



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA CIVIL

DESARROLLO DE UN MODELO DE ELECCIÓN DE RUTA EN METRO

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL

CRISTIAN HERNÁN HERRERA ROJAS

PROFESORA GUÍA:
MARCELA MUNIZAGA MUÑOZ

MIEMBROS DE LA COMISIÓN:
ALEJANDRO ALDEA SALAZAR
DANIEL SCHWARZ SANZ

SANTIAGO DE CHILE
ABRIL 2014

RESUMEN DE LA MEMORIA PARA OPTAR AL
TÍTULO DE: Ingeniero Civil
POR: Cristian Hernán Herrera Rojas
FECHA: 09-04-2014
PROFESORA GUÍA: Marcela Munizaga Muñoz

DESARROLLO DE UN MODELO DE ELECCIÓN DE RUTA EN METRO

El presente trabajo aborda el problema de modelar la elección de ruta en el Metro de Santiago para construir una herramienta capaz de capturar la comparación que realizan los usuarios de sus rutas disponibles. La motivación de este trabajo es el uso de información automatizada generada por las transacciones bip! y las variables operacionales de Metro para incorporarlas a modelos de elección de ruta en modalidad predictiva, en desmedro de metodologías tradicionales más costosas tales como las encuestas EOD, tomando en cuenta que la red aumentará su extensión en 37% para el año 2018 con la puesta en marcha de las Líneas 3 y 6. El enfoque del modelo de elección propuesto es mixto, ya que se modelan usuarios con comportamiento semicompensatorio, en que existe tradeoff entre los atributos de algunas rutas y en que hay descarte de las demás rutas disponibles.

En este trabajo se utiliza un modelo con umbrales de tolerancia de los atributos, para modelar que si una ruta posee uno o varios atributos que sobrepasan el umbral, la alternativa correspondiente se descarta del conjunto de rutas disponibles. En particular, este trabajo ocupa el modelo Logit Multinomial Restringido CMNL, que emplea la heurística de incorporación de funciones de penalización o cutoffs a las funciones de utilidad para modelar implícitamente el descarte de rutas que superan uno o más valores umbrales de los atributos, por lo que no se requiere conocer a priori el conjunto de elección de cada individuo.

Se propone un modelo liviano que emplee variables sencillas de obtener, tanto para calibración como para su uso en modalidad predictiva, con el objetivo de observar si la elección de ruta se explica apropiadamente, o bien si requiere de atributos más sofisticados. Para este estudio, se consideran las variables explicativas tradicionales de tiempo de viaje y el número de transbordos, además de la variable de costo angular para medir cuán directa es una ruta. Las técnicas de calibración del CMNL están basadas en el principio de máxima verosimilitud.

Los resultados muestran que los usuarios poseen un comportamiento semicompensatorio en la decisión sobre las rutas en Metro, por lo que se cumple el objetivo de este trabajo. Dada la estructura de la red de Metro y los resultados de las calibraciones, se concluye que el transbordo es un factor penalizador, mientras que el costo angular y el tiempo de viaje son factores compensatorios. De cierta manera, este resultado confirma que el modelo CMNL reproduce el modelo Logit Multinomial MNL, o sea, el modelo CMNL delata si existe o no comportamiento compensatorio en vez de hacer dicho supuesto a priori.

Agradecimientos

Este trabajo es el punto culminante de mi paso por la gloriosa Universidad de Chile. Si bien el trabajo, las enseñanzas y las vicisitudes del camino han contribuido a mi formación como Ingeniero, nada de esto hubiera sido posible sin las personas que han aportado con su grano de arena para que yo haya llegado hasta acá.

En primer lugar, agradezco a toda mi familia por apoyarme siempre. Gracias por todo. Gracias por soportar mis frustraciones varias. Gracias a mi madre y a mi padre, Miriam y Hernán, por inculcarme la importancia del estudio, de incentivar a seguir adelante, y a darme todas las condiciones necesarias para desempeñarme en mis estudios y desenvolverme en la vida. Ustedes lograron que haya estudiado en la mejor Universidad del país. Aunque mi padre ya no está con nosotros, siento que me está apoyando desde algún lugar, procurando que saque lo mejor de cada situación a la que me toque enfrentar.

Muchas gracias a la profesora Marcela Munizaga por darme la oportunidad de desarrollar mi Trabajo de Título en el marco del proyecto FONDEF, siempre dispuesta a ayudarme y orientarme con todos mis planteamientos e inquietudes. Agradezco a Daniel Schwarz por darme la oportunidad de realizar mi práctica profesional en Metro y por la posibilidad de desarrollar este tema vinculado a Metro, que me tuvo trabajando con mucho entusiasmo. También doy las gracias al profesor Alejandro Aldea por sus comentarios sobre este trabajo.

A todos los compañeros con que compartí en los periodos del Plan Común y del plan común de Ingeniería Civil: aprovecho la oportunidad de agradecer a cada uno de ustedes por ayudarme muchísimo en desenvolverme con más soltura y confianza en mi paso por esta Universidad, sobre todo en el aspecto social.

Gracias a la gente de Transporte por ayudarme a mejorar aún más mi vida social, ya que me hicieron pasar muchos ratos entretenidos, tanto en la hora de almuerzo que a menudo se pasa del horario como en las juntas de fin de año, happy hours, titulaciones, congresos y otros eventos. Me permitieron que pudiera explotar mi faceta culinaria en postres y dulces, siempre recibieron las frutas de Chada con mucho agrado, y en el último tiempo han contestado todas mis trivias con entusiasmo.

Finalmente, doy las gracias especialmente a aquellas personas que siempre han creído en mí y que me han dado cabida en sus vidas, apoyándome en mis desafíos y emprendimientos, y estando presentes para escuchar y comprender mis aprehensiones e inseguridades, orientándome con sus puntos de vista para superarlas.

Tabla de contenido

Resumen	ii
Agradecimientos	iii
Índice de tablas	v
Índice de figuras	vi
1. Introducción.....	1
2. Revisión Bibliográfica	3
2.1. Modelos de elección de ruta en redes de Metro.....	3
2.2. El problema de la elección.....	5
2.3. Modelos de elección discreta y la formación del conjunto de alternativas disponibles ...	7
2.4. El modelo Logit Multinomial Restringido CMNL.....	10
3. Análisis de Datos.....	13
3.1. Análisis de los viajes de la EOD de Metro 2011	13
3.2. Análisis de las rutas	17
4. Aplicación del Modelo.....	23
4.1. Calibración del modelo.....	23
4.2. Resultados.....	28
4.3. Discusión	32
5. Conclusiones.....	37
Bibliografía.....	39
Anexos.....	41
Anexo A: Código en MATLAB para calcular el costo angular de segmentos de rutas con origen en Línea 1	41
Anexo B: Código en GAUSS para el modelo MNL 6	42
Anexo C: Código en GAUSS para el modelo CMNL 3	43
Anexo D: Código en GAUSS para el modelo CMNL 4	45

Índice de tablas

Tabla 3.1: Pares O-D dentro de cada línea	13
Tabla 3.2: Clasificación de los pares O-D según el número de rutas registradas.....	14
Tabla 3.3: Número de rutas razonables de Metro.....	18
Tabla 3.4: Estadística descriptiva de las rutas en que existe un proceso de elección.....	20
Tabla 3.5: Correlaciones entre los atributos del modelo de elección de ruta	21
Tabla 3.6: Estadística de las diferencias de tiempo entre rutas extremas.....	22
Tabla 4.1: Simbología de los parámetros para los modelos de elección de ruta calibrados.....	29
Tabla 4.2: Calibración de modelos MNL	30
Tabla 4.3: Calibración de modelos CMNL	30
Tabla 4.4: Tests de razón de verosimilitud para modelos anidados	31
Tabla 4.5: Rango máximo del tiempo de viaje en modelos con tv^2	33

Índice de figuras

Figura 2.1: Penalizaciones de la utilidad del Modelo de Swait.....	9
Figura 2.2: Forma funcional de la penalización	12
Figura 2.3: Forma funcional del logaritmo de la penalización.....	12
Figura 3.1: Gráfico del Coeficiente de Gini – Ruta más elegida para pares O-D con dos rutas ...	16
Figura 3.2: Gráfico del Coeficiente de Gini – Ruta más elegida para pares O-D con tres rutas ...	16
Figura 3.3: Gráfico del Coeficiente de Gini – Ruta más elegida para pares O-D con cuatro rutas	16
Figura 3.4: Cálculo del costo angular.....	20
Figura 3.5: Estructura de la red de Metro.....	21
Figura 3.6: Diferencias de tiempo de viaje entre rutas extremas.....	22
Figura 4.1: Gráfico de las utilidades de modelos con tv^2	32
Figura 4.2: Análisis del cutoff de transbordos ($\theta_{tr} = -1,5819$).....	35
Figura 4.3: Análisis del cutoff del costo angular ($\theta_{cang} = -0,2602$).....	36
Figura 4.4: Análisis del cutoff de la diferencia de tiempo de viaje ($\theta_{tv} = -0,0769$).....	36

Capítulo 1

Introducción

La tecnología disponible en los sistemas de Transporte Público permite recolectar datos de transacciones de pago de los usuarios así como datos de posición de los buses. Dichos datos permiten obtener el horario, la ubicación y el modo de transporte en que se realizan las transacciones. Las principales ventajas de esta forma de recolección de datos son el bajo costo y la alta calidad de la obtención de la información, lo que permite enfocar el tiempo y los recursos a mejorar las técnicas de modelación del comportamiento de los usuarios. Dicha información ha permitido desarrollar modelos sofisticados para predecir el comportamiento de los usuarios frente a cambios de operación y gestión del Transporte Público, los que se manifiestan en cambios de las estructuras de demanda y de viajes. Particularmente en la ciudad de Santiago, donde la modalidad de pago del sistema integrado de buses y Metro mediante tarjeta bip! presenta uso en más del 97% de los casos (Beltrán et al., 2011), se presenta una gran oportunidad para aplicar metodologías sofisticadas para el modelamiento del comportamiento de los usuarios de Transporte Público. El presente trabajo propone un nuevo método de modelación de la estructura de viajes de los usuarios en la red de Metro usando información automatizada, que complementa a las herramientas que se han desarrollado, sin embargo es posible adaptarlo para su uso en otros modos del sistema de Transporte Público.

Metro realiza periódicamente encuestas origen-destino con el objetivo de estimar la estructura de demanda y de viajes de los usuarios de la red de Metro. Éstas se realizan mediante encuestas en terreno a una muestra de usuarios, preguntando por origen y destino del viaje, transbordos realizados, modos de acceso y egreso de la red de Metro, además de variables socioeconómicas tales como el sexo, la edad y el motivo del viaje. Con la puesta en marcha de las Líneas 3 y 6 en el año 2018 la red aumentará su extensión en un 37%, lo cual implica un incremento considerable de los costos de mano de obra y tiempo para el levantamiento de encuestas si es que se sigue con el procedimiento tradicional, además de aumentar la complejidad de recoger fehacientemente la ruta de cada viaje registrado. Por ello, se vuelve necesario desarrollar una nueva metodología que privilegie el uso de información automatizada en desmedro de la metodología actual. Esta metodología debiera estar basada en el uso de las transacciones bip! registradas en el sistema y en la medición localizada de flujos, especialmente en las estaciones de transbordo.

La idea entonces es modelar la elección de ruta de los usuarios de Metro para cada par origen-destino de estaciones a través de un modelo de elección discreta basado en los atributos de las rutas. Se propone un modelo liviano que emplee variables sencillas de obtener, tanto para calibración como para su uso en modalidad predictiva, con el objetivo de observar si la elección de ruta se explica apropiadamente, o bien si requiere de atributos más sofisticados, como por ejemplo, la comodidad, la confiabilidad, y la capacidad de las líneas y coches. Las variables incorporadas en el modelo son las tradicionales, tales como el tiempo de viaje y el número de transbordos, y se propone incluir el costo angular, que es una variable que describe si las rutas son directas, o sea, sin desvíos importantes. Hay varios enfoques para modelar elecciones discretas, pero las más destacables son los modelos de utilidad aleatoria (MUA) descritos por

McFadden (1968) y los modelos de eliminación por aspectos (EBA) descritos por Tversky (1972). En este trabajo se propone un modelo que de alguna forma integra ambos enfoques, a través de un modelo semicompensatorio que considera umbrales de tolerancia de los atributos de las rutas.

El objetivo principal de este Trabajo de Título consiste en desarrollar un modelo de elección de ruta en la red de Metro según los atributos que poseen las rutas considerando el efecto de umbrales, que complementa el trabajo de Munizaga y Palma (2012) de obtención de matrices origen-destino de viajes y que permita construir perfiles de carga en cada una de las líneas de la red de Metro. El modelo de Munizaga y Palma asigna todos los viajes en cada par origen-destino de la red de Metro a la ruta de menor tiempo, no tomando en cuenta la existencia de otras alternativas posibles de rutas para realizar dichos viajes. Este Trabajo de Título viene a incorporar esas alternativas con un modelo que compara implícitamente las rutas con la ayuda de una heurística cuyo resultado es el descarte de alternativas cuando uno o más atributos superan los niveles de tolerancia establecidos por los usuarios.

El resultado esperado de este Trabajo de Título es obtener un modelo de elección de ruta en Metro que capte el fenómeno de comparación que realizan los usuarios al cotejar alternativas de ruta según los atributos de ellas y los niveles aceptables de dichos atributos que toleren los usuarios. La idea es desarrollar un modelo predictivo para Metro, lo que significa que esta herramienta se puede utilizar tanto con los datos de la estructura de la red y de tiempos de viaje que se registran actualmente como con los datos que se registrarán en la red ampliada en 2018.

El siguiente capítulo de este informe presenta la Revisión Bibliográfica, que recopila la literatura existente sobre los principales modelos desarrollados de elección de ruta en redes de Metro, para después introducirse en un análisis más profundo sobre el problema de la elección discreta que justifica el uso de modelos de elección con umbrales, y que finalmente describe el modelo de elección discreta utilizado en este trabajo: el Logit Multinomial Restringido (CMNL). En el tercer capítulo se estudia el fenómeno de reparto de la elección de las rutas que los usuarios de Metro declararon realizar en la Encuesta Origen-Destino de 2011 de Metro. Posteriormente, se encuentran las rutas razonables para los pares origen-destino con sus respectivos atributos para emplear en el modelo. El cuarto capítulo describe las metodologías de calibración empleadas para el CMNL y reporta los resultados obtenidos con su respectiva discusión. Este trabajo finaliza con el capítulo de Conclusiones sobre el trabajo desarrollado y las extensiones que se pueden seguir a partir del modelo de elección calibrado.

Capítulo 2

Revisión Bibliográfica

En el presente capítulo se reportan los trabajos más destacados sobre modelos de elección y desarrollos específicos de éstas en redes de Metro. Este capítulo comienza por la revisión de los trabajos más destacados relacionados con la predicción de elección de ruta en sistemas de Transporte Público, en particular sobre Metro. Después, se desarrolla un análisis más profundo del problema de elección discreta con el objetivo de justificar el uso de un modelo de elección de ruta con umbrales, y se mencionan los modelos de elección discreta más destacados que incorporan la formación del conjunto efectivo de alternativas que un individuo considera. Finalmente, se estudia con detalle uno de los modelos de elección discreta con umbrales, el Logit Multinomial Restringido CMNL, el cual se utiliza en este trabajo para modelar la elección de ruta en Metro.

2.1. Modelos de elección de ruta en redes de Metro

El uso de sistemas de pago automatizados a través de tarjetas inteligentes en los sistemas de Transporte Público ha permitido no solo recolectar las tarifas que pagan los usuarios, sino también obtener bases de datos que son útiles para la planificación de dichos sistemas por parte de las autoridades de Transporte, para mejorar el desempeño estratégico, táctico y operacional de los sistemas de transporte. Pelletier et al. (2011) realizan una revisión de las distintas experiencias de estos sistemas de recolección automatizada de datos presentes en los desarrollos de Transporte Público alrededor del mundo. En esa revisión describen las tecnologías empleadas, los hardwares y softwares utilizados, las implementaciones, el procesamiento de datos, aspectos legales y privados de la información recolectada, entre otros temas.

Munizaga y Palma (2012) desarrollan una herramienta para el Sistema de Transporte Público de Santiago que predice los paraderos de buses y estaciones de Metro de destino de un usuario dada la cadena de transacciones con la tarjeta bip! que realiza durante un periodo dado y más los datos de GPS de los buses. El objetivo es construir una matriz origen-destino de los viajes en Transporte Público en Santiago. En el caso de la estimación de la estación de bajada en Metro, es la estación que se encuentra más cercana a la siguiente transacción dentro de una circunferencia definida por el tiempo de caminata, cuyo radio es de 1.000 metros en el estudio de Munizaga y Palma. La ruta dentro de la red de Metro entre la estación de origen conocida y la estación de destino estimada se obtiene utilizando el criterio de ruta mínima con el algoritmo de Dijkstra.

Una aplicación desarrollada para la elección de ruta en el Metro de Santiago es la propuesta por Raveau et al. (2011), en la cual incluyen las variables topológicas de la red complementarias a las variables tradicionales basadas en los niveles de servicio (tiempo de viaje, costo, transferencias, etc.). Los supuestos detrás de ese modelo son que los usuarios tienden a usar rutas más directas al destino final, tienden a usar rutas más conocidas o en las que viajen con mayor frecuencia, y que los usuarios consideran también factores como la comodidad, confiabilidad o características físicas de los coches y estaciones. Para describir cuán directa es

una ruta, definen el costo angular como un indicador de penalidad de cuán indirecta es una ruta. Para describir el conocimiento de la red, utilizan variables proxy que miden el nivel de uso de las diferentes rutas. El modelo de elección de ruta que emplean es un Logit Multinomial en que la función de utilidad contiene las variables descritas anteriormente, tanto las tradicionales como las de costo angular y las del conocimiento de la red. En su análisis, Raveau et al. consideran solo los pares origen-destino que posean dos o más alternativas de ruta. Calibran el modelo considerando las distancias geográficas reales y las distancias distorsionadas del plano esquemático de red que se muestra a los usuarios. Concluyen que cuando no están presentes las variables no tradicionales, los parámetros estimados están sesgados. También muestran que las distorsiones en el mapa esquemático de la red inducen asignaciones de viajes que reducen los niveles de servicio en la red.

Ke et al. (2013) proponen un modelo de elección de ruta y un algoritmo de cálculo de las proporciones de elección basados en la clasificación de los pasajeros aplicados a la red de metro de Pekín, con el objetivo de considerar la familiaridad del pasajero en la red. Así, los pasajeros se clasifican en dos tipos: familiares, los que poseen experiencia de viaje y eligen su ruta por la experiencia; y ocasionales, los que rara vez toman la ruta y eligen su ruta al observar el mapa. Para cada tipo de pasajero, se calcula la impedancia integrada del viaje: para los tipos familiares es el tiempo de viaje generalizado, y para los tipos ocasionales es la composición de la distancia de recorrido, la distancia en línea recta, el número de estaciones y el número de transbordos. Para la elección de ruta utilizaron el método de corrección de proporción del modelo de distribución multi-ruta. Los resultados fueron corregidos por el coeficiente de hacinamiento, el coeficiente de tiempos de transbordos y el coeficiente de demanda por asientos.

Cuando ya se ha establecido algún modelo de elección, se requiere de un algoritmo capaz de asignar los viajes a las rutas. En general los modelos de elección de ruta asignan el flujo en una red minimizando el costo total y considerando restricciones de capacidad de los arcos, asumiendo que hay información perfecta de los costos de viaje. Prashker y Bekhor (2004) realizan una revisión de los modelos de elección de ruta, donde se analizan los supuestos del comportamiento de los usuarios, y de los algoritmos de asignación a la red en el contexto del problema de equilibrio de usuario estocástico. El presente Trabajo de Título no contempla un modelo de asignación que considere la congestión ni el hacinamiento en los trenes ni en las vías como variables que induzcan a los usuarios a optar por otras rutas en la red de Metro, simplificando el método de asignación. La inclusión de estas variables requeriría la resolución de un problema recursivo, lo que se escapa del objetivo del presente desarrollo.

A pesar de que las rutas de Metro están caracterizadas por atributos (algunos de los cuales son percibidos directamente por los usuarios y otros no) que se combinan entre sí con distintas valoraciones de tal forma que todas las alternativas de un par origen-destino (O-D) sean utilizadas, a menudo no se considera que hay alternativas de rutas de un par O-D que los usuarios simplemente no escogen, debido a que los niveles de los atributos que caracterizan a dichas rutas no resultan atractivos para los usuarios, provocando un aumento de ocupación de las rutas más atractivas con la consecuente congestión en trenes y estaciones, sobre todo en horarios punta. Como un paso previo a la incorporación de las restricciones de capacidad de los trenes y de las vías en la asignación, es necesario estudiar qué atributos poseen valores umbrales que hacen que los usuarios opten por una u otra ruta para viajar en cada par O-D. Esto se traduce en un descarte de alternativas, lo que los modelos de elección discreta más difundidos no toman en cuenta, arrojando predicciones erradas. Por lo tanto, la motivación del presente trabajo es modelar la elección de ruta como un procesamiento de información más complejo, y existen líneas de

investigación que incorporan el descarte en sus modelos de elección. En la siguiente sección, se discute el proceso que realiza un individuo para tomar una decisión sobre la elección de una alternativa, y después se describen los modelos más destacados que incorporan el descarte de alternativas.

2.2. El problema de la elección

La elección consiste en el proceso de toma de decisiones que realiza un individuo o un grupo de personas al tener que seleccionar una alternativa. Así, cada usuario establece el problema de elección (en este caso la elección de una ruta entre dos estaciones de Metro), identifica las alternativas asociadas (las rutas posibles), evalúa los atributos de cada una de éstas, elige una alternativa e implementa la elección realizada (realiza el viaje a través de la ruta escogida). La elección se realiza desde un conjunto universal de alternativas no vacío que contiene todas las alternativas existentes. El tomador de decisiones restringe dicho conjunto según factores económicos, sociales, culturales, físicos y/o ambientales, generando así un subconjunto de alternativas llamado conjunto de alternativas disponibles o conjunto de elección, que contiene todas las alternativas que son factibles y que se supone que el individuo conoce cuando realiza la elección.

En el caso de un par origen-destino en la red de Metro, un individuo elige una ruta principalmente por el tiempo que demora en utilizar una opción u otra, pero también considera el periodo del viaje, el nivel de hacinamiento en los coches o si una ruta es más directa para llegar, entre otros factores. Así, el individuo descarta alternativas según criterios bien definidos hasta obtener una o más rutas factibles a elegir. Existen varios enfoques para modelar elecciones discretas, pero hay dos modelos destacables que se describen a continuación.

Los modelos de utilidad aleatoria MUA (McFadden, 1968) plantean que, dado un segmento de población N y un conjunto A_n de alternativas disponibles para el individuo n , la valoración del individuo $n \in N$ por la alternativa $i \in A_n$ está dada por el valor de la utilidad U_{ni} , de tal forma que la probabilidad de que el individuo n escoja la alternativa i es:

$$P_{ni} = \mathbb{P}(U_{ni} > U_{nj}, \forall j \in A_n) \quad \forall i \in A_n \quad (2.1)$$

En esta formulación, las utilidades son variables aleatorias, ya que reflejan la falta de información con respecto a las características de las alternativas o de los individuos desde el punto de vista del modelador. Esto se representa por una descomposición de la función de utilidad en dos partes: la componente determinística V_{ni} , conocida por el modelador y que es función de un vector de atributos X_i que define la alternativa i ; y la componente aleatoria ε_{ni} , que representa tanto la incapacidad del modelador en apreciar todos los atributos y variaciones que influyen en el comportamiento de los individuos, como también los errores de medición, la imperfección de la información disponible y la forma funcional de la componente determinística. Así, la función de utilidad es la suma de la componente determinística y la componente aleatoria:

$$U_{ni} = V_{ni} + \varepsilon_{ni} \quad (2.2)$$

Según los supuestos que se hagan de la componente aleatoria, se obtienen distintos modelos de tipo MUA. Los más conocidos son el modelo Logit Multinomial MNL, que asume que los errores distribuyen idéntica e independientemente Gumbel (iid Gumbel), y el modelo Probit, en que el error sigue una distribución Normal con una matriz de covarianza general (Σ).

Los modelos de eliminación por aspectos EBA (Tversky, 1972) constituyen la representación de la elección como un proceso estocástico en el que las alternativas son eliminadas secuencialmente hasta que quede solo una. Las alternativas están descritas por un vector de características o atributos, que definen si una alternativa posee o no tales características. Se asume que los individuos realizan una categorización de los atributos y a cada uno le asignan un valor umbral máximo aceptable, en función de sus gustos y restricciones personales. El proceso comienza con el atributo mejor categorizado, se compara dicho atributo de todas las alternativas con el valor umbral asociado, por lo que todas las alternativas que sobrepasan el umbral se descartan. El proceso se repite con las alternativas que quedan y con los atributos que siguen en orden de importancia hasta que quede una alternativa que satisfaga todas las restricciones. Si quedan varias alternativas, ellas son elegidas con igual probabilidad. Es posible que existan diferentes secuencias de eliminación, por lo que la probabilidad de elegir una alternativa es la suma de las probabilidades de todas las secuencias que finalizan con esa alternativa. Los umbrales son un criterio para especificar el conjunto de alternativas de un individuo implícitamente, ya que postula un mecanismo de aceptación-rechazo de los atributos.

Los modelos MUA asumen que los individuos poseen un comportamiento compensatorio entre los atributos¹ (tradeoff entre los atributos) y que el conjunto de elección es conocido y fijo, además se supone que los atributos de cada opción se encuentran en el rango aceptable para que los individuos los consideren. Los modelos EBA, en cambio, no consideran comportamiento compensatorio entre los atributos, a través de un proceso de descarte de alternativas, y el conjunto de elección es variable. La estructura de elección de modelos EBA es rígida, por lo que es más difícil su implementación; mientras que los modelos MUA flexibilizan el método de elección debido al comportamiento compensatorio. La desventaja de los modelos MUA es que la calibración de parámetros es más complicada cuando el número de alternativas disponible es grande, requiriendo mayor capacidad computacional. Otro problema es que si un atributo posee un valor muy alto se puede salir del rango de comportamiento en la función de utilidad, dado que no contempla el descarte de una o más alternativas si es que el atributo supera cierto nivel tolerable, por lo que deriva en un modelo poco realista, invalidando las predicciones hechas en él.

Los modelos MUA son los más utilizados para representar la elección de modo de Transporte, y requieren estudiar el conjunto de alternativas disponibles y las cotas de los atributos, ya que se asume que el conjunto de elección está predefinido y que los atributos se encuentran en el rango aceptable. La hipótesis en este Trabajo de Título es que la elección de ruta en Metro en el caso en que existan dos o más para un par origen-destino seguirá un comportamiento compensatorio si la diferencia de tiempos entre una alternativa y las restantes no sobrepasa el valor umbral de la diferencia de tiempos, en caso contrario, los usuarios descartarán dicha ruta y todos viajarán por una única ruta (la de menor tiempo).

¹ El comportamiento compensatorio se puede observar en la elección de modo de viaje. Por ejemplo, si los modos disponibles son auto y bus, y los atributos conocidos son el costo monetario del viaje y el tiempo del viaje, se tiene que el modo auto posee menor tiempo de viaje que el bus, mientras que el modo bus posee menor costo monetario que el auto. Al valorar ambos modos en la función de utilidad, se obtiene que tanto el auto como el bus tienen utilidades parecidas, por lo que hay un reparto en la elección entre ambos modos en la muestra. En cambio, el comportamiento no compensatorio descarta una alternativa dado que se supera el umbral en algún(ambos) atributo(s), por lo que toda la muestra elige un único modo.

2.3. Modelos de elección discreta y la formación del conjunto de alternativas disponibles

La definición de disponibilidad de alternativas es un tema que en la práctica los modeladores rara vez toman en cuenta, porque consideran que el conjunto de alternativas disponibles se genera de manera exógena (ver discusión en Castro, 2009). Asimismo, no consideran que el individuo realiza un proceso de elección más complejo, en el cual la determinación del conjunto factible de alternativas está sujeta a las restricciones que el individuo perciba en el proceso de elección. Estas restricciones pueden estar relacionadas con su ambiente (por ejemplo, el ingreso) o pueden ser autoimpuestas por el mismo individuo según sus características físicas, sociales o culturales, las que son más difíciles de identificar por el modelador. Por esta razón es que las restricciones pueden ser determinísticas o probabilísticas, lo que depende del grado de conocimiento del modelador sobre el proceso de elección de los individuos en estudio.

Una de las herramientas más usadas para generar el conjunto de elección son los modelos con umbrales, que son cotas impuestas por los individuos a cada atributo considerado para la elección de una alternativa. La idea es que si una alternativa posee uno o varios atributos que sobrepasan el umbral, la alternativa correspondiente se descarta del conjunto de alternativas disponibles.

La justificación para ocupar modelos con umbrales es que los individuos, al tomar una decisión, combinan comportamientos compensatorios y no compensatorios simultáneamente, de tal manera que las alternativas que no cumplan con las restricciones no sean consideradas en la elección. También se desea captar que las personas no realizan un análisis exhaustivo de cada alternativa, modelando a los individuos como procesadores de información (Castro, 2009). La idea de esta herramienta es integrar los enfoques de modelos MUA, a través de la valoración de las alternativas con la función de utilidad en función de sus atributos, y de los modelos EBA, a través de un proceso de aceptación-rechazo de las alternativas disponibles discriminando por sus atributos. La ventaja de los umbrales es que son más fáciles de implementar en modelos de elección que un proceso de descarte de las alternativas.

Hay dos maneras para modelar el conjunto de alternativas disponibles: generar el conjunto con un modelo externo o incluir en el modelo de elección la disponibilidad de alternativas implícitamente. Se describen ambos procedimientos a continuación.

Manski (1977) plantea un modelo externo de generación del conjunto de alternativas, en el cual el proceso de elección ocurre en dos etapas. La primera etapa consiste en que el individuo n escoge mediante fuerzas exógenas el conjunto de alternativas disponibles o conjunto de elección A_n a considerar para su proceso de elección. Este conjunto pertenece al espacio de conjuntos de elecciones Γ , que corresponde a todos los subconjuntos posibles que se pueden formar del conjunto universal de alternativas. La segunda etapa consiste en la elección de la alternativa i perteneciente al conjunto escogido A_n . Bajo esta teoría, la probabilidad de que el individuo n escoja la alternativa i es:

$$P_{ni} = \sum_{A_n \in \Gamma} \mathbb{P}_n(i|A_n) \mathbb{P}_n(A_n|\Gamma) \quad (2.3)$$

donde $\mathbb{P}_n(A_n|\Gamma)$ es la probabilidad de que el individuo n escoja el conjunto de alternativas A_n del espacio de conjuntos de elecciones Γ y $\mathbb{P}_n(i|A_n)$ es la probabilidad de que el individuo n escoja la alternativa i condicional al conjunto de elección A_n . A pesar de su generalidad, la expresión (2.3) se vuelve difícil de implementar computacionalmente, dado que si el conjunto universal de alternativas posee m elementos, el número de subconjuntos pertenecientes a Γ es $2^m - 1$, el cual crece si el número de alternativas es grande, por lo que la sumatoria (2.3) tendrá como máximo $2^m - 1$ términos, por lo que se requieren heurísticas para calcular las probabilidades (Morikawa, 1995).

Los modelos implícitos consisten en incluir dentro del proceso de elección de una alternativa las restricciones que limiten el número de alternativas consideradas, por lo que se representa a un individuo que combina al mismo tiempo tanto la elección como la generación del conjunto de alternativas. Esta modelación implica realizar en una etapa el proceso de elección descrito en el párrafo anterior, lo que resulta más fácil y conveniente de implementar computacionalmente, debido a que se puede calibrar con métodos clásicos. Sin embargo, pueden surgir errores de especificación cuando los atributos poseen doble rol, no pudiendo identificar las variables que modifican la utilidad de las que definen el conjunto de alternativas (Cascetta y Papola, 2001). La modelación implícita del conjunto de alternativas disponibles es el enfoque empleado en este Trabajo de Título.

Swait (2001) plantea el modelo de elección como un problema de optimización en el que el individuo escoge la alternativa que maximiza su utilidad sujeto a restricciones (umbrales) en los atributos. En este contexto el individuo puede violar los umbrales, pero con un costo que disminuye el nivel de utilidad alcanzado cuando se violan las restricciones, lo que se modela con la introducción de penalizaciones o cutoffs a la función de utilidad. Así, el problema de elección de la alternativa i , que pertenece al conjunto universal de alternativas C , se modela de la siguiente manera:

$$\begin{aligned}
 \max U &= \sum_{i \in C} \delta_i U(X_i) + \sum_{i \in C} \sum_k \delta_i (w_k \lambda_{ik} + v_k \kappa_{ik}) \\
 \text{s. a.} \quad &\sum_{i \in C} \delta_i = 1 \\
 &\delta_i (a_i - X_i) - \lambda_i \leq 0 \quad \forall i \in C \\
 &\delta_i (X_i - b_i) - \kappa_i \leq 0 \quad \forall i \in C \\
 &\delta_i \in \{0,1\} \quad \forall i \in C \\
 &\lambda_i, \kappa_i \geq 0 \quad \forall i \in C
 \end{aligned} \tag{2.4}$$

donde:

- δ_i es una variable binaria que vale uno si la alternativa i es escogida, y cero si no;
- X_i es el vector de atributos de la alternativa i ;
- a_i y b_i son, respectivamente, los vectores de umbrales mínimos y máximos del vector de atributos X_i de la alternativa i ;
- λ_{ik} y κ_{ik} son las cantidades en que se violan respectivamente los umbrales mínimos y máximos del atributo k cuando se elige la alternativa i ;
- w_k y v_k son parámetros a calibrar que representan la desutilidad marginal de violar la restricción de violar el umbral mínimo y máximo del atributo k , respectivamente.

El problema (2.4) posee tres variables de decisión: δ_i , asociado a la elección de la alternativa i ; y λ_{ik} y κ_{ik} , asociados a la decisión de violar las restricciones impuestas con el correspondiente costo que ello implica, dado que se eligió la alternativa i . Como se observa en la Figura 2.1, en el intervalo entre las cotas superior e inferior de los atributos, la función de utilidad es compensatoria, mientras que fuera del intervalo las penalizaciones de la utilidad se activan, lo que provoca que la función de utilidad disminuya su valor. Se observa que mientras mayor es la violación a la restricción, menor es el valor de la utilidad y mayor es la diferencia con respecto a la función de utilidad compensatoria.

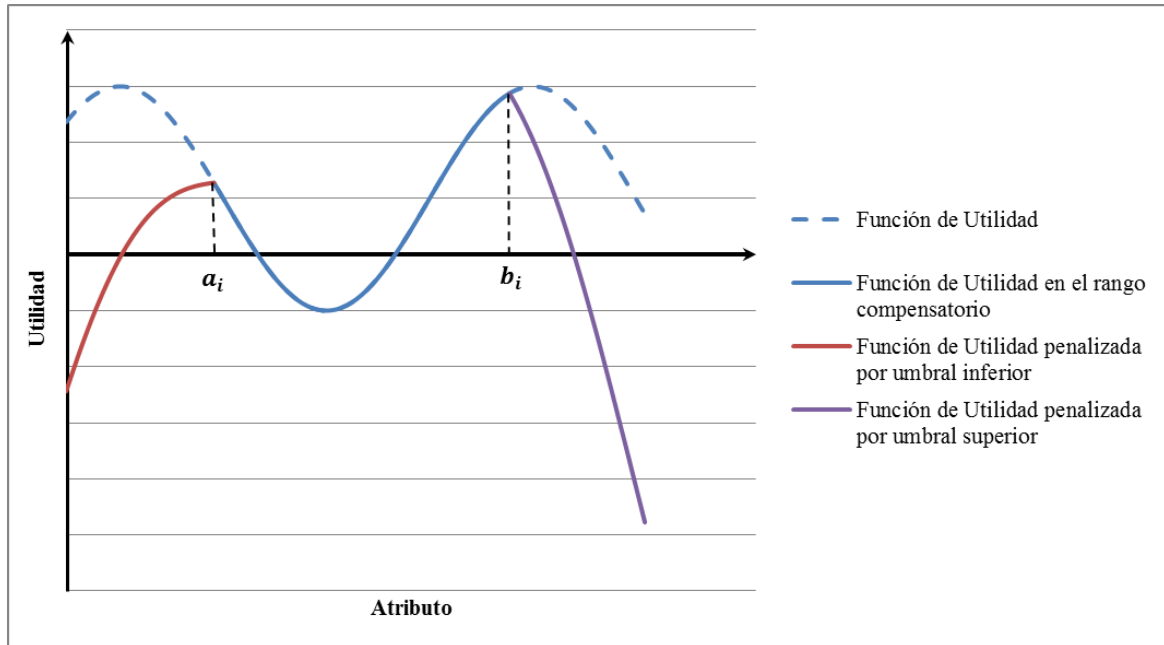


Figura 2.1: Penalizaciones de la utilidad del Modelo de Swait
(Fuente: Elaboración propia)

Así, el modelo (2.4) es capaz de capturar comportamientos compensatorios, no compensatorios, además de comportamientos mixtos en los individuos. Swait calibra los parámetros de las funciones de utilidad y las penalizaciones para un caso de elección de arriendo de automóviles, concluyendo que la introducción de penalizaciones contribuye significativamente a explicar las elecciones de los individuos y la sensibilidad a los atributos. Ya que esta formulación es un problema de optimización determinístico, no permite que las penalizaciones posean error (no son aleatorias). Además, las funciones de utilidad no son derivables en los umbrales por la inclusión de umbrales lineales, por lo que la curva no es suave, lo que implica que la utilidad es inestable cerca de los umbrales en los cálculos en problemas de equilibrio, provocando problemas computacionales.

Cantillo y Ortúzar (2005) plantean un modelo semicompensatorio en dos etapas, en el cual proponen expresiones para las probabilidades de la formulación de Manski (2.3). Las consideraciones son que la probabilidad de que se incluya una alternativa en el conjunto de elección es igual a la probabilidad de que los atributos de la alternativa no sobrepasen el vector de umbrales para el individuo n . También consideran que si dos o más alternativas son obtenidas de la primera etapa, se asume un comportamiento compensatorio en la segunda etapa de elección. En la segunda etapa de elección, asumen un modelo MUA, en el cual la expresión obtenida para la probabilidad de elección de una alternativa depende, entre otras variables, de la distribución de la componente aleatoria. Su análisis empírico es con datos simulados sobre un modelo de

elección de modo de transporte y con datos reales sobre la elección de ruta entre Santiago y Valparaíso. Concluyen que si hay evidencia de la existencia de umbrales en la población, el uso de un modelo totalmente compensatorio produce sesgos en la estimación de parámetros y se pueden producir errores significativos en la predicción.

El trabajo de Kaplan et al. (2012) relaja el supuesto de que los términos de error sean idénticos e independientemente distribuidos a través de las alternativas en la etapa de elección. Se incorpora una estructura de error flexible en la etapa de elección basada en la utilidad. Considerando la formulación de Manski, la representación de la formación del conjunto de elección toma en cuenta múltiples umbrales correlacionados, mientras que la elección basada en la utilidad acomoda los datos con patrones correlacionados anidados a través de las alternativas y la heterogeneidad de los gustos a través de la población, utilizando las formulaciones del mixed logit. El análisis empírico es acerca de una encuesta de elección de departamentos realizada a estudiantes de la ciudad de Haifa, Israel, correlacionando tanto las características de los departamentos; de los vecindarios en donde su ubican; de la cercanía a los sectores de servicios, estudios y empleos con las características socioeconómicas de los estudiantes. Los resultados mostraron mejoras en la bondad de ajuste y en la capacidad predictiva del modelo al compararlo con un modelo con la estructura de error iid. Kaplan et al. plantean que el trabajo futuro debiera orientarse a considerar modelos de clase latente y segmentación de la población en la elección semicompensatoria; reglas alternativas de decisión debido a que hay alternativas dominantes; y el uso de modelos dinámicos en que se integren aspectos históricos, oferta de mercado no estacionaria, y el aspecto temporal.

Con el objetivo de captar el fenómeno de comparación y descarte de rutas posibles de los pares O-D de Metro, en el presente trabajo se emplea una heurística en que la utilidad (expresión 2.2) se le incorpora una función de penalización, que depende del valor de los atributos de una alternativa y de los valores umbrales que los usuarios poseen de dichos atributos. La siguiente sección presenta los detalles del modelo Logit Multinomial Restringido CMNL, el cual sigue la heurística mencionada aquí y el que se calibra como parte de este Trabajo de Título.

2.4. El modelo Logit Multinomial Restringido CMNL

Martínez et al. (2009) plantean el Modelo Logit Multinomial Restringido CMNL, el cual se utiliza en este trabajo. Con la idea de definir un modelo de una etapa en donde se imponen umbrales a los individuos en la función de utilidad, se plantea un modelo semicompensatorio con una función de utilidad aleatoria restringida, con la incorporación de una función de penalización o cutoff, con el fin de disminuir fuertemente el valor de la función de utilidad cuando los atributos de una alternativa no cumplan las restricciones a las que están sujetas a los individuos. La función de utilidad restringida U_{ni} de un individuo n asociada a la alternativa i es:

$$U_{ni}(X_i) = V_{ni}(X_i) + \frac{1}{\mu} \ln(\varphi_{ni}(X_i)) + \varepsilon_{ni} \quad (2.5)$$

donde:

- V_{ni} es la componente determinística compensatoria de la función de utilidad;
- X_i es el vector de atributos de la alternativa i ;
- ε_{ni} es la componente aleatoria de la función de utilidad;
- μ es el parámetro de escala de la distribución Gumbel;
- φ_{ni} corresponde a la penalización o cutoff impuesto por el individuo n a la alternativa i .

El cutoff está multiplicado por el inverso del parámetro de escala de la distribución Gumbel, el que aumenta la penalización cuando la dispersión de la utilidad aumenta, lo que implica que el cutoff en el CMNL es más relevante cuando la variabilidad de la utilidad aumenta, o sea, cuando μ disminuye.

Suponiendo que el error distribuye idéntica e independientemente Gumbel (iid Gumbel), la probabilidad de que el individuo n elija la alternativa i es:

$$P_{ni} = \frac{\varphi_{ni} \exp(\mu V_{ni})}{\sum_{j \in C_n} \varphi_{nj} \exp(\mu V_{nj})} \quad (2.6)$$

El cutoff está construido por la composición de penalidades de todos los K atributos que describen una alternativa, tanto en el umbral mínimo como en el umbral máximo:

$$\varphi_{ni} = \prod_k^K \varphi_{nik}^L \cdot \varphi_{nik}^U \quad (2.7)$$

donde φ_{nik}^L es la penalización del umbral inferior del atributo k de la alternativa i que impone el individuo n y φ_{nik}^U es la penalización del umbral superior del atributo k de la alternativa i que impone el individuo n . Las formas funcionales de estas penalizaciones son de un logit binomial, las que se especifican en las expresiones (2.8) y (2.9):

$$\varphi_{nik}^L = \frac{1}{1 + \exp(\omega_k (a_{nk} - x_{ik} + \rho_k))} \quad (2.8)$$

$$\varphi_{nik}^U = \frac{1}{1 + \exp(\omega_k (x_{ik} - b_{nk} + \rho_k))} \quad (2.9)$$

donde:

- a_{nk} y b_{nk} son los valores de las cotas o umbrales inferior y superior respectivamente del atributo k que restringen la elección del individuo n ;
- x_{ik} es el valor del atributo k de la alternativa i ;
- ω_k es el parámetro de escala del logit binomial;
- ρ_k es un parámetro de ajuste dado por la expresión (2.10), en la cual η_k es la proporción de la población que viola la restricción impuesta al atributo k .

$$\rho_k = \frac{1}{\omega_k} \ln \left(\frac{1 - \eta_k}{\eta_k} \right) \quad (2.10)$$

En las Figuras 2.2 y 2.3 se grafican la forma funcional de la penalización de la cota superior y el logaritmo de dicha penalización, respectivamente, en función de la diferencia entre el valor del atributo k de la alternativa i y la cota superior. En la Figura 2.3 se observa que dado que la penalización adopta valores entre cero y uno, el logaritmo de la penalización es negativo, y se observa que cuando el atributo de la alternativa supera el valor umbral, el logaritmo del cutoff se vuelve más negativo, por lo que hace disminuir el valor de la función de utilidad restringida cuando se supera el umbral. Cabe destacar que la formulación (2.5), con la incorporación de un logaritmo en el cutoff, permite que la transición entre la zona compensatoria y la no compensatoria (ver Figura 2.1) sea suave.

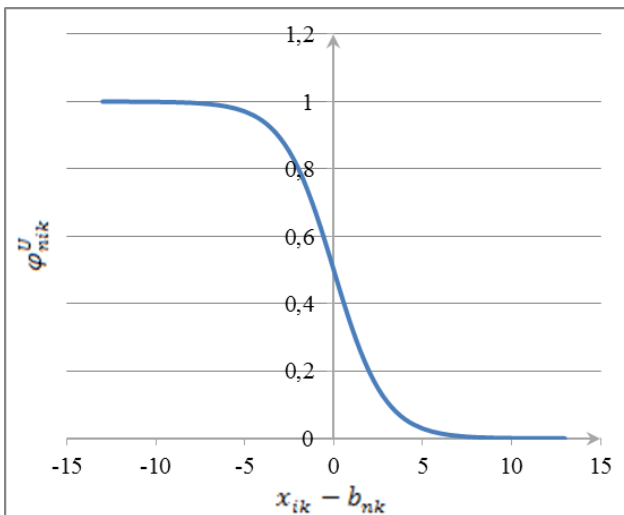


Figura 2.2: Forma funcional de la penalización
(Fuente: Elaboración propia)

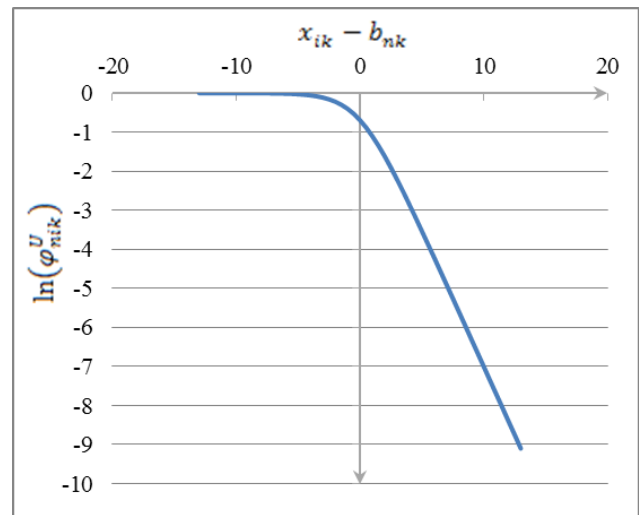


Figura 2.3: Forma funcional del logaritmo de la penalización
(Fuente: Elaboración propia)

Castro et al. (2013) desarrollan métodos de calibración del modelo CMNL, utilizando el método de máxima verosimilitud. Especifican la función de verosimilitud para el CMNL y derivan e interpretan las condiciones de optimalidad, discuten sobre la identificabilidad de los parámetros y proponen calibraciones con cotas exógenas (umbrales conocidos por el modelador) y endógenas (umbrales estimados por el modelador) al cutoff. Obtienen que las condiciones de primer orden del CMNL reproducen las condiciones del MNL dentro del dominio de las variables. Concluyen que el modelo CMNL reproduce el modelo MNL, pero al revés no. También obtienen que la estimación del valor subjetivo del tiempo varía significativamente entre el modelo CMNL y MNL, incluso dentro del dominio compensatorio, el modelo MNL subestima los valores del modelo CMNL. El detalle de las metodologías de calibración se explican en Castro (2009).

Bierlaire et al. (2010) realizan un ejemplo sencillo de elección y una calibración de una encuesta de un tren de alta velocidad hipotético en Suiza, en que analizan cuán bien se aproxima la generación implícita del conjunto de elección del modelo CMNL a la generación explícita del conjunto de alternativas disponibles de Manski (1977). Muestran que el CMNL no captura el proceso de generación del conjunto de elección, ya que la complejidad del modelo Martínez et al. (2009) es lineal con el número de alternativas mientras que en el enfoque de Manski la complejidad es exponencial. El modelo CMNL, para que sea un mecanismo eficiente de generación de conjunto de elección, requiere del supuesto de que la probabilidad de considerar una alternativa crece con su utilidad, lo que significa que el conjunto de elección sólo depende de las preferencias del individuo. Sin embargo, las alternativas con alta utilidad pueden ser descartadas en presencia de restricciones presupuestarias o físicas.

Se propone utilizar el CMNL con el objetivo de modelar el reparto entre las distintas rutas posibles de los pares origen-destino en la red de Metro, a través de la introducción de umbrales que expliquen que los usuarios descartan ciertas rutas dado que sobrepasan niveles umbrales. Así por ejemplo, una posible ruta que supere una cota de tiempo de viaje esperado o deseado, será menos probable de ser escogida por los usuarios. Con el modelo de elección escogido, el siguiente paso es recopilar los datos disponibles y procesarlos para la calibración del modelo CMNL. El siguiente capítulo reporta la información disponible sobre las rutas empleadas por los usuarios en la red de Metro y las variables explicativas a incluir en el modelo de elección.

Capítulo 3

Análisis de Datos

En este capítulo se realiza un análisis tanto del fenómeno de la elección de ruta que los usuarios de Metro declaran realizar como de los datos empleados en la calibración del modelo de elección de ruta. Los objetivos de este capítulo son caracterizar a la muestra con que se calibra el modelo para explicar el fenómeno de las elecciones de ruta y vislumbrar las variables que expliquen la elección de ruta de los usuarios.

Para analizar el reparto en la elección las rutas de los distintos pares origen-destino (O-D) de la red de Metro se utiliza la Encuesta Origen-Destino (EOD) que realizó Metro en el año 2011. Con respecto a las variables explicativas a proponer para el modelo, éstas se obtienen de los variables de operación de Metro y de modelos de ruta implementados por la Universidad de Chile. Con esta información, se efectúa un análisis estadístico descriptivo y se indaga en las correlaciones entre dichas variables. Para la estadística y la posterior calibración del modelo, se emplean los tiempos de viaje y las observaciones de la EOD del periodo fuera de punta.

3.1. Análisis de los viajes de la EOD de Metro 2011

En esta sección se realiza un análisis de la Encuesta Origen-Destino de Metro del año 2011, en la que se registran los viajes que realizaron los usuarios de la red al momento de la interceptación. La información usada para el análisis consiste en las estaciones de origen y de destino y los transbordos entre líneas que cada encuestado declaró realizar. Esta encuesta se efectuó en los periodos punta mañana, punta tarde y fuera de punta de los días laborales, y en días sábados y festivos. Dado el fenómeno que se desea estudiar, en esta sección se analizan los viajes con transbordo para determinar en qué pares O-D se registra un reparto en la elección de ruta, o bien se registra una predominancia de una ruta, o simplemente no hay reparto y todos los usuarios eligieron una única ruta.

La red de Metro posee 108 estaciones, de las cuales hay 16 estaciones correspondientes a estaciones de transferencia entre las líneas, contando la estación de transbordo como dos estaciones, una para cada línea. Así, hay 11.540 pares de estaciones posibles en que se pueden realizar viajes. La Tabla 3.1 muestra el número de estaciones por línea y los pares de estaciones dentro de cada línea. Considerando los antecedentes, se desprende que hay 8.970 pares de estaciones de la red que requieren de transbordo, correspondientes al 77,7% de los pares posibles.

Línea	Número de Estaciones	Pares O-D dentro de una misma línea
L1	27	702
L2	22	462
L4	23	506
L4A	6	30
L5	30	870
Total	108	2.570

Tabla 3.1: Pares O-D dentro de cada línea
(Fuente: Elaboración propia)

En la EOD 2011 se registraron 374.467 observaciones, de las cuales 135.316 corresponden a viajes con transbordos, los que equivalen al 36,1% de las observaciones. Estos viajes con transbordos se registraron en 5.039 pares O-D, correspondientes al 43,7% de los 11.540 pares O-D posibles, y correspondientes al 56,2% de los 8.970 pares O-D con transbordos posibles. La EOD 2011 registra que los usuarios llegan a ocupar hasta cuatro rutas posibles para un mismo par O-D. En la Tabla 3.2 se muestran la cantidad de pares O-D que registran una, dos, tres y cuatro rutas en la encuesta EOD 2011.

	Número de Pares O-D	%
Pares O-D con una ruta registrada	2.876	57%
Pares O-D con dos rutas registradas	1.716	34%
Pares O-D con tres rutas registradas	404	8%
Pares O-D con cuatro rutas registradas	43	1%
Total de pares O-D registrados	5.039	100%

Tabla 3.2: Clasificación de los pares O-D según el número de rutas registradas
(Fuente: Elaboración propia)

Para clasificar los pares O-D según el reparto en la elección de ruta, se considera en primer lugar expresar los repartos en términos porcentuales y usar la ruta con mayor porcentaje como criterio de clasificación. Originalmente se estableció que hay *reparto* cuando la ruta más usada no supera el 50% de las preferencias, que hay *predominancia de ruta* cuando dicha ruta está entre 50% y menos de 90% de las preferencias, y que hay *ausencia de reparto* cuando la ruta más usada registra preferencia igual o mayor al 90%. El problema de esta clasificación es que los porcentajes dependen del número de alternativas y del número de observaciones de forma tal que no siempre sintetizan la información de forma clara. Por ejemplo, repartos equitativos de 50% entre dos rutas de pares O-D pueden representar que en cada ruta se haya hecho un solo viaje, o bien, 47 viajes en cada ruta. Por otra parte, un reparto del 47% y 53% entre dos rutas puede indicar que se registraron 8 y 9 viajes, respectivamente, o bien 66 y 73 viajes. Sin embargo, a la luz del análisis estas rutas registran en realidad *reparto* cuando se las está clasificando en *predominancia de ruta*. Por estas razones, se utiliza otra medida de dispersión que captura tanto los repartos como los casos límite.

Dados los problemas expuestos anteriormente, se utiliza el Coeficiente de Gini (Glasser, 1962) para capturar la dispersión del par de estaciones del Metro O-D. Éste refleja el grado de dispersión de la elección entre R rutas registradas que ocuparon los usuarios. El cálculo del coeficiente γ_{O-D} para un par O-D se efectúa con las fórmulas (3.1) y (3.2):

$$\gamma_{O-D} = \frac{1}{2\mu_{O-D}R(R-1)} \sum_{i=1}^R \sum_{j=1}^R |V_i^{O-D} - V_j^{O-D}| \quad i \neq j \quad (3.1)$$

$$\mu_{O-D} = \frac{1}{R} \sum_{i=1}^R V_i^{O-D} \quad (3.2)$$

donde:

- R es el número de rutas registradas que los usuarios hicieron en el par O-D analizado;
- V_i^{O-D} y V_j^{O-D} son el número de viajes que se registraron en la ruta i y j respectivamente en el par O-D analizado;
- μ_{O-D} es el promedio de los viajes en las rutas registradas del par O-D analizado.

Para $R \geq 2$, el Coeficiente de Gini γ_{O-D} varía entre cero y uno, donde cero indica que todas las rutas registradas del par O-D poseen la misma cantidad de usuarios, y uno indica que una única ruta fue escogida por todos (o casi todos) los usuarios. Cuando $R = 1$, es decir, cuando solo se registra que se utilizó una ruta en el par O-D analizado, no hay reparto entre las posibles rutas para viajar y los usuarios escogieron sólo una ruta, por lo que el coeficiente γ_{O-D} se asume igual a uno.

Luego de clasificar los pares O-D según el número de rutas registradas, se clasifican los pares de estaciones según el grado de reparto que registran con el Coeficiente de Gini, con criterios de repartos diferenciados según la categoría de número de rutas a la que pertenezca cada par. En las Figuras 3.1-3.3 se grafican los pares O-D según el Coeficiente de Gini versus el porcentaje de viajes registrados en la ruta más usada en cada par. Se observa la tendencia en la que mientras más viajes registra una ruta posible en cada par, aumenta el coeficiente de Gini, por lo que disminuye la equidad en el reparto del uso de las rutas posibles de cada par de estaciones.

Para los pares O-D con dos rutas registradas, se observa que para los pares O-D con $\gamma_{O-D} < 0,45$ hay *reparto* en la elección de las rutas, y para los pares con $0,45 \leq \gamma_{O-D} \leq 0,99$ hay *predominancia de ruta*. Cuando las proporciones de elección entre las dos rutas son de 73% y 27%, comienza a haber *predominancia de ruta*. Hay 1.176 pares que registran *predominancia de ruta*, correspondientes al 68,5% de los pares O-D con dos rutas; y hay 540 pares que registran *reparto*, correspondientes al 31,5% de los 1.716 pares O-D con dos rutas.

Los pares O-D con tres y cuatro rutas registradas presentan alta dificultad para determinar criterios según el Coeficiente de Gini que distingan desde qué valor del coeficiente se empieza a registrar *predominancia de ruta* en vez de *reparto*, debido a que no es posible distinguir límites claros correlacionados entre el Coeficiente de Gini y el reparto de viajes de las rutas registradas para distinguir los casos de *reparto* de los de *predominancia de ruta*. Otro fenómeno observado es que en varios pares O-D se registra *predominancia de dos rutas* en desmedro de la(s) otra(s) ruta(s) registrada(s), lo que dificulta aún más distinguir desde qué valor del Gini se puede clasificar los pares O-D con *predominancia de ruta*, porque tampoco se observan límites definidos para distinguir desde qué valor del Gini hay *predominancia de dos rutas* ni desde qué valor del Gini hay *predominancia de una ruta*, es decir, no se observa un orden estricto de *predominancia* a medida que el Coeficiente de Gini aumenta. Para ilustrar este problema, al analizar el grupo de pares O-D con tres rutas registradas, se observa que hay un par O-D con $\gamma_{O-D} = 0,38$ con repartos de 7, 4 y 2 viajes (54%, 31% y 15%, respectivamente), por lo que se clasificaría en *reparto*; pero a continuación el siguiente par O-D con $\gamma_{O-D} = 0,39$ registra 22, 20 y 4 viajes (48%, 43% y 9%, respectivamente), por lo que se clasificaría en *predominancia de dos rutas*; y más adelante se observa un par O-D con $\gamma_{O-D} = 0,40$ con repartos de 2, 5 y 8 viajes (13%, 33% y 53%, respectivamente), lo que se clasificaría en *reparto*. Esto muestra que no se presenta relación de orden en el reparto de las rutas registradas según el Coeficiente de Gini.

A pesar de los problemas en la clasificación, cabe destacar que en todas las clasificaciones de pares O-D según el número de rutas registradas, el Coeficiente de Gini recoge satisfactoriamente los grados de equidad en los casos extremos, tanto de reparto equitativo de las rutas registradas como en el predominio de una ruta con reparto de viajes cercano al 99%, registrando el valor cero cuando hay reparto equitativo entre rutas y registrando valores de γ_{O-D} entre 0,95 y 0,99 cuando una ruta registra uso del 98% y del 99% frente a las otras alternativas registradas.

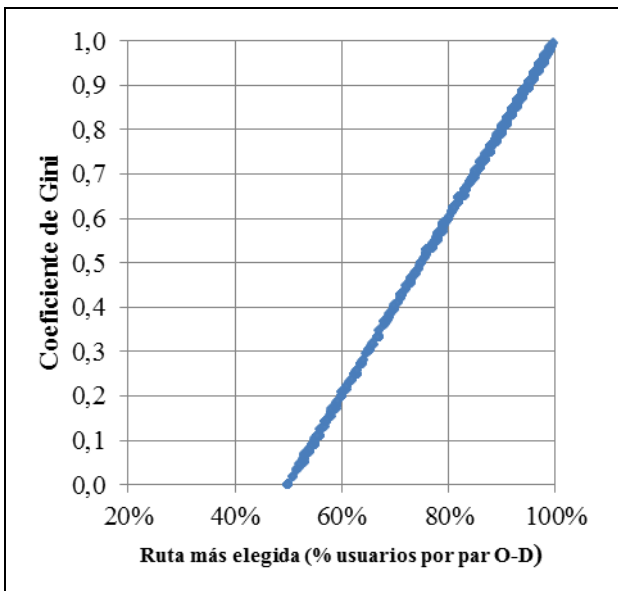


Figura 3.1: Gráfico del Coeficiente de Gini – Ruta más elegida para pares O-D con dos rutas
(Fuente: Elaboración propia)

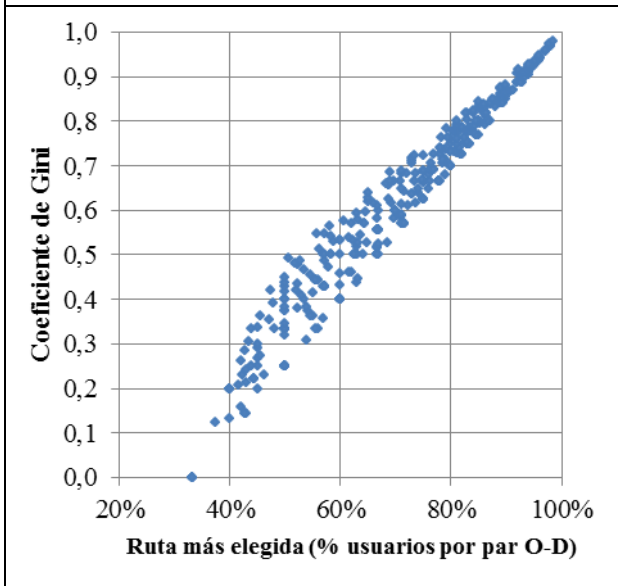


Figura 3.2: Gráfico del Coeficiente de Gini – Ruta más elegida para pares O-D con tres rutas
(Fuente: Elaboración propia)

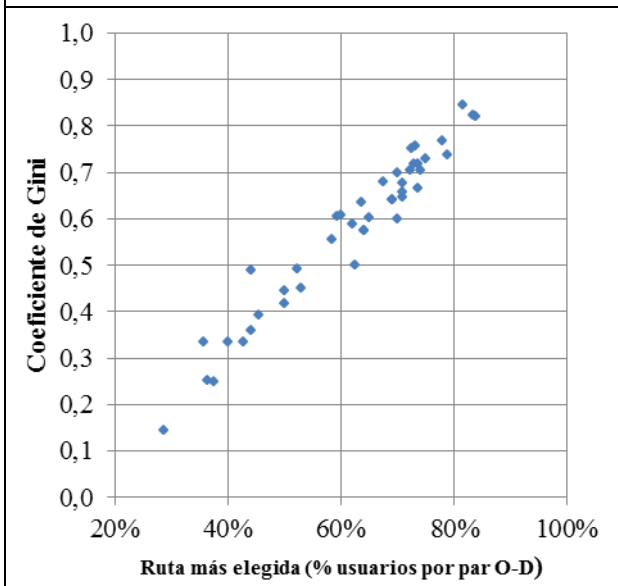


Figura 3.3: Gráfico del Coeficiente de Gini – Ruta más elegida para pares O-D con cuatro rutas
(Fuente: Elaboración propia)

3.2. Análisis de las rutas

Para desarrollar el modelo de elección de ruta, se deben identificar todas las rutas posibles y razonables que poseen cada uno de los 11.540 pares de estaciones, con el objetivo de conocer el tiempo de viaje y el número de transbordos de cada ruta. La información disponible para la calibración del modelo consiste en la base de datos de la encuesta EOD 2011 de Metro, que registra las rutas que los usuarios declararon realizar en su viaje en la red. Para caracterizar las rutas con atributos, se dispone de las variables de operación de Metro en junio de 2011, de las cuales se utilizan los tiempos de viaje entre estaciones en el periodo fuera de punta². Por ende, se debe tener consistencia en que las observaciones para la calibración correspondan a viajes en periodo fuera de punta (ver sección 4.1). Finalmente, también se tienen disponibles los tiempos de caminata en las estaciones de transbordo que la Universidad de Chile midió entre abril y mayo de 2013, y las coordenadas UTM³ de las estaciones de Metro.

El primer paso es establecer todas las rutas posibles para cada par O-D de estaciones de Metro. Para ello, se realiza un procedimiento llamado k-Dijkstra que ha sido implementado por la División de Transporte de la Universidad de Chile, el cual recoge todas las k rutas posibles de menor tiempo de un par O-D, estableciendo la secuencia de estaciones que posee cada ruta y los tiempos de viaje entre estaciones consecutivas. Los datos ingresados al procedimiento son los tiempos de viaje entre estaciones, y los tiempos de caminata y de espera en cada estación de transbordo multiplicados por tres⁴, debido a que la valoración del tiempo de caminata y de espera es mayor que la del tiempo de viaje. Para el presente trabajo, se obtienen las mejores cinco rutas posibles para cada par O-D. Dado que el algoritmo arroja rutas no razonables entre los pares O-D, es necesario filtrar las rutas según una serie de criterios y supuestos que distingan entre rutas razonables y no razonables.

Supuestos y criterios:

- Todas las rutas registradas en la encuesta EOD 2011 de Metro son razonables, dado que son rutas que los usuarios declararon realizar.
- Los usuarios que viajan entre pares O-D de una misma línea no realizan transbordos.
- El número máximo de transbordos que realiza un usuario es tres, dado que en la EOD 2011 se registra que los usuarios declararon realizar hasta tres transbordos.
- Los usuarios no realizan viajes en sentido contrario con tal de lograr abordar un tren en alguna estación aguas arriba.
- Los usuarios no escogen rutas con ciclos o loops.

² Si se hubiera analizado algún periodo punta (mañana o tarde), se tendría que considerar los servicios expresos de las Líneas 2, 4 y 5, donde hay cambios en los arcos de la red, lo que por un lado aumenta de manera considerable la velocidad de operación. Por otro lado, la mayor afluencia de pasajeros en los periodos punta reduce la velocidad de operación. Estos dos efectos influyen directamente en el tiempo de viaje en la red, lo que implicaría realizar modelos de elección de ruta más sofisticados, como por ejemplo un modelo dinámico, o bien incorporando la velocidad de operación como variable de decisión, lo que se escapa al objetivo del presente desarrollo.

³ Sistema de Coordenadas Universal Transversal de Mercator.

⁴ Hay modelos de elección de ruta en Transporte Público calibrados por Fernández y De Cea (2003) en que los parámetros de tiempo de viaje, tiempo de caminata y espera son respectivamente 1, 1,93 y 3,63. El uso del valor tres en el presente trabajo se debe más bien a que corresponde a información proveniente de un algoritmo de búsqueda de rutas como lo es el k-Dijkstra, donde está implementado que los parámetros de tiempo de caminata y de espera en los transbordos son iguales a tres. Sin embargo, es posible obtener parámetros propios para este desarrollo separando los tiempos de viaje en las líneas de los de caminata y de espera.

- Se incluyen rutas con desvíos importantes con respecto a las alternativas de rutas más directas⁵ dentro de un par O-D para que el modelo sea capaz de predecir la probabilidad de elección de este tipo de rutas.
- Se incluyen rutas con diferencias de tiempo de viaje considerables con respecto a las alternativas de menor tiempo de viaje dentro de un par O-D para que el modelo sea capaz de predecir la probabilidad de elección de este tipo de rutas.
- Si la estación de origen corresponde a una estación de transbordo, el usuario ingresa directamente a la estación de transferencia de la línea que le sirve para realizar su viaje.
- Si la estación de destino corresponde a una estación de transbordo, el usuario egresa de la estación por la salida correspondiente a la estación de transbordo de la línea por la cual llegó al destino.

La Tabla 3.3 muestra el número de pares de estaciones de Metro que poseen una, dos, tres, cuatro y cinco rutas razonables. En total se registran 21.359 rutas razonables. Por un lado, en el número de pares O-D que poseen solo una ruta están incluidos tanto los pares pertenecientes a una misma línea como aquellos entre distintas líneas. Según la Tabla 3.1, hay 2.570 pares O-D de una misma línea, por lo que hay 2.694 pares de estaciones de distintas líneas que solo poseen una ruta razonable, correspondientes al 51,2% de los pares con una ruta razonable. Por otra parte, hay 6.276 pares O-D que poseen dos o más rutas razonables, correspondientes al 54,4% de todos los pares O-D posibles. De la Tabla 3.3 se desprende que hay 16.095 rutas (75,4% de las 21.359 rutas razonables) que están enfrentadas a una o más alternativas, es decir, donde se presenta un proceso de elección de ruta.

Número de Rutas Razonables	Número de Pares O-D	% O-D	Número de Rutas	% Rutas
1	5.264	45,6%	5.264	24,6%
2	3.416	29,6%	6.832	32,0%
3	2.186	18,9%	6.558	30,7%
4	665	5,8%	2.660	12,5%
5	9	0,1%	45	0,2%
Total	11.540	100%	21.359	100%

Tabla 3.3: Número de rutas razonables de Metro
(Fuente: Elaboración propia)

Una vez obtenidas las rutas razonables con sus respectivas secuencias de estaciones, se caracterizan con una serie de atributos o variables descriptivas. Para este estudio, se consideran las variables tradicionales de tiempo de viaje y el número de transbordos, además de la variable de costo angular para medir cuán directa es una ruta. Estas variables son sencillas de obtener y no requieren de supuestos fuertes sobre la estructura de la red, ni de las estaciones ni de los trenes, sobre todo cuando se modela en modalidad predictiva. A continuación se explican con más detalle en qué consisten y cómo están constituidas las variables incluidas en este modelo:

- Tiempo de viaje⁶ (tv): El tiempo de viaje de una ruta dentro de una misma línea es la suma de los tiempos de viaje entre estaciones consecutivas de la secuencia. El tiempo de

⁵ Una ruta es directa cuando es en línea recta entre el origen y el destino. Una ruta indirecta posee desvíos importantes a lo largo de la ruta que tienden a alejarse del destino, lo que hace recorrer distancias mucho mayores que la euclidiana.

viaje de una ruta con transbordos, además de sumar los tiempos de viaje entre estaciones consecutivas de la secuencia, incluye el tiempo de caminata entre las líneas de las estaciones de transbordo y el tiempo de espera en la estación de transbordo (que es la mitad del intervalo de tiempo entre trenes), ambos multiplicados por tres. No se consideran en este cálculo el tiempo de acceso a la estación de origen ni el tiempo de egreso de la estación de destino. Cabe destacar que los tiempos de viaje corresponden al periodo de fuera de punta y están expresados en minutos.

- b) Diferencia de tiempo de viaje ($d = tv - tmin$): Para cada par O-D, se busca la ruta con el menor tiempo de viaje ($tmin$). Luego, para cada alternativa de ruta del par O-D (incluida la de menor tiempo) se calcula la diferencia entre el tiempo de viaje de la alternativa en cuestión y la alternativa de menor tiempo de viaje. Esta diferencia es igual a cero cuando la alternativa es la de menor tiempo y el máximo valor de dicha diferencia se obtiene con la ruta de mayor tiempo de viaje. Esta variable, medida en minutos, se pretende incluir dentro de la penalización o cutoff de la función de utilidad del CMNL para modelar el fenómeno de comparación entre las otras alternativas disponibles para un par O-D.
- c) Número de transbordos (tr): Para cada ruta, se contabiliza el número de transbordos que un usuario debe realizar para viajar por cada ruta. Se cuentan los transbordos que implican descender de un tren para posteriormente subir a otro en la estación de transbordo. Siguiendo los criterios de rutas razonables, no se considera que se realizan transbordos cuando las estaciones de origen y/o de destino son estaciones de transbordo. Tampoco se hacen transbordos si los pares O-D están dentro de una misma línea. Esta variable es discreta, y sus valores varían entre cero y tres.
- d) Costo angular ($cang$): Esta variable es un indicador de penalidad que mide cuán indirecta es una alternativa de ruta de un par O-D (ver discusión en Raveau et al., 2011). Para el cálculo de esta variable, primero se obtienen los segmentos de ruta constituidos por estaciones consecutivas de la secuencia de estaciones que se obtienen con el algoritmo de k-Dijkstra. Después, se calcula la distancia euclidiana de cada segmento de ruta D_{seg} con los datos de las coordenadas UTM de las estaciones de Metro, que se transforman en kilómetros. Se calcula el costo angular del segmento de ruta $cang_{seg}$ con la fórmula (3.3), en la cual θ_{seg} es el ángulo formado entre el rayo que une la estación de origen del segmento y la estación de destino de la ruta (ver Figura 3.4), y el rayo que une la estación de origen del segmento y la estación consecutiva.

$$cang_{seg} = D_{seg} \text{sen} \left(\frac{\theta_{seg}}{2} \right) \quad (3.3)$$

La penalidad del segmento es mínima cuando θ_{seg} es igual a 0° (cuando el segmento llega directo a la estación de destino), y la penalidad es máxima cuando θ_{seg} es igual a 180° (cuando el segmento es en dirección opuesta). El costo angular de una ruta es la suma de los costos angulares de los segmentos de ruta. La Figura 3.4 ilustra el cálculo del costo angular para la ruta del par O-D que posee la secuencia de estaciones Origen \rightarrow T1 \rightarrow T2 \rightarrow Destino y la expresión (3.4) muestra el cálculo para el ejemplo. Cabe destacar que el costo angular es mayor o igual a cero.

⁶ En realidad, se debería hablar de tiempo de viaje generalizado, porque es la suma de los tiempos de viaje en las líneas, y de los tiempos de caminata y espera en estaciones de transbordo ponderados por tres.

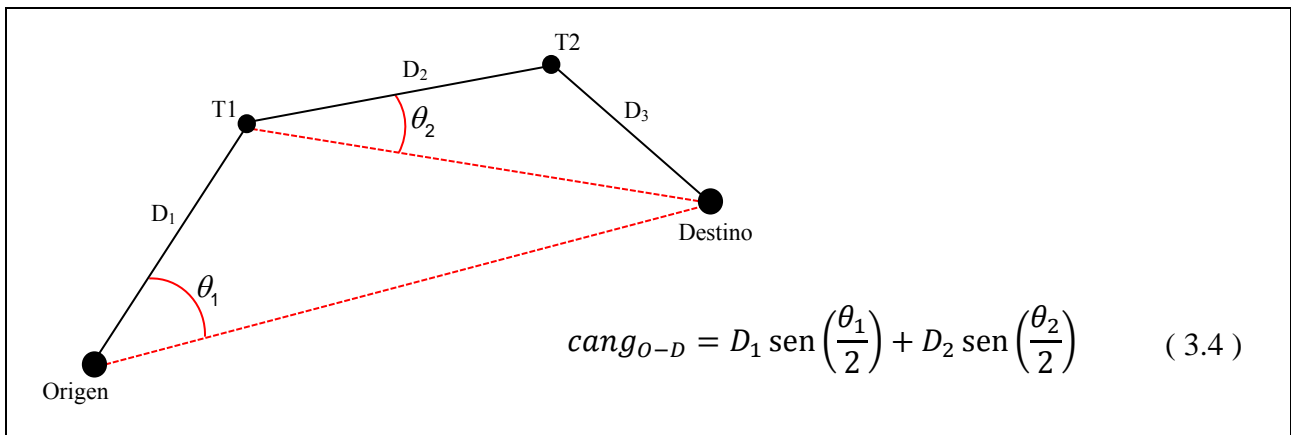


Figura 3.4: Cálculo del costo angular
(Fuente: Raveau et al., 2011)

A partir de los datos disponibles para calcular las variables descriptivas de las 21.359 rutas, se construye una base de datos que caracteriza a todas las rutas razonables de la red. En lo que resta de este capítulo, se muestra la estadística descriptiva de los pares O-D que poseen dos o más rutas razonables, porque en esos pares ocurre un proceso de elección discreta que interesa modelar en este trabajo. Por la misma razón, la información que se utiliza para la calibración del modelo de elección corresponde a la de los pares con dos o más rutas disponibles (razonables).

La Tabla 3.4 muestra la estadística descriptiva de las variables de tiempo de viaje, costo angular y número de transbordos de las 16.095 rutas en que existe un proceso de elección. Se observa que el tiempo de viaje promedio en estas rutas de 43,11 minutos, en que están involucrados el tiempo de transbordo entre estaciones. El tiempo de viaje máximo registrado en la base de datos es de 90,37 minutos, correspondiente al par Plaza de Maipú-Plaza de Puente Alto por la ruta con transbordos en San Pablo, Baquedano y Vicente Valdés. El costo angular posee un valor promedio relativamente bajo de 5,938 comparado con el valor máximo registrado, aunque es difícil establecer a través de los números desde qué valor la ruta es indirecta o no, pero se espera que sea una variable relevante en los pares de estaciones localizados relativamente cerca entre sí, donde predominan rutas indirectas. Llama la atención el valor promedio relativamente alto de los transbordos de 1,7, ya que los transbordos provocan molestia en la experiencia de viaje de los usuarios, por lo que se espera que el modelo prediga que mientras más transbordos requiera una ruta, menos usuarios la ocuparán.

Tiempo de viaje (min)		Costo Angular		Transbordos	
Media	43,11	Media	5,938	Media	1,7
Mediana	42,93	Mediana	5,161	Mediana	2
Moda	47,93	Moda	4,581	Moda	2
Desviación estándar	14,57	Desviación estándar	3,673	Desviación estándar	0,67
Rango	88,43	Rango	26,859	Rango	3
Mínimo	1,93	Mínimo	0,005	Mínimo	0
Máximo	90,37	Máximo	26,864	Máximo	3

Tabla 3.4: Estadística descriptiva de las rutas en que existe un proceso de elección
(Fuente: Elaboración propia)

Cabe comentar que el número mínimo de transbordos es cero, lo que resulta extraño debido a que estos pares O-D son entre distintas líneas, por lo que se espera que el mínimo de transbordos sea uno. Esto se explica por la doble codificación de las estaciones de pares O-D con

origen y/o destino en estaciones de transbordo (recordar que se cuenta la estación de transbordo como dos estaciones), lo que involucra el supuesto adoptado aquí sobre los ingresos y egresos de las estaciones de transbordo, lo que genera que pueda existir más de una alternativa para realizar viajes en dichos pares. Por ejemplo, el par Las Rejas – Baquedano L5 posee dos rutas razonables: la primera, que es el viaje directo por Línea 1 hasta la estación Baquedano L1, en donde se asume que egresa desde Baquedano L1 y por lo tanto no transborda a Baquedano L5 para después egresar, por lo que la variable de transbordos de esta ruta es cero; y la segunda, que incluye el transbordo a la Línea 5 en la estación San Pablo y por donde llega directo a estación Baquedano L5 para directamente egresar desde ahí, por lo que la variable de transbordos de esta ruta es uno.

La Tabla 3.5 muestra las correlaciones entre los atributos de las rutas en que existe un proceso de elección. Se observa que la correlación entre el tiempo de viaje y el costo angular es alta, lo que indica que las rutas más indirectas se realizan en más tiempo. La correlación entre el tiempo de viaje y el número de transbordos no resulta ser tan alta, debido a que la red registra muchas rutas con tramos largos en una misma línea (ver Figura 3.5), lo que provoca que la incidencia del tiempo de transbordo no sea tan relevante en los tiempos de viaje, en otras palabras, no hay una gran densidad de líneas de Metro en la ciudad de Santiago, lo que hace que el número de transbordos necesarios para un viaje no sea tan alto. La correlación entre el número de transbordos y el costo angular es baja, lo que se explica porque en la mayoría de los pares O-D se registran rutas bastante directas con un bajo número de transbordos necesarios, lo que atenúa el efecto de los pares de estaciones más cercanos entre sí (sobre todo en la zona del centro de Santiago) en donde sí predominan las rutas indirectas donde hay tendencia de alejarse del destino en el transcurso del recorrido de dichas rutas.

CORRELACIONES	Tiempo de viaje (min)	Costo Angular	Transbordos
Tiempo de viaje (min)	1		
Costo Angular	0,831	1	
Transbordos	0,527	0,324	1

Tabla 3.5: Correlaciones entre los atributos del modelo de elección de ruta
(Fuente: Elaboración propia)

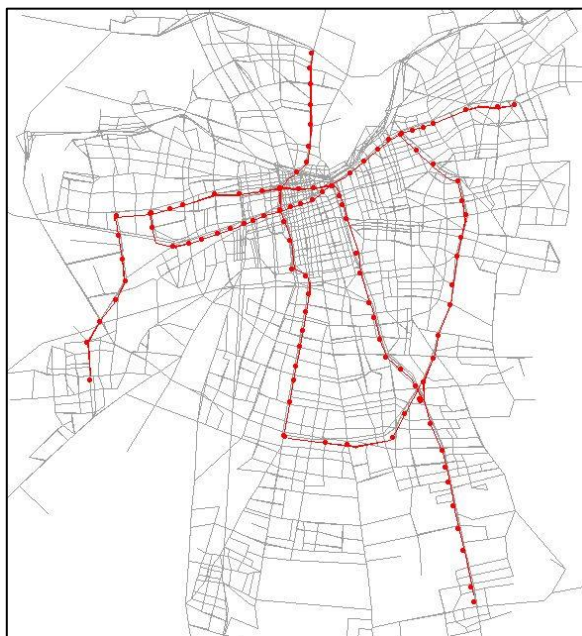


Figura 3.5: Estructura de la red de Metro
(Fuente: Metro S.A.)

La Tabla 3.6 muestra las estadísticas de las diferencias de tiempo de viaje entre las alternativas de mayor y menor tiempo de viaje de los 6.276 pares O-D en que existe un proceso de elección de ruta. La media de diferencias de tiempo es de 14,37 minutos y en la Figura 3.6 se observa que en 4.797 pares O-D (76,4% de los 6.276 pares) se registran diferencias menores o iguales a 20 minutos entre las alternativas extremas, lo que sugiere que los usuarios llegarían a elegir rutas con diferencias entre 15 a 20 minutos aproximadamente con respecto a la ruta de menor tiempo. Se espera que el modelo prediga reparto entre las alternativas de pares O-D que posean diferencias de tiempo poco significativas (comportamiento compensatorio), mientras que en los pares O-D que registran diferencias de tiempo significativas entre las alternativas se espera que el modelo descarte las alternativas con mayor tiempo de viaje.

Diferencia de tiempo de viaje (min)	
Media	14,37
Mediana	11,72
Moda	22,03
Desviación estándar	10,07
Rango	63,17
Mínimo	0,02
Máximo	63,18
Cuenta	6.276

Tabla 3.6: Estadística de las diferencias de tiempo entre rutas extremas
(Fuente: Elaboración propia)

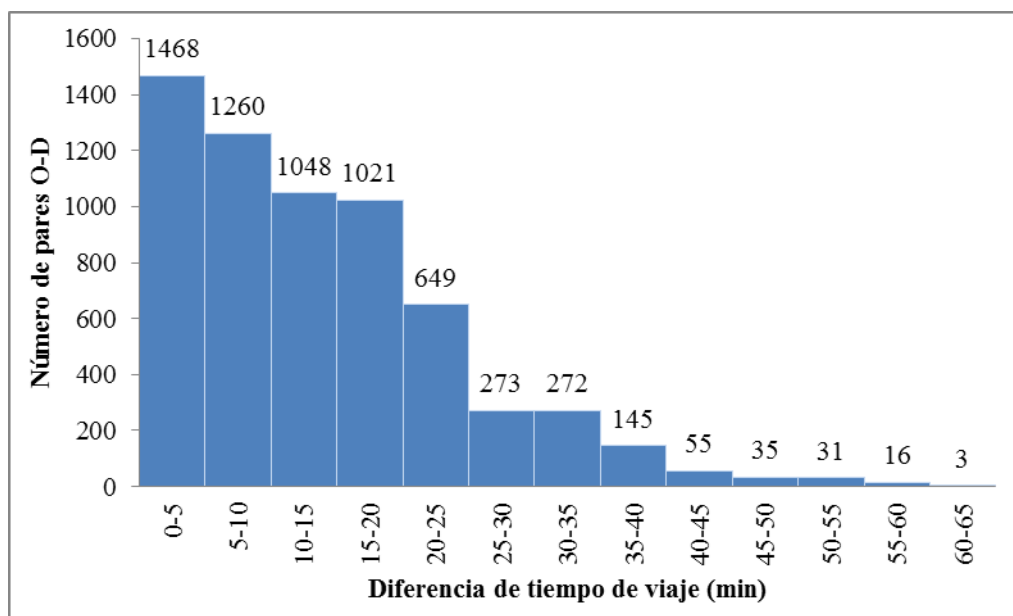


Figura 3.6: Diferencias de tiempo de viaje entre rutas extremas
(Fuente: Elaboración propia)

Con la información disponible de los viajes registrados en la EOD 2011 y con los atributos de las rutas descritos en este capítulo, se efectúa la calibración del modelo de elección de ruta que emplea el Logit Multinomial Restringido CMNL, descrito en el Capítulo 2. El siguiente capítulo profundiza en las especificaciones de las funciones de utilidad a calibrar y describe las metodologías de calibración empleadas para el CMNL con el objetivo de comparar los modelos y discutir la validez de los resultados.

Capítulo 4

Aplicación del Modelo

En este capítulo se muestra la calibración del modelo Logit Multinomial Restringido CMNL para modelar la elección de ruta con umbrales en Metro. Para ello, se emplean las variables descriptivas descritas en el capítulo anterior y la información de los viajes registrados por la EOD 2011 de Metro. La primera sección describe las especificaciones del CMNL a calibrar y las metodologías de calibración utilizadas. Además de los modelos CMNL, se calibran modelos Logit Multinomial MNL con el propósito de contrastar ambos modelos y visualizar los aportes de los modelos CMNL al problema de elección de ruta desarrollado en este trabajo. En la segunda sección se muestran los resultados obtenidos con la calibración de modelos MNL y CMNL, se discute la validez de los valores de los parámetros obtenidos y se reportan los test estadísticos de los modelos. Finalmente, en la última sección se discute el rango de validez de los modelos obtenidos y se comparan entre sí, para finalmente evaluar si el CMNL modela satisfactoriamente la elección de ruta en Metro.

4.1. Calibración del modelo

Con el objetivo de reproducir el comportamiento de los usuarios de Metro, la calibración del modelo CMNL requiere de los viajes registrados en la Encuesta Origen-Destino de Metro de 2011. Recordando la información presentada en el capítulo anterior, se registran 135.316 observaciones que presentan viajes con transbordos, los que se realizaron en cualquier día tipo (día laboral, sábado y festivo) y en todos los periodos (punta mañana, punta tarde y fuera de punta). Dado que se utiliza la información de los tiempos de viaje en periodo fuera de punta, se emplean las observaciones registradas en este mismo periodo en día laboral para ser consistente, evitando así posibles sesgos en las estimaciones de los parámetros del modelo.

En la base de la EOD 2011, se buscan los viajes con transbordos registrados en día laboral en el periodo fuera de punta. La EOD 2011 registra 38.957 viajes con transbordos en periodo fuera de punta (28,8% de 135.316) en los horarios 10:00-11:00, 13:00-14:00, 16:00-17:00 y 17:00-18:00. Se revisa y modifica la base de datos para ser consistente con los criterios de rutas razonables establecidos en el capítulo anterior. Debido a la metodología de procesamiento de datos y a que la EOD registra rutas que el modelo k-Dijkstra no recoge, solo fueron posibles de rescatar y caracterizar con atributos a 22.556 viajes con transbordos en fuera de punta (57,9% de 38.957), lo que de todas formas resulta una cantidad de observaciones con bastante riqueza de información, permitiendo una buena reproducción del comportamiento de los usuarios. Cada una de estas observaciones está descrita por la alternativa de ruta escogida y por los atributos (tiempo de viaje, tiempo de viaje de la ruta mínima, costo angular y número de transbordos) de todas las alternativas disponibles para su viaje en el par O-D, incluida la alternativa escogida.

El siguiente paso es reproducir el comportamiento de los usuarios a través de la estimación de parámetros de la funciones de utilidad de los modelos CMNL y MNL. Para ello, se emplea el método de máxima verosimilitud (ver discusión en Castro, 2009), que corresponde a un método estadístico usado para calcular los parámetros que mejor se ajustan a los datos

disponibles. Así, se define la función de verosimilitud $L(\beta)$ como la función de densidad conjunta de un conjunto de variables aleatorias (en este caso, las utilidades):

$$L(\beta) = \prod_n^N \prod_{i \in C_n} P_{ni}^{\delta_{ni}} \quad (4.1)$$

donde:

- β es el vector de parámetros de la función de utilidad a calibrar;
- N es el número de observaciones empeladas en la calibración;
- C_n es el conjunto de rutas disponibles que posee el individuo asociado a la observación n ;
- i es la alternativa de ruta escogida por el individuo n ;
- P_{ni} es la probabilidad de que el individuo n escoja la ruta i ;
- δ_{ni} es una variable binaria que vale uno si el individuo n escoge la ruta i y cero si no.

La técnica de máxima verosimilitud consiste en encontrar los parámetros definidos en el vector β que maximicen la probabilidad de obtener los valores que se dieron en la muestra, esto es, se maximiza la función (4.1) derivándola con respecto a cada parámetro, obteniendo las condiciones de primer orden (CPO) para despejar dichos parámetros. Para facilitar los cálculos de las CPO, se calcula el logaritmo de $L(\beta)$, con lo que se obtiene la función de log-verosimilitud $l(\beta)$ definida por la expresión (4.2), de la cual se obtienen las CPO. Castro (2009) calcula e interpreta las CPO para el modelo CMNL. Más adelante se hace referencia a las CPO del CMNL que son más relevantes para el problema abordado en este trabajo.

$$l(\beta) = \sum_n^N \sum_{i \in C_n} \delta_{ni} \ln(P_{ni}) \quad (4.2)$$

El siguiente paso es establecer las especificaciones de las funciones de utilidad para los modelos MNL y CMNL de elección de ruta, las cuales están asociadas a las probabilidades de elección P_{ni} . La idea es probar con distintas especificaciones a través de distintas combinaciones de variables explicativas que mejor reproduzcan las elecciones de ruta en Metro. Se obtienen distintas expresiones para las probabilidades dependiendo de si se trabaja con modelos MNL o CMNL.

Para la función de utilidad del MNL, se opta por una componente determinística de la función de utilidad (ver expresión (2.2) de la sección 2.2) que es lineal en parámetros, es decir,

$$V_{ni} = \sum_k^K \beta_k x_{ik} \quad (4.3)$$

donde x_{ik} es el valor del atributo k para la alternativa i y β_k es el parámetro del atributo k a calibrar. Así, en el modelo MNL la probabilidad de que el individuo n elija la alternativa i es:

$$P_{ni} = \frac{\exp(\mu V_{ni})}{\sum_{j \in C_n} \exp(\mu V_{nj})} \quad (4.4)$$

donde μ es el parámetro de escala de la distribución Gumbel. Como μ no es identificable, se debe estimar el parámetro de cada atributo y el parámetro de escala en conjunto, es decir, $\theta_k = \mu\beta_k$. Entonces, para el modelo MNL empleado aquí, la función de utilidad a calibrar está dada por la expresión (4.5), en donde se deben estimar los parámetros θ_k de los K atributos. Finalmente, la expresión de las probabilidades de elección del MNL que se incorpora a la log-verosimilitud está dada por la expresión (4.6).

$$\tilde{V}_{ni} = \mu V_{ni} = \sum_k^K \mu\beta_k x_{ik} = \sum_k^K \theta_k x_{ik} \quad (4.5)$$

$$P_{ni} = \frac{\exp(\tilde{V}_{ni})}{\sum_{j \in C_n} \exp(\tilde{V}_{nj})} \quad (4.6)$$

En el caso de la función de utilidad del CMNL (ver sección 2.4) que está dada por la expresión (2.5), se emplean las metodologías propuestas por Castro (2009). Se define la componente determinística de la utilidad del CMNL como \bar{U}_{ni} , que es la suma entre la componente compensatoria V_{ni} y la función de penalización o cutoff $\ln(\varphi_{ni})$ multiplicada por el inverso de μ :

$$\bar{U}_{ni} = V_{ni} + \frac{1}{\mu} \ln(\varphi_{ni}) \quad (4.7)$$

donde φ_{ni} es el cutoff definido por la composición de penalizaciones superiores e inferiores (expresiones 2.7, 2.8 y 2.9) y V_{ni} tiene una especificación lineal en parámetros dada por (4.3). Para este modelo de elección de ruta en Metro, sólo se emplean penalizaciones superiores (ver expresión 2.9 para φ_{nik}^U), y se asume que los umbrales superiores b_{nk} son universales para toda la muestra, es decir, $b_{nk} = b_k, \forall n$. De esta forma, se obtiene la expresión (4.8) para \bar{U}_{ni} :

$$\bar{U}_{ni} = \sum_k^K \beta_k x_{ik} + \frac{1}{\mu} \ln \left[\prod_k^K \frac{1}{1 + \exp(\omega_k(x_{ik} - b_k + \rho_k))} \right] \quad (4.8)$$

Desarrollando el logaritmo del producto de las penalizaciones o cutoffs superiores de los K atributos, se obtiene que:

$$\bar{U}_{ni} = \sum_k^K \beta_k x_{ik} - \frac{1}{\mu} \sum_k^K \ln[1 + \exp(\omega_k(x_{ik} - b_k + \rho_k))] \quad (4.9)$$

Así, en el modelo CMNL empleado en este trabajo la probabilidad de que el individuo n elija la alternativa i es:

$$P_{ni} = \frac{\exp(\mu\bar{U}_{ni})}{\sum_{j \in C_n} \exp(\mu\bar{U}_{nj})} \quad (4.10)$$

donde μ es el parámetro de escala de la distribución Gumbel. Nuevamente, como μ no es identificable, se debe estimar el parámetro de cada atributo de la componente compensatoria y el parámetro de escala en conjunto, es decir, $\theta_k = \mu\beta_k$. Ahora bien, dado que el cutoff está multiplicado por el inverso de μ , los parámetros del cutoff que se calibran son puros, sin factores

adicionales. Entonces, para el modelo CMNL empleado aquí, la función de utilidad a calibrar está dada por la expresión (4.11), donde se deben estimar los parámetros de la componente compensatoria θ_k y los parámetros de los cutoffs ω_k , b_k y ρ_k de los K atributos. Finalmente, la expresión de las probabilidades de elección del CMNL que se incorpora a la log-verosimilitud está dada por la expresión (4.12).

$$\bar{V}_{ni} = \mu \bar{U}_{ni} = \sum_k^K \theta_k x_{ik} - \sum_k^K \ln[1 + \exp(\omega_k(x_{ik} - b_k + \rho_k))] \quad (4.11)$$

$$P_{ni} = \frac{\exp(\bar{V}_{ni})}{\sum_{j \in C_n} \exp(\bar{V}_{nj})} \quad (4.12)$$

Cabe destacar que ninguna de las especificaciones anteriores en la componente compensatoria posee constante modal⁷, ya que se trata de la elección de ruta en un mismo modo de transporte y además se asume que los usuarios escogen rutas por los atributos que registren cada una de ellas y no por sus características propias, es decir, son alternativas de rutas no rotuladas⁸.

Al calcular las CPO del modelo CMNL, Castro (2009) muestra los problemas de identificabilidad de los parámetros a calibrar. Los problemas más relevantes para el modelo de elección de ruta de este trabajo son los siguientes:

- Identificabilidad de θ_k y ω_k del atributo k : cuando el atributo k está presente tanto en la componente compensatoria como en el cutoff, y dado que en el cutoff superior de k se tiene que si x_{ik} tiende a b_k , entonces $\exp(\omega_k(x_{ik} - b_k + \rho_k)) \gg 1$, por lo que el cutoff $-\ln\{1 + \exp[\omega_k(x_{ik} - b_k + \rho_k)]\}$ adopta la forma lineal $-\omega_k(x_{ik} - b_k + \rho_k)$, por lo que x_{ik} estará multiplicado por la diferencia de parámetros $(\theta_k - \omega_k)$, provocando un problema de correlación potencial entre θ_k y ω_k . Esto muestra que es más conveniente que el atributo k se encuentre en la componente compensatoria o bien en el cutoff, lo que depende de dónde hace más sentido conceptualmente especificar cada variable.
- Identificabilidad de b_k y ρ_k del atributo k : al derivar la log-verosimilitud con respecto a b_k y ρ_k e igualar a cero, se obtiene que la CPO de ρ_k es combinación lineal de la CPO de b_k , lo que indica que la información para estimar dichos parámetros es la misma, por lo que nunca va a ser posible estimar ambos parámetros por separado.

⁷ Cabe recordar que las constantes modales recogen todo lo no observado por el modelador, incluidas las características intrínsecas de una alternativa que la distinguen de las demás opciones disponibles.

⁸ Este supuesto es cuestionable, ya que en Metro hay rutas que están caracterizadas por atributos intrínsecos tales que permiten distinguir entre una ruta y otra. Por ejemplo, puede ser que los usuarios tiendan a escoger rutas que pasen por estaciones de la Línea 4 debido a que emplea trenes más espaciosos y a que está provista de una tecnología más reciente y desarrollada que mejora la comodidad del viaje en contraste con otras líneas más antiguas. En ese caso, además de las constantes modales, otra forma de especificación sería introducir variables dummy que indiquen si una ruta posee o no dichas características.

El primer problema de identificabilidad entre θ_k y ω_k ocurre dependiendo de la especificación de la componente compensatoria, de los datos usados y de los atributos penalizados. En cambio, el segundo problema de identificabilidad entre b_k y ρ_k ocurre siempre, lo que implica que es necesario fijar exógenamente uno de los dos parámetros o bien calibrarlos en conjunto ($-b_k + \rho_k$). Debido a este inconveniente, Castro propone varios métodos de calibración para el CMNL, de los cuales dos de ellos se emplean con los datos empleados en este trabajo:

- a) Calibración del CMNL con cota exógena: cuando el umbral b_k para el atributo k es conocido o es impuesto por el modelador, se puede calcular la proporción de la muestra η_k que supera el valor de dicho atributo con el registro en la EOD de las rutas escogidas por los usuarios encuestados. Para reducir el trabajo computacional de la calibración, el parámetro ρ_k está relacionado con la proporción η_k a través de la expresión (2.10), por lo que se reemplaza en la expresión (4.11), obteniendo la función de utilidad (4.13). Así, esta metodología solo requiere calibrar θ_k y ω_k .

$$\bar{V}_{ni} = \sum_k^K \theta_k x_{ik} - \sum_k^K \ln \left\{ 1 + \left(\frac{1 - \eta_k}{\eta_k} \right) \exp[\omega_k(x_{ik} - b_k)] \right\} \quad (4.13)$$

La literatura dice que empíricamente se ha establecido que una condición necesaria para obtener parámetros para la metodología con cota exógena es que la proporción de la muestra que viola la restricción debe ser muy baja, es decir, la mayor parte de la población debe respetar la restricción (elegir alternativas cuyos atributos penalizados se encuentran bajo el umbral) y por lo tanto los atributos de la alternativa elegida deben estar dentro del espacio compensatorio correspondiente. En otras palabras, empíricamente se ha visto que valores altos de η_k dificultan recuperar los parámetros. Para este trabajo, se calibraron modelos con cotas exógenas para el cutoff de la diferencia de tiempo de viaje d sin llegar a buenos resultados, ya que para identificar los parámetros asociados a d la cota debe ser a lo menos de $b_d = 35$ minutos, lo que es poco realista porque asume un comportamiento compensatorio en los pares O-D con diferencias de tiempo entre las alternativas extremas que son menores a esa cota, atenuando posibles descartes de rutas indirectas y/o con muchos transbordos. A pesar de que se calibraron modelos con cutoff del número de transbordos con cota $b_{tr} = 2$ que sí reportaron buenos resultados, en todos los modelos con cota exógena se registra un ajuste pobre de los datos frente a otras especificaciones.

- b) Calibración del CMNL con cota endógena: debido al problema de identificabilidad de b_k y ρ_k , Castro propone calibrar ambos valores en conjunto. La ventaja de esta metodología es que no es necesario incorporar ninguna cota en forma arbitraria (como lo propone la metodología de calibración anterior), por lo que los parámetros obtenidos son los que maximicen el ajuste según el comportamiento de los usuarios. Así, se obtiene la función de utilidad (4.14) en que se calibran los parámetros θ_k , ω_k y $\omega_k(\rho_k - b_k)$. Se calibra este último producto como un único parámetro en vez de calibrar dos parámetros que están multiplicados entre sí para facilitar el trabajo computacional.

$$\bar{V}_{ni} = \sum_k^K \theta_k x_{ik} - \sum_k^K \ln \{ 1 + \exp[\omega_k x_{ik} + \omega_k(\rho_k - b_k)] \} \quad (4.14)$$

Una vez establecido el marco teórico de la calibración y las especificaciones de las funciones de utilidad, la siguiente etapa consiste en obtener los parámetros de las especificaciones propuestas para los modelos MNL y CMNL. Para ello, se emplea el método de máxima verosimilitud descrito anteriormente con la metodología de calibración con cota endógena. El cálculo de los parámetros se realiza mediante el programa computacional GAUSS 6.0 y su herramienta MAXLIK. La siguiente sección, junto con las especificaciones y los resultados de los parámetros de los modelos propuestos, describe las herramientas estadísticas que permiten evaluar la validez de los modelos.

4.2. Resultados

Las funciones de utilidad que se calibran para los modelos de elección de ruta en Metro empleados en este Trabajo de Título son las expresiones (4.5) para el MNL y (4.14) para el CMNL. Éstas son expresiones generales en las que todos los K atributos aparecen tanto en la componente compensatoria como en el cutoff. La discusión anterior sobre la identificabilidad de los parámetros y las metodologías de calibración muestra que es necesario calibrar varios modelos con distintas cantidades de variables y con distintas interpretaciones de los atributos, ya sea como factores penalizadores de las utilidades o como factores compensatorios en la toma de decisiones. Esta sección muestra las calibraciones de las especificaciones con los mejores resultados según los criterios estadísticos y conceptuales que se detallan a continuación.

Para probar la hipótesis de este trabajo con respecto al descarte de rutas por las diferencias de tiempo de viaje, se calibran modelos CMNL con cutoffs de dicha variable. Asimismo, se explora si los cutoffs de transbordos y de costo angular explican la elección de ruta. Dado que existe evidencia empírica de que en las rutas con tiempo de viaje alto se tiene que un minuto más de viaje importa menos que en las rutas con bajo tiempo de viaje, se incorpora el cuadrado del tiempo de viaje en la componente compensatoria para capturar ese efecto y para observar si logra funcionar como una forma más sencilla de modelar cutoffs relacionados con el tiempo de viaje. Como se menciona al inicio de este capítulo, también se calibran modelos MNL para contrastarlos con los CMNL para evaluar si la incorporación de umbrales modela satisfactoriamente la elección de ruta y para verificar la existencia de sesgos en los parámetros.

Con respecto a los signos de los parámetros del cutoff, se espera que ω_k sea estrictamente positivo, ya que es un parámetro de escala de los logit binomiales correspondientes a los cutoffs de cada atributo. El parámetro $\omega_k(\rho_k - b_k)$ no tiene restricción de signo porque es la diferencia entre el parámetro de ajuste del cutoff de un atributo y su valor umbral, y debido al problema de identificabilidad no es posible saber cuál parámetro es mayor que el otro. Con respecto a los signos de los parámetros de la componente compensatoria, por un lado se espera que sean negativos los del tiempo de viaje θ_{tv} , del costo angular θ_{cang} y de los transbordos θ_{tr} , porque son las utilidades marginales de sus respectivos atributos, y para estas variables se tiene que mientras mayor sea el valor del atributo menor es la utilidad. Por otro lado, se espera que el parámetro del tiempo de viaje al cuadrado θ_{tv^2} sea positivo, porque se incorpora el efecto de que mientras mayor es el tiempo de viaje menos importa un minuto adicional, y esto se logra con un término de segundo orden para el tiempo de viaje cuyo coeficiente o parámetro es positivo⁹.

⁹ Para ver este efecto, se puede considerar la forma funcional de la utilidad con el tiempo de viaje como variable independiente. Entonces, se ve que la utilidad adopta la forma de una parábola que pasa por el origen y disminuye hasta un mínimo, o sea, posee concavidad positiva, por lo que el valor del coeficiente del término cuadrático del tiempo de viaje debe ser positivo. La siguiente sección analiza el efecto de incluir tv^2 .

Así como interesa conocer los valores de cada parámetro para la especificación de cada modelo propuesto, también interesa saber si efectivamente cada modelo se ajusta bien a los datos de la muestra, si las variables explican significativamente el problema de elección, y también interesa comparar estadísticamente entre los modelos para discriminar cuál es el más apropiado.

Para el cálculo de parámetros y los valores de los test estadísticos, se utiliza el programa GAUSS 6.0 y su herramienta MAXLIK, que a partir de un vector inicial de parámetros y con cálculos iterativos de los parámetros que maximizan la función log-verosimilitud (ver discusión en Castro, 2009), entrega el valor de los parámetros asociados a cada variable de la especificación con que se trabaja, el valor del estadístico del test-t t_0 de cada parámetro (recordar que si $|t_0| > 1,96$ el parámetro es significativamente distinto de cero) y el valor de la log-verosimilitud máxima $l(\theta^*)$. También se calcula la log-verosimilitud para el modelo equiprobable $l(0)$ (cuyo valor para este modelo es de -19.730,4) para calcular el indicador de bondad de ajuste de cada modelo como $\rho^2 = 1 - \frac{l(\theta^*)}{l(0)}$.

Para decidir entre dos modelos anidados (por ejemplo: entre un MNL, que es un modelo restringido, y un CMNL, que es un modelo general, y que comparten la misma componente compensatoria), se realiza el test de razón de verosimilitud LR para decidir si el modelo restringido es apropiado, en el que se calcula el valor $LR = -2\{l(\theta_r) - l(\theta)\}$, donde $l(\theta_r)$ es la log-verosimilitud del modelo restringido, $l(\theta)$ es la log-verosimilitud del modelo general y r es el número de restricciones lineales. Entonces, el valor de LR se compara con el valor de la distribución chi-cuadrado $\chi_{r,95\%}^2$, y se rechaza la hipótesis nula (concluyendo entonces que el modelo restringido no es apropiado) si $LR > \chi_{r,95\%}^2$.

Para el presente trabajo, se calibraron 31 especificaciones de la función de utilidad de los modelos MNL y CMNL, de las cuales se muestran a continuación los resultados de 12 modelos que registran los mejores indicadores estadísticos. Cabe recordar que dichos modelos se calibran con la base de datos de $N = 22.556$ observaciones. La Tabla 4.1 muestra la simbología de los parámetros calibrados asociados a las variables utilizadas en los modelos (ver descripción de las variables en la sección 3.2).

Parámetro	Símbolo
Parámetro del tiempo de viaje	θ_{tv}
Parámetro del tiempo de viaje al cuadrado	θ_{tv^2}
Parámetro del costo angular	θ_{cang}
Parámetro del número de transbordos	θ_{tr}
Parámetro de escala del cutoff de la diferencia de tiempo de viaje	ω_d
Cota superior del cutoff de la diferencia de tiempo de viaje	b_d
Parámetro de ajuste del cutoff de la diferencia de tiempo de viaje	ρ_d
Parámetro de escala del cutoff del número de transbordos	ω_{tr}
Cota superior del cutoff del número de transbordos	b_{tr}
Parámetro de ajuste del cutoff del número de transbordos	ρ_{tr}
Parámetro de escala del cutoff del costo angular	ω_{cang}
Cota superior del cutoff del costo angular	b_{cang}
Parámetro de ajuste del cutoff del costo angular	ρ_{cang}

Tabla 4.1: Simbología de los parámetros para los modelos de elección de ruta calibrados
(Fuente: Elaboración propia)

La Tabla 4.2 muestra los valores resultantes de las calibraciones de los modelos MNL y la Tabla 4.3 muestra los valores resultantes de las calibraciones de los modelos CMNL. Junto con cada parámetro se registran entre paréntesis los valores de los estadísticos resultantes t_0 del test-t. Para cada modelo, se entregan los valores de la log-verosimilitud del modelo equiprobable y de la máxima log-verosimilitud calculada con los parámetros resultantes, y el indicador de bondad de ajuste ρ^2 .

Parámetro	Modelo MNL 1	Modelo MNL 2	Modelo MNL 3	Modelo MNL 4	Modelo MNL 5	Modelo MNL 6	Modelo MNL 7
θ_{tv}	-0,2058 (-88,560)	-0,3702 (-43,220)	-0,2486 (-78,312)	-0,3687 (-43,342)	-	-0,0787 (-17,935)	-0,1758 (-18,970)
θ_{tv^2}	-	0,0020 (21,080)	-	0,0016 (15,993)	-	-	0,0012 (12,216)
θ_{cang}	-	-	0,1960 (21,835)	0,1525 (16,424)	-0,4238 (-63,521)	-0,2560 (-19,852)	-0,2742 (-21,248)
θ_{tr}	-	-	-	-	-	-1,5819 (-48,112)	-1,5512 (-47,092)
$l(\mathbf{0})$	-19.730,4	-19.730,4	-19.730,4	-19.730,4	-19.730,4	-19.730,4	-19.730,4
$l(\theta^*)$	-11.834,0	-11.613,3	-11.606,7	-11.479,7	-16.109,9	-10.249,7	-10.175,7
ρ^2	0,4002	0,4114	0,4117	0,4182	0,1835	0,4805	0,4843

Tabla 4.2: Calibración de modelos MNL
(Fuente: Elaboración propia)

Parámetro	Modelo CMNL 1	Modelo CMNL 2	Modelo CMNL 3	Modelo CMNL 4	Modelo CMNL 5
θ_{tv}	-0,0769 (-17,185)	-0,1788 (-16,789)	-	-0,0769 (-18,331)	-0,1788 (-19,311)
θ_{tv^2}	-	0,0013 (11,015)	-	-	0,0013 (12,888)
θ_{cang}	-	-	-0,2602 (-23,600)	-0,2602 (-20,415)	-0,2789 (-23,009)
θ_{tr}	-	-	-	-	-
ω_d	-	-	0,0769 (21,387)	-	-
$\omega_d(\rho_d - b_d)$	-	-	13,2278 (0,260)	-	-
ω_{tr}	1,9066 (26,703)	1,9681 (81,491)	1,9067 (28,894)	1,9065 (26,260)	1,9681 (25,529)
$\omega_{tr}(\rho_{tr} - b_{tr})$	-1,1671 (-3,717)	-1,5726 (-36,346)	-1,1679 (-4,235)	-1,1670 (-3,761)	-1,5726 (-5,417)
ω_{cang}	0,2602 (22,358)	0,2789 (24,981)	-	-	-
$\omega_{cang}(\rho_{cang} - b_{cang})$	14,7717 (0,840)	81,8506 (28,929)	-	-	-
$l(\mathbf{0})$	-19.730,4	-19.730,4	-19.730,4	-19.730,4	-19.730,4
$l(\theta^*)$	-10.232,6	-10.152,0	-10.232,6	-10.232,6	-10.152,0
ρ^2	0,4814	0,4855	0,4814	0,4814	0,4855

Tabla 4.3: Calibración de modelos CMNL
(Fuente: Elaboración propia)

Las calibraciones mostradas anteriormente, tanto de los modelos MNL como de los modelos CMNL, arrojan resultados satisfactorios para las variables descriptivas desde el punto de vista estadístico, ya que prácticamente todos los parámetros son significativamente distintos de cero, por lo que las variables descriptivas sí explican la elección de ruta en Metro en cada uno de los modelos. A excepción de los modelos MNL 3 y MNL 4, donde el parámetro θ_{cang} del costo angular es positivo, los parámetros calibrados poseen los signos esperados. A pesar de que en los modelos CMNL 1 y CMNL 3 se registra que los respectivos parámetros $\omega_{cang}(\rho_{cang} - b_{cang})$ y $\omega_d(\rho_d - b_d)$ no son significativos, el signo de ellos es correcto, por lo que se sigue la recomendación de la literatura de mantener estos parámetros.

Con respecto a los valores de la log-verosimilitud, se observa que todos los modelos registran valores de $l(\theta^*)$ mayores que el del modelo equiprobable $l(0)$, lo que indica que todos los modelos mejoran la estimación de la probabilidad de que se reproduzcan las elecciones entre las rutas registradas en la EOD 2011. Ahora bien, con respecto al tema de qué modelos ajustan mejor a los datos, los ρ^2 de los modelos CMNL registran valores mayores que los ρ^2 de los modelos MNL, lo que da indicios de que el CMNL funciona bien para el problema de elección de este trabajo. Los modelos CMNL registran ajustes de 0,4814 y 0,4855; mientras que los modelos MNL 1 al MNL 5 registran menor ajuste, dado que no tienen incorporados todas las variables explicativas trabajadas aquí. En contraste, los modelos MNL 6 y MNL 7 (que incluyen los tres atributos) sí registran ajustes comparables con los modelos CMNL, con valores de 0,4805 y 0,4843, respectivamente. De todas formas, los modelos CMNL calibrados poseen un ajuste levemente mejor que los modelos MNL 6 y MNL 7.

Para una comparación más rigurosa entre los modelos calibrados, en la Tabla 4.4 se muestra el resultado de los tests de razón de verosimilitud entre los modelos anidados que se presentan en este trabajo. En todos los casos, los modelos restringidos se rechazan, concluyendo que no son adecuados para el problema de elección de ruta en Metro.

Modelo General	$l(\theta)$	Modelo Restringido	$l(\theta_r)$	LR	r	$\chi^2_{r,95\%}$	Resultado Test
Modelo CMNL 1	-10.232,6	Modelo MNL 1	-11.834,0	3.202,907	4	9,488	modelo general
Modelo CMNL 2	-10.152,0	Modelo MNL 2	-11.613,3	2.922,581	4	9,488	modelo general
Modelo CMNL 2	-10.152,0	Modelo CMNL 1	-10.232,6	161,095	1	3,841	modelo general
Modelo CMNL 3	-10.232,6	Modelo MNL 5	-16.109,9	11.754,653	4	9,488	modelo general
Modelo CMNL 4	-10.232,6	Modelo MNL 3	-11.606,7	2.748,178	2	5,991	modelo general
Modelo CMNL 5	-10.152,0	Modelo MNL 4	-11.479,7	2.655,337	2	5,991	modelo general
Modelo CMNL 5	-10.152,0	Modelo CMNL 4	-10.232,6	161,095	1	3,841	modelo general

Tabla 4.4: Tests de razón de verosimilitud para modelos anidados
(Fuente: Elaboración propia)

Dado que los indicadores estadísticos del CMNL son mejores que los del MNL, se concluye que el modelo CMNL es adecuado para modelar el problema tratado en este trabajo. Ahora bien, como los modelos CMNL presentados tienen prácticamente los mismos ajustes entre sí, hay que elegir cuál especificación modela mejor la elección de ruta. Sin embargo, al mirar las especificaciones y los parámetros obtenidos, hay dos temas que tomar en cuenta: el primero, sobre el rango de validez de los atributos para los modelos CMNL obtenidos; y el segundo, sobre la igualdad de los valores absolutos de los parámetros asociados a algunos atributos en distintas especificaciones. La siguiente sección abarca dichos temas y finalmente decide qué especificación de los CMNL modela mejor la elección de ruta en Metro.

4.3. Discusión

Esta sección tiene como objetivo escoger la especificación entre los modelos CMNL de la Tabla 4.3 que modela mejor el fenómeno de la elección de ruta en Metro. La sección anterior muestra que todos los modelos CMNL calibrados son adecuados desde el punto de vista estadístico, pero deja pendiente la elección del modelo más adecuado para el problema abordado en este Trabajo de Título. Como se explica en la sección anterior, hay que prestar atención a dos temas que potencialmente pueden restar validez a las especificaciones calibradas. Estos temas, que en principio representan falencias al enfoque empleado en este trabajo, se convierten en una oportunidad para obtener argumentos sólidos para escoger el modelo más apropiado.

El primer tema consiste en el rango de validez de los modelos MNL y CMNL calibrados anteriormente. A excepción de los modelos MNL 3 y MNL 4, todos los modelos poseen parámetros con los signos esperados, lo que permite decir que en esos modelos mientras mayores son los valores de los atributos de una ruta, menor es la utilidad percibida, por lo que menos usuarios la escogen. Sin embargo, hay que tener cuidado con los modelos que incluyen el tiempo de viaje al cuadrado (tv^2), porque se obtienen predicciones erradas cuando el tiempo de viaje de una ruta es alto. Esto puede observarse en la Figura 4.1, porque al graficar las utilidades considerando sólo los atributos tv y tv^2 de los modelos MNL y CMNL correspondientes, se ve que los modelos MNL 7, CMNL 2 y CMNL 5 alcanzan un valor mínimo en tv^* , y más allá de ese valor la utilidad comienza a aumentar. También se grafica la componente compensatoria de los modelos CMNL 1 y CMNL 4 considerando sólo el tiempo de viaje.

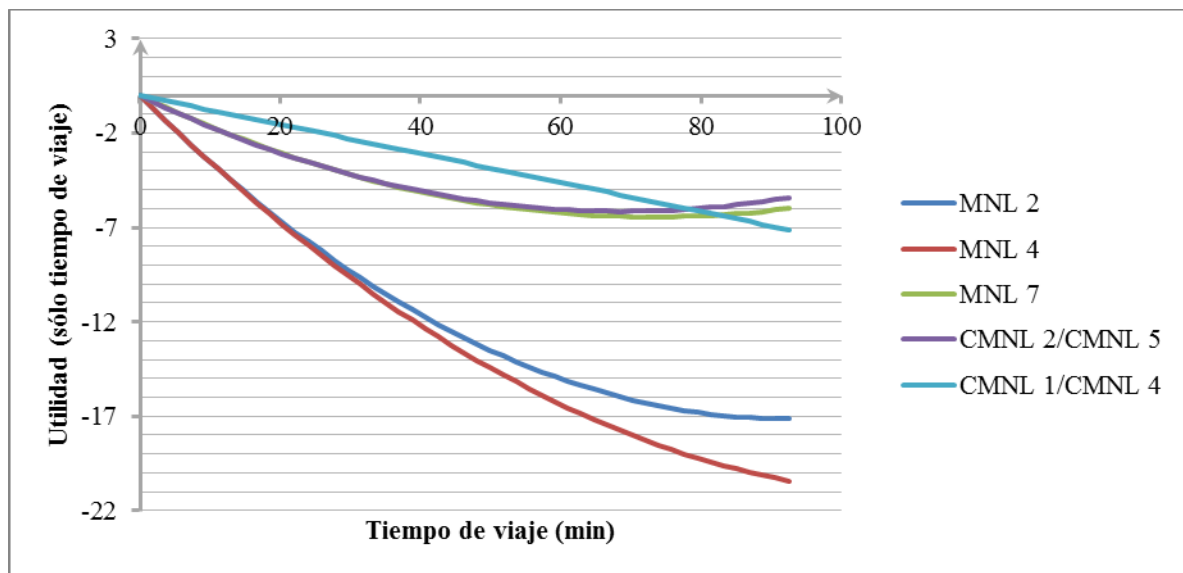


Figura 4.1: Gráfico de las utilidades de modelos con tv^2
(Fuente: Elaboración propia)

Para cada modelo con tv^2 , tanto MNL como CMNL, se deriva la utilidad con respecto al tiempo de viaje, obteniendo la utilidad marginal del tiempo de viaje, y luego se iguala a cero para obtener el valor en que se minimiza la utilidad. La Tabla 4.5 registra los valores de tv^* en que la utilidad es mínima. Para valores superiores a tv^* , la utilidad aumenta, lo que lleva a malas predicciones al utilizar los modelos con tv^2 , ya que se obtiene utilidad alta para rutas con tiempo de viaje alto, lo que contradice el supuesto de que rutas con alto tiempo de viaje son poco atractivas. Cabe recordar de la Tabla 3.4 que existen rutas con tiempos de viaje de hasta 90,37 minutos, lo que supera los valores de tv^* de los modelos MNL 7, CMNL 2 y CMNL 5.

	Modelo	tv^* (min)
MNL	Modelo MNL 2	92,55
	Modelo MNL 4	115,22
	Modelo MNL 7	73,25
CMNL	Modelo CMNL 2	68,77
	Modelo CMNL 5	68,77

Tabla 4.5: Rango máximo del tiempo de viaje en modelos con tv^2
(Fuente: Elaboración propia)

A pesar de lo anterior, en el rango de validez para el tiempo de viaje (entre cero y tv^*), las especificaciones de la utilidad con tv^2 modelan satisfactoriamente la elección de ruta, porque consideran el fenómeno de que un minuto adicional importa menos cuando el tiempo de viaje es alto, es decir, si existe una ruta con una diferencia de pocos minutos (y si la ruta de referencia posee alto tiempo de viaje), es posible que el modelo prediga un nivel de preferencia no despreciable por aquella ruta. No obstante este hecho, es preferible emplear un modelo con validez en todo el rango de las variables explicativas, porque es más práctico no tener que prestar atención al rango de validez de las variables.

Como se observa en la Figura 4.1, los modelos CMNL 1 y CMNL 4 son válidos en todo el rango de tv porque siempre se tiene que a mayor tiempo de viaje menor es la utilidad. Así, los modelos CMNL 2 y CMNL 5 quedan descartados, pero cabe recalcar que en rango de validez para el tiempo de viaje funcionan bien.

La discusión anterior deja solo a tres modelos: CMNL 1, CMNL 3 y CMNL 4. Ahora bien, a pesar que los modelos CMNL 2 y CMNL 5 poseen los mejores ajustes ($\rho^2 = 0,4855$), curiosamente son los primeros en descartarse debido al argumento del rango de validez. Los CMNL restantes poseen un ajuste levemente menor ($\rho^2 = 0,4814$), pero registran un ajuste levemente mejor que el MNL 6 con $\rho^2 = 0,4805$ (el MNL 7, con $\rho^2 = 0,4843$, se descarta por el argumento del rango de validez). El modelo MNL 6 se debe tomar en cuenta porque registra un ajuste similar comparado con los CMNL.

Mientras el MNL asume comportamiento compensatorio, el CMNL no asume dicho comportamiento al incorporar penalizaciones de atributos dados los umbrales de la muestra. Castro (2009) y Castro et al. (2013) concluyen que el CMNL reproduce el MNL, lo que indica que el CMNL posee la ventaja de que delata cuando hay tradeoff entre los atributos, en contraste con el MNL, que asume a priori un comportamiento compensatorio de los individuos. El estudio del segundo tema, relacionado con la igualdad de parámetros entre las distintas especificaciones, pretende observar si existe un comportamiento compensatorio en la elección de ruta, o bien si es semicomplementario.

Al observar la Tabla 4.3, se observa que entre las distintas especificaciones de los modelos CMNL se registran valores iguales de algunos parámetros de las componentes compensatorias y de los parámetros de los cutoffs. Los casos en que ocurren estos fenómenos son los siguientes:

- a) El valor del parámetro de escala $\omega_{cang} = 0,2602$ en el CMNL 1 es igual al valor absoluto del parámetro compensatorio $\theta_{cang} = -0,2602$ de los modelos CMNL 3 y CMNL 4.
- b) El valor del parámetro de escala $\omega_d = 0,0769$ en el CMNL 3 es igual al valor absoluto del parámetro compensatorio $\theta_{tv} = -0,0769$ de los modelos CMNL 1 y CMNL 4.
- c) El valor del parámetro de escala $\omega_{cang} = 0,2789$ en el CMNL 2 es igual al valor absoluto del parámetro compensatorio $\theta_{cang} = -0,2789$ del modelo CMNL 5.
- d) Los parámetros $\theta_{tv} = -0,1788$, $\theta_{tv^2} = 0,0013$, $\omega_{tr} = 1,9681$ y $\omega_{tr}(-b_{tr} + \rho_{tr}) = -1,5726$ son idénticos entre los modelos CMNL 2 y CMNL 5.

Lo anterior hace pensar que los modelos CMNL 1, CMNL 3 y CMNL 4 son los mismos, y que los modelos CMNL 2 y CMNL 5 son los mismos también. Esto implicaría que para este desarrollo, la aproximación lineal $\theta_k x_{ik}$ modela lo mismo que el cutoff $-\ln\{1 + \exp[\omega_k(x_{ik} - b_k + \rho_k)]\}$ calibrando un parámetro menos. Esto corresponde al principio de parsimonia, el cual consiste en que si una especificación de la función utilidad es capaz de modelar lo mismo que otra especificación con menos parámetros por calibrar, entonces es preferible escoger la especificación que posee menos parámetros por calibrar.

Para escoger la especificación del CMNL que mejor modele la elección de ruta, se aplica el principio de parsimonia a través de los análisis de los cutoffs de diferencia de tiempo de viaje, costo angular y de transbordos. Dichos análisis consisten en graficar y comparar los cutoffs con las aproximaciones lineales, o sea, se revisa en qué casos el cutoff $-\ln\{1 + \exp[\omega_k(x_{ik} - b_k + \rho_k)]\}$ se aproxima a la forma funcional $-\omega_k x_{ik} = \theta_k x_{ik}$. Asimismo, se varía el valor de los parámetros de los cutoffs relacionados con $(-b_k + \rho_k)$ para observar qué cambios se producen en la utilidad.

La Figura 4.2 grafica el cutoff del número de transbordos tr del modelo CMNL 4, el cual se compara con la componente compensatoria modelo MNL 6 que posee los transbordos (donde $\theta_{tr} = -1,5819$) y con la aproximación lineal del cutoff $-\omega_{tr}(tr_i - b_{tr} + \rho_{tr})$. También se varía la componente $(-b_{tr} + \rho_{tr})$, cuyo valor en el CMNL 4 es de $-0,6121$, y se grafica el cutoff para valores de $(-b_{tr} + \rho_{tr})$ iguales a 0 y $-0,27$. A pesar de que tr es una variable discreta, se grafica como si fuera continua para facilitar el análisis.

Se observa que el cutoff de transbordos tiende a una aproximación lineal más bien cuando el número de transbordos es alto, lo que indica que el valor de $\exp[\omega_{tr}(tr_i - b_{tr} + \rho_{tr})]$ no es tan grande. Se observa que el cutoff se encuentra sobre la curva de $\theta_{tr} tr_i$, lo que indica que la variable de transbordos representa más bien un efecto penalizador de la utilidad en vez de un factor compensatorio, porque las rutas de la red de Metro no requieren de una cantidad importante de transbordos por la baja densidad de líneas (ver discusión sobre las correlaciones en la sección 3.2). Si se considerara $\theta_{tr} tr_i$ en vez del cutoff, se sobrestimaría el efecto de los transbordos en la elección de ruta. Cabe destacar que si $(-b_{tr} + \rho_{tr}) = -0,34$, el cutoff es aproximadamente asintótico a la curva $\theta_{tr} tr_i$.

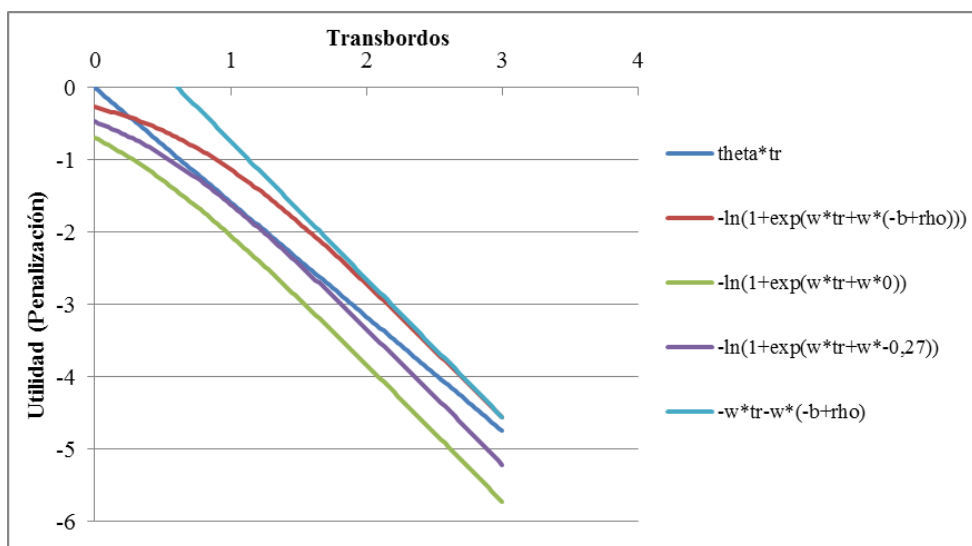


Figura 4.2: Análisis del cutoff de transbordos ($\theta_{tr} = -1,5819$)
(Fuente: Elaboración propia)

La Figura 4.3 grafica el cutoff del costo angular *cang* del modelo CMNL 1, el cual se compara con la aproximación lineal de las componentes compensatorias de los modelos CMNL 3 y CMNL 4, considerando solo la componente de costo angular (en los CMNL 3 y CMNL 4, $\theta_{cang} = -0,2602$). También, se varía la componente $(-b_{cang} + \rho_{cang})$, cuyo valor en el CMNL 1 es igual a 56,7706 y se grafica el cutoff para valores de $(-b_{cang} + \rho_{cang})$ iguales a -10, 0 y 10.

Se observa que el cutoff de costo angular de viaje es igual a la aproximación lineal $\theta_{cang}cang_i$ desplazado por un valor constante. Esto indica que $\exp[\omega_{cang}(cang_i - b_{cang} + \rho_{cang})] \gg 1$, por lo que el cutoff es igual a $-\omega_k(x_{ik} - b_k + \rho_k)$, por lo que el valor constante es igual a $\omega_k(-b_{cang} + \rho_{cang})$. A pesar de que ambas curvas son diferentes, están diferenciadas solo por el término constante $\omega_{cang}(-b_{cang} + \rho_{cang})$ que está presente en todas las alternativas, por lo que al evaluar las probabilidades utilizando el CMNL 3 y el CMNL 4, se obtienen los mismos resultados, por lo que es lo mismo modelar el costo angular como un factor penalizador (a través de un cutoff) que como un factor compensatorio.

La Figura 4.4 grafica el cutoff de la diferencia de tiempo de viaje *d* del modelo CMNL 3, el cual se compara con la aproximación lineal de la componente compensatoria del modelo CMNL 4, considerando solo la componente de tiempo de viaje (en el CMNL 4, $\theta_{tv} = -0,0769$). También, se varía la componente $(-b_d + \rho_d)$, cuyo valor en el CMNL 3 es de 172,103 y se grafica el cutoff para valores de $(-b_d + \rho_d)$ iguales a 0 y 100.

Nuevamente, se observa que el cutoff de diferencia de tiempo de viaje es igual a la aproximación lineal $\theta_{tv}tv_i$ desplazado por un valor constante. El cutoff de *d* adopta la forma lineal $-\omega_d(d_i - b_d + \rho_d)$, porque el valor $\exp[\omega_d(d_i - b_d + \rho_d)] \gg 1$. A pesar de que las curvas $\theta_{tv}tv_i$ y $-\ln\{1 + \exp[\omega_d(d_i - b_d + \rho_d)]\}$ son diferentes, están diferenciadas solo por el término constante $\omega_d(-b_d + \rho_d)$ que está presente en todas las alternativas, por lo que al evaluar las probabilidades utilizando el CMNL 3 y el CMNL 4, se obtienen los mismos resultados, por lo que es lo mismo modelar el tiempo de viaje como un factor penalizador (a través de un cutoff) que como un factor compensatorio.

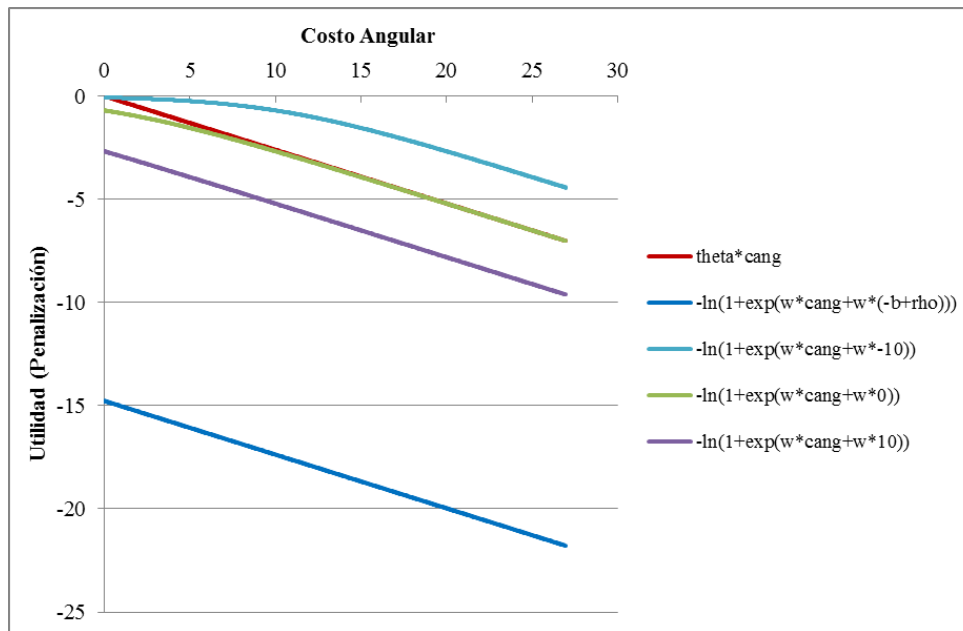


Figura 4.3: Análisis del cutoff del costo angular ($\theta_{cang} = -0,2602$)
(Fuente: Elaboración propia)

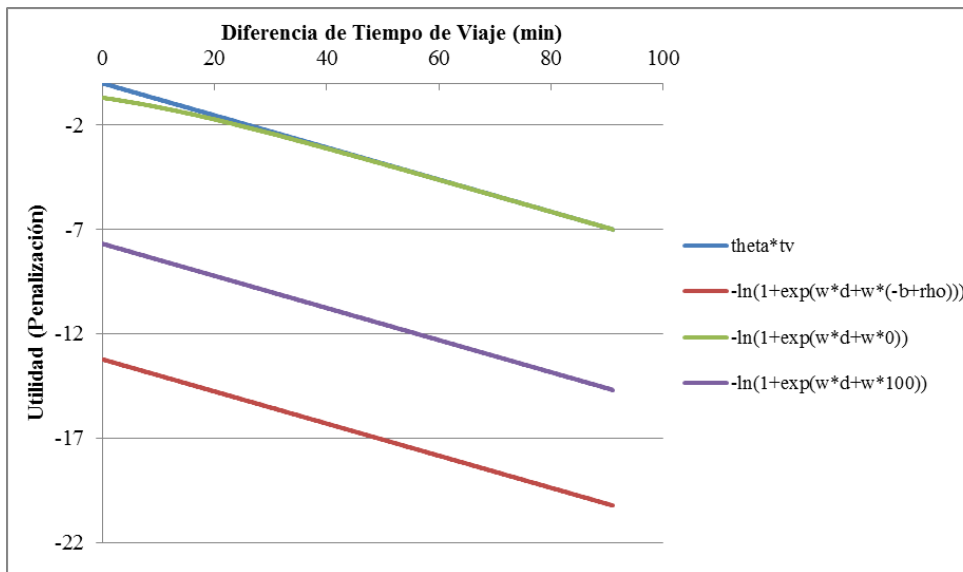


Figura 4.4: Análisis del cutoff de la diferencia de tiempo de viaje ($\theta_{tv} = -0,0769$)
(Fuente: Elaboración propia)

El análisis de los cutoffs muestra que el número de transbordos es un factor penalizador en la elección de rutas, mientras que el costo angular es un factor compensatorio, por lo que la elección del mejor modelo está entre CMNL 3 y CMNL 4, los cuales poseen el mismo ajuste. En principio, estos modelos son idénticos, con la salvedad de que mientras el CMNL 3 incorpora el tiempo de viaje como factor penalizador (como diferencia de tiempo de viaje), el CMNL 4 lo incorpora como variable de la componente compensatoria. Según la hipótesis sostenida aquí de que los usuarios procesan la información de los tiempos de viaje de las rutas y que comparan las alternativas entre sí para realizar su elección, se decide finalmente que el modelo que propone este trabajo es el CMNL 4, porque recoge el mismo fenómeno de comparación que el modelo CMNL 3. Por parsimonia, siempre se prefiere un modelo con menos parámetros a calibrar que explique lo mismo, por lo que se decide que el tiempo de viaje es un factor compensatorio.

Capítulo 5

Conclusiones

El presente Trabajo de Título analiza el fenómeno de la elección de ruta en el Metro de Santiago con el fin de entregar una herramienta capaz de capturar la comparación que realizan los usuarios de sus rutas disponibles. La motivación de este trabajo es el uso de información automatizada generada por las transacciones bip! y las variables de operación de Metro, para incorporarla a modelos de elección de ruta en modalidad predictiva, en desmedro de metodologías tradicionales más costosas tales como las encuestas EOD, tomando en cuenta que la red aumentará su extensión en 37% para el año 2018 con la puesta en marcha de las Líneas 3 y 6. El enfoque del modelo de elección propuesto es mixto, ya que se modelan usuarios con comportamiento semicompensatorio, en que existe tradeoff entre los atributos de algunas rutas y que hay descarte de las demás rutas disponibles. Los resultados muestran que efectivamente los usuarios poseen un comportamiento semicompensatorio en la decisión sobre las rutas en Metro, por lo que se cumple el objetivo de este trabajo.

El desarrollo del trabajo considera en primer lugar examinar el fenómeno de elección de rutas que los usuarios declararon realizar en la Encuesta Origen-Destino 2011 de Metro, con el propósito de detectar el fenómeno de descarte de rutas entre los pares origen-destino que dicha encuesta recoge. Se observa que en el 91% de los pares O-D registrados en la EOD 2011 se registran una y dos rutas, y que en el 9% restante, correspondientes a los pares con tres y cuatro rutas registradas, se observa en la mayoría de ellas que existe predominancia de una y dos rutas. A pesar de que no es posible realizar una clasificación de los pares O-D registrados que permita contar con certeza los pares en que existe reparto entre rutas o bien predominancia de rutas, se puede concluir que los usuarios consideran una o dos alternativas de rutas para su viaje y que el resto de las posibles opciones son descartadas del conjunto de elección. La tarea entonces es explicar qué atributos de las rutas hacen que los usuarios no consideren algunas de las alternativas disponibles.

En este trabajo se ha utilizado un modelo con umbrales de tolerancia de los atributos: el Logit Multinomial Restringido CMNL, que emplea la heurística de incorporación de funciones de penalización o cutoffs a las funciones de utilidad para modelar implícitamente el descarte de rutas que superan uno o más valores umbrales de los atributos. En primer lugar se buscan todas las rutas razonables de los 11.540 pares O-D de Metro según una lista de criterios bien definidos y flexibles, para posteriormente describir dichas rutas con variables sencillas de obtener que sean capaces de explicar el comportamiento de los usuarios. Para los modelos calibrados aquí, los atributos utilizados son tiempo de viaje, el costo angular y el número de transbordos. Se prueban modelos CMNL que penalicen las utilidades de las rutas por el costo angular, el número de transbordos y/o la diferencia de tiempo de viaje. Desde el punto de vista estadístico, en todos los modelos dichas variables explican bastante bien la elección de ruta, al obtener parámetros significativamente distintos de cero. Asimismo, los modelos CMNL poseen mejores ajustes que los modelos MNL calibrados en este trabajo, y según el test de razón de verosimilitud, son más apropiados que los modelos restringidos MNL con los que están anidados.

El efecto de la inclusión de las penalizaciones en las funciones de utilidad se confunde bastante con el comportamiento compensatorio, ya que en varios modelos los valores de los parámetros de escala de los cutoffs y de los parámetros de las componentes compensatorias son idénticos en valor absoluto entre sí. Dada la estructura de la red de Metro y los resultados de las calibraciones, se concluye que el transbordo es un factor penalizador, mientras que el tiempo de viaje y el costo angular son factores compensatorios, por lo que se propone modelar con el modelo CMNL 4, ya que resulta ser un modelo que explica del mismo modo el fenómeno de comparación entre rutas que el modelo CMNL 3. El modelo CMNL 4 posee mejor ajuste que el MNL 6 y requiere calibrar un parámetro menos que el CMNL 3. De cierta manera, este resultado confirma las conclusiones de Castro et al. (2013) de que el modelo CMNL reproduce el MNL, o sea, el modelo CMNL delata si existe o no comportamiento compensatorio en vez de hacer dicho supuesto a priori.

Para la investigación futura de los modelos de elección de ruta, se propone considerar un efecto dinámico en la elección que tome en cuenta el hacinamiento y la capacidad de las líneas y de los trenes, es decir, modelar que las personas escogen otras alternativas dado que algunas rutas se encuentran congestionadas, sobre todo si se modelan los viajes en horario punta.

Una propuesta para mejorar la calibración consiste en separar del tiempo de viaje en ruta los tiempos de caminata y de espera en los transbordos, con el objetivo de que la calibración indique la valoración que los usuarios dan a dichos tiempos en vez de suponer que están valorados por el triple del tiempo de viaje, obteniendo así parámetros propios de tiempos de viaje, caminata y espera.

A pesar de que los parámetros asociados a las variables explican bastante bien la elección y el ajuste resulta bastante bueno, queda propuesto incluir otras variables no tradicionales descritas por Raveau et al. (2011) para mejorar el poder explicativo de la elección. De hecho, este trabajo incluye el costo angular como una manera de incorporar el conocimiento de la red de Metro para distinguir las rutas directas de las indirectas. Si se incluyen las otras variables no tradicionales descritas por Raveau et al., se requiere hacer supuestos más fuertes sobre la estructura de la red y de las estaciones, sobre todo si se desea proyectar y planificar futuras líneas y modificaciones en la red de Metro.

A partir de este modelo de elección de ruta, las líneas de acción son el desarrollo de una aplicación práctica para comparar perfiles de cargas observados y predichos en la red de Metro, y la implementación computacional que mejore el modelo de obtención para obtener matrices origen-destino en el sistema de Transporte Público basado en las cadenas de transacciones bip!

Bibliografía

- Beltrán, P., Cortés, C., Gschwender, A., Ibarra, R., Munizaga, M. A., Ortega, M., Palma, C. y Zúñiga, M. (2011). Obtención de información valiosa a partir de datos de Transantiago. *XV Congreso Chileno de Ingeniería de Transporte, 03-06 Octubre*.
- Bierlaire, M., Hurtubia, R. y Flötteröd, G. (2010). An analysis of the implicit choice set generation using the Constrained Multinomial Logit model. *Proceedings of the Transportation Research Board 2010 Annual Meeting January 10-14, 2010*.
- Cantillo, V. y Ortúzar, J. D. D. (2005). A semi-compensatory discrete choice model with explicit attribute thresholds of perception. *Transportation Research Part B: Methodological*, 39(7), 641-657.
- Cascetta, E. y Papola, A. (2001). Random utility models with implicit availability/perception of choice alternatives for the simulation of travel demand. *Transportation Research Part C: Emerging Technologies*, 9(4), 249-263
- Castro, M. (2009). Calibración de un modelo de elección discreta logit con restricción del conjunto de alternativas. *Tesis presentada para la obtención del grado de Magíster en Ingeniería de Transporte de la Universidad de Chile*. Santiago de Chile, Enero 2009.
- Castro, M., Martínez, F. y Munizaga, M. A. (2013). Estimation of a constrained multinomial logit model. *Transportation*, 1-19.
- Fernández y de Cea Ingenieros Ltda. (2003). Análisis Modernización de Transporte Público, V Etapa. *Biblioteca SECTRA*, <http://www.sectra.gob.cl/biblioteca>
- Glasser, G. J. (1962). Variance formulas for the mean difference and coefficient of concentration. *Journal of the American Statistical Association*, 57(299), 648-654.
- Kaplan, S., Shiftan, Y. y Bekhor, S. (2012). Development and estimation of a semi-compensatory model with a flexible error structure. *Transportation Research Part B: Methodological*, 46(2), 291-304.
- Ke, Q., Peng, Z. y Zhi-peng, Q. (2013). Model of Passenger Route Choice in the Urban Rail Transit Network. *LISS 2012*, 463-468. Springer Berlin Heidelberg.
- Manski, C. F. (1977). The structure of random utility models. *Theory and decision*, 8(3), 229-254.
- Martínez, F., Aguila, F. y Hurtubia, R. (2009). The constrained multinomial logit: A semi-compensatory choice model. *Transportation Research Part B: Methodological*, 43(3), 365-377.
- Morikawa, T. (1995). A hybrid probabilistic choice det Model with compensatory and noncompensatory choice rules. *Proceedings of the 7th World Conference on Transport Research*.

- Munizaga, M. A. y Palma, C. (2012). Estimation of a disaggregate multimodal public transport Origin–Destination matrix from passive smartcard data from Santiago, Chile. *Transportation Research Part C: Emerging Technologies*, 24, 9-18.
- Prashker, J. N. y Bekhor, S. (2004). Route choice models used in the stochastic user equilibrium problem: a review. *Transport reviews*, 24(4), 437-463.
- Pelletier, M. P., Trépanier, M. y Morency, C. (2011). Smart card data use in public transit: A literature review. *Transportation Research Part C: Emerging Technologies*, 19(4), 557-568.
- Raveau, S., Muñoz, J. C. y De Grange, L. (2011). A topological route choice model for metro. *Transportation Research Part A: Policy and Practice*, 45(2), 138-147.
- Swait, J. (2001). A non-compensatory choice model incorporating attribute cutoffs. *Transportation Research Part B: Methodological*, 35(10), 903-928.

Anexos

Anexo A: Código en MATLAB para calcular el costo angular de segmentos de rutas con origen en Línea 1

```
function costo_angular_L1()

tic

coordenadas=xlsread('Ángulo Viaje Dijkstra k=5 L1.xlsx','Coordenadas','B2:D109');
base_linea_1=xlsread('Ángulo Viaje Dijkstra k=5 L1.xlsx','Dijkstra k=5
L1','D2:H387850');
angulo=zeros(387849,1);
costo=zeros(387849,1);

for i=1:387848

    if(base_linea_1(i,5)==base_linea_1(i,2))
        angulo(i)=0;
        costo(i)=0;
    else
        xi=coordenadas(base_linea_1(i,5),2);
        yi=coordenadas(base_linea_1(i,5),3);
        xi2=coordenadas(base_linea_1(i+1,5),2);
        yi2=coordenadas(base_linea_1(i+1,5),3);
        xd=coordenadas(base_linea_1(i,2),2);
        yd=coordenadas(base_linea_1(i,2),3);
        ax=xi2-xi;
        ay=yi2-yi;
        bx=xd-xi;
        by=yd-yi;
        angulo(i)=acos((ax*bx+ay*by)/(sqrt(ax^2+ay^2)*sqrt(bx^2+by^2)));
        costo(i)=sqrt((xi2-xi)^2+(yi2-yi)^2)*sin(angulo(i)*0.5);
    end

end

alpha=xlswrite('Ángulo Viaje Dijkstra k=5 L1.xlsx',angulo,'Dijkstra k=5
L1','K2:K387850');
beta=xlswrite('Ángulo Viaje Dijkstra k=5 L1.xlsx',costo,'Dijkstra k=5
L1','L2:L387850');

toc

end
```

Nota: Para el resto de las líneas de la red de Metro, se cambia el archivo de lectura y se modifican las celdas de información que el software debe ingresar.

Anexo B: Código en GAUSS para el modelo MNL 6

```
screen on;
library maxlik;
#include maxlik.ext;
maxset;

/* Lectura de archivo */
base=xlsreadm("Base FP Definitiva.xlsx","A3:AB22558",1,"");
n=rows(base);
m=cols(base);

/* Definición de variables */
proc lpr(x,z);
local v;

v=zeros(n,5);

/* x son los parámetros, z los atributos */
/* x[1] parámetro de tiempo de viaje, x[2] de costo angular, x[3] de transbordos */

v[.,1]=x[1]*z[.,3] + x[2]*z[.,14] + x[3]*z[.,19];
v[.,2]=x[1]*z[.,4] + x[2]*z[.,15] + x[3]*z[.,20];
v[.,3]=x[1]*z[.,5] + x[2]*z[.,16] + x[3]*z[.,21];
v[.,4]=x[1]*z[.,6] + x[2]*z[.,17] + x[3]*z[.,22];
v[.,5]=x[1]*z[.,7] + x[2]*z[.,18] + x[3]*z[.,23];

retp((z[.,2].==1).*v[.,1]+(z[.,2].==2).*v[.,2]+(z[.,2].==3).*v[.,3]+
(z[.,2].==4).*v[.,4]+(z[.,2].==5).*v[.,5]-ln((exp(v).*z[.,24:28])*ones(5,1))));
endp;

x0={1,1,1};

__title = "Modelo MNL 6";
_max_Algorithm = 2;
_max_MaxTime = 360;
_max_CovPar = 0;

varnam = "tviaje"|"cang"|"#trans";
_max_ParNames = varnam;

output file=output_mnl_6_fp.txt reset;

{x,f,g,cov,ret}=maxlik(base,0,&lpr,x0);
call maxprt(x,f,g,cov,ret);
```


Anexo C: Código en GAUSS para el modelo CMNL 3

```
screen on;
library maxlik;
#include maxlik.ext;
maxset;

/* Lectura de archivo */
base=xlsreadm("Base FP Definitiva.xlsx","A3:AB22558",1,"");
n=rows(base);
m=cols(base);

/* Definición de variables */
proc lpr(x,z);
local v,lnphi;

v=zeros(n,5);

/*Cálculo de lnphi: cutoff de la diferencia de tiempo de viaje y del número de
transbordos*/
v=zeros(n,5);
lnphi=ones(n,5);

for q(1,n,1);
    lnphi[q,1]=ln( 1 + ( exp(x[2]*(z[q,3]-z[q,13])+x[3]) ) )
        + ln( 1 + ( exp(x[4]*z[q,19]+x[5]) ) ) );
    lnphi[q,2]=ln( 1 + ( exp(x[2]*(z[q,4]-z[q,13])+x[3]) ) )
        + ln( 1 + ( exp(x[4]*z[q,20]+x[5]) ) ) );
    lnphi[q,3]=ln( 1 + ( exp(x[2]*(z[q,5]-z[q,13])+x[3]) ) )
        + ln( 1 + ( exp(x[4]*z[q,21]+x[5]) ) ) );
    lnphi[q,4]=ln( 1 + ( exp(x[2]*(z[q,6]-z[q,13])+x[3]) ) )
        + ln( 1 + ( exp(x[4]*z[q,22]+x[5]) ) ) );
    lnphi[q,5]=ln( 1 + ( exp(x[2]*(z[q,7]-z[q,13])+x[3]) ) )
        + ln( 1 + ( exp(x[4]*z[q,23]+x[5]) ) ) );
endfor;

/* x son los parámetros, z los atributos */
/* Compensatoria: x[1] costo angular
Cutoff de dif tiempo de viaje: x[2] w de dif. de tiempo viaje,
x[3] es B_dif=w*(-bnd+rhod),
Cutoff de transbordos: x[4] w de tranbordos, x[5] es B_tr=w*(-bntr+rhotr) */

v[.,1]= x[1] * z[.,14] - lnphi[.,1];
v[.,2]= x[1] * z[.,15] - lnphi[.,2];
v[.,3]= x[1] * z[.,16] - lnphi[.,3];
v[.,4]= x[1] * z[.,17] - lnphi[.,4];
v[.,5]= x[1] * z[.,18] - lnphi[.,5];

retp((z[.,2].==1).*v[.,1]+(z[.,2].==2).*v[.,2]+(z[.,2].==3).*v[.,3]+
(z[.,2].==4).*v[.,4]+(z[.,2].==5).*v[.,5]-ln((exp(v).*z[.,24:28])*ones(5,1)));
endp;

x0={0,1,-1,1,-1};

__title = "Modelo CMNL 3";
_max_Algorithm = 2;
_max_MaxTime = 360; _max_CovPar = 0;
```

```
varnam = "cang"|"w_dif"|"B_dif"|"w_trans"|"B_trans";  
_max_ParNames = varnam;  
  
output file=output_cmn1_3_fp.txt reset;  
  
{x,f,g,cov,ret}=maxlik(base,0,&lpr,x0);  
call maxprt(x,f,g,cov,ret);
```

Anexo D: Código en GAUSS para el modelo CMNL 4

```
screen on;
library maxlik;
#include maxlik.ext;
maxset;

/* Lectura de archivo */
base=xlsreadm("Base FP Definitiva.xlsx","A3:AB22558",1,"");
n=rows(base);
m=cols(base);

/* Definición de variables */
proc lpr(x,z);
local v,lnphi;

v=zeros(n,5);

/*Cálculo de lnphi, el cutoff del número de transbordos*/
v=zeros(n,5);
lnphi=ones(n,5);

for q(1,n,1);
    lnphi[q,1]=ln( 1 + ( exp(x[3]*z[q,19]+x[4]) ) );
    lnphi[q,2]=ln( 1 + ( exp(x[3]*z[q,20]+x[4]) ) );
    lnphi[q,3]=ln( 1 + ( exp(x[3]*z[q,21]+x[4]) ) );
    lnphi[q,4]=ln( 1 + ( exp(x[3]*z[q,22]+x[4]) ) );
    lnphi[q,5]=ln( 1 + ( exp(x[3]*z[q,23]+x[4]) ) );
endfor;

/* x son los parámetros, z los atributos */
/* Compensatoria: x[1] tiempo de viaje, x[2] costo angular
Cutoff de transbordos: x[3] w de transbordos, x[4] es B_tr=w*(-bntr+rhotr) */

v[.,1]= x[1] * z[.,3] + x[2] * z[.,14] - lnphi[.,1];
v[.,2]= x[1] * z[.,4] + x[2] * z[.,15] - lnphi[.,2];
v[.,3]= x[1] * z[.,5] + x[2] * z[.,16] - lnphi[.,3];
v[.,4]= x[1] * z[.,6] + x[2] * z[.,17] - lnphi[.,4];
v[.,5]= x[1] * z[.,7] + x[2] * z[.,18] - lnphi[.,5];

retp((z[.,2].==1).*v[.,1]+(z[.,2].==2).*v[.,2]+(z[.,2].==3).*v[.,3]+
(z[.,2].==4).*v[.,4]+(z[.,2].==5).*v[.,5]-ln((exp(v).*z[.,24:28])*ones(5,1)));
endp;

x0={0.5,0.5,0,0};

__title = "Modelo CMNL 4";
_max_Algorithm = 2;
_max_MaxTime = 360; _max_CovPar = 0;

varnam = "tviaje"|"cang"|"w_trans"|"B_trans";
_max_ParNames = varnam;

output file=output_cmnl_4_fp.txt reset;
{x,f,g,cov,ret}=maxlik(base,0,&lpr,x0);
call maxprt(x,f,g,cov,ret);
```

