{3.4}$$

$$Cov(\zeta_i, \zeta_j) = Cov(\varepsilon_i, \varepsilon_j) = 0 \qquad \forall \quad i \neq j \tag{3.5}$$

Además, se utilizará la siguiente notación:

$$E(\varepsilon, \varepsilon') = \Phi \text{ (matriz diagonal de varianza-covarianza de } \varepsilon) \tag{3.6}$$

$$E(\zeta, \zeta') = \Psi \text{ (matriz diagonal de varianza-covarianza de } \zeta) \tag{3.7}$$

3.1.2 Identificabilidad

Para que el modelo MIMIC sea identificable, además de que no hayan más coeficientes que ecuaciones en (3.1) y (3.2), se debe fijar la escala de las variables latentes (esto se denomina *normalizar*). Stapleton (1978) recomienda fijar en uno, cualquier elemento λ distinto de cero[9] de cada columna de Λ_y.

Supongamos que tenemos 5 variables indicadoras (y) y 2 variables latentes (η). En este caso la ecuación (3.2) tomaría la siguiente forma:

$$\begin{bmatrix} \boldsymbol{y_1} \\ \boldsymbol{y_2} \\ \boldsymbol{y_3} \\ \boldsymbol{y_4} \\ \boldsymbol{y_5} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda_{1\,1}} & \boldsymbol{\lambda_{1\,2}} \\ \boldsymbol{\lambda_{2\,1}} & \boldsymbol{\lambda_{2\,2}} \\ \boldsymbol{\lambda_{3\,1}} & \boldsymbol{\lambda_{3\,2}} \\ \boldsymbol{\lambda_{4\,1}} & \boldsymbol{\lambda_{4\,2}} \\ \boldsymbol{\lambda_{5\,1}} & \boldsymbol{\lambda_{5\,2}} \end{bmatrix} \cdot \begin{bmatrix} \boldsymbol{\eta_1} \\ \boldsymbol{\eta_2} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\varepsilon_1} \\ \boldsymbol{\varepsilon_2} \\ \boldsymbol{\varepsilon_3} \\ \boldsymbol{\varepsilon_4} \\ \boldsymbol{\varepsilon_5} \end{bmatrix} \tag{3.2b}$$

[9] Un elemento puede ser cero si es que se postula *a priori* que no existe relación entre dicha variable latente y dicho indicador. Por ejemplo, si en la ecuación (3.2b) se postula que no existe relación entre η_1 e y_2, entonces $\lambda_{1\,2}$ será cero.

Si elegimos (arbitrariamente, pero tomando en cuenta que sean distintos de cero) $\lambda_{2\,1}$ y a $\lambda_{4\,2}$ como los elementos a fijar en uno, obtendríamos:

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \end{bmatrix} = \begin{bmatrix} \lambda_{1\,1} & \lambda_{1\,2} \\ 1 & \lambda_{2\,2} \\ \lambda_{3\,1} & \lambda_{3\,2} \\ \lambda_{4\,1} & 1 \\ \lambda_{5\,1} & \lambda_{5\,2} \end{bmatrix} \cdot \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \end{bmatrix} \tag{3.2c}$$

En general, este es el camino tomado por la mayoría de los investigadores (Bollen, 1989, pp. 18).

Alternativamente, Stapleton (1978) demuestra que el modelo también es identificable si se normalizan los elementos diagonales de la matriz de covarianzas Ψ.

Supongamos que tenemos 2 variables latentes (η) y 4 variables objetivas (x). La ecuación (3.1) tomaría la siguiente forma:

$$\begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} = \begin{bmatrix} \gamma_{1\,1} & \gamma_{1\,2} & \gamma_{1\,3} & \gamma_{1\,4} \\ \gamma_{2\,1} & \gamma_{2\,2} & \gamma_{2\,3} & \gamma_{2\,4} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} + \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \tag{3.1b}$$

Si se normalizan en "uno" los elementos diagonales de la matriz de varianza covarianza Ψ, obtenemos[10]:

$$\Psi = \begin{bmatrix} \zeta_1 & 0 \\ 0 & \zeta_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{3.6b}$$

[10] No debemos olvidar los supuestos planteados por la ecuación 3.5.

A diferencia de la normalización de los indicadores, en que la mayoría de los investigadores fija en "uno" cualquier elemento distinto de cero de cada columna de Λ_y (Bollen, 1989, pp. 18), para el caso de las varianzas no existe tal convención. Así, para este ejemplo, ζ_1 y ζ_2 podrían haber tomado cualquier valor real positivo.

Sin embargo, Sapleton (1978) recalca que aunque la normalización no tiene efectos en los signos ni en los test estadísticos resultantes asociados a las variables, sí hace difícil dar una interpretación cuantitativa a los coeficientes estimados. Esta discusión será retomada en la sección 5.3.2 y en el capítulo 5.4.

3.1.3 Estimación

Para la estimación de los modelos anteriores, existen diversos caminos posibles, todos ellos direccionados a minimizar la distancia entre la matriz de covarianzas muestral (S) y aquélla resultante del modelo propuesto ($\sum(\theta)$).

La función de distancia debe cumplir con las siguientes condiciones (Bollen, 1989, pp. 106):

- $F(S, \sum(\theta))$ es un escalar.
- $F(S, \sum(\theta)) \geq 0$
- $F(S, \sum(\theta))$ es continua en S y $\sum(\theta)$
- $F(S, \sum(\theta)) = 0$ si y solo si $S = \sum(\theta)$

Luego, el problema a resolver consiste en encontrar un vector de coeficientes θ que minimice F. La función de verosimilitud cumple con los requisitos de $F(S, \sum(\theta))$, por lo que es posible estimar θ a través de máxima verosimilitud. Para los modelos MIMIC podemos escribir S y $\sum(\theta)$ de la siguiente forma:

$$S=\begin{bmatrix}\sum_{yy} & \sum_{yx}\\ \sum_{xy} & \sum_{xx}\end{bmatrix} \tag{3.8}$$

en que:

$\sum_{yy}$ = es la matriz de varianza covarianza muestral de los indicadores “y”.

$\sum_{xx}$ = es la matriz de varianza covarianza muestral de las variables objetivas “x”.

$\sum_{xy}$ = es la matriz de covarianza muestral entre las variables objetivas “x” y los indicadores “y”.

$\sum_{yx}$ = es la matriz de covarianza muestral entre los indicadores “y” y las variables objetivas “x” ($\sum_{yx} = \sum_{xy}'$).

Por otra parte:

$$\sum(\theta)=\begin{bmatrix}\Lambda_y(\Gamma\Phi\Gamma'+\Psi)\Lambda_y'+\phi & \Gamma\Phi\Lambda_y\\ \Phi\Gamma'\Lambda_y' & \Phi\end{bmatrix} \tag{3.9}$$

en que todos los términos de (3.9) están descritos en la sección 3.1.1.

Si *x* e *y* distribuyen Normal multivariado, la función de log verosimilitud para modelos MIMIC es la siguiente (Bollen 1989, pp. 133-134):

$$L_1 = \frac{-N(p+q)}{2} ln|2\pi| - \frac{N}{2} ln|\sum(\theta)| - \frac{1}{2}\left(\sum_{i=1}^{N} z_i' \cdot \sum(\theta)^{-1} \cdot z_i\right) \tag{3.10}$$

en que N es el tamaño de la muestra y z_i corresponde al vector de desviaciones con respecto a la media de las variables objetivas “*x*” y de los indicadores “*y*”, es decir:

$$z_i = (x_1 - \overline{x}_1, x_2 - \overline{x}_2, \dots\dots, x_q - \overline{x}_q, y_1 - \overline{y}_1, \dots\dots, y_p - \overline{y}_p) \tag{3.11}$$

La ecuación (3.10) puede reescribirse de la siguiente forma (Bollen 1989, pp. 133-134):

$$L_1 = \frac{-N(p+q)}{2} ln|2\pi| - \frac{N}{2} ln|\sum(\theta)| - \frac{1}{2}\left(\sum_{i=1}^{N} traza[z_i' \cdot \sum(\theta)^{-1} \cdot z_i]\right)$$

$$= \frac{-N(p+q)}{2} ln|2\pi| - \frac{N}{2} ln|\sum(\theta)| - \frac{N}{2}\left(\sum_{i=1}^{N} traza[N^{-1} \cdot z_i \cdot z_i' \cdot \sum(\theta)^{-1}]\right)$$

$$= \frac{-N(p+q)}{2} ln|2\pi| - \frac{N}{2} ln|\sum(\theta)| - \frac{N}{2}\left(\sum_{i=1}^{N} traza[N^{-1} \cdot z_i \cdot z_i' \cdot \sum(\theta)^{-1}]\right)$$

$$= \frac{-N(p+q)}{2} ln|2\pi| - \frac{N}{2} ln|\sum(\theta)| - \frac{N}{2}\ (traza\ [S \cdot \sum(\theta)^{-1}])$$

$$= constante - \frac{N}{2} ln|\sum(\theta)| - \frac{N}{2}\ (traza\ [S \cdot \sum(\theta)^{-1}]) \qquad (3.10b)$$

Sin pérdida de generalidad, Bollen (1989) pp. 134, plantea que es posible reemplazar el término "constante" que aparece en la ecuación 3.10b por otro término constante (*-ln|S|-(p+q)*) sin afectar la maximización de L_1. Similarmente (Bollen (1989) pp. 134), propone amplificar (3.10b) por $(-\frac{2}{N})$, para obtener la función de log verosimilitud utilizada en la práctica:

$$L = ln|\sum(\theta)| + \ traza\ (S \cdot \sum(\theta)^{-1}) - \ln|S| - \ (p+q) \qquad (3.12)$$

Dado que la maximización se hace con respecto a *θ,* los últimos dos sumandos de la ecuación (3.12) no tienen relevancia para el problema.

3.2 Modelos Logit de elección discreta

A continuación se presenta un breve resumen de las principales características de estos modelos (Ortúzar y Willumsen, 2001).

En la teoría económica clásica, se asume que el individuo se comporta de forma racional, en el sentido de elegir la alternativa o camino que maximiza su utilidad (satisfacción personal). Sin embargo, el investigador no posee toda la información relevante para comprender el proceso de decisión del individuo; por esto y para hacer posible la modelación, se plantea que la función de utilidad tiene una componente medible por el investigador (V) y otra aleatoria (ε), que no lo es.

Así, la función de utilidad que le brinda la alternativa i al individuo n, viene dada por:

$$U_{in} = V_{in} + \varepsilon_{in} \tag{3.13}$$

Si los errores ε_{in} distribuyen IID Gumbel con media cero, se obtiene la típica expresión del modelo Logit simple multinomial (MNL).

$$P_i = \frac{e^{\beta V_i}}{\sum e^{\beta V_j}} \tag{3.14}$$

$$A_j \in A(n)$$

En que P_i es la probabilidad de escoger la alternativa i, β es el factor de escala de la distribución Gumbel (Ortúzar y Willumsen, 2001), A_j representa a la alternativa j dentro del conjunto de alternativas A_n del individuo n.

Si la muestra utilizada para estimar los modelos está basada en la elección, y si todos los individuos tienen disponibles todas las opciones, es decir, A_n es igual para todos, se

deben corregir todas las constantes modales (incluida aquella de referencia que se hace cero) de la siguiente forma (Manski y Lerman 1977):

$$K_i' = K_i - \ln\left(\frac{q_i}{Q_i}\right) \tag{3.15}$$

en que K'_i son las nuevas constantes, K_i son las constantes a corregir, q_i es la proporción de la alternativa A_i en la muestra y Q_i es la proporción de mercado de la alternativa A_i.

Para calibrar (3.14) es necesario maximizar la siguiente función de log verosimilitud:

$$L_2 = \sum_{n=1}^{N} \sum_{A_j \in A(n)} g_{jn} \cdot ln\left(P_{jn}\right) \tag{3.16}$$

en que g_{jn} vale uno si el individuo n escoge la alternativa j (0 en otro caso).

3.3 Modelos Logit-MIMIC integrados

En la sección 3.1 se introdujo el tema de variables latentes y ecuaciones estructurales. Recordemos que por variable latente se entiende un concepto abstracto no medible directamente. En nuestro caso, estas variables latentes conforman percepciones individuales sobre la comodidad y la seguridad, determinadas por elementos objetivos (x) y por errores estocásticos (ζ), que permiten la variación de estas percepciones entre individuos. Por otra parte, dada la condición abstracta de las variables latentes, éstas sólo pueden ser expresadas por el individuo a través de indicadores (y). Por último, la aleatoriedad de este proceso es capturada por los errores ε.

Las variables latentes (η) construidas, pueden ser añadidas a la función indirecta de utilidad (8) de la siguiente forma:

$$U_{in} = z_i \cdot \alpha + \eta_{in} \cdot \beta + \varepsilon_{in} \tag{3.17}$$

donde z_i es un vector de atributos (e.g. costo, tiempo de viaje, ingreso, etc.), α y β son coeficientes a estimar que explican el efecto de z_i y de η_{in} en la función de utilidad.

La estimación de los modelos resultantes puede resolverse de dos formas: simultánea o secuencial. Ambos caminos serán explicados en las siguientes secciones.

3.3.1 Estimación simultánea de modelos Logit-MIMIC integrados

La probabilidad conjunta de observar la elección d_i y el conjunto de indicadores "y" viene dada por:

$$P\left(d_i = 1, y|z, x, \alpha, \beta, \Lambda_y, \Gamma\right) = \int_{\eta} P(d_i = 1|z, \eta, \alpha, \beta) g\left(y|\eta, \Lambda_y\right) h(\eta|x, \Gamma) d\eta \quad (3.18)$$

Luego, para cada individuo, la función de verosimilitud puede escribirse como:

$$L_n = \prod_{i \in A_n} P\left(d_{in} = 1, y_n|z, x_n, \alpha, \beta, \Lambda_y, \Gamma\right)^{d_i}$$

$$= \int_{\eta} \prod_{i \in A_n} P(d_{in} = 1|z, \eta, \alpha, \beta) g\left(y_n|\eta, \Lambda_y\right) h(\eta|x_n, \Gamma) d\eta \quad (3.19)$$

Por lo tanto, la función de verosimilitud para el total de observaciones (N) es:

$$L_N = \prod_{n=1}^{N} \prod_{i \in A_n} P\left(d_{in} = 1, y_n|z, x_n, \alpha, \beta, \Lambda_y, \Gamma\right)^{d_{in}}$$

$$= \prod_{n=1}^{N} \int_{\eta} \prod_{i \in A_n} P(d_{in} = 1|z, \eta, \alpha, \beta) g\left(y_n|\eta, \Lambda_y\right) h(\eta|x_n, \Gamma) d\eta \quad (3.20)$$

Para obtener el conjunto de coeficientes ($\alpha,\beta,\Lambda_y,\Gamma$) que maximiza (3.20) se puede aplicar una transformación logarítmica. Luego, la función a optimizar resulta:

$$\max_{\alpha,\beta,\Lambda y,\Gamma} \sum_{n=1}^{N} \sum_{i \in A_n} d_{in} \cdot ln\left(P\left(d_{in} = 1, y_n | z, x_n, \alpha, \beta, \Lambda_y, \Gamma\right)\right) \tag{3.21}$$

El problema recién expuesto puede ser complejo de resolver en términos computacionales. Por lo general, conlleva la resolución de integrales multidimensionales que dependen del número de variables latentes consideradas. En la siguiente sección se plantea un método alternativo de resolución.

3.3.2 Estimación secuencial de modelos Logit-MIMIC integrados

En este caso se estima primeramente el modelo MIMIC. Con el modelo estructural[11] estimado, se pueden construir las variables latentes para cada individuo de acuerdo a (3.1). Las variables latentes construidas, alimentan al modelo de elección discreta, de la forma descrita en (3.17). Por lo general, existen paquetes estadísticos diseñados para resolver cada etapa del problema, por lo que este proceso resulta bastante conveniente.

Sin embargo, este enfoque no se hace cargo del hecho que las variables latentes construidas poseen errores de medición. Esta situación provoca que los errores estándar de los coeficientes estimados en la segunda etapa, resulten menores a los que realmente serían si el modelo de la segunda etapa "supiese" que algunos de los regresores que lo alimentan poseen errores de medición asociados.

Murphy y Topel (1985) proponen un método para corregir este problema y calcular la verdadera matriz de covarianzas de los coeficientes de la segunda etapa. Además, indican que por lo general los procesos de estimación secuencial conllevan estimadores consistentes para los coeficientes estimados en la segunda etapa. Este método se describe en detalle en el Anexo 1.

[11] Corresponde a la ecuación 3.1 (ver sección 3.1.1).

4 DESCRIPCIÓN DE LA ENCUESTA

4.1 Metodología

La encuesta se basó en el formulario diseñado por Rizzi et al. (2005), al que se le introdujeron diversas modificaciones basadas en las experiencias de las encuestas pre-piloto y piloto (descritas en los siguientes párrafos). Todas las encuestas fueron realizadas en la Estación Central de trenes y en el Terminal Alameda, que alberga a las compañías de buses Pullman Bus y Tur Bus. Los destinos considerados fueron Curicó, Talca, Linares y Chillán.

El principal objetivo de la encuesta pre-piloto fue determinar el tiempo aproximado de duración de cada entrevista y la forma correcta de aproximarse a los pasajeros; además se probó la formulación de preguntas de los indicadores[12]. Se realizaron 20 encuestas (10 usuarios de buses y 10 usuarios de trenes). Se encontró que la duración promedio de cada entrevista era de 10 minutos y que las personas comprendían bien el diseño y sentido general de las preguntas.

La etapa piloto consideró 50 encuestas para usuarios de buses y 50 para usuarios de tren. Dado que en la etapa pre-piloto algunas personas tuvieron dificultades contestando en una escala de 1 a 7 las preguntas relacionadas con los indicadores (ver sección 2.2), la mitad de las encuestas se realizaron con una escala de 1 a 7 y la otra mitad, con una escala semántica que consideraba 5 niveles desde “muy malo” hasta “muy bueno”. El resultado fue que aunque algunas personas tenían dificultades con la escala de 1 a 7, para la gran mayoría resulta mucho más familiar que la escala semántica propuesta, por lo que se optó por mantener la escala de 1 a 7 para la etapa final. Con la encuesta piloto se estimaron modelos de elección discreta preliminares, para probar la calidad de los datos. Estos modelo incluyeron “variables latentes” construidas de la forma utilizada por

[12] En todas las etapas de la encuesta (pre piloto, piloto y final) se utilizaron exactamente los mismos indicadores diseñados por Rizzi et al. (2005) introducidos en la sección 2.2.

Rizzi et al. (2005) (mediante análisis factorial). Los resultados fueron correctos, por lo que se creyó conveniente dar paso a la encuesta final.

La encuesta final se realizó en la semana del 3 al 7 de noviembre del 2008 y consistió en 600[13] encuestas válidas (libres de errores), 301 encuestas a usuarios de buses y 299 a usuarios de trenes. Las encuestas fueron desarrolladas por un equipo de cuatro personas (además del tesista) que fueron entrenadas con este fin.

La jornada de trabajo comenzaba a las 6:30 A.M. en el terminal de trenes (para lograr encuestar a las personas que viajaban en el servicio expreso de las 7:30 A.M. con destino a las ciudades de Talca y Chillán) y finalizaba alrededor de las 19:30[14] en el terminal de buses. Las encuestas se realizaban en las boleterías y andenes de la Estación Central. A lo largo de la jornada, el equipo de encuestadores rotaba entre las estaciones de buses y la estación de trenes, para que todos los integrantes del equipo encuestasen de forma más o menos homogénea la misma cantidad de usuarios de bus y tren. Para la encuesta final, se consideró apropiado agregar preguntas con respecto a:

- Experiencias previas relativas a accidentes o robos sufridos en bus y/o tren ocurridas en viajes anteriores. Se esperaba que las vivencias previas influyeran en las percepciones de seguridad de los viajeros con respecto al modo.
- Si el usuario viajaba o no con computador portátil (*notebook*). Se observó que las personas que viajaban con computador personal, por lo general trabajan durante el viaje y además tienen una preocupación extra con respecto a la seguridad. Esta condición, podría resultar causal de las variables latentes consideradas.
- Siguiendo las recomendaciones de Lyons y Urry (2005) y de Lyons et al. (2007) se incluyó preguntas relacionadas con la actividad que la persona realizaba

[13] Para decidir el tamaño muestral se tomó como experiencia el trabajo de Rizzi et al. (2005). En dicho estudio, el tamaño fue de 600 encuestas y los resultados obtenidos fueron razonables.

[14] Esta jornada laboral permitía encuestar a pasajeros de todas las salidas de trenes, dado que el último tren con destino al sur partía a las 18:30.

durante el viaje (aquella actividad que le dedicará la mayor cantidad de tiempo). Se esperaba que las actividades realizadas afectaran la percepción de la comodidad.

4.2 Análisis de la muestra

Como se menciona en la sección anterior, la muestra consta de 600 observaciones válidas: 301 usuarios de bus y 299 usuarios de tren. Para el caso del Bus los destinos se distribuyen en forma más o menos homogénea, sin embargo, en el tren la gran mayoría de los viajes se realiza a Chillán. Esto se debe principalmente a la existencia de tres servicios expresos al día, que se desplazan sin interrupciones a las ciudades de Talca y Chillán, situación bastante conveniente para los usuarios. En la Figura 4-1 se puede observar la distribución de destinos por modo.

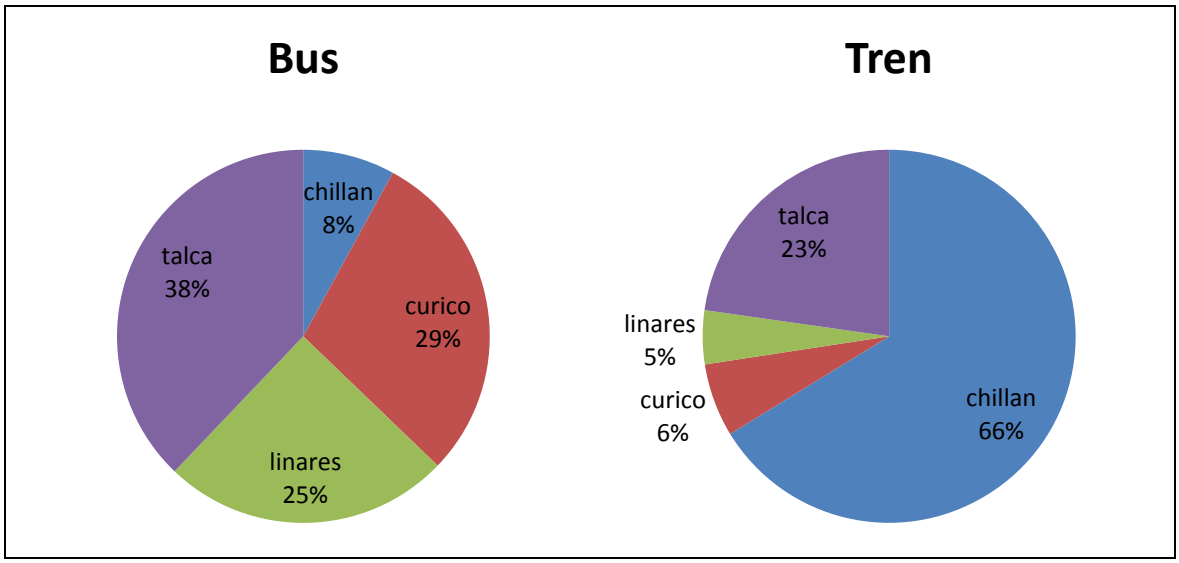

Figura 4-1: Distribución de destinos por modo

Similarmente, los motivos de viaje no presentan grandes diferencias por modo. Puede observarse, sin embargo, que las personas prefieren utilizar el tren para motivos recreacionales y el bus para visitas personales. La Figura 4-2 resume la información de

motivos de viaje por modo. Por otra parte, la distribución del ingreso[15] presenta bastantes diferencias entre ambos modos. Por lo general, los usuarios de tren perciben mayores rentas que los usuarios de bus. La Figura 4-3 da cuenta de esta situación.

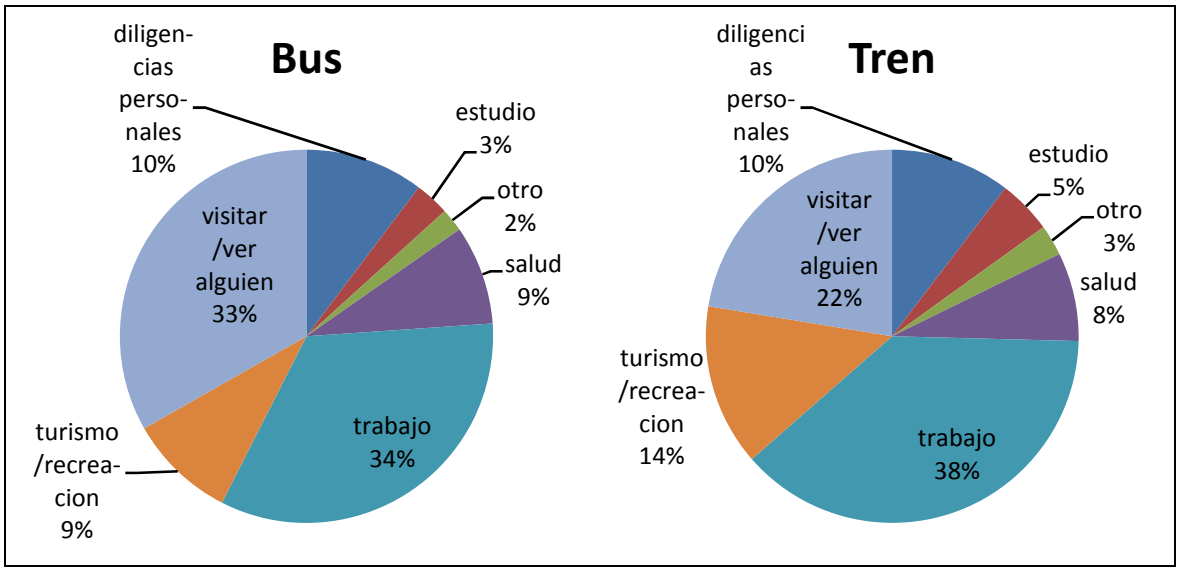

Figura 4-2: Distribución de motivos de viaje por modo

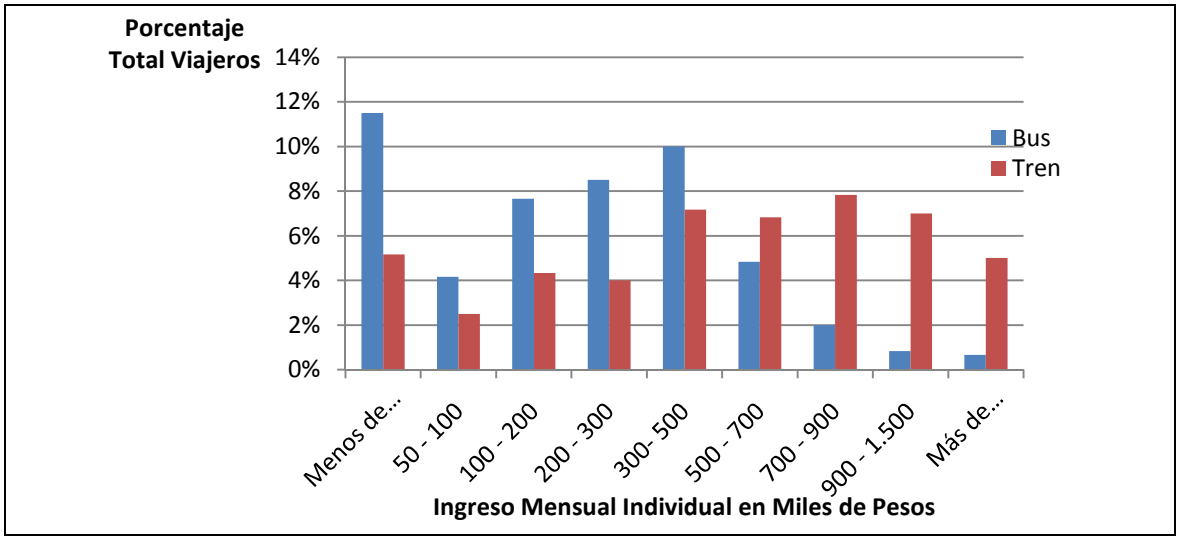

Figura 4-3: Distribución del ingreso por modo

[15] Ingreso mensual individual.

Similarmente al caso del ingreso, existe una marcada diferencia entre los usuarios de bus y de tren con respecto al nivel de estudios que poseen. Por ejemplo, un 52,2% de usuarios de tren declaran poseer educación universitaria completa, comparado con un 14,3% de usuarios de bus con esta condición. La Tabla 4-1 resume esta información.

Tabla 4-1: Nivel de Educación por tipo usuario.

Educación	Bus	Tren
Básica incompleta	6.0%	1.7%
Básica completa	7.3%	2.3%
Media incompleta	9.6%	2.7%
Media completa	34.2%	18.4%
Técnica incompleta	3.0%	1.0%
Técnica completa	12.6%	12.0%
Universitaria incompleta	13.0%	9.7%
Universitaria completa	14.3%	52.2%
Total	100%	100%

Por otra parte, si se realiza una descripción de la actual ocupación de los encuestados (Tabla 4-2), se observa que los jubilados y empleados utilizan más el tren y que las dueñas de casa y estudiantes utilizan más el bus.

Tabla 4-2: Ocupación por tipo de usuario.

Ocupación	Bus	Tren
Cesante	1.7%	0.0%
Dueña de casa	19.9%	11.7%
Empleado	32.6%	47.2%
Estudiante	12.6%	8.7%
Independiente	25.9%	24.1%
Jubilado	3.7%	7.4%
Sin ocupación	1.7%	0.7%
Otro	2.0%	0.3%
Total	100%	100%

Con respecto a la actividad que los viajeros declaran realizar durante el viaje, se observa que los usuarios de tren desarrollan principalmente actividades relacionadas a la lectura y el trabajo y por otra parte, los pasajeros de tren aprovechan el tiempo para dormir y leer. La Tabla 4-3 contiene esta información.

Tabla 4-3: Actividad desarrollada durante el viaje por tipo de usuario.

Actividad	Bus	Tren
aburrirse	0.7%	0.7%
comer/beber	1.0%	2.7%
dormir	29.9%	11.4%
hablar por celular	1.7%	1.0%
hablar con otro pasajero	6.6%	9.0%
jugar con hijos o nietos	1.3%	2.3%
leer por placer	26.6%	40.8%
mirar paisaje/personas	16.9%	11.0%
escuchar música o radio	13.6%	7.0%
trabajar/estudiar	1.7%	13.4%
otro	0.0%	0.7%
Total	100%	100%

Un análisis de las variables vivenciales contenidas en la encuesta, nos muestra que una gran cantidad de usuarios (independientemente del modo elegido), declaran haber visto o sufrido un accidente o robo en el bus. Como se mostrará en el Capítulo 5, estas variables impactan directamente las percepciones subjetivas de los usuarios con respecto a la seguridad.

La Tabla 4-4: Variables vivenciales por tipo de usuario. resumen esta información.

Tabla 4-4: Variables vivenciales por tipo de usuario.

Variable vivencial	Bus	Tren	Total
¿Ha visto o sufrido algún accidente en tren?	21	43	64
¿Ha visto o sufrido algún accidente en bus?	72	82	154
¿Ha visto o sufrido algún robo en tren o terminal de tren?	23	12	35
¿Ha visto o sufrido algún robo en bus o terminal de bus?	89	87	176

Las variables indicadoras presentan importantes diferencias por modo. En general, el tren obtiene mejores notas en todos los atributos, exceptuando el ítem "conveniencia de horarios de viaje". Sin embargo, la diferencia porcentual de la nota en este caso (8,9%) es bastante baja en comparación con las diferencias observadas en los otros atributos (por ejemplo "seguridad al viajar" con 28,1% de diferencia), sugiriendo que los usuarios no perciben grandes diferencias de conveniencia para el tren y el bus. Esta información se encuentra detallada en el Anexo 2. Una revisión desagregada por usuario muestra tendencias casi idénticas. Los usuarios de Bus evalúan mejor al tren en todos los indicadores, exceptuando la "conveniencia de los horarios de viaje". Esta información se encuentra en el Anexo 3. Por otra parte, los usuarios de Tren evalúan mejor al tren en todos los indicadores sin excepciones (ver Anexo 4). Además podemos decir que en promedio, los usuarios de tren evalúan mejor al tren que los usuarios de bus (30,0% contra un 7,8% de diferencia porcentual, ver Anexo 4 y Anexo 3 respectivamente).

En el Anexo 5 y el Anexo 6 podemos observar las respuestas a las variables indicadoras desagregadas por género, con respecto al bus y al tren respectivamente. No se observan diferencias importantes, lo que sugiere que el género no afecta mayormente a las variables indicadoras.

Por último, en los Anexos Anexo 7 y Anexo 8 se muestran las respuestas a las variables indicadoras desagregadas por tramo de edad, con respecto al bus y al tren respectivamente. Resulta difícil realizar comentarios a través de la inspección los resultados expuestos, ya que no se observan grandes diferencias en las respuestas de los usuarios de distintas edades para un mismo modo. No obstante lo anterior, Johansson et al. (2008) y Temme et al. (2007) obtienen que la edad influye positivamente en las percepciones de comodidad y seguridad.

5 ESTIMACIÓN DE MODELOS

En este capítulo se presentan los distintos pasos efectuados para llegar a los modelos finales. En la primera sección (5.1), se explican los procedimientos que determinaron el número de variables latentes a considerar y los indicadores con que se vinculan. En la segunda sección (5.2), se expone el modelo MIMIC calibrado. Por último, en la sección 5.3 se presentan los modelos de elección discreta que servirán de punto de comparación con el modelo final (que se detalla en la sección 5.3.2). Vale destacar que para la resolución de los modelos MIMIC se utilizó el *software* AMOS de SPSS (Arbuckle J, 2006) y para la calibración de los modelos de elección discreta, se utilizó el programa ALOGIT (Daly, 1995).

5.1 Análisis factorial exploratorio

Quizás las preguntas más importantes que el investigador debe responder cuando se enfrenta a problemas que incluyan variables latentes son: "¿cuántas variables incluir?" y "¿qué variables incluir?".

Parte fundamental de la solución viene dada por los objetivos del estudio y los resultados obtenidos durante las sesiones de grupo focal. Durante esta etapa el investigador va conociendo aquellas cosas que las personas consideran relevantes en sus decisiones. Una vez que se logra identificar un conjunto de variables importantes para un problema dado, se procede a construir preguntas indicadoras de estas variables. Por ejemplo, si se determina que la "conciencia ecológica" es una variable que las personas incluyen en sus decisiones, uno podría construir preguntas enfocadas a conocer esta inclinación, por ejemplo: "en una escala de 1 a 7 qué importancia le da usted al reciclaje", o "en una escala de 1 a 7 cómo encuentra el nivel de ruido en la ciudad". Qué preguntas incluir y cómo formularlas, es un proceso delicado que debe considerar tanto la experiencia del grupo focal, los objetivos del estudio y aspectos lingüísticos entre otros.

Una vez resueltos estos temas, se debe considerar la realización de un análisis factorial (Bollen, 1989). Esta es una técnica estadística ampliamente utilizada en diversas disciplinas, como el marketing y la sicología, cuyo objetivo es reducir y condensar la información existente. El primer paso consiste en determinar el número de "dimensiones" o variables latentes a considerar. Con este fin, se realizó un análisis de componentes principales (Greene, 1997). El análisis de componentes principales es una técnica que busca reducir las dimensiones de los datos. Así, el objetivo es determinar el número de factores subyacentes a los datos que expliquen la mayor cantidad de la variabilidad de éstos.

La realización de este análisis sugirió conservar las dos primeras componentes (aquéllas mayores a uno[16]), ya que explicaban alrededor del 57% de la varianza total (ver Anexo 9). El posterior estudio de la matriz de soluciones rotada, nos permitió interpretar las dimensiones consideradas como "comodidad" y "seguridad"[17]. Una explicación detallada de lo realizado en esta fase se presenta en el Anexo 9.

A continuación, se realizó un análisis factorial exploratorio, para determinar el conjunto de indicadores que "cargan" las dos variables latentes encontradas ($\eta_{seguridad}$ y $\eta_{comodidad}$).

En términos simples, el análisis factorial corresponde a estimar la ecuación (3.2), es decir, postular relaciones causales que vinculen a las variables latentes (η) y a los indicadores (y).

---

[16] En el análisis de componentes principales por lo general se conservan las componentes mayores a uno.

[17] La tercera componente, "conveniencia horaria" fue desechada pues produjo malos resultados en los modelos MIMIC principalmente a nivel de signos. Por ejemplo, a medida que aumentaba la frecuencia de servicios por modo (variable asociada a la conveniencia horaria), empeoraba la percepción de la conveniencia horaria, lo que lógicamente no parece razonable.

El principal criterio para seleccionar modelos consistió en verificar que los efectos de las variables latentes sobre los indicadores fuese el esperado. Por ejemplo, a medida que aumenta la "seguridad" se esperaría que aumentara el indicador asociado a "seguridad en cuanto a robo en los terminales". En términos prácticos, esto quiere decir que si $\eta_{seguridad}$ aumenta en una unidad, el indicador (y) "seguridad en cuanto a robo en los terminales" debe aumentar λ_y unidades[18] (ver ecuación 3.2). Cabe destacar que durante esta etapa del proceso se probaron muchísimas formas para la ecuación 3.2, sin embargo, las que arrojaron mejores resultados tanto a nivel de ajuste general como de signos, fueron aquellas más intuitivas y simples.

5.2 Modelo MIMIC

En la sección anterior se presentó el procedimiento utilizado durante el análisis factorial. En la práctica, el análisis factorial fue realizado no sólo con las variables indicadoras (y) (ecuación 3.2), sino que incluyendo además las variables objetivas (x) como causales de las variables latentes, es decir, resolviendo conjuntamente las ecuaciones 3.1 y 3.2. Esto se debe a que en diversas ocasiones, formas funcionales correctas resultantes del análisis factorial (signos y ajuste general apropiados), pueden generar resultados inconsistentes (principalmente en signos) al incorporar las variables objetivas (x). De esta forma, el proceso de búsqueda de un modelo correcto se vuelve un asunto más complejo y de carácter iterativo (y que además, posteriormente, debe considerar el modelo de elección discreta, como se explica en la sección 5.3.2).

En la Tabla 5-1 se resume la información de la forma en que variables latentes e indicadores fueron relacionados. En la primera línea, se encuentran las variables latentes construidas (Comodidad Bus, Seguridad Bus, Comodidad Tren y Seguridad Tren). En la columna de la izquierda, se encuentran los 10 indicadores introducidos en la sección 2.2

[18] λ_y es el elemento de Λ_y que vincula a la variable latente $\eta_{seguridad}$ con el indicador "seguridad en cuanto a robo en los terminales".

Si la variable latente está relacionada con el indicador, es decir, que el elemento de Λ_y[19] asociado a esa variable latente y ese indicador es distinto de cero, en la Tabla 5-1 se lee "Sí". Por otra parte, si no existe relación entre la variable latente y el indicador, se presenta un espacio en blanco.

Análogamente, en la Tabla 5-2 se presentan las relaciones entre variables objetivas y variables latentes, es decir, una versión esquemática de los elementos de Γ[20]. Algunas de las variables que aparecen en la Tabla 5-2 merecen mayor explicación. Por ejemplo, "Desarrollo de actividades intelectuales" es una variable de tipo dicotómica que agrupa aquellas personas que van a leer, trabajar o estudiar durante el viaje; toma valor uno si desarrollan estas actividades y cero en otro caso.

"Ingreso Alto" agrupa a aquellas personas que perciben un ingreso mensual individual sobre los $400.000; toma valor uno si pertenece a este rango y cero en otro caso.

La variable "Desarrollo de actividades sociales" es de tipo dicotómica y vale uno para aquellas personas que van a conversar con otros pasajeros o jugar con sus niños durante el viaje y cero en otro caso.

El "Índice de fatalidades" es una variable que pretende hacerse cargo del nivel de riesgo al que se ven enfrentados los usuarios de la encuesta. Es esperable que el nivel de accidentes del modo repercuta en las percepciones de seguridad de los viajeros. Para construir este índice se utilizó información contenida en un banco de datos proporcionado por Conaset[21], que detalla el número de fallecidos en accidentes que involucraron buses de transporte público en las regiones en cuestión (Metropolitana, VI, VII y VIII) entre los años 2000 y 2007.

---

[19] Ver ecuación 3.2.

[20] Ver ecuación 3.1.

[21] Conaset (2007).

Tabla 5-1: Relación entre Indicadores y variables latentes

Indicador	Comodidad Bus	Seguridad Bus	Comodidad Tren	Seguridad Tren
La facilidad para viajar con niños o bultos pesados	Sí		Sí	
El aprovechamiento del tiempo en el viaje para leer, escribir, trabajar o realizar otras actividades	Sí		Sí	
La comodidad durante el viaje (calidad de los asientos, espacios disponible, entre otras)	Sí		Sí	
La puntualidad en el cumplimiento del horario del viaje	Sí		Sí	
La relajación durante el viaje (capacidad para disfrutar el viaje)	Sí	Sí	Sí	Sí
La seguridad al viajar (sensación de que ocurra un accidente en el transcurso del viaje)		Sí		Sí
La seguridad en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)		Sí		Sí
La seguridad en cuanto a robo/hurto al interior del vehículo/coche tren		Sí		Sí
La seguridad en cuanto a robo/hurto en Terminales		Sí		Sí

Nota: Si la variable latente está relacionada con el indicador, es decir, que el elemento de Λ_y[22] asociado a esa variable latente y ese indicador es distinto de cero, en la Tabla 5-1 se lee "Sí". Por otra parte, si no existe relación entre la variable latente y el indicador, se presenta un espacio en blanco.

[22] Ver ecuación 3.2.

Estos datos se encuentran desagregados a nivel de caminos, situación bastante conveniente dado que para todos los destinos (Curicó, Talca, Chillán y Linares) casi la totalidad del viaje se realiza en la Ruta 5 Sur.

Así, para cada destino se construyó el índice de la siguiente forma:

$$\text{Índice } fatalidades\ j =$$

$$\frac{1}{8}\frac{\sum_{i=2000}^{2007}\ (fatalidades\ pasajeros\ bus\ Ruta\ 5\ Sur\ \text{año}\ i)}{\frac{km-bus}{\text{año}}\cdot\frac{pax}{bus}}\cdot distancia\ j \qquad (5.1)^{23}$$

$$En\ que\ \ j = Curicó, Talca, Chillán, Linares.$$

En el numerador se calcula el promedio de fatalidades de los últimos ocho años ocurridos en el trayecto de la Ruta 5 Sur que abarca desde Santiago a Chillán y luego se pondera por la distancia entre Santiago y el destino j. Esto implica que el nivel de riesgo es directamente proporcional a la distancia viajada. Por último, el denominador se hace cargo de distribuir las fatalidades entre los pasajeros-kilómetros promedio que recorre un bus al año.

Este índice presenta varias deficiencias, dentro de las cuales se puede destacar el suponer que la tasa de fatalidades es homogénea a lo largo de la ruta, es decir, todos los kilómetros recorridos tienen igual nivel de peligrosidad. Sin embargo (luego de probar muchas formas alternativas), permitió obtener resultados consistentes en el modelo MIMIC.

[23] Los términos $\frac{km-bus}{\text{año}}$ promedio y $\frac{pax}{bus}$ promedio se obtienen de Sánchez et al. (2006) y MIDEPLAN (2008) respectivamente.

Tabla 5-2: Relación entre Variables objetivas y variables latentes

Variable Objetiva	Comodidad Bus	Seguridad Bus	Comodidad Tren	Seguridad Tren
¿Ha visto o sufrido algún robo en bus o terminal de bus?		Sí		
¿Ha visto o sufrido algún accidente en bus?		Sí		
Desarrollo de Actividades Intelectuales	Sí		Sí	
Edad	Sí	Sí	Sí	Sí
¿Viaja en Grupo?	Sí		Sí	
Ingreso Alto	Sí			
¿Viaja con *notebook*?		Sí		Sí
¿Ha visto o sufrido algún accidente en tren?				Sí
Índice de Fatalidades		Sí		Sí
Desarrollo de Actividades Sociales	Sí		Sí	
¿Ha visto o sufrido algún robo en tren o terminal de tren?				Sí

Nota: Si la variable objetiva está relacionada con la variable latente, es decir, que el elemento de Γ[24] asociado a esa variable objetiva y esa variable latente es distinto de cero, en la Tabla 5-2 se lee "Sí". Por otra parte, si no existe relación entre la variable latente y la variable objetiva, se presenta un espacio en blanco.

[24] Ver ecuación 3.1.

Dado que no existen reportes sobre pasajeros de tren fallecidos entre el período 2000 y 2007, en el modelo estructural se probó la hipótesis de que la tasa de accidentes de bus afectaba no sólo la percepción de seguridad del bus, sino también la del tren. Así, se esperaría que a medida que aumenta el índice de fatalidades disminuya la sensación de seguridad del bus, pero a la vez el tren sea percibido comparativamente más seguro.

En la Tabla 5-3 se presentan los resultados[25] para la ecuación 3.1, es decir, el detalle de los elementos que conforman Γ. A continuación se explican dichos resultados.

Tal como se esperaría, el hecho de haber sido víctima o testigo de un accidente o robo, afecta negativamente la percepción de seguridad de los usuarios. El viajar con *notebook* afecta negativamente la sensación de seguridad del bus, ya que las personas temen ser víctimas de un robo. Este efecto es contrario en el tren. El realizar algún tipo de actividad intelectual que requiera concentración durante el viaje, impacta negativamente en la sensación de comodidad del bus y positivamente en el tren. Esto se debe a que el tren presenta mejores condiciones para el desarrollo de estas actividades: asientos más amplios, movimiento suave y constante del carro, electricidad para poder conectar el *notebook* (disponible en primera clase), entre otras. Por otra parte, el desarrollo de actividades sociales influye negativamente en la sensación de comodidad de ambos modos, sin embargo, este impacto es mayor en el caso del bus.

Por otro lado, el índice de fatalidades tuvo el efecto esperado: a medida que aumentan los accidentes de buses interurbanos con fallecidos para las rutas en cuestión, disminuye la sensación de seguridad del bus y simultáneamente, el tren se percibe como un modo comparativamente más seguro. La variable Viaja en Grupo influye positivamente en la sensación de comodidad de los usuarios, tanto de tren como de bus. Es de esperar que la sensación de comodidad de los usuarios sea mayor si la persona que se sienta al lado es alguien conocido. El “Ingreso Alto” influye negativamente en la sensación de

[25] Estos resultados corresponden a aquéllos que alimentan el modelo final (Modelo 1) que se detalla en la sección 4.3.2.

comodidad del bus (las personas de ingresos más altos son "usuarios más exigentes"). Por último, al igual que los resultados reportados por Johansson et al. (2005) y Temme et al. (2007), el modelo indica que la edad influye positivamente en las sensaciones de comodidad y de seguridad del bus y del tren[26].

Tabla 5-3: Resultados variables objetivas (ecuación 3.1)

Variable Objetiva	Comodidad Bus	Seguridad Bus	Comodidad Tren	Seguridad Tren
¿Ha visto o sufrido algún robo en bus o terminal de bus?		-0,1794 (-1,79)		
¿Ha visto o sufrido algún accidente en bus?		-0,4438 (-4,28)		
Desarrollo de Actividades Intelectuales	-0,1985 (-2,03)		0,1833 (2,78)	
Edad	0,0064 (2,3)	0,0087 (3,26)	0,0034 (1,78)	0,0059 (2,23)
¿Viaja en Grupo?	0,3216 (2,84)		0,2992 (3,86)	
Ingreso Alto	-0,4103 (-4,25)			
¿Viaja con *notebook*?		-0,2542 (-2,03)		0,2421 (2,91)
¿Ha visto o sufrido algún accidente en tren?				-0,1875 (-1,33)
Índice de Fatalidades		-0,2088 (-1,35)		0,4738 (3,00)
Desarrollo de Actividades Sociales	-0,2315 (-1,45)		-0,1072 (-0,99)	
¿Ha visto o sufrido algún robo en tren o terminal de tren?				-0,4135 (-2,23)

Nota: test t entre paréntesis-

[26] Es difícil dar una interpretación a este hecho, pues se podría argumentar que las personas más jóvenes son menos aversas al riesgo a a la vez requieren menos comodidades que las personas de más edad.

Por otra parte, en la Tabla 5-4 se presentan los resultados para la ecuación 3.2, es decir, el detalle de los elementos que conforman Λ_y (ver ecuación 3.2).Podemos observar que todos los efectos son consistentes. Un aumento en las variables latentes, implica un aumento en las variables indicadoras. Cabe destacar que aquellos coeficientes "*1*" corresponden a los indicadores normalizados (ver sección 3.1.2). Para estos casos no es posible calcular el test t, por lo que bajo los indicadores normalizados se reporta "n.a.", es decir, "no aplica".

Tabla 5-4: Resultados variables indicadoras (ecuación 3.2)

Indicador	Comodidad Bus	Seguridad Bus	Comodidad Tren	Seguridad Tren
La facilidad para viajar con niños o bultos pesados	0,7429 (12,75)		0,9877 (11,91)	
El aprovechamiento del tiempo en el viaje para leer, escribir, trabajar o realizar otras actividades	0,81 (12,91)		0,8427 (10,86)	
La comodidad durante el viaje (calidad de los asientos, espacios disponible, entre otras)	1 (n.a.)		1 (n.a.)	
La puntualidad en el cumplimiento del horario del viaje	0,7736 (13,51)		0,7723 (10,93)	
La relajación durante el viaje (capacidad para disfrutar el viaje)	0,7384 (12,09)	0,8315 (9,78)	0,775 (11,52)	0,5024 (8,06)
La seguridad al viajar (sensación de que ocurra un accidente en el transcurso del viaje)		0,9986 (10,79)		0,7872 (9,12)
La seguridad en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)		1 (n.a.)		0,9934 (11,15)
La seguridad en cuanto a robo/hurto al interior del vehículo/coche tren		0,9621 (11,33)		1 (n.a.)
La seguridad en cuanto a robo/hurto en Terminales		0,2976 (4,53)		0,1834 (4,09)

Nota: test t entre paréntesis. Para los indicadores normalizados se presenta (n.a.), es decir, “no aplica”.

5.3 Modelo de elección discreta

En esta sección se presentan los modelos estimados. Cabe destacar que todas las constantes modales presentan la corrección estipulada por la ecuación (3.15). Para las particiones modales de mercado, se utilizó la información contenida en Astra Ltda. (1999).

5.3.1 Modelos base

En una primera etapa, se calibraron varios modelos Logit sin considerar variables latentes. La idea era obtener un punto de referencia para comparar con modelos que sí incluyeran variables latentes. El modelo finalmente elegido incluye las siguientes variables (ver Tabla 5-5):

- Costo_de_viaje $*$ Categoría: producto entre el costo del pasaje y la categoría[27] (uno si es primera clase y cero si es clase normal).
- Costo_de_viaje $*$ Empresa_paga_el_pasaje: producto entre el costo del pasaje y la variable dummy "Empresa o institución paga el pasaje" (vale uno si la empresa o institución paga el pasaje y cero en otro caso).
- Tiempo_de_viaje $*$ Sexo: producto entre el tiempo de viaje y el sexo del encuestado (uno si es hombre).
- Costo_de_viaje: expresado en miles de pesos.
- Tiempo_de_viaje: expresado en minutos.

[27] Se asume que el bus semi-cama y el tren preferente son equivalentes y corresponden a "primera clase". El bus clásico y el tren salón corresponden a "clase normal".

Tabla 5-5: Coeficientes modelo Logit Base

Coeficiente θ	Estimado
Costo_de_viaje * Categoría	0,4820 (9,8)
Costo_de_viaje * Empresa_paga_el_pasaje	0,3018 (3,3)
Tiempo_de_viaje * Sexo	-0,003897 (-1,1)
Costo_de_viaje	-0,8030 (-9,1)
Tiempo_de_viaje	-0,02104 (-3,9)
Constante_modal_Bus	0,9499 (-0,7)
Constante_modal_Tren	-3,9087 (-0,7)
Log-verosimilitud en convergencia	-287,02
Rho^2	0,3099

Nota: test t entre paréntesis-

Puede apreciarse que los signos son correctos en los coeficientes del costo y tiempo. En el caso de que se viaje en mejor categoría y/o que la empresa pague el pasaje, la utilidad del modo aumenta, lo que es razonable. Además, una inspección directa del coeficiente asociado al tiempo y al género del usuario sugiere que los hombres valoran más su tiempo de viaje que las mujeres. Por otra parte, se observa en forma menos directa el siguiente fenómeno: para aquellas personas a quienes la empresa o institución paga el pasaje[28] y escogen viajar en primera clase, el coeficiente el costo del pasaje resulta irrelevante.

[28] Tomando en cuenta que tienen la posibilidad de elegir libremente en qué categoría viajar.

Para muestras “grandes”, esto se puede testear de la siguiente forma:

$$\frac{\theta_{\text{Costo_de_viaje}} + \theta_{\text{Costo_de_viaje} * \text{Categoría}} + \theta_{\text{Costo_de_viaje} * \text{Empresa_paga_el_pasaje}}}{\sqrt{Var\left(\theta_{\text{Costo_de_viaje}} + \theta_{\text{Costo_de_viaje} * \text{Categoría}} + \theta_{\text{Costo_de_viaje} * \text{Empresa_paga_el_pasaje}}\right)}}, \quad (5.2)$$

Donde:

$$\begin{aligned} Var\left(\theta_{\text{Costo_de_viaje}} + \theta_{\text{Costo_de_viaje} * \text{Categoría}} + \theta_{\text{Costo_de_viaje} * \text{Empresa_paga_el_pasaje}}\right) = \\ Var\left(\theta_{\text{Costo_de_viaje}}\right) + Var\left(\theta_{\text{Costo_de_viaje} * \text{Categoría}}\right) \\ + Var\left(\theta_{\text{Costo_de_viaje} * \text{Empresa_paga_el_pasaje}}\right) \\ + 2Cov\left(\theta_{\text{Costo_de_viaje}}, \theta_{\text{Costo_de_viaje} * \text{Categoría}}\right) \\ + 2Cov\left(\theta_{\text{Costo_de_viaje}}, \theta_{\text{Costo_de_viaje} * \text{Empresa_paga_el_pasaje}}\right) \\ + 2Cov\left(\theta_{\text{Costo_de_viaje} * \text{Categoría}}, \theta_{\text{Costo_de_viaje} * \text{Empresa_paga_el_pasaje}}\right) \end{aligned} \quad (5.3)$$

Lo que nos da:

$$\frac{-0{,}0192}{0{,}018451} = -0{,}14135$$

Como este valor es inferior, en valor absoluto, al valor de tabla (-1,96) para un nivel de confianza del 95%, se puede concluir que la suma de los coeficientes no es estadísticamente distinta de cero, es decir, que para este tipo de usuarios, el costo del pasaje resulta irrelevante.

5.3.2 Modelo final

La búsqueda del modelo final es producto de un proceso iterativo que se debe llevar a cabo simultáneamente con la búsqueda de modelos MIMIC. Esto se debe a que modelos MIMIC que resultaron consistentes en la primera etapa (principalmente con los signos correctos), al ser combinados con los modelos de elección discreta pueden generar resultados incoherentes en términos de signos (por ejemplo, que a medida que aumente la seguridad disminuye la utilidad del modo).

Como se menciona en la sección 3.1.1, es necesario incluir algún tipo de información (restricción) adicional al modelo MIMIC para que este pueda ser identificable. Stapleton (1978) señala que la normalización escogida no tiene impacto sobre los test estadísticos ni sobre los signos de los coeficientes, pero sí en los estimadores puntuales de los coeficientes. Sin embargo, las normalizaciones que se realizan en el modelo de medición[29] impactan no sólo los estimadores de los coeficientes en el modelo de elección discreta, sino también el ajuste general y sus test estadísticos asociados.

A continuación se presentan cuatro modelos cuyas únicas diferencias radican en la elección de los coeficientes normalizados en el modelo de medición. De acuerdo a lo explicado en la sección 3.1.2, para los Modelos 1, 3 y 4 se normalizan elementos λ de la matriz Λ_y. En cambio, para el Modelo 2 se normalizan los elementos diagonales de la matriz Ψ. Los resultados obtenidos sugieren que estas variaciones no pueden ser desestimadas. Además, se incluye las correcciones recomendadas por Murphy y Topel (1985).

Para definir qué elementos de Λ_y (indicadores) normalizar en el Modelo 1, se realizó lo siguiente[30]:

I. En primer lugar se estimó el modelo MIMIC con una normalización cualquiera de los elementos Λ_y (teniendo en cuenta las condiciones de Stapleton, 1978, explicadas en la sección 3.1.1).
II. En segundo lugar se analizaron las columnas de la matriz Λ_y recientemente estimada; cada una contiene los coeficientes λ estimados y exactamente un "uno" producto de la normalización utilizada.
III. Se escoge el mayor λ en valor absoluto de cada columna.

[29] Corresponde a la ecuación 3.2 (ver sección 3.1.1).

[30] Este procedimiento lo desarrollé en forma individual, sin embargo, es el mismo procedimiento utilizado por Álvarez-Daziano, R. (2009).

IV. Finalmente, se realiza una nueva estimación del modelo MIMIC normalizando en "uno" aquellos λ escogidos en III.

Siguiendo este procedimiento, todos los elementos de la matriz Λ_y tendrán como valor máximo "uno". La idea de utilizar esta normalización es permitir una interpretación un poco más clara de la matriz Λ_y resultante.

Para el Modelo 2, se normalizan en "uno" las varianzas de las variables latentes; así, la matriz Ψ es una matriz identidad de 4x4 (tiene cuatro elementos, porque hay cuatro variables latentes: comodidad bus, comodidad tren, seguridad bus y seguridad tren).

Finalmente, para los Modelos 3 y 4, se escogen valores de λ al azar de la matriz Λ_y (teniendo en cuenta las condiciones de Stapleton, 1978) y se normalizan en "uno". El detalle de los λ que fueron normalizados, puede obtenerse en el Anexo 10.

Se pueden hacer múltiples observaciones de los resultados expuestos en la Tabla 5-6. En primer lugar, no existen grandes diferencias entre los test t antes y después de las correcciones propuestas por Murphy y Topel (1985). Es difícil intentar dar una explicación a este hecho. En particular, puede deberse al hecho que aunque ambos modelos fueron calibrados con la misma muestra, la información utilizada por cada uno prácticamente no se traslapa; es decir, las variables utilizadas en cada uno no se utilizan en el otro. También puede deberse a la baja varianza de los errores asociados a las variables latentes calibradas.

Otra explicación posible es el amplio uso de variables dicotómicas tanto en el modelo estructural como en el de elección discreta. Esto implica que los elementos de la matriz R_3 (ver Anexo 1) sean "pequeños", al ser el promedio de la multiplicación de vectores que contienen muchos ceros. Este fenómeno se repite con R_4. De todas formas, no es posible determinar *a priori* la magnitud del efecto de las correcciones sobre los test t. En Murphy y Topel (1985), se observan casos donde las correcciones van desde el uno hasta cerca del 300%.

Tabla 5-6: Coeficientes de modelos con variables latentes

Coeficiente θ	Modelo 1[31] Estimado	Test t	Test t corregido	Modelo 2 Estimado	Test t	Test t corregido
Costo_de_viaje * Categoría	0,429	8,641	8,580	0,428	8,619	8,508
Costo_de_viaje * Empresa_paga_el_pasaje	0,218	2,357	2,356	0,218	2,335	2,292
Tiempo_de_viaje * Sexo	0,000	-0,035	-0,035	0,000	-0,065	-0,064
Seguridad	1,115	2,551	2,548	1,005	2,544	2,521
Comodidad	1,612	5,085	5,065	1,524	5,562	5,508
Costo_de_viaje	-0,741	-8,382	-8,284	-0,740	-8,396	-8,241
Tiempo_de_viaje	-0,020	-3,528	-3,524	-0,020	-3,496	-3,484
Constante_modal_Bus	1,144	1,168	1,167	1,383	1,732	1,704
Constante_modal_Tren	-3,909			-3,909		
Log-Likelihood	-267,265			-264,004		
Rho^2	0,357			0,365		

Coeficiente θ	Modelo 3 Estimado	Test t	Test t corregido	Modelo 4 Estimado	Test t	Test t corregido
Costo_de_viaje * Categoría	0,428	8,627	8,590	0,429	8,669	8,510
Costo_de_viaje * Empresa_paga_el_pasaje	0,217	2,343	2,328	0,217	2,319	2,314
Tiempo_de_viaje * Sexo	0,000	-0,036	-0,036	0,000	-0,120	-0,119
Seguridad	1,112	2,533	2,512	1,683	1,153	1,133
Comodidad	2,180	5,240	5,226	2,224	5,574	5,533
Costo_de_viaje	-0,740	-8,382	-8,320	-0,741	-8,399	-8,346
Tiempo_de_viaje	-0,020	3,518	3,454	-0,020	-3,641	-3,606
Constante_modal_Bus	1,211	1,330	1,321	1,058	0,979	0,973
Constante_modal_Tren	-3,909			-3,909		
Log-Likelihood	-266,236			-267,569		
Rho^2	0,360			0,357		

Nota: el lector puede notar que la variable "Tiempo_de_viaje_*Sexo" no resultó significativa en los modelos con variables latentes. Sin embargo, se mantiene su uso porque en el Modelo Base que no incluye variables latentes resultó significativa al 70% de confianza. "

[31] Los resultados del modelo MIMIC que alimenta este modelo se presentan en la Tabla 5-3.

Los estimadores de todos aquellos coeficientes que no acompañan a las variables latentes como el costo, tiempo de viaje, etc., presentan cambios mínimos y en muchos casos, son idénticos en los distintos modelos. No obstante, los cambios en los estimadores de los coeficientes asociados a la comodidad y a la seguridad, sí presentan cambios más considerables. Por ejemplo, entre los modelos 2 y 4 los coeficientes asociados a la seguridad y a la comodidad presentan diferencias del 67% y 45% respectivamente. Además, la discrepancia en el test t asociado a la seguridad tampoco es despreciable y bordea el 55%.

Adicionalmente, se observa que nuevamente el coeficiente del costo de viaje para las personas que viajan en primera clase y les pagan el pasaje pareciera ser irrelevante. Haciendo un cálculo similar al realizado para el modelo base en la sección 5.3.1 se volvió a constatar el mismo resultado.

Por último, es necesario indicar lo siguiente. Durante la etapa de estimación de modelos de elección discreta que incluyen variables latentes, se observó que ante pequeños cambios en el modelo MIMIC, principalmente debidos a la inclusión u omisión de alguna variable objetiva "x" o indicadora "y", se generaban grandes cambios en el modelo de elección discreta (signos incorrectos en los parámetros asociados a la comodidad y/o seguridad). Es decir, el modelo de elección resulta bastante sensible a la especificación del modelo estructural. Este hecho implicó un extenso proceso de modelación, que culminó con las especificaciones de los modelos recién expuestos.

5.3.3 Justificación de las técnicas utilizadas

Las secciones anteriores han intentado explicar y mostrar la inclusión de variables latentes en modelos de elección discreta. Sin embargo, es legítimo preguntarse si todo esto es necesario: ¿qué hubiese sucedido si en vez de haber estimado modelos MIMIC se hubiese utilizado esas mismas variables directamente en la función de utilidad del modelo logit?

La respuesta es que siguiendo ese enfoque los resultados no resultan satisfactorios. Muchas de las variables presentan signos contrarios a la intuición (como las relativas a robos y accidentes en el bus, que aumentan la utilidad del modo bus, además de algunas variables indicadoras) y otras resultan difíciles de interpretar (¿qué significado, por ejemplo, tiene la variable "viaja con *notebook*" si se ingresa directamente en la función de utilidad?). Por otra parte, se pierde la capacidad de diferenciar el efecto de estas variables por modo, como es el caso de las variables socioeconómicas[32] y de características de los usuarios. En cambio, el uso de ecuaciones estructurales sí permite utilizar estas variables para hacer diferencias en las percepciones subjetivas de cada modo por parte de los usuarios (e.g. en las percepciones de comodidad y de seguridad). Adicionalmente, los modelos estructurales permiten obtener un conocimiento más acabado de las relaciones causales que influyen en las percepciones de los viajeros, lo que no es posible lograr introduciendo directamente las variables utilizadas en el modelo MIMIC en el modelo de elección discreta. En el Anexo 11 se muestran los resultados de un modelo Logit calibrado con todas las variables utilizadas en el Modelo 1 (Modelo Final) donde se pueden apreciar los fenómenos recién discutidos.

El mismo análisis es realizado incluyendo directamente las variables indicadoras al Modelo Base. En este caso, nuevamente se presentan inconsistencias a nivel de signos para algunas variables indicadoras, por ejemplo las variables indicadoras "La *comodidad* durante el viaje (calidad de los asientos, espacios disponible, entre otras)" y "La seguridad en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)" presentan signos negativos en la función de utilidad, lo que es claramente incorrecto. Estos resultados pueden ser revisados en el Anexo 12.

[32] Dado que el modelo Logit trabaja en base a diferencia de los atributos, no es posible utilizar exactamente la misma variable en ambos modos.

5.4 Análisis y Discusiones

5.4.1 Elección del mejor modelo

A continuación se presentan los resultados de comparar el Modelo Base (ver Tabla 5-5) y los modelos que incluyen variables latentes (Tabla 5-6), a través del test de razón de verosimilitud.

Recordemos que el test LR o razón de verosimilitud, se calcula de la siguiente forma (Ortúzar y Willumsen, 2001):

$$LR = -2 \cdot \left\{ l\left(\hat{\theta}_r\right) - l(\hat{\theta}) \right\} \tag{5.4}$$

en que $l\left(\hat{\theta}_r\right)$ es la log-verosimilitud del Modelo Base (Tabla 5-5) o restringido y $l(\hat{\theta})$ la del modelo no restringido (con variables latentes). Ambos valores se reportan en las Tabla 5-5 y Tabla 5-6 respectivamente.

Tabla 5-7: Test LR

Modelo Base v/s	LR	Valor Tabla[33]	¿Rechazo Modelo Base?
Modelo 1	39,5138	5,99	Sí
Modelo 2	46,0358	5,99	Sí
Modelo 3	41,5718	5,99	Sí
Modelo 4	38,9058	5,99	Sí

Nota: el Modelo Base se considera una versión restringida de los Modelos 1 al 4, dado que no contiene las dos variables latentes (comodidad y seguridad).

[33] El valor de tabla se calcula para una distribución χ^2 con dos grados de libertad a un nivel de confianza del 95%.

En todos los casos, los modelos que incluyen variables latentes resultan superiores que el Modelo Base que no las incluye.

La Tabla de Éxito Predictivo (Ortúzar y Willumsen, 2001) es una forma más intuitiva de obtener una medida de bondad de ajuste. Recordemos que se calcula de la siguiente forma:

Tabla 5-8: Tabla Éxito Predictivo

Tabla de Éxito Predictivo	Modelo Base		Total Fila
	Bus	Tren	
Bus	N_{BB}	N_{BT}	$N_{BB}+N_{BT}$
Tren	N_{TB}	N_{TT}	$N_{TT}+N_{TB}$
Total columna	$N_{BB}+N_{TB}=N_{\bullet B}$	$N_{BT}+N_{TT}=N_{\bullet T}$	$N_{BB}+N_{BT}+N_{TT}+N_{TB}=N$
Proporción predicha	$N_{\bullet B}/N$	$N_{\bullet T}/N$	1
Prop. correctamente predicha	$N_{BB}/N_{\bullet B}$	$N_{TT}/N_{\bullet T}$	$(N_{BB}+N_{TT})/N$

donde N_{ij} es el número esperado de individuos que elijen "*i*" (bus o tren), pero para los cuales el modelo predice que elegirían "*j*" (bus o tren).

En las tablas Tabla 5-9 y Tabla 5-10 se muestran los resultados de este cálculo para el modelo base y para todos los modelos que incluyen variables latentes. En la Tabla 5-9 se observa que el modelo base predice correctamente y en forma relativamente balanceada para ambos modos.

Tabla 5-9: Tabla Éxito Predictivo Modelo Base

Tabla de Éxito Predictivo	Modelo Base		Total Fila
	Bus	Tren	
Bus	211,5119	89,4881	301
Tren	89,3746	209,6254	299
Total columna	300,8865	299,1135	600
Proporción predicha	0,5015	0,4985	1
Prop. correctamente predicha	0,7030	0,7008	0,7019

Tabla 5-10: Tablas Éxito Predictivo Modelos con Variables Latentes

Tabla de Éxito Predictivo	Modelo 1		Total Fila
	Bus	Tren	
Bus	218,0655	82,9345	301
Tren	82,8999	216,1001	299
Total columna	300,9654	299,0346	600
Proporción predicha	0,5016	0,4984	1
Prop. correctamente predicha	0,7246	0,7227	0,7236
Tabla de Éxito Predictivo	Modelo 2		Total Fila
	Bus	Tren	
Bus	218,0238	82,9762	301
Tren	80,7951	218,2049	299
Total columna	298,8189	301,1811	600
Proporción predicha	0,4980	0,5020	1
Prop. correctamente predicha	0,7296	0,7245	0,7270
Tabla de Éxito Predictivo	Modelo 3		Total Fila
	Bus	Tren	
Bus	217,4153	83,5847	301
Tren	81,5281	217,4719	299
Total columna	298,9434	301,0566	600
Proporción predicha	0,4982	0,5018	1
Prop. correctamente predicha	0,7273	0,7224	0,7248
Tabla de Éxito Predictivo	Modelo 4		Total Fila
	Bus	Tren	
Bus	218,2083	82,7917	301
Tren	83,6193	215,3807	299
Total columna	301,8276	298,1724	600
Proporción predicha	0,5030	0,4970	1
Prop. correctamente predicha	0,7230	0,7223	0,7226

Como se puede ver en las tablas, en todos los casos los modelos con variables latentes mejoran sus proporciones correctamente predichas en comparación con el modelo

base[34]. Si bien estas mejoras son del orden de dos puntos porcentuales absolutos, son prácticamente inexistentes entre los cuatro modelos que incluyen variables latentes. Esto nos permite inferir que la elección de normalización pareciera no tener impacto en términos predictivos.

Así, como las diferencias entre los distintos modelos con variables latentes son tan pequeñas, se optó por escoger el Modelo 1 como modelo final, debido a que la normalización usada en ese caso permite interpretar más fácilmente los efectos de las variables latentes sobre los indicadores (el mayor efecto posible es "uno").

5.4.2 Cálculo de valores subjetivos

Uno de los principales usos dados a los modelos Logit guarda relación con el cálculo de valores subjetivos del tiempo. En este trabajo se pretende, además, calcular los valores subjetivos por mejoras en la comodidad y la seguridad. En la Tabla 5-11 se presentan los valores subjetivos por reducción de tiempo de viaje, mejoras en la seguridad y la comodidad (VST, VSS y VSC respectivamente).

Al observar los intervalos de confianza (Armstrong et al. 2001) para el valor subjetivo del tiempo del Modelo Base y de los Modelos 1 al 4, podemos ver que no existen diferencias estadísticas para un nivel del 95% de confianza. Sin embargo, las estimaciones puntuales de estos valores, son menores para los modelos que incluyen variables latentes (Modelos 1 al 4). Esto es de esperar, dado que el valor de reducción de tiempo de viaje del Modelo Base está compuesto no sólo por el valor de reducir el tiempo de viaje en sí mismo, sino por la reducción del las incomodidades e inseguridades asociadas al tiempo de viaje. El efecto recién descrito, es capturado por las variables latentes en los modelos 1 al 4. Por otra parte, la normalización escogida pareciera no afectar las estimaciones puntuales de los VST para los modelos con variables latentes.

[34] Ver Tabla 5-9.

Tabla 5-11: Disposiciones al Pago

Modelo	Valor Subjetivo		
	VST ($/hora)	VSS ($/u.d.s)	VSC ($/u.d.c)
Base	2045 [921,13 3647,18]	*	*
Modelo1	1857 [779,23 3292.12]	1760 [370,81 3475,09]	2540 [1488,41 4028,83]
Modelo2	1904 [799,19 3373,47]	1590 [334,51 3137,17]	2400 [1457,67 3769,87]
Modelo3	1899 [796,88 3366,73]	1750 [370,17 3469,75]	3440 [2013,23 5459]
Modelo4	1911 [802,24 3388,01]	2650 [559,53 5243,05]	3500 [2051,41 5559,76]

Notas: los intervalos de confianza (Armstrong et al. 2001) se presentan entre corchetes para un nivel de 95% de confianza. (*) Como el Modelo Base no incluye variables latentes no se pueden calcular disposiciones al pago por mejoras en la comodidad o seguridad.
u.d.s equivale a unidad de seguridad.
u.d.c equivale a unidad de comodidad.

Si se realiza un análisis similar y observamos los intervalos de confianza de los VSS podríamos decir que no existen diferencias apreciables entre estos valores para un nivel de confianza del 95%, por lo que todos los caminos de normalización planteados serían estadísticamente equivalentes (lo mismo es válido para los VSC). Sin embargo, esta conclusión no es del todo cierta. No debemos olvidar que los Modelos 1 al 4 son producto de distintos tipos de normalizaciones, por lo que es necesario destacar lo siguiente. Una vez escogidos los indicadores o varianzas a normalizar, los cambios en las magnitudes de las normalizaciones (es decir, si se normaliza en “1” ó “10”) sólo tienen impactos en la escala de los coeficientes estimados en el modelo estructural. Sin embargo, esto no implica que los VSS y VSC se mantengan constantes para modelos cuyas normalizaciones difieran únicamente en magnitud, ya que sólo los coeficientes del modelo Logit asociados a las variables latentes (y no todos) se ajustan a estos cambios de escala. Este hecho es bastante importante, dado que agrega un grado de libertad más a

la gama de posibles VSS y VSC que podrían obtenerse para una misma especificación del modelo. Por ejemplo (como se detalla en la Tabla 5-12) si en el Modelo 1 hubiésemos normalizado los indicadores en "10" en vez de "1" hubiésemos obtenido VSS y VSC 10 veces menores. Por otra parte, los VST de estos modelos se mantienen invariantes.

Tabla 5-12: Sensibilidad ante cambio en magnitud de normalización Modelo 1

Modelo	Valor Subjetivo		
	VST ($/hora)	VSS ($/u.d.s)	VSC ($/u.d.c)
Modelo1_normalizado_en_1	1856,57 [779,23 3292.12]	1760 [370,81 3475,09]	2540 [1488,41 4028,83]
Modelo1_normalizado_en_10	1856,57 [779,23 3292.12]	176 [37,08 347,51]	254 [148,84 402,88]

Notas: los intervalos de confianza (Armstrong et al. 2001) se presentan entre corchetes para un nivel de 95% de confianza.
u.d.s equivale a unidad de seguridad.
u.d.c equivale a unidad de comodidad.

En este caso, existirían diferencias significativas al 95% de confianza entre los VSS y VSC que surgen de distintas magnitudes en las normalizaciones del Modelo 1. Con este ejercicio se pretende mostrar que el criterio estadístico no resulta tan adecuado para realizar conclusiones con respecto a las normalizaciones utilizadas. Lamentablemente no es posible determinar con criterios objetivos y contundentes qué camino de normalización es el más apropiado. Sin embargo, tal como se explica en la sección 3.1.2, el camino tomado por la mayoría de los investigadores (lo que no quiere decir que sea el "verdadero") es fijar en "1" algún elemento de la columna de Λ_y (Bollen, 1989, pp. 18).

5.4.3 Cálculo de las principales elasticidades Modelo 1 (modelo final)

Las elasticidades puntuales con respecto al atributo "*k*" de la alternativa "*i*" para el individuo "*n*" (X_{ikn}) del modelo Logit, vienen dadas por la siguiente expresión (Ortúzar y Willumsen, 2001):

$$E_{X_{ikn}}^{P_{in}} = \theta_{ik} \cdot X_{ikn} \cdot \ (1 - P_{in}) \tag{5.5}$$

en que θ_{ik} corresponde al coeficiente del modelo Logit para el atributo X_{ik} y P_{in} es la probabilidad de que el individuo "*n*" escoja la alternativa "*i*". Esta medida resulta útil para comparar los cambios porcentuales de las probabilidades de elección, ante cambios porcentuales de los atributos. En la Tabla 5-13 se presentan las principales elasticidades para el Modelo Base y para el Modelo 1.

Tabla 5-13: Elasticidades puntuales Modelo Base y Modelo 1

Variable	Elasticidades Modelo Base		Elasticidades Modelo 1	
	Bus	Tren	Bus	Tren
Seguridad	(n.a.)	(n.a.)	0,0592	1,0120
Comodidad	(n.a.)	(n.a.)	0,2213	1,0236
Costo	-8,1291	-10,1740	-7,5023	-12,1511
Tiempo de viaje	-13,1693	-8,0269	-9,4691	-7,4690

Nota: Dado que el Modelo Base no contiene variables latentes, el cálculo de elasticidades en este caso no aplica (n.a.).

Como se puede observar, en la mayoría de los casos, las elasticidades para el Modelo Base son mayores a las elasticidades del Modelo 1. Esto es un resultado razonable, dado que se esperaría que la omisión de variables latentes genere una sobrevaloración de los efectos de las variables de costo y tiempo en la demanda de viajes. Sería interesante contar con más estudios para observar si existe una tendencia en la sobreestimación de las elasticidades puntuales en modelos que omitan variables latentes.

Por otra parte, para el Modelo 1 los efectos en la partición modal ante cambios porcentuales en las variables latentes son mucho menores a los efectos que tienen cambios porcentuales en el costo y el tiempo de viaje en la probabilidad de elección. En el caso del bus, estas diferencias son mucho más dramáticas. Sin embargo, dado que las elasticidades dependen de los parámetros del modelo Logit, no debemos olvidar que son sensibles a la normalización escogida. Afortunadamente, esta dependencia se debe únicamente a la estrategia de normalización utilizada[35] y no es fruto de las magnitudes de dichas normalizaciones[36], dado que tanto el producto $\theta_{ik} \cdot X_{ikn}$ como las probabilidades de elección P_{in} son invariantes ante cambios de escala en X_{ikn}.

[35] Normalizar elementos de Λ_y o elementos de Ψ, como se explica en la sección 3.1.1.
[36] A diferencia del caso del cálculo de VSS y VSC, como se muestra en la sección 5.4.2.

6 CONCLUSIONES

El resultado más importante de este trabajo y que no ha sido adecuadamente reportado en la literatura, es que las normalizaciones o restricciones adicionales utilizadas en el modelo MIMIC para permitir su identificabilidad, tienen efectos importantes en los coeficientes asociados a las variables latentes del modelo de elección discreta.

Aunque la mayoría de los investigadores decide normalizar en "1" alguno de los indicadores (Bollen, 1989, pp. 18), esto no implica que éste camino sea demostrablemente superior a otro. Podría argumentarse que normalizar en "1" algún indicador es razonable si se supone que dicho indicador es una medida sensata de la variable latente. Por otra parte, en el caso de que se decida normalizar las varianzas, no es posible utilizar el mismo argumento. ¿Qué medida de error es razonable? ¿1? ¿10? ¿Cómo decidir por una u otra? Estas son algunas de las preguntas que surgen y que no tienen una respuesta clara. Como camino de investigación alternativo, sería interesante ver cuál es el efecto de las normalizaciones en los coeficientes de los modelos MIMIC-Logit integrados si éstos se estiman mediante el método simultáneo descrito en la sección 3.3.1.

Esto hace confusa la interpretación de los valores económicos por mejoras en la seguridad y comodidad: ¿qué normalización es la que nos lleva a los resultados "verdaderos" para los valores subjetivos de las variables latentes? Una salida posible a este problema es incluir, dentro de la encuesta, un indicador que sea -en sí mismo- una variable latente. Por ejemplo, si se pretende modelar una variable que sea "flexibilidad", se podría considerar un indicador denominado "flexibilidad". Así, se postularía a priori que la variable latente "flexibilidad" afecta en 1 al indicador "flexibilidad" más una medida de error. En términos prácticos, esto implica que en la encuesta se debe incluir una pregunta como la siguiente: "qué nota le pone usted a la flexibilidad del modo". De esta forma, se podría suponer razonablemente que ése indicador afectaría en "1" a la variable latente "flexibilidad".

Siguiendo esta misma línea, el hecho que los valores económicos asociados a las variables latentes dependan, además, de la magnitud de dichas normalizaciones, hace difícil que esta metodología pueda ser utilizada con confianza en evaluación social de proyectos, dado que se pueden generar valores subjetivos tan altos o tan bajos como el investigador lo desee, simplemente cambiando la escala de las normalizaciones. Por otra parte, vale la pena recalcar que las elasticidades no son sensibles a las magnitudes de las normalizaciones realizadas (aunque sí al camino de normalización). Esto implica que si se logra encontrar el camino de normalización "verdadero", se podrían utilizar con mayor confianza las elasticidades calculadas.

Por otra parte, durante la etapa de estimación se encontró que los modelos de elección discreta eran altamente sensibles ante pequeños cambios de especificación del modelo estructural, arrojando resultados inconsistentes (por ejemplo, a medida que aumenta la seguridad, disminuye la utilidad del modo). Esto implica que aun cuando la normalización utilizada sea "la correcta", existirán dudas acerca de la especificación misma del modelo, lo que impide realizar conjeturas fiables sobre los coeficientes calibrados.

Sin embargo, en esta tesis se encuentra que los modelos estructurales con variables latentes enriquecen los modelos de elección discreta. En todos los casos analizados los modelos con variables latentes resultaron mejores, tanto en términos explicativos como de ajuste que aquellos que no las consideraban. Por otra parte, las correcciones a los errores estadísticos para modelos secuenciales sugeridas por Murphy y Topel (1985) no produjeron cambios sustantivos. No fue posible determinar con claridad las razones de este hecho, sin embargo se sospecha que tienen que ver con la gran cantidad de variables dicotómicas utilizadas en el modelo estructural, a la baja variabilidad de los errores estimados para las variables latentes y al hecho de que la información utilizada para estimar los modelo MIMIC y de elección discreta, prácticamente no se traslapan. De todas formas en Murphy y Topel (1985), se presentan ejemplos en que las correcciones estadísticas tienen alto impacto y otros en que el efecto es casi nulo.

Adicionalmente, los modelos de elección discreta que incluyen variables latentes permiten lograr un entendimiento más acabado del proceso mental de los usuarios de transporte público, presentando además mejoras en el ajuste y en la predicción. Por ejemplo, se descubren relaciones interesantes entre las percepciones de comodidad y seguridad asociadas al tipo de actividad que se desarrolla durante el viaje, a las experiencias con respecto a robos o accidentes en los modos, entre otras. Y, como se explica en la sección 5.3.3, esta información adicional no produjo buenos resultados cuando fue ingresada directamente en la función de utilidad del modelo de elección discreta, sin embargo, los modelos estructurales sí permitieron su utilización.

BIBLIOGRAFÍA

Álvarez-Daziano, R. (2009). Comunicación privada.

Arbuckle, James L. (2006). *Amos 7.0 User´s Guide.*

Armstrong P., Garrido R.A., Ortúzar J. de D. (2001). Confidence intervals to bound the value of time. *Transportation Research* 37E, 143-161.

Ashok, K., Dillon, W.R., y Yuan, S. (2002). Extending discrete choice models to incorporate attitudinal and other latent variables. *Journal of Marketing Research* 39, 31–46.

Astra Ltda. (1999). Análisis y desarrollo evaluación sistema de transporte interurbano: IV Etapa. ESTRASUR. Informe para SECTRA.

Ben-Akiva, M., D. McFadden, T. Gärling, D. Gopinath, J. Walker, D. Bolduc, A. Börsch-Supan,P. Delquié, O. Larichev, T. Morikawa, A. Polydoropoulou, and V. Rao (1999) "Extended Framework for Modeling Choice Behavior", *Marketing Letters* 10(3), 187-203.

Bollen, K.A. (1989). *Structural Equations with Latent Variables*. Nueva York: John Wiley and Sons.

Conaset (2007). Registro de Accidentes Viales de Chile, Serie 2000-2007. Dirección de Fronteras y Servicios Especializados, Departamento de Servicios de Tránsito y Carreteras (OS2), Carabineros de Chile, Santiago.

Daly, A. J, (1995). *Alogit 4.2: User`s Guide*. Hage Consoulting Group, La Haya.

Golob,T. (2003). Structural equation modeling for travel behavior research. *Transportation Research* 37B, 1-25.

Greene, W. (1997). *Econometric Analysis.* New Jersey: Prentice Hall.

Iragüen, P. y Ortúzar, J. de D. (2004). Willingness-to-pay for reducing fatal accidents risks in urban areas: an Internet-based web page stated preference survey. *Accident Analysis and Prevention 36*, 513-524.

Jöreskog, K. G. (1973). A general Method for estimating a linear structural equation system. *Structural Equation Models in the Social Sciences*. New York: Academic Press, pp. 85-112.

Jöreskog K.G. y D. Sörbom (1986). PRELIS: A preprocessor for LISREL. Mooresville, IN : Scientific Software, Inc.

Lyons, G. y Urry, J. (2005). Travel time use in the information age. *Transportation Research* 39A, 257-276.

Lyons, G., Jain, J. y Holley, D. (2007). The use of travel time by rail passengers in Great Britain. *Transportation Research* 41A, 107-120.

Manski, C. y Lerman, S.R. (1977). The estimation of choice probabilities from choice based samples. *Econometrica* 45 (8), 1977-1988.

McFadden, D. (1986). The choice theory approach to marketing research. *Marketing Science* 5, 275-97.

MIDEPLAN (2008). Sistema de estadísticas básicas de inversión (Sebi) 2008.

Murphy, K. y Topel, H. (1985). Estimation and inference in two-step econometric models. *Journal of Business and Economic Statistics* 3, 370-379.

Ortúzar, J. de D. y Willumsen, L.G. (2001). *Modelling Transport.* Tercera Edición, Chichester: John Wiley and Sons.

Sánchez, R., Nazif, J., Rojas, D. y Velasco, A. (2006). Instrumentos Para La Toma De Decisiones En Políticas De Seguridad Vial En América Latina: El Índice De Seguridad De Transito (Insetra). United Nations Publications.

Rizzi, L.I. y Ortúzar, J. de D. (2003). Stated preference in the valuation of interurban road safety. *Accident Analysis and Prevention* 35, 9-22.

Rizzi, L.I., Cantillo, V. y Bronfman, N. (2005). Influencia de la percepción del nivel de servicio en la elección modal para viajes interurbanos de transporte público hacia el sur de Chile. *Actas del XII Congreso Chileno de Ingeniería de Transporte*. Valparaíso, 17 al 21 de Octubre 2005.

Stapleton, D. (1978). Analyzing political participation data with a Mimic model. *Sociological Methodology* 9, 52-74.

Temme, D., Paulssen, M. y Dannewald, Ti. (2007). Integrating latent variables in discrete choice models: how higher-order values and attitudes determine consumer choice. SFB 649 Discussion Papers SFB649DP2007-065, Sonderforschungsbereich 649, Humboldt University, Berlin, Germany.

Vredin Johansson, M., Heldt, T. y Johansson, P. (2005). Latent variables in a travel mode choice model: attitudinal and behavioral indicator variables. Working Paper 2005:5, Uppsala University.

Walker, J. (2001). *Extended Discrete Choice Models: Integrated Framework, Flexible Error Structures and Latent Variables*. PhD Thesis, MIT.

ANEXOS

Anexo 1: Descripción corrección errores estándar

A continuación, se presenta el método descrito por Murphy y Topel (1985) para corregir la matriz de covarianzas de los errores de los coeficientes finales, en el caso en que los modelos de ambas etapas sean estimados vía máxima verosimilitud.

Sean $L_1(\theta_1)$ y $L_2(\theta_1,\theta_2)$ las funciones de log verosimilitud de los estimadores de la primera y segunda etapa respectivamente. Θ_1 es el vector de coeficientes de la primera etapa y θ_2 es el vector de coeficientes de la segunda etapa. L_2 depende tanto de los coeficientes estimados durante la primera etapa, como de aquellos a estimar en la segunda etapa.

Recordemos que:

$$L_1 = \frac{-N(p+q)}{2} ln|2\pi| - \frac{N}{2} ln|\sum(\theta)| + \frac{1}{2}\left(\sum_{i=1}^{N} z_i' \cdot \sum(\theta)^{-1} \cdot z_i\right)$$

$$L_2 = \sum_{n=1}^{N} \sum_{A_j \in A(n)} g_{jn} \cdot ln\left(P_{jn}\right)$$

Para continuar, es necesario calcular:

$$R_1(\theta_1) = E\frac{\partial L_1}{\partial \theta_1}\left(\frac{\partial L_1}{\partial \theta_1}\right)' = -E\frac{\partial^2 L_1}{\partial \theta_1 \partial \theta_1{}'}$$

$$R_2(\theta_2) = E\frac{\partial L_2}{\partial \theta_2}\left(\frac{\partial L_2}{\partial \theta_2}\right)' = -E\frac{\partial^2 L_2}{\partial \theta_2 \partial \theta_2{}'}$$

$$R_3(\theta) = E\frac{\partial L_2}{\partial \theta_1}\left(\frac{\partial L_2}{\partial \theta_2}\right)' = -E\frac{\partial^2 L_2}{\partial \theta_1 \partial \theta_2{}'}$$

$$R_4(\theta) = E\frac{\partial L_1}{\partial \theta_1}\left(\frac{\partial L_2}{\partial \theta_2}\right)'$$

Por último, para calcular la matriz de varianza covarianza corregida ∑, debemos hacer:

$$\sum \quad = R_2^{-1} + R_2^{-1}\left[R_3' R_1^{-1} R_3 \; - R_4' R_1^{-1} R_3 \; - R_3' R_1^{-1} R_4 \;\right] R_2^{-1}$$

Vale destacar, que tanto las matrices R_1 como R_2 pueden ser obtenidas de las salidas de los programas Amos y Alogit respectivamente, dado que corresponden a las inversas de las matrices de varianza covarianza de los coeficientes de cada etapa.

Anexo 2: Notas Variables Indicadoras por Modo

Indicador	Bus	Tren	Diferencia Porcentual
La *conveniencia* de los horarios de viaje (mayor disponibilidad de viajes diarios)	5,82	5,30	-8,9%
La *facilidad* para viajar con niños o bultos pesados	4,82	6,04	25,3%
El *aprovechamiento del tiempo* en el viaje para leer, escribir, trabajar o realizar otras actividades	5,14	6,12	19,1%
La *comodidad* durante el viaje (calidad de los asientos, espacios disponible, entre otras)	5,03	6,20	23,3%
La *puntualidad* en el cumplimiento del horario del viaje	5,56	6,25	12,4%
La *relajación* durante el viaje (capacidad para disfrutar el viaje)	5,15	6,22	20,8%
La *seguridad* al viajar (sensación de que no ocurra un accidente en el transcurso del viaje)	4,77	6,11	28,1%
La *seguridad* en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)	4,22	5,30	25,6%
La *seguridad* en cuanto a robo/hurto al interior del vehículo/coche tren	4,77	5,66	18,7%
La *seguridad* en cuanto a robo/hurto en Terminales	4,21	5,38	27,8%
Promedio	4,95	5,86	18,4%

Anexo 3: Promedio respuestas Variables Indicadoras para usuarios de bus

Indicador	Bus	Tren	Diferencia Porcentual
La *conveniencia* de los horarios de viaje (mayor disponibilidad de viajes diarios)	5,96	4,90	-17,8%
La *facilidad* para viajar con niños o bultos pesados	5,11	5,76	12,7%
El *aprovechamiento del tiempo* en el viaje para leer, escribir, trabajar o realizar otras actividades	5,53	5,92	7,0%
La *comodidad* durante el viaje (calidad de los asientos, espacios disponible, entre otras)	5,34	6,03	13,0%
La *puntualidad* en el cumplimiento del horario del viaje	5,76	6,07	5,4%
La *relajación* durante el viaje (capacidad para disfrutar el viaje)	5,53	6,01	8,8%
La *seguridad* al viajar (sensación de que no ocurra un accidente en el transcurso del viaje)	5,11	5,94	16,3%
La *seguridad* en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)	4,36	5,03	15,6%
La *seguridad* en cuanto a robo/hurto al interior del vehículo/coche tren	4,85	5,33	9,9%
La *seguridad* en cuanto a robo/hurto en Terminales	4,26	4,87	14,2%
Promedio	5,18	5,59	7,8%

Anexo 4: Promedio respuestas Variables Indicadoras para usuarios de tren

Indicador	Bus	Tren	Diferencia Porcentual
La *conveniencia* de los horarios de viaje (mayor disponibilidad de viajes diarios)	5,67	5,70	0,4%
La *facilidad* para viajar con niños o bultos pesados	4,54	6,31	39,2%
El *aprovechamiento del tiempo* en el viaje para leer, escribir, trabajar o realizar otras actividades	4,75	6,33	33,3%
La *comodidad* durante el viaje (calidad de los asientos, espacios disponible, entre otras)	4,71	6,38	35,5%
La *puntualidad* en el cumplimiento del horario del viaje	5,36	6,43	20,1%
La *relajación* durante el viaje (capacidad para disfrutar el viaje)	4,76	6,42	34,8%
La *seguridad* al viajar (sensación de que no ocurra un accidente en el transcurso del viaje)	4,42	6,28	41,8%
La *seguridad* en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)	4,08	5,57	36,5%
La *seguridad* en cuanto a robo/hurto al interior del vehículo/coche tren	4,69	5,98	27,5%
La *seguridad* en cuanto a robo/hurto en Terminales	4,16	5,89	41,7%
Promedio	4,71	6,13	30,0%

Anexo 5: Promedio respuestas Variables Indicadoras Bus por género

Indicador	Mujeres	Hombres	Diferencia Porcentual
La *conveniencia* de los horarios de viaje (mayor disponibilidad de viajes diarios)	5,86	5,78	-1,4%
La *facilidad* para viajar con niños o bultos pesados	4,89	4,76	-2,6%
El *aprovechamiento del tiempo* en el viaje para leer, escribir, trabajar o realizar otras actividades	5,19	5,09	-2,0%
La *comodidad* durante el viaje (calidad de los asientos, espacios disponible, entre otras)	5,10	4,95	-2,9%
La *puntualidad* en el cumplimiento del horario del viaje	5,67	5,44	-4,1%
La *relajación* durante el viaje (capacidad para disfrutar el viaje)	5,25	5,04	-4,1%
La *seguridad* al viajar (sensación de que no ocurra un accidente en el transcurso del viaje)	4,84	4,69	-3,2%
La *seguridad* en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)	4,19	4,25	1,6%
La *seguridad* en cuanto a robo/hurto al interior del vehículo/coche tren	4,91	4,63	-5,7%
La *seguridad* en cuanto a robo/hurto en Terminales	4,09	4,34	5,9%
Promedio	5,00	4,90	-2,0%

Anexo 6: Promedio respuestas Variables Indicadoras Tren por género

Indicador	Mujeres	Hombres	Diferencia Porcentual
La *conveniencia* de los horarios de viaje (mayor disponibilidad de viajes diarios)	5,36	5,23	-2,4%
La *facilidad* para viajar con niños o bultos pesados	6,07	6,00	-1,2%
El *aprovechamiento del tiempo* en el viaje para leer, escribir, trabajar o realizar otras actividades	6,11	6,13	0,4%
La *comodidad* durante el viaje (calidad de los asientos, espacios disponible, entre otras)	6,23	6,18	-0,9%
La *puntualidad* en el cumplimiento del horario del viaje	6,27	6,23	-0,7%
La *relajación* durante el viaje (capacidad para disfrutar el viaje)	6,22	6,22	0,0%
La *seguridad* al viajar (sensación de que no ocurra un accidente en el transcurso del viaje)	6,13	6,08	-0,8%
La *seguridad* en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)	5,24	5,36	2,3%
La *seguridad* en cuanto a robo/hurto al interior del vehículo/coche tren	5,66	5,65	-0,1%
La *seguridad* en cuanto a robo/hurto en Terminales	5,23	5,53	5,7%
Promedio	5,85	5,86	0,1%

Anexo 7: Promedio respuestas Variables Indicadoras Bus por tramo de edad.

Indicador	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89
La *conveniencia* de los horarios de viaje (mayor disponibilidad de viajes diarios)	5,77	5,78	5,89	5,90	5,81	5,64	6,16	4,36
La *facilidad* para viajar con niños o bultos pesados	4,66	4,69	4,66	4,77	4,96	5,06	5,29	4,50
El *aprovechamiento del tiempo* en el viaje para leer, escribir, trabajar o realizar otras actividades	5,27	5,23	4,88	5,17	5,18	5,24	5,21	4,79
La *comodidad* durante el viaje (calidad de los asientos, espacios disponible, entre otras)	5,23	4,89	4,86	4,98	5,20	5,10	5,37	4,64
La *puntualidad* en el cumplimiento del horario del viaje	5,39	5,43	5,45	5,55	5,79	5,64	5,77	4,93
La *relajación* durante el viaje (capacidad para disfrutar el viaje)	5,37	5,10	4,82	5,13	5,28	5,17	5,80	4,93
La *seguridad* al viajar (sensación de que no ocurra un accidente en el transcurso del viaje)	5,06	4,62	4,57	4,66	4,77	5,01	5,43	4,93
La *seguridad* en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)	4,26	4,08	4,05	4,28	4,14	4,44	4,71	4,93
La *seguridad* en cuanto a robo/hurto al interior del vehículo/coche tren	4,65	4,72	4,62	4,76	4,85	4,83	5,30	4,50
La *seguridad* en cuanto a robo/hurto en Terminales	4,21	3,91	4,10	4,18	4,28	4,38	5,06	5,00
Promedio	4,99	4,84	4,79	4,94	5,03	5,05	5,41	4,75

Anexo 8: Promedio respuestas Variables Indicadoras Tren por tramo de edad

Indicador	10-19	20-29	30-39	40-49	50-59	60-69	70-79	80-89
La *conveniencia* de los horarios de viaje (mayor disponibilidad de viajes diarios)	5,47	5,09	5,16	5,58	5,13	5,55	5,47	5,64
La *facilidad* para viajar con niños o bultos pesados	5,76	5,80	6,24	6,14	5,94	6,29	6,01	5,86
El *aprovechamiento del tiempo* en el viaje para leer, escribir, trabajar o realizar otras actividades	5,87	6,06	6,32	6,23	6,00	6,18	5,96	6,00
La *comodidad* durante el viaje (calidad de los asientos, espacios disponible, entre otras)	6,11	6,07	6,23	6,25	6,23	6,39	6,10	6,21
La *puntualidad* en el cumplimiento del horario del viaje	6,37	6,26	6,18	6,29	6,23	6,30	6,23	6,07
La *relajación* durante el viaje (capacidad para disfrutar el viaje)	5,90	6,15	6,27	6,31	6,13	6,32	6,40	6,36
La *seguridad* al viajar (sensación de que no ocurra un accidente en el transcurso del viaje)	6,26	6,00	6,17	6,20	6,09	6,08	6,06	5,86
La *seguridad* en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)	5,31	5,22	5,29	5,44	5,15	5,43	5,33	5,93
La *seguridad* en cuanto a robo/hurto al interior del vehículo/coche tren	5,84	5,37	5,67	5,77	5,64	5,79	5,80	6,14
La *seguridad* en cuanto a robo/hurto en Terminales	5,34	4,94	5,44	5,47	5,49	5,52	5,77	6,07
Promedio	5,82	5,70	5,90	5,97	5,80	5,99	5,91	6,01

Anexo 9: Resultados análisis componentes principales

Puede observarse que las dos primeras componentes, cuyos valores son 4,68 y 1,05 (mayores a uno) dan cuenta del 57,29% de la varianza total.

Componente	Valores Propios Iniciales			Factores Seleccionados			Rotación Factores Seleccionados		
	Total	% de la Varianza	Acumulado	Total	% de la Varianza	Acumulado	Total	% de la Varianza	Acumulado
1	4,68	46,79	46,79	4,68	46,79	46,79	3,39	33,92	33,92
2	1,05	10,50	57,29	1,05	10,50	57,29	2,18	21,76	55,68
3	0,93	9,25	66,55	0,93	9,25	66,55	1,09	10,86	66,55
4	0,63	6,34	72,89						
5	0,61	6,15	79,04						
6	0,51	5,13	84,16						
7	0,47	4,71	88,87						
8	0,41	4,09	92,96						
9	0,38	3,83	96,79						
10	0,32	3,21	100,00						

Anexo 9: Resultados análisis componentes principales (continuación)

Sin embargo, para interpretar de mejor forma los componentes, resulta conveniente conservar los tres primeros. Así, los resultados de la matriz rotada son:

Indicador		Componente 1	Componente 2	Componente 3
(i)	La *conveniencia* de los horarios de viaje (mayor disponibilidad de viajes diarios)	0,090	0,038	0,962
(ii)	La facilidad para viajar con niños o bultos pesados	0,694	0,295	0,095
(iii)	El aprovechamiento del tiempo en el viaje para leer, escribir, trabajar o realizar otras actividades	0,752	0,102	0,103
(iv)	La comodidad durante el viaje (calidad de los asientos, espacios disponible, entre otras)	0,788	0,277	0,066
(v)	La puntualidad en el cumplimiento del horario del viaje	0,610	0,238	0,308
(vi)	La relajación durante el viaje (capacidad para disfrutar el viaje)	0,791	0,266	-0,018
(vii)	La seguridad al viajar (sensación de que ocurra un accidente en el transcurso del viaje)	0,699	0,396	-0,112
(viii)	La seguridad en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)	0,359	0,653	0,012
(ix)	La seguridad en cuanto a robo/hurto al interior del vehículo/coche tren	0,219	0,814	0,164
(x)	La seguridad en cuanto a robo/hurto en Terminales	0,230	0,792	-0,039

La primera componente correspondería a "comodidad", la segunda a "seguridad" y la tercera a "conveniencia horaria". Las magnitudes de los efectos de los indicadores sobre las componentes permiten realizar estas interpretaciones. Las variables indicadoras asociadas a "comodidad" (los indicadores (*ii*), (*iii*), (*iv*), (*v*) y (*vi*)) presentan una mayor magnitud asociada a la primera componente (componente 1). Las variables indicadoras

Anexo 9: Resultados análisis componentes principales (continuación)

asociadas a la seguridad (variables indicadoras (*vii*), (*viii*), (*ix*) y (*x*)) presentan una mayor magnitud a la segunda componente (componente 2). Por último, la variable indicadora (*i*) asociada a la conveniencia horaria, carga principalmente a la tercera componente (componente 3) lo que nos permitiría interpretar esta componente como "conveniencia"

Anexo 10: Indicadores normalizados Modelos 3 y 4

Normalizaciones Modelo 3

Indicador	Comodidad Bus	Seguridad Bus	Comodidad Tren	Seguridad Tren
La facilidad para viajar con niños o bultos pesados	1			
El aprovechamiento del tiempo en el viaje para leer, escribir, trabajar o realizar otras actividades			1	
La comodidad durante el viaje (calidad de los asientos, espacios disponible, entre otras)				
La puntualidad en el cumplimiento del horario del viaje				
La relajación durante el viaje (capacidad para disfrutar el viaje)				
La seguridad al viajar (sensación de que ocurra un accidente en el transcurso del viaje)				
La seguridad en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)		1		
La seguridad en cuanto a robo/hurto al interior del vehículo/coche tren				1
La seguridad en cuanto a robo/hurto en Terminales				

Anexo10: Indicadores normalizados Modelos 3 y 4 (continuación)

Normalizaciones Modelo 4

Indicador	Comodidad Bus	Seguridad Bus	Comodidad Tren	Seguridad Tren
La facilidad para viajar con niños o bultos pesados	1		1	
El aprovechamiento del tiempo en el viaje para leer, escribir, trabajar o realizar otras actividades				
La comodidad durante el viaje (calidad de los asientos, espacios disponible, entre otras)				
La puntualidad en el cumplimiento del horario del viaje				
La relajación durante el viaje (capacidad para disfrutar el viaje)		1		1
La seguridad al viajar (sensación de que ocurra un accidente en el transcurso del viaje)				
La seguridad en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)				
La seguridad en cuanto a robo/hurto al interior del vehículo/coche tren				
La seguridad en cuanto a robo/hurto en Terminales				

Anexo 11: Inclusión directa de variables del modelo MIMIC en el modelo de elección discreta[37]

Coeficiente θ	Estimador
Costo_de_viaje * Categoría	0,3375 (5,5)
Costo_de_viaje* Empresa_paga_el_pasaje	0,2671 (2,5)
Tiempo_de_viaje * Sexo	-0,0058 (-1,1)
Costo_de_viaje	-0,696 (-6,4)
Tiempo_de_viaje	0,00043 (0,1)
Constante_modal_Bus	8,57 (7,7)
Constante_modal_Tren	-3,91 (7,7)
¿Ha visto o sufrido algún robo en bus o terminal de bus?	0,3332 (0,9)
¿Ha visto o sufrido algún accidente en bus?	0,6306 (1,6)
Desarrollo de Actividades Intelectuales	-0,9968 (-3,0)
Edad	-0,0212 (-2,1)
¿Viaja en Grupo?	-0,8739 (-2,1)
Ingreso Alto	-1,005 (-3,1)
¿Viaja con *notebook*?	-0,7061 (-1,5)
¿Ha visto o sufrido algún accidente en tren?	0,8868 (1,6)
Índice de Fatalidades	-11,38 (-7,9)
Desarrollo de Actividades Sociales	-1,337 (-2,3)

Nota: test t entre paréntesis

[37] Todas las variables socioeconómicas y de características de los usuarios fueron incluidas en la función de utilidad del bus.

Anexo11: Inclusión directa de variables del modelo MIMIC en modelo de elección discreta (continuación)

Coeficiente θ	Estimador
¿Ha visto o sufrido algún robo en tren o terminal de tren?	-0,0364 (-0,1)
La facilidad para viajar con niños o bultos pesados	0,04312 (0,3)
El aprovechamiento del tiempo en el viaje para leer, escribir, trabajar o realizar otras actividades	0,4786 (3,2)
La comodidad durante el viaje (calidad de los asientos, espacios disponible, entre otras)	-0,02984 (-0,2)
La puntualidad en el cumplimiento del horario del viaje	-0,1887 (-1,3)
La relajación durante el viaje (capacidad para disfrutar el viaje)	0,3396 (2,1)
La seguridad al viajar (sensación de que ocurra un accidente en el transcurso del viaje)	0,1989 (1,4)
La seguridad en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)	-0,092 (-0,6)
La seguridad en cuanto a robo/hurto al interior del vehículo/coche tren	0,04663 (0,3)
La seguridad en cuanto a robo/hurto en Terminales	0,622 (4,4)
Log-verosimilitud	-149,7476
Rho^2	0,6349

Nota: test t entre paréntesis-

Anexo 12: Inclusión directa de variables indicadoras en el modelo de elección discreta

Coeficiente θ	Estimador
Costo_de_viaje * Categoría	0,4051 (7,9)
Costo_de_viaje* Empresa_paga_el_pasaje	0,2873 (2,9)
Tiempo_de_viaje * Sexo	-0,003476 (-0,8)
Costo_de_viaje	-0,6874 (-7,2)
Tiempo_de_viaje	-0,01257 (-2,1)
Constante_modal_Bus	1,40 (1,6)
Constante_modal_Tren	0,68 (1,6)
La *conveniencia* de los horarios de viaje (mayor disponibilidad de viajes diarios)	0,3163 (3,8)
La facilidad para viajar con niños o bultos pesados	0,1535 (1,5)
El aprovechamiento del tiempo en el viaje para leer, escribir, trabajar o realizar otras actividades	0,3105 (2,6)
La comodidad durante el viaje (calidad de los asientos, espacios disponible, entre otras)	-0,03038 (-0,3)
La puntualidad en el cumplimiento del horario del viaje	-0,1130 (-1,0)
La relajación durante el viaje (capacidad para disfrutar el viaje)	0,4025 (3,0)
La seguridad al viajar (sensación de que ocurra un accidente en el transcurso del viaje)	0,1080 (0,9)
La seguridad en el vehículo/coche tren (nivel de protección brindado por el vehículo en caso de accidente)	-0,06892 (-0,5)
La seguridad en cuanto a robo/hurto al interior del vehículo/coche tren	0,01491 (0,1)
La seguridad en cuanto a robo/hurto en Terminales	0,4566 (4,1)
Log-Likelihood	0,4569
Rho^2	-225,86

Nota: test t entre paréntesis-