



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

MODELACIÓN DEL PROCESO DE ELECCIÓN DE UNIVERSIDAD

ALEJANDRO SCHMIDT GARIN

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

Profesor Supervisor:
JUAN DE DIOS ORTÚZAR SALAS
RICARDO PAREDES MOLINA

Santiago de Chile, Enero, 2014

© 2014, Alejandro Schmidt Garin



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

MODELACIÓN DEL PROCESO DE ELECCIÓN DE UNIVERSIDAD

ALEJANDRO SCHMIDT GARIN

Tesis presentada a la Comisión integrada por los profesores:

JUAN DE DIOS ORTÚZAR

RICARDO PAREDES MOLINA

JUAN ENRIQUE COEYMANS AVARIA

MARÍA VERÓNICA SANTELICES ETCHEGARAY

MIGUEL NUSSBAUM VOEHL

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

Santiago de Chile, Enero, 2014

A mi familia.

AGRADECIMIENTOS

En primer lugar, quiero dar las gracias a mis profesores supervisores, Juan de Dios Ortúzar y Ricardo Paredes, por estar siempre disponibles para darme sus consejos y guía. Sin su apoyo, mi trabajo no hubiese sido el mismo. Quiero darle un especial agradecimiento al profesor Ortúzar por motivarme a hacer el Magíster y por darme la oportunidad y libertad de investigar en un tema que me apasiona.

Quisiera agradecer también a la profesora Verónica Santelices de la Facultad de Educación, por el tiempo que dedicó a leer mi tesis, a ver mis presentaciones y a darme sus comentarios, los que fueron de gran utilidad para la elaboración de la versión final de este documento.

A los profesores, funcionarios y alumnos de postgrado del Departamento de Transportes y Logística de la Pontificia Universidad Católica de Chile, por todas las instancias que compartimos y que hicieron que mi paso por el Magíster fuera una muy buena experiencia.

Finalmente quisiera agradecer a mi familia y amigos, por estar siempre conmigo, por darme ánimo en los momentos más difíciles y por apoyarme de manera incondicional durante toda mi vida universitaria.

ÍNDICE GENERAL

	Pág.
DEDICATORIA	ii
AGRADECIMIENTOS	iii
ÍNDICE DE TABLAS	viii
ÍNDICE DE FIGURAS.....	xii
RESUMEN.....	xiv
ABSTRACT	xvi
1. INTRODUCCIÓN.....	1
1.1 Motivación	1
1.2 Objetivos	1
1.3 Alcances	2
1.4 Contenido	2
2. REVISIÓN DE LITERATURA	4
3. MARCO TEÓRICO	18
3.1 Modelos de Elección Discreta.....	18
3.2 Teoría de la Utilidad Aleatoria.....	18

3.3	Modelo Logit Multinomial (MNL)	20
3.4	Modelo Logit Jerárquico (HL)	23
3.5	Modelo Logit de Clases Latentes (LCL)	25
4.	CONTEXTO DE LA INVESTIGACIÓN	28
4.1	Descripción General	28
4.2	Acreditación	29
4.3	Rankings	30
4.4	Matrícula y Cobertura	31
4.5	Prueba de Selección Universitaria	32
4.6	Sistema de Postulaciones	33
5.	INFORMACIÓN RECOPIADA PARA LA MODELACIÓN	35
5.1	Datos Disponibles	35
5.1.1	Base de datos Prueba de Selección Universitaria (PSU)	35
5.1.2	Base de datos de las universidades (BDU)	35
5.1.3	Ranking revista Qué Pasa (RQP)	36
5.2	Banco de Datos de Estimación	36
5.2.1	De carreras a universidades	36

5.2.2	Filtros realizados.....	37
5.3	Análisis del Banco de Datos	39
5.3.1	Variables socioeconómicas de los individuos	39
5.3.2	Atributos de las alternativas.....	48
6.	ESTIMACIÓN DE MODELOS Y RESULTADOS	54
6.1	Constantes Específicas	54
6.2	Choice-set.....	55
6.3	Modelo Logit Multinomial (MNL)	55
6.4	Modelo Logit Jerárquico (HL)	56
6.4.1	Estructura de nidos.....	56
6.4.2	Modelo estimado.....	57
6.5	Modelo Logit de Clases Latentes (LCL).....	61
6.5.1	Modelo con dos clases latentes (2LCL).....	61
6.5.2	Modelo con tres clases latentes (3LCL)	67
6.6	Comparación de Modelos	73
6.6.1	Disposición al pago por atributos	73
6.6.2	Capacidad predictiva de los modelos.....	77

6.7	Aplicación Práctica	81
7.	CONCLUSIONES Y RECOMENDACIONES	84
	BIBLIOGRAFÍA.....	87
A.	ANEXO 1: Análisis de Datos	96
B.	ANEXO 2: Modelación	102

ÍNDICE DE TABLAS

	Pág.
Tabla 4.1: Cantidad de alumnos según tipo de institución.....	29
Tabla 4.2: Puntajes promedio PSU	33
Tabla 5.1: Expectativas de obtener una beca e ingreso.....	46
Tabla 5.2: Estadísticas de postulaciones por estrato de ingreso.....	47
Tabla 5.3: Atributos de las alternativas.....	49
Tabla 6.1: Estimación modelo HL basado en Paredes <i>et al.</i> (2012)	58
Tabla 6.2: Parámetros promedio según quintil de ingreso	60
Tabla 6.3: Composición de clases modelo 2LCL	62
Tabla 6.4: Estimación modelo 2LCL	63
Tabla 6.5: Parámetros promedio según quintil modelo 2LCL	65
Tabla 6.6: Composición de clases modelo 3LCL	68
Tabla 6.7: Estimación del modelo 3LCL	69
Tabla 6.8: Parámetros promedio según quintil modelo 3LCL	71

Tabla 6.9: Disposición al pago (miles de \$/año) por años de acreditación según ingreso	73
Tabla 6.10: Arancel promedio (millones de \$/año) según años de acreditación	74
Tabla 6.11: Disposiciones al pago (miles de \$/año) por promedio PSU por ingreso	75
Tabla 6.12: Promedio PSU según años de acreditación.....	76
Tabla 6.13: Disposiciones al pago (miles de \$/año) por ranking Qué Pasa según ingreso	76
Tabla 6.14: Ranking Qué Pasa según años de acreditación	77
Tabla 6.15: Características socioeconómicas muestras de estimación y validación.....	78
Tabla 6.16: Proporciones de mercado observada y modeladas.....	79
Tabla 6.17: Capacidad predictiva y test chi-cuadrado FPR	80
Tabla 6.18: Test de razón de verosimilitud directa	81
Tabla 6.19: Universidades con mayor proporción de mercado modelada.	82
Tabla A.1: Listado de carreras	96
Tabla A.2: Categorías de ingreso e ingreso promedio CASEN.	97
Tabla A.3: Definición quintiles de ingreso per cápita.....	97

Tabla A.4: Definición categorías de ingreso.....	98
Tabla A.5: Individuos de acuerdo a Tipo de Dependencia Escolar.	100
Tabla A.6: Estadísticas descriptivas Promedio NEM y Ranking.....	100
Tabla A.7: Estadísticas descriptivas Puntajes PSU.....	100
Tabla B.1: Universidades y Constantes Modales.....	105
Tabla B.2: Modelo MNL Base.....	107
Tabla B.3: Indicadores de Ajuste Modelo MNL Base.....	107
Tabla B.4: Modelo MNL con Variaciones Sistemáticas en los Gustos.	108
Tabla B.5: Indicadores de ajuste modelo MNL con Variaciones sistemáticas en los gustos.....	109
Tabla B.6: Categorías Universidades basada en Paredes et al. (2012).	109
Tabla B.7: Categorías Universidades basada en Reyes y Rosso (2012).	110
Tabla B.8: Universidades y categorías modelo HL.	111
Tabla B.9: Estimación modelo HL basado en Reyes y Rosso (2012).	113
Tabla B.10: Indicadores de ajuste modelo HL basado en Reyes y Rosso (2012).....	114
Tabla B.11: Comparación parámetros de modelos sin y con Proporción Mismo NSE.	115

Tabla B.12: Test de significancia estadística conjunta de parámetros del modelo HL..116

Tabla B.13: Test de significancia estadística conjunta de parámetros del modelo 2LCL.
.....117

Tabla B.14: Test de significancia estadística conjunta de parámetros del modelo 3LCL.
.....118

ÍNDICE DE FIGURAS

	Pág.
Figura 4.1: Evolución de la matrícula educación superior en Chile	31
Figura 5.1: Educación de los padres e ingreso	42
Figura 5.2: Tipo de dependencia escolar e ingreso	43
Figura 5.3: Promedio NEM y Ranking según ingreso	44
Figura 5.4: Proporción del total de puntajes por tramo e ingreso	45
Figura 5.5: Numero de postulaciones según rango de puntaje.....	48
Figura 5.6: Distribución de universidades según promedio PSU	51
Figura 5.7: Distribución de universidades según arancel.....	52
Figura 6.1: Estructura modelo HL basado en Paredes <i>et al.</i> (2012)	57
Figura A.1: Proporción de individuos según quintil de ingreso.....	98
Figura A.2: Composición de la muestra en términos de categorías de ingreso utilizadas.	99
Figura A.3: Nivel educacional de los padres.	99

Figura A.4: Dispersión Ranking y Puntaje PSU.	101
Figura A.5: Proporción de categorías de ingreso según número de postulaciones.	101
Figura B.1: Estructura modelo HL basado en Reyes y Rosso (2012).....	112

RESUMEN

El objetivo general de esta tesis es estudiar los principales elementos en que se basa la elección de universidad en Chile y cuáles son las características individuales que determinan cómo varían las preferencias a lo largo de la población.

Se cuenta con información disponible sobre la jerarquización realizada por los egresados de la educación secundaria al momento de postular al sistema universitario; esta se utiliza para construir un banco de datos que contiene, además, sus características socio-económicas, y datos descriptivos de las universidades (*Años de Acreditación, Ranking Qué Pasa y Localización*) y de las carreras (*Arancel, Promedio PSU y Duración*).

Con este banco de datos, se estimó un modelo Logit Jerárquico que permite considerar la estructura de correlación existente entre distintos grupos de universidades. También se estimó modelos Logit de Clases Latentes, con el objetivo de capturar la heterogeneidad existente entre los individuos de la muestra e identificar grupos con patrones de comportamiento similares. El mejor modelo estimado (presenta mejores indicadores en la estimación y se ajusta de mejor manera a la base de datos de validación) tiene tres clases latentes que están asociadas a: (i) individuos de ingreso alto, (ii) individuos de ingreso bajo/medio de alto rendimiento académico, y (iii) individuos de ingreso bajo/medio de bajo rendimiento académico.

Se encuentra que las personas eligen de forma racional, aumentando la probabilidad de escoger una universidad si disminuye su costo o aumenta su calidad, y prefiriendo – *ceteris paribus* – instituciones localizadas en la misma provincia que sus hogares. También se confirma la existencia de categorías de universidades (en cuanto a la percepción individual de su nivel de complejidad/calidad) en que algunas son claramente preferidas sobre otras.

Palabras Clave: Elección Discreta, Elección de Universidad, Logit de Clases Latentes

ABSTRACT

The main goal of this thesis is to study the main elements behind the choice of college process in Chile and which are the individual characteristics determining the variation of preferences among the population.

We had information about the ranking made by high school graduates when they apply to the university system (i.e. different college alternatives); this data was used to generate an estimation sample databank containing the revealed preferences of individuals, their socioeconomic characteristics, and descriptive data about the various college alternatives (*Years of Accreditation, Ranking and Location*) and the different careers offered (*Cost, Average Test Score of Enrolled Students and Duration*).

Using this databank, we estimated a Hierarchical Logit model that allowed us to introduce the correlation structure among different groups of universities. We also estimated a Latent Class Logit model, with the aim of capturing the heterogeneity among individuals in the sample and identify groups of individuals exhibiting similar behaviour. The best model (with best goodness-of-fit indicators and with better adjustment to a validation sample) has three latent classes which are related to: (i) high income individuals; (ii) individuals with low/mid income but high academic achievement, and (iii) individuals with low/mid income and low academic achievement.

We found that individuals are rational choice makers, that the odds of choosing a given university increases with a reduction in cost or with an increase in quality, and that – all other things being equal – people tend to prefer colleges located in the same province of their home. We could also confirm the existence of categories of universities (in terms of the individual perception of their complexity/quality level) with a clear preference for some of them.

Key Words: Discrete Choice, University Choice, College Choice, Latent Class Logit

1. INTRODUCCIÓN

1.1 Motivación

La matrícula de la educación superior en Chile ha ido en un constante aumento en los últimos años. En el período 1990-2009 presentó un aumento superior al 240% (SIES, 2010), llegando a poco más de 835.000 alumnos en total, alcanzando una cobertura del 30,8% para el rango entre 18 y 24 años.

Esta mayor cobertura ha ido de la mano con más oportunidades pero éstas difieren enormemente. Más carreras, distintos precios y distinta calidad, son rasgos distintivos del sistema chileno y es también la tendencia mundial. En un contexto en el que muchas de las políticas de desarrollo de la educación superior no diferencian entre universidades, donde la calidad es medida - en lo fundamental- a través de indicadores objetivables, y donde las percepciones y, mejor, las preferencias reveladas por parte de los estudiantes pueden contribuir enormemente a entender el concepto amplio de calidad, es relevante entender la forma en que los egresados de la educación media toman decisiones.

Los estudiantes que egresan de la educación media se ven enfrentados, en primer lugar, a la decisión de si seguir estudiando (y pasar a formar parte del 30,8% que es parte del sistema de educación superior), o si empezar a trabajar pasando a formar parte del mercado laboral. Una segunda decisión, una vez que se ha tomado la opción de seguir estudios superiores, es qué carrera elegir, decisión que tiene directa relación con la vocación del individuo, pero también con sus capacidades intelectuales y económicas. Finalmente, los individuos se ven enfrentados a la elección entre las distintas universidades que componen el sistema de educación superior.

1.2 Objetivos

Esta tesis busca estudiar los determinantes de la elección de universidad y si, en particular, la decisión de los individuos se relaciona con variables observables como

costo, calidad y localización. Es nuestra preocupación analizar también el efecto que pueden tener diferentes características socioeconómicas en la decisión, a fin de diseñar políticas públicas relativas al acceso, la información y distintas políticas sectoriales.

Como señalamos, determinar si los individuos consideran grupos o categorías de universidades en su proceso de elección es relevante no solo para entender el proceso sino para el diseño institucional. Para modelar la decisión asociada a “en qué universidad estudiar”, se han utilizado generalmente modelos de elección discreta bastante sencillos, como el logit multinomial y el logit jerárquico (Carrasco y Ortúzar, 2002). En esta investigación se comparan los resultados de este tipo de modelos estándar con los de modelos logit de clases latentes, que permiten incorporar en mejor forma la heterogeneidad de la población.

1.3 Alcances

La disponibilidad y calidad de la información suelen ser un insumo crítico en la estimación de modelos; en este caso estimamos modelos utilizando información proveniente de fuentes de fácil acceso y que se encuentran disponibles en forma permanente en el tiempo, a fin que la metodología desarrollada también sea una herramienta disponible a lo largo del tiempo.

1.4 Contenido

La tesis está estructurada en ocho capítulos. En el siguiente se presenta una revisión bibliográfica de las principales investigaciones que han utilizado modelos de elección discreta para modelar la elección de universidad. En el tercer capítulo se presenta el marco teórico de los modelos de elección discreta, se desarrollan los aspectos centrales de la teoría de la utilidad aleatoria, y se definen los modelos logit multinomial (MNL), logit jerárquico (HL) y logit de clases latentes (LCL).

En el cuarto capítulo se describe el sistema de educación superior chileno, a fin de especificar el contexto en que se desarrolla la situación de elección analizada. En el quinto se describen los datos disponibles, se desarrolla el proceso de construcción del banco de datos y se realiza un análisis descriptivo de las principales variables socioeconómicas de los individuos y atributos de las universidades.

En el sexto capítulo se describe el proceso de estimación de modelos, se analizan en detalle los principales resultados obtenidos y se hace una comparación entre modelos a partir de las disposiciones al pago calculadas y la capacidad predictiva de cada uno mediante una muestra de validación. En el séptimo se presenta una aplicación práctica de los modelos para estudiar el efecto de distintas políticas en la decisión de los individuos y cómo cambian las proporciones de mercado de cada universidad, en cuanto a la primera preferencia de los individuos. Finalmente, en el octavo capítulo se presentan las principales conclusiones obtenidas a partir de los resultados de esta investigación.

2. REVISIÓN DE LITERATURA

La decisión de asistir (o no) a la universidad ha sido ampliamente estudiada (Becker, 1964; Jackson y Weatherby, 1975; Christensen *et al.*, 1975; Leslie y Brinkman, 1987; Becker, 1990; Heller, 1997; Ellwood y Kane, 2000; Cabrera y La Nasa, 2000), pero la decisión de a qué universidad asistir ha tenido menos atención. Entre quienes han estudiado esta segunda decisión, se pueden apreciar dos corrientes. Por una parte se encuentran aquellos que basan sus investigaciones en datos de demanda agregados, relacionando la cantidad total de individuos que ingresan a una universidad con características de la misma (Bezman y Depken, 1998; Spence, 1973; Lehr y Newton, 1978; Wetzel *et al.*, 1998). Otra línea se centra en modelar la elección individual como una función de los atributos de las universidades y características socioeconómicas del individuo. Chapman (1981), por ejemplo, desarrolla un modelo conceptual sobre los principales determinantes de la elección de universidad, definiendo cuatro categorías de variables relevantes: (i) características individuales, (ii) personas influyentes, (iii) características de las universidades y (iv) esfuerzos de las instituciones por comunicarse con los individuos. El modelo asume que a partir de estos cuatro grupos de variables el individuo construye una “expectativa general de la vida universitaria” que es finalmente lo que determina su elección de universidad.

Entre los autores que han estudiado el proceso de elección universitaria mediante el uso de herramientas estadísticas, el enfoque más utilizado, es el de modelación de elecciones discretas (McFadden, 1973).

Radner y Miller (1970) presentan la primera aproximación al estudio de este tipo de elecciones mediante modelos de elección discreta. Usando una base pequeña (190 individuos) y alternativas de manera agregada (no asistir, *college* público, universidad privada, etc.). Esta es una representación simplificada de las opciones de los estudiantes al terminar su etapa escolar; sus conclusiones son a nivel agregado. Consideran como

variables explicativas de la elección una interacción entre el nivel de habilidad del individuo (puntaje SAT¹) y la selectividad de la alternativa (definida como el SAT promedio de los alumnos), y la razón entre el costo de la alternativa y el ingreso de la familia. Además, “no asistir” se caracterizó como una alternativa adicional sin costo y se le asignó una “selectividad” igual al SAT promedio de los individuos que no asisten a la educación superior; lo que es cuestionable ya que optar por no ser parte del sistema debiera significar algún beneficio a cambio. A pesar de todas estas limitaciones, Radner y Miller (1970) es relevante por haber introducido el enfoque de elección discreta en la modelación de la elección de universidad.

Kohn *et al.* (1976) desarrollan un modelo secuencial que considera tres etapas: (i) la decisión de continuar o no a la educación superior, (ii) la elección de la universidad condicionada a haber tomado la decisión de continuar estudiando y (iii) la decisión de vivir en el campus o viajar diariamente a estudiar, dada la elección de universidad. El modelo especificado utiliza variables específicas para cada alternativa, tales como los costos de matrícula, alojamiento y comida, la calidad de la universidad (la *selectividad*, como en Radner y Miller, 1970), la amplitud de la oferta educacional y los ingresos por alumno de la universidad (como una *proxy* del nivel de gasto por alumno). Además, considera características individuales tales como ingreso familiar, nivel de habilidad (Radner y Miller, 1970) y distancia desde la casa a la universidad. También incorporan una variable que busca medir la afinidad del individuo con la alternativa, definida como el cuadrado de la diferencia entre su nivel de habilidad y la calidad de la universidad.

En relación a Radner y Miller (1970), este trabajo cuenta con un banco de datos mucho más rico, tanto en términos de la cantidad de observaciones como de las alternativas

¹ *Scholastic Aptitude Test.*

posibles, ya que considera a cada opción universitaria como una alternativa independiente. Además, considera un conjunto de variables para describir cada alternativa más completo.

Sin embargo, una de las principales limitaciones del estudio está asociada al manejo del conjunto de alternativas disponibles (*choice set*) para cada individuo. Como éste no se conoce debe ser estimado, existiendo la posibilidad de introducir sesgos en los resultados si se incluye alguna alternativa superior a las que el individuo realmente considera (Swait y Ben-Akiva, 1986; Horowitz y Louviere, 1995). Niu y Tienda (2008), al estimar una variedad de modelos de elección de universidad, demuestran que diferentes criterios en la conformación del *choice-set* pueden llevar a diferencias significativas en los parámetros estimados. La metodología utilizada en la construcción del *choice-set*, se basa en un modelo que permite calcular la probabilidad de ser admitido en cada alternativa.

Punj y Staelin (1978) realizan un estudio de elección de programa de postgrado en negocios (MBA), con datos de 177 individuos que fueron admitidos en Carnegie-Mellon, Estados Unidos. Entre las variables consideradas se encuentran costo, acceso a becas y créditos, calidad, variables asociadas al ajuste entre individuo y universidad (en términos de calidad), distancia y cantidad de vacantes ofrecidas. Entre los resultados obtenidos, destaca que el costo de la universidad tiene un efecto mayor sobre los estudiantes que solicitan becas, lo cual es consistente con lo esperado (individuos que no tienen la necesidad de ayuda estudiantil debieran tener menor sensibilidad respecto al costo). Se encuentra también que el valor asociado a un aumento de US\$ 1.000 en términos de becas, tiene un impacto positivo mayor que un aumento de la misma magnitud en acceso a créditos; esto es de esperarse dadas las diferencias asociadas a cada tipo de ayuda. Los autores también encuentran un efecto positivo al aumentar la calidad de la universidad, pero un efecto negativo al aumentar la diferencia existente

entre ésta y el nivel de habilidad del individuo, sugiriendo que los postulantes prefieren estudiar en universidades con un nivel de calidad similar a su nivel de habilidad.

Nuevamente, la principal desventaja de este estudio está asociada a la limitada cantidad de observaciones disponibles. Adicionalmente, y por limitaciones computacionales, los autores decidieron considerar solo 20 de las 120 alternativas existentes. Por otra parte, al considerar solamente individuos admitidos en una universidad determinada, se puede suponer que existe un sesgo en los resultados obtenidos si quienes postulan a ella tienen alguna característica especial, teniendo en cuenta – además – que del total de la muestra más del 50% eligió esa universidad.

El artículo de Fuller *et al.* (1982) aparece explícitamente como una refinación del trabajo de Kohn *et al.* (1976). Los autores presentan la elección de universidad basándose en la teoría del capital humano (Mincer, 1974; Freeman, 1986), en que la decisión de asistir a la educación superior y elegir universidad, están relacionadas con los ingresos futuros esperados en contraposición con los costos asociados e ingresos que se dejan de percibir debido a estar estudiando. La variable utilizada para medir los ingresos futuros está relacionada tanto con la calidad de la universidad como con el nivel de habilidad del individuo, ambas medidas en términos del puntaje SAT. Las variables de costo están representadas por el valor del arancel, el costo de vida y el acceso a becas; todas ingresan al modelo divididas por el ingreso familiar del individuo. El ingreso percibido (no percibido) se estima en función de la habilidad personal y de variables sociodemográficas como sexo y etnicidad.

En este estudio, la decisión de ir o no a la universidad y a qué universidad asistir, se trata de manera conjunta, en lugar de secuencialmente como en el caso de Kohn *et al.* (1976). Los autores utilizan un modelo MNL, en que los individuos debían elegir entre las siguientes opciones: (i) todas las alternativas universitarias en que el individuo fue

admitido; (ii) la universidad de acceso libre más cercana a su hogar; (iii) ingresar a la fuerza laboral, y (iv) ingresar a las fuerzas armadas.

Entre los principales resultados se encuentra que para un alumno con un determinado nivel de habilidad, la utilidad asociada a asistir a universidades de mayor calidad aumenta hasta un punto en que empieza a disminuir, cuando la calidad de la universidad supera por mucho la habilidad individual. Hallan también que el efecto de un dólar adicional en el valor del arancel tiene un efecto negativo mayor al de un dólar adicional en el ingreso no percibido, mientras que un dólar adicional en becas tiene un efecto prácticamente igual al de una disminución del mismo tamaño en el arancel.

Oosterbeek *et al.* (1992) estudian la elección entre las cinco escuelas de negocios universitarias en Holanda, basándose en la teoría del capital humano y considerando que el nivel de ingreso de un individuo puede tener una relación directa con la universidad de la que egresa. Los autores formulan y estiman un modelo MNL, considerando como principal variable la estimación del valor presente de los ingresos futuros del individuo (definido como función de sus características socioeconómicas, con parámetros específicos para cada alternativa) luego de haber salido de alguna de las cinco universidades consideradas en el estudio. La metodología utilizada en esta investigación no parece adecuada para modelar una elección como la universitaria, ya que ésta tiene relación directa con las características de cada institución (calidad, amplitud de oferta, vida universitaria, etc.) y éstas no fueron consideradas en el estudio.

En su trabajo de tesis, Varela (1996) modela la elección conjunta de universidad y carrera en el sistema chileno en base a un MNL estimado con un banco de datos correspondiente a cuatro universidades y un conjunto de carreras. Las principales determinantes de la elección son el valor del arancel, la cantidad de vacantes de la carrera (que tiene un impacto negativo) y el porcentaje de profesores con Ph.D., además de un conjunto de variables *dummy* para caracterizar las distintas universidades (y que se

interpreta como el prestigio de la misma). Si bien esta tesis es el primer avance del estudio de elecciones en educación en Chile el conjunto de universidades considerado es sólo una pequeña proporción del total de alternativas en el sistema, el modelo estimado considera en forma simultánea la elección de universidad y carrera, en lugar de haber planteado la elección como un proceso jerárquico de dos etapas: elección de universidad y elección de carrera, y la muestra disponible considera solamente individuos que postularon a la Pontificia Universidad Católica de Chile. Finalmente la experiencia en EE.UU. sugiere incluir variables asociadas a la calidad de distintas alternativas universitarias, tales como el nivel académico promedio de sus estudiantes, y la diferencia entre ésta y el nivel de habilidad del individuo.

Montgomery (2002) estudia la elección de programa de post-grado en negocios, utilizando un modelo logit jerárquico (HL); en un primer nivel plantea la decisión de seguir o no un programa y en qué modalidad seguirlo (i.e. asistir *full-time*, *part-time* o no asistir), y en el segundo nivel, las alternativas de universidad para cada caso. Las variables utilizadas en el estudio se clasificaron en cuatro grupos: costo, calidad de la alternativa, conveniencia de atender, y retornos esperados por atender. Los resultados muestran que la elasticidad de la probabilidad de elegir una determinada alternativa con respecto a la matrícula, si bien negativa es bastante pequeña en términos absolutos lo que muestra poca sensibilidad a variaciones en el costo del programa. Ello también se ve reflejado en que la existencia de programas de ayuda económica parece no tener incidencia en la elección. Este es el primer estudio que incorpora información de un ranking en la función de utilidad, mostrando luego que la elasticidad de la probabilidad de elección con respecto a estar entre los primeros 20 lugares del ranking es positiva y significativamente distinta de cero. Finalmente, el estudio encuentra una gran preferencia por universidades ubicadas en la misma región de residencia del individuo, siendo ésta una de las variables que tienen mayor incidencia en la elección.

Long (2004a) en base a observaciones para tres décadas distintas (70's, 80's y 90's) analiza cómo han cambiado las preferencias individuales a lo largo del tiempo. Estima modelos para cada corte temporal considerado y luego realiza una comparación de los resultados. El estudio toma en cuenta variables relacionadas con la calidad universitaria tales como la razón entre la cantidad de alumnos y la cantidad de profesores; la cantidad de profesores con Ph.D.; el nivel de gasto institucional, y la diferencia entre el percentil SAT del alumno y el de la institución. Para incorporar el costo asociado a asistir a una determinada alternativa universitaria, se incluye como variables el valor del arancel y la distancia. El principal resultado de este estudio es que, con el paso del tiempo, la relevancia del valor del arancel ha disminuido. Con un aumento en US\$ 1.000 la probabilidad de asistir a una determinada universidad caía a menos de la mitad en la década de los 70, en los 90 ésta cae solamente un 30%. En contraparte, las variables asociadas a la calidad universitaria presentan un aumento en su importancia en la decisión. Lo anterior refleja cómo las prioridades de los individuos han pasado de estar centradas en el costo a estarlo en atributos relacionados con la calidad.

En relación al acceso a distintos tipos de becas y créditos, Avery y Hoxby (2004) consideran cómo los distintos formatos de ayuda financiera inciden de manera diferente en las elecciones de los individuos. Avery y Hoxby (2004) muestran que mientras un aumento de mil dólares en becas genera un aumento de 11% en la probabilidad de elección, una disminución de la misma magnitud en el arancel genera un aumento de la probabilidad de elección cercano al 2%. Ello sugiere que los alumnos prefieren recibir un monto mayor en becas. Los autores argumentan que dicho efecto se puede deber a que una disminución en el arancel tiene un impacto mayor en términos de la calidad universitaria (porque implica una disminución del presupuesto), mientras que el aumento en el monto de las becas no tiene necesariamente ese impacto. Otro resultado interesante del estudio, es que la valoración individual de los créditos, y opciones de trabajo y estudio, son cercanas a la valoración de las becas, aun cuando las primeras tienen costos mucho mayores (pago del crédito e inversión en tiempo respectivamente).

Los autores testean también si es que existen aspectos de las ayudas económicas que los individuos debiesen ignorar, como si la beca tiene o no algún nombre, si el individuo recibe la beca concentrada en el primer año, y el porcentaje de la beca respecto al total de costos en que debe incurrir el individuo.

Los resultados muestran que los alumnos priorizan que la beca tenga un nombre y que se concentre en los primeros años, por sobre el monto total de la ayuda percibida. Además, los alumnos se fijan en el porcentaje de sus costos que representa la beca, lo que permite que - bajo ciertas circunstancias - cambien su decisión óptima basándose en el porcentaje recibido más allá del monto total de la beca; esto parece ser una decisión irracional salvo que se tengan restricciones de crédito. Por otro lado, más allá de los impactos que tienen las ayudas estudiantiles sobre la elección, encuentran que si la universidad es la más selectiva (entre las que fue admitido), la probabilidad de elegirla aumenta considerablemente. Lo mismo ocurre si la universidad es la misma en que estudió su padre o algún hermano. Finalmente, si bien encuentran que la distancia, incluida de forma lineal y no lineal, no tiene un efecto sobre la elección, los resultados sugieren que si la universidad está en el mismo estado que el hogar del individuo, esto tiene un efecto positivo en la probabilidad de elección.

Long (2004b), por su parte, compara cómo eligen los individuos en estados con distintos niveles de subsidio a la educación superior, diferenciando tanto por nivel de habilidad como por ingreso. Estima un modelo que incorpora el costo neto que enfrenta cada alumno (arancel), la distancia, el nivel de gasto por alumno, la razón entre profesores y estudiantes, el porcentaje de profesores con Ph.D. y una variable relacionada con la cantidad de aporte estatal que recibe cada universidad.

Concluyen que a medida que aumenta el nivel de subsidio a las instituciones públicas aumenta el porcentaje de individuos que se matricula en este tipo de instituciones. Además, el Estado con mayor subsidio (California) es aquel en que los individuos se

matriculan en universidades de mayor calidad en términos del percentil SAT de sus alumnos y del nivel de gasto por alumno. En contraposición los individuos del estado de Massachusetts deben pagar casi tres veces el valor pagado por un alumno en California para recibir menos (tanto en términos del percentil SAT como del nivel de gasto por alumno).

Sugieren también que al pasar de un sistema de subsidio a las instituciones públicas a un sistema sin ayuda estatal, la cantidad de alumnos que asisten a la educación superior se mantiene prácticamente inalterada (cae en un 1,9%), pero la cantidad de individuos que deciden asistir a una universidad de dos años, en lugar de una de cuatro, aumenta en un 14%. Adicionalmente, la cantidad de individuos estudiando en instituciones públicas cae en un 29%, lo que se ve incrementado en el caso de alumnos de bajos recursos (38%). Long (2004b) concluye que distintos formatos de aporte estatal pueden incidir directamente en la calidad de educación que eligen los individuos, siendo el formato de aportes directos a instituciones públicas el que logra un mayor nivel de “inversión” en capital humano.

Mark *et al.* (2004) desarrollan un experimento de preferencias declaradas (Louviere *et al.*, 2000) sobre la elección de programas de postgrado en Economía Agraria, con el objetivo de obtener la valoración de algunos de los principales atributos considerados en la elección de este tipo de programas. El formato escogido para el experimento es el de jerarquización o ranking (Ortúzar y Garrido, 1994), en que los encuestados deben ordenar un total de nueve alternativas en orden de preferencia, luego de lo cual se procede a estimar modelos de elección discreta. En su análisis, consideran los siguientes cuatro atributos: distancia, ranking, instalaciones disponibles, y estipendio, cada uno con tres niveles de variación definidos en el diseño del experimento. Los resultados en términos de la disposición al pago por los atributos no monetarios, muestran que los individuos estarían dispuestos a pagar US\$ 4.848 para pasar de una institución con un ranking entre 11 y 20 a otra con un ranking entre 1 y 10, mostrando una fuerte

preferencia por las universidades de mayor ranking. Además, la disposición a pagar por acceder a programas que dispongan de oficina y computador, puede ser mayor a lo que costaría a las instituciones proveerlos, por lo que dar acceso a este tipo de facilidades puede permitir una disminución en el estipendio, sin necesariamente disminuir cuan atractiva es la institución.

Niu *et al.* (2006) estudian el impacto de la selectividad de las instituciones universitarias sobre la elección individual en Texas. Los autores buscan, además, analizar si los alumnos del 10% superior de cada colegio tienen una forma diferente de elegir (en el contexto de una nueva política de admisión universitaria). El efecto de la selectividad se mide mediante una escala de cinco puntos, que se incluyen en la estimación mediante un conjunto de variables dicotómicas. Utilizan como variables explicativas de la elección al costo, el porcentaje de ayuda financiera recibido, la cantidad de vacantes, la distancia y variables *dummy* asociadas al tipo de institución (privada, no en Texas). Encuentran que al aumentar el nivel de selectividad de una institución su probabilidad de ser elegida aumenta, constatándose una considerable diferencia entre las universidades más selectivas y el resto. Este efecto es aún mayor para los estudiantes que se encuentran dentro del 10% superior de sus colegios. Por otra parte, se encuentra que alumnos provenientes de diferentes tipos de colegio tienen distinta sensibilidad frente al nivel de selectividad de las universidades. Por ejemplo, los alumnos de colegios pobres tienen una valoración muy inferior a la de los alumnos provenientes de colegios ricos y de colegios asociados a instituciones universitarias.

Drewes y Michael (2006) utilizan un modelo MNL para estudiar la elección universitaria de individuos en la provincia de Ontario, Canadá. Dado que el sistema de postulación a las universidades requiere que los alumnos realicen un ranking de sus preferencias, los autores utilizan la técnica de explosión del ranking para obtener la mayor cantidad de información posible a partir de las observaciones disponibles (Chapman y Staelin, 1982). En el sistema de Ontario, el promedio de postulaciones por

persona es 2,77 pues si se quiere inscribir más de tres opciones se debe pagar un valor adicional por cada postulación extra.

Para generar el conjunto de alternativas disponibles, los autores supusieron que, dado el formato de postulación, las alternativas incluidas en el ranking eran superiores a todas las demás, por lo que consideraron un conjunto completo, con todas las alternativas disponibles, en lugar de uno restringido por las postulaciones realizadas. Dada la posibilidad de sesgo por inclusión de alternativas, reestiman el modelo considerando solamente a los alumnos de más alto rendimiento, argumentando que éstos no debiesen autoexcluirse de algunas alternativas, presentando su ranking real de alternativas y no uno condicionado por las expectativas de admisión. Dado que en el sistema de Ontario los aranceles son fijos, los autores decidieron incluir tres variables relacionadas con el costo de asistir a una determinada universidad: (i) costos no obligatorios asociados a la vida universitaria; (ii) gasto por alumno de la universidad en becas, y (iii) distancia entre el hogar del individuo y la universidad. Además, incluyeron un conjunto de variables asociadas a la calidad de la universidad, entre las que se encuentran el gasto en salarios, el porcentaje de las clases del primer semestre que tienen menos de 25 alumnos, el promedio de notas de los alumnos que ingresan y una variable asociada a la calidad de investigación de la universidad. Por último agregaron un ranking que se publica anualmente en una revista de circulación nacional.

Entre los resultados se observa que al aumentar los costos asociados a la vida universitaria aumenta la probabilidad de elegir una determinada opción. Los autores argumentan que esto, más que una inconsistencia, está asociado a la calidad de los servicios ofrecidos (suponiendo que a mayor costo, mejor es el servicio recibido). Las elasticidades obtenidas para la distancia y gasto en becas son las esperadas. Se aprecia que la muestra completa es más sensible a la distancia y menos sensible a las becas que la muestra de alumnos con alto nivel académico; esto puede reflejar que alumnos de menor nivel académico le den mayor relevancia a la localización de la universidad y, a

su vez, que tengan menos expectativas de ser los receptores de las becas que entregan las universidades.

Con respecto a la calidad de los alumnos que ingresan, al estimar el modelo utilizando la muestra completa se obtuvo una elasticidad negativa, pero al reestimarlos utilizando solamente los alumnos de buen rendimiento la elasticidad resultó positiva, como era de esperarse, y de una magnitud bastante mayor al de la muestra completa. Esto puede indicar que los individuos de bajo rendimiento académico tienen una aversión importante a postular a universidades con alumnos de alto nivel. Finalmente, con respecto a la influencia del ranking, los autores argumentan que éste solo es relevante en el caso de universidades pequeñas o con pocos años de existencia (dado que se conoce poco de ellas), mientras que para las universidades más tradicionales no sería un factor relevante en la decisión.

Carter y Curry (2011) analizan la elasticidad de la probabilidad de elección de una determinada universidad con respecto al valor del arancel. La metodología incluye el uso de encuestas de preferencias declaradas (PD), cuya principal diferencia con estudios anteriores es que el conjunto de elecciones se construye a partir de las alternativas realmente consideradas por cada individuo. Consideran cuatro niveles de universidad en el conjunto de elecciones individual: (i) una universidad focal, en la que se centra el estudio; (ii) dos universidades primarias; (iii) una universidad secundaria, y (iv) dos universidades terciarias. La primera es fija y las demás se incorporan a partir de información entregada por los individuos. La encuesta de PD presentaba siete alternativas, de las cuales seis eran universidades, descritas únicamente por el nombre de la institución y el valor del arancel; la última era una opción asociada a no elegir ninguna alternativa. Los encuestados debían responder 19 situaciones de elección, de las cuales la primera no se utilizó en la estimación, por ser considerada “precalentamiento” del experimento, y dos de las dieciocho restantes fueron utilizadas como muestra de validación.

El modelo estimado considera dos variables principales, una asociada al arancel y otra asociada al nivel de la alternativa (focal, primaria o secundaria). Adicionalmente se incorporó interacciones asociadas a las características individuales. Los resultados del estudio incluyen elasticidades de la matrícula inusualmente altas en comparación con resultados obtenidos en estudios anteriores (Leslie y Brinkman, 1987; Heller, 1997; Montgomery, 2002). Los autores los atribuyen al tipo de experimento utilizado, en que se cuenta con una universidad focal en que se centra la investigación y luego con alternativas a ésta. Otra explicación podría ser el diseño utilizado, dado que el experimento se construyó a partir de una alternativa que aparecía en los conjuntos de elecciones de todos los individuos de la muestra y a la que además asistían todos los encuestados (aunque no necesariamente la eligían en la encuesta).

Sá *et al.* (2012) analizan de manera conjunta las distintas modalidades de vivienda que tienen los alumnos (vivir en su hogar o fuera de éste) y la elección de universidad en Holanda. Estiman un modelo HL con dos nidos. Por una parte, todas las opciones de universidad considerando vivir en la casa de sus padres y, por otra, todas las alternativas considerando vivir fuera del hogar. Entre sus variables descriptivas, utilizan dos medidas diferentes relacionadas con la calidad; por una parte, un indicador agregado de calidad construido a partir de una encuesta en que alumnos de diferentes programas evaluaban distintos aspectos relacionados con la calidad (currículo, instalaciones y docencia entre otros) y, por otra, la cantidad de profesores por alumno, estimando modelos separados con cada una. Adicionalmente, incorporan variables como la diversidad de programas ofrecidos, ubicación de la universidad, la distancia entre el hogar y la universidad, si ésta es religiosa, la proporción de sus compañeros de colegio que tienen intención de ir a cada universidad, el costo de arriendo para cada modalidad de vivienda y un conjunto de variables que describen las posibilidades de consumo de los individuos. No incorporan el arancel de las universidades, porque en Holanda éste es fijo, existiendo diferencias con instituciones de educación superior no universitarias que ofrecen carreras de un perfil más técnico, pero que no fueron incluidas en el estudio. Los autores decidieron

utilizar solamente a los alumnos que postulaban a universidades y que acababan de terminar su educación secundaria para la estimación de los modelos.

Los autores encuentran que los indicadores de calidad utilizados, si bien tienen el signo esperado, no son significativos (lo que los autores atribuyen a que en Holanda la calidad de las universidades es muy similar). La distancia, por su parte, tiene un efecto negativo como era de esperar y un alto nivel de significancia estadística, con un impacto mucho mayor para las alternativas pertenecientes al nido asociado a vivir en el hogar. La proporción de compañeros que tiene intención de asistir a cada universidad aparece como relevante en la elección, y lo mismo ocurre con la variable asociada a si las universidades son religiosas.

Considerando el tipo de datos utilizados, los distintos modelos estimados y los diferentes sistemas de admisión universitaria presentados en la literatura revisada, y teniendo en consideración los datos disponibles para el sistema de educación superior en Chile, se considera que: (i) dadas las características del sistema de admisión universitaria chileno es posible realizar estudios de elección de universidad aplicados a la realidad nacional; (ii) se puede modelar el proceso de elección mediante diversos tipos de modelos utilizados en la literatura, y (iii) es posible incorporar en la modelación los principales atributos presentados en la literatura (calidad, costo, localización), ya que las distintas variables asociadas a éstos se encuentran disponibles para el caso chileno.

3. MARCO TEÓRICO

3.1 Modelos de Elección Discreta

Los modelos de elección discreta permiten tratar situaciones en las que individuos se ven enfrentados a elegir una alternativa entre un conjunto finito de opciones. Son de gran utilidad para la toma de decisiones ya que permiten estimar la valoración de los atributos de las distintas alternativas por parte de los individuos y sirven también como herramienta para realizar predicciones de demanda.

En sus distintas variantes, los modelos de elección discreta han sido ampliamente utilizados en áreas tan diversas como: (i) estudiar la demanda potencial de autos eléctricos (Beggs, 1981; Jensen *et al.*, 2012); (ii) modelar la elección de hora de inicio de viaje (Hess *et al.*, 2007; Arellana *et al.*, 2012); (iii) estudiar las características de los vinos (Lockshin *et al.*, 2006; Palma *et al.*, 2013) entre muchas otras. En particular, estos modelos también han tenido aplicaciones a modelar la elección de universidad (Kohn *et al.*, 1976; Fuller *et al.*, 1982; Montgomery, 2002; Drewes y Michael, 2006), pero en su mayoría se trata de casos en Norteamérica y Europa, con sistemas de admisión universitaria muy diferentes al chileno.

3.2 Teoría de la Utilidad Aleatoria

Esta teoría (McFadden, 1974) tiene como supuesto principal que todos los individuos (q) se comportan de manera racional, es decir, buscan maximizar su utilidad (U_q) sujetos a sus restricciones. También se supone que todos los individuos cuentan con información completa respecto a las alternativas que tienen disponibles. Bajo estos dos supuestos se suele caricaturizar el comportamiento individual modelado como “*homo economicus*”. Así, un individuo q elegirá la alternativa A_i , entre un conjunto $A(q)$ de alternativas, si y sólo si la utilidad de dicha alternativa (U_{iq}) es igual o mayor que la utilidad asociada a las demás alternativas del conjunto de elección.

Desde el punto de vista del modelador, se asume que éste no cuenta con información completa y, por lo tanto, no puede observar la verdadera utilidad que le entrega cada alternativa a un individuo en particular. Para incorporar este efecto se supone que la utilidad se descompone en dos: (i) una utilidad observable o representativa (V_{iq}), asociada a aquellos atributos que el modelador puede observar, y (ii) un término de error (ε_{iq}) asociado a la variabilidad no capturada por la utilidad representativa.

De esta manera se tiene que:

$$U_{iq} = V_{iq} + \varepsilon_{iq} \quad (3-1)$$

y, como se mencionó anteriormente, el individuo q elegirá la alternativa A_i siempre que se cumpla que $U_{iq} \geq U_{jq}$, lo que es equivalente a:

$$V_{iq} + \varepsilon_{iq} \geq V_{jq} + \varepsilon_{jq}, \forall A_j \in A(q) \quad (3-2)$$

lo que puede ser reescrito como (Ortúzar y Willumsen, 2011):

$$V_{iq} - V_{jq} \geq \varepsilon_{jq} - \varepsilon_{iq}, \forall A_j \in A(q) \quad (3-3)$$

De esta expresión se puede desprender que un individuo elegirá la alternativa A_i cuando la diferencia en términos de la utilidad representativa con respecto a todas las demás alternativas disponibles sea mayor que la diferencia entre los errores. Ahora bien, dada la naturaleza aleatoria del término $\varepsilon_{jq} - \varepsilon_{iq}$ el modelador sólo puede plantear una probabilidad de elección:

$$P_{iq} = \Pr \{V_{iq} - V_{jq} \geq \varepsilon_{jq} - \varepsilon_{iq}, \forall A_j \in A(q)\} \quad (3-4)$$

Para calcular P_{iq} es necesario realizar algunos supuestos sobre el término de error y su distribución las cuales darán como resultado distintos modelos de elección discreta (Ortúzar y Willumsen, 2011; Train, 2009).

3.3 Modelo Logit Multinomial (MNL)

El modelo MNL se genera al asumir que los errores (ε_{iq}) distribuyen independiente e idénticamente Gumbel con media cero y desviación estándar σ (Ortúzar, 2000). Con esto, la probabilidad que un individuo elija una alternativa determinada, condicional al conjunto de elección, puede ser expresada como (Domencich y McFadden, 1975):

$$P_{iq} = \frac{e^{\lambda \cdot V_{iq}}}{\sum_{j \in A(q)} e^{\lambda \cdot V_{jq}}} \quad (3-5)$$

donde λ , conocido como parámetro de escala del modelo, está relacionado con la varianza de los errores (ε_{iq}) mediante la siguiente ecuación (McFadden, 1974):

$$\lambda = \frac{\pi}{\sigma\sqrt{6}} \quad (3-6)$$

Por su parte, la utilidad representativa (V_{iq}) se construye a partir de un conjunto de variables observables ($X_{k iq}$), que representan atributos de las alternativas o características socioeconómicas de los individuos, y los parámetros (θ_{ki}) asociados a éstas, deben ser estimados. Además estas variables pueden ser genéricas, si aparecen en las utilidades de todas las alternativas con el mismo parámetro, o específicas, si aparecen

solamente en la función de utilidad de una alternativa. Así, una expresión general de la función de utilidad representativa será:

$$V_{iq} = \sum_k \theta_{ki} \cdot X_{k iq} \quad (3-7)$$

donde k son los distintos atributos que describen las alternativas.

De la ecuación anterior se puede desprender que el parámetro de escala (λ) no puede ser estimado, puesto que se encuentra multiplicando a los parámetros de la función de utilidad representativa (V_{iq}). Como lo que realmente interesa es obtener los parámetros θ_{ki} normalmente se realiza la estimación conjunta de $(\lambda \cdot \theta_{k,i})$, asumiendo implícitamente que λ vale uno y que, por lo tanto, los coeficientes estimados son las utilidades representativas marginales de los atributos.

Para estimar los parámetros θ_{ki} se utiliza el método de máxima verosimilitud. Este requiere, como información de entrada para la estimación, datos observados de los atributos ($X_{k iq}$) para cada alternativa, el conjunto de alternativas disponible ($A(q)$) y la elección observada de cada individuo (Ortúzar, 2000). Así, la función de verosimilitud (L) se define como:

$$L = \prod_q \prod_{A_j \in A(q)} P_{jq}^{g_{jq}} \quad (3-8)$$

donde g_{jq} es una variable *dummy* que vale uno si el individuo q escoge la alternativa A_j y cero en otro caso.

Por lo tanto, la función de verosimilitud es igual al producto de las probabilidades de elección de las alternativas escogidas para los individuos en la muestra. Por simplicidad se suele trabajar con el logaritmo de la función de verosimilitud (log-verosimilitud, l):

$$l = \ln L(\theta) = \sum_q \sum_{A_j \in A(q)} g_{jq} \cdot \ln(P_{jq}) \quad (3-9)$$

esta función tiene un rango de valores entre menos infinito y cero, mostrando los modelos un mejor ajuste a medida que ésta crece.

Una propiedad relevante de los modelos MNL es la de independencia de alternativas irrelevantes (IAI), bajo la cual la razón entre las probabilidades de elección de dos alternativas solamente depende de las utilidades representativas asociadas a éstas y no al resto de las alternativas disponibles:

$$\frac{P_{iq}}{P_{jq}} = \frac{\frac{e^{\lambda \cdot V_{iq}}}{\sum_{k \in A(q)} e^{\lambda \cdot V_{kq}}}}{\frac{e^{\lambda \cdot V_{jq}}}{\sum_{k \in A(q)} e^{\lambda \cdot V_{kq}}}} = \frac{e^{\lambda \cdot V_{iq}}}{e^{\lambda \cdot V_{jq}}} \quad (3-10)$$

Esta propiedad representa una limitación del modelo ya que si se trabaja con alternativas correlacionadas puede inducir a resultados contra-intuitivos (McFadden, 1974).

Variaciones Sistemáticas de los Gustos (VSG)

El modelo MNL, tal como se ha planteado, no permite incorporar diferencias en las valoraciones de atributos entre distintos individuos de la muestra, es decir, se considera que todos los individuos tienen la misma valoración de los atributos.

Una forma de introducir la posibilidad que distintos tipos de individuo tengan distintas valoraciones de los atributos, es a través de las denominadas Variaciones Sistemáticas en los Gustos (Ortúzar y Willumsen, 2011, p. 279). Esta técnica consiste en hacer que los parámetros de cada atributo sean una función de las características individuales. Así por ejemplo, el parámetro θ_{ki} definido anteriormente, pasaría a ser un parámetro θ_{kij} definido de la siguiente manera:

$$\theta_{kij} = \theta_{ki} + \sum_m \theta_{kimq} \cdot X_{m,q}, \forall k, i, q \quad (3-11)$$

donde m son las distintas variables socioeconómicas que caracterizan a los individuos de la base de datos. De esta forma cada individuo que cuente con niveles diferentes de las variables m incluidas en la ecuación, tendrá un parámetro distinto.

3.4 Modelo Logit Jerárquico (HL)

El modelo HL (Williams, 1977; Daly y Zachary, 1978) permite introducir estructuras de correlación entre las alternativas, eliminando el efecto generado por la propiedad de Independencia de Alternativas Irrelevantes (IAI, Ortúzar y Willumsen, 2011). La diferencia entre este modelo y el MNL planteado en el punto anterior, es que las alternativas correlacionadas se agrupan en nidos, de manera tal que los errores (ε_{ij}) asociados a las alternativas pertenecientes a cada nido sean independientes e idénticamente distribuidos. Se postula así que, dado que una alternativa pertenece al nido N , su probabilidad de ser elegida entre las alternativas del nido es un MNL:

$$P(i|N) = P_{ij|N} = \frac{e^{\Delta_N \cdot V_{ij}}}{\sum_{j \in N} e^{\Delta_N \cdot V_{jq}}} \quad (3-12)$$

y la utilidad de las alternativas que no están correlacionadas, así como la utilidad conjunta de las alternativas anidadas, cuentan con un error adicional que es también independiente e idénticamente distribuido. Por ejemplo, suponiendo la existencia de un único nido N , la probabilidad de elegirlo está dada por:

$$P(N) = P_{N,q} = \frac{e^{\lambda \cdot V_{Nq}}}{\sum_{j \notin N} e^{\lambda \cdot V_{jq}} + e^{\lambda \cdot V_{Nq}}} \quad (3-13)$$

donde V_{Nq} , la utilidad conjunta de todas las alternativas anidadas, se define como la multiplicación entre el inverso del parámetro Δ_N , a estimar, y la utilidad máxima esperada (EMU) de todas las alternativas pertenecientes al nido (Ortúzar, 2000):

$$EMU = \ln \sum_{i \in N} e^{\Delta_N \cdot V_{iq}} \quad (3-14)$$

Así, la probabilidad de elegir una alternativa A_i , perteneciente al nido N , está dada por el producto de la probabilidad de elegir el nido (P_{Nq}) y la probabilidad de elegir la alternativa dado que ya se ha elegido el nido:

$$P_{iq} = P_{Nq} \cdot P_{iq|N} \quad (3-15)$$

El parámetro Δ_N , que se obtiene de la estimación del modelo, está relacionado con la correlación ρ_N existente entre las alternativas del nido de la siguiente forma (Carrasco y Ortúzar, 2002):

$$1/\Delta_N = \sqrt{1 - \rho_N} \quad (3-16)$$

3.5 Modelo Logit de Clases Latentes (LCL)

El modelo LCL (Gopinath, 1995; Greene y Hensher, 2003; Walker y Li, 2007; Shen, 2009) tiene como base la teoría de clases latentes (Goodman, 1974; Heckman y Singer, 1984) que plantea que el comportamiento individual está asociado a la pertenencia a una clase, dentro de un conjunto finito de clases existentes en una población. La pertenencia a dicha clase es una variable no observable (latente) por el modelador y, por lo tanto, si se quiere incorporar el efecto de pertenecer a alguna clase, esta debe ser estimada.

En este modelo la pertenencia a alguna de las clases latentes se puede traducir tanto en restricciones en la disponibilidad de alternativas (por ejemplo: individuos que, por sus características, son usuarios cautivos de un modo de transporte) como en diferentes valoraciones de los atributos (por ejemplo: diferencias en la valoración del tiempo de viaje). Estos segmentos de la población - desconocidos por el modelador - pueden estar asociados a estilos de vida, actitudes o hasta a visiones ideológicas o políticas del individuo (Hurtubia *et al.*, 2013).

El modelo LCL se compone de dos sub-modelos: uno de pertenencia a clases y otro de elección específico de cada clase. El segundo determina el comportamiento asociado a cada clase, permitiendo obtener las probabilidades de elección de cada alternativa condicional a la pertenencia a una clase determinada:

$$P(i|s) = P_{iq|s} = \frac{e^{\Delta_s \cdot V_{iq}}}{\sum_{j \in A(q)} e^{\Delta_s \cdot V_{jq}}}, \forall s \quad (3-17)$$

donde s caracteriza a cada uno de los S segmentos existentes en la población.

Por su parte, dado que la clase a la que pertenece cada individuo es desconocida, el modelo de pertenencia permite obtener las probabilidades de que cada individuo

pertenezca a cada una de las S clases. Para esto es necesario definir una función de pertenencia (F_{sq}) de la siguiente forma:

$$F_{sq} = \sum_l \delta_{ls} \cdot X_{ql} + \xi_{sq}, \forall s \quad (3-18)$$

donde X_{ql} es un conjunto de atributos de los individuos (que explican la pertenencia a las distintas clases), δ_{ls} es un conjunto de parámetros asociados al peso relativo que tiene cada atributo en la pertenencia a cada clase y ξ_{sq} es un término de error; si se asume que éste distribuye Gumbel independiente e idéntico, se puede plantear la probabilidad de pertenencia a cada clase como (Walker y Li, 2007):

$$P(s) = P_{sq} = \frac{e^{\sum_l \delta_{ls} \cdot X_{ql}}}{\sum_{m \in S} e^{\sum_l \delta_{lm} \cdot X_{ql}}} \quad (3-19)$$

Finalmente se puede obtener la probabilidad de elección de cada alternativa A_i como:

$$P_{iq} = \sum_s P_{sq} \cdot P_{iq|s} \quad (3-20)$$

Ambas formulaciones se estiman de manera conjunta, por máxima verosimilitud, en forma equivalente al modelo MNL. La información requerida para estimar este modelo es la siguiente: (i) atributos de las alternativas, (ii) características socioeconómicas de los individuos y (iii) cantidad de clases latentes en la población.

La cantidad de clases latentes que contiene un modelo se determina como parte de un proceso iterativo de estimación, de manera tal de obtener el número de clases que presente un mejor ajuste estadístico manteniendo la correcta consistencia económica

tanto de los parámetros estimados como de su relación con la composición socioeconómica de las clases a las que pertenecen.

4. CONTEXTO DE LA INVESTIGACIÓN

Este capítulo se ha construido basándose principalmente en el documento de la Comisión Nacional de Financiamiento de la Educación Superior en Chile (2012).

4.1 Descripción General

La educación superior en Chile se divide en tres tipos de instituciones: universidades, que están facultadas para entregar tanto títulos profesionales como técnicos; institutos profesionales (IP), que al igual que las universidades están capacitados para entregar títulos profesionales y técnicos; y centros de formación técnica (CFT), que sólo pueden entregar títulos técnicos de nivel superior. El sistema se estructura de esta manera a partir de una reforma realizada el año 1981, la que tuvo como principal cambio permitir la creación y funcionamiento de instituciones de educación superior privadas sin financiamiento estatal, Esto generó la formación de dos grupos de universidades; por una parte aquellas que existían antes de 1981, las que pasaron a formar parte del Consejo de Rectores de las Universidades Chilenas (CRUCH) y, por otra, todas aquellas que se crearon luego de la reforma, que son denominadas universidades privadas.

Luego de la reforma aumento el número de instituciones existentes, llegando el año 2011 a haber 60 universidades, 45 IP y 73 CFT. Del total de universidades, 25 forman parte del CRUCH (pudiendo así acceder a fuentes de financiamiento estatal) y 35 conforman el grupo de universidades privadas. En la Tabla 4.1 se puede apreciar que, en promedio, los CFT son las instituciones con un menor número de vacantes ofrecidas, mientras que las universidades del CRUCH ofrecen la mayor cantidad de vacantes en promedio. Es interesante notar que el mínimo número de alumnos en una universidad del CRUCH es de 3.511, lo que indica que todas las universidades de este grupo tienen una masa crítica de estudiantes considerable, cosa que no ocurre para las demás instituciones de educación superior. Además, las universidades del CRUCH presentan el

menor coeficiente de variación entre todas las instituciones, por lo que la dispersión de la cantidad de alumnos entre universidades es baja.

Tabla 4.1: Cantidad de alumnos según tipo de institución

	Promedio	Desviación estándar	Mínimo	Máximo
CFT	2.320	6.716	52	44.386
IP	6.780	12.690	67	56.700
U. Privada	10.173	9.868	246	39.193
U. CRUCH	12.429	7.506	3.511	33.001

Fuente: SIES (2011)

4.2 Acreditación

El sistema de educación superior chileno cuenta, a partir del año 2006, con un Sistema Nacional de Aseguramiento de la Calidad de Educación Superior, que cumple con el objetivo de verificar y promover la calidad en las instituciones de educación superior a través del licenciamiento de instituciones nuevas, la acreditación tanto de las distintas instituciones como de las carreras que éstas imparten y la entrega de información para el funcionamiento del sistema. El proceso de licenciamiento está a cargo del Consejo Nacional de Educación (CNED) y es obligatorio para las nuevas instituciones que aspiran a tener autonomía. De no sortearse con éxito, el organismo a cargo del proceso debe solicitar al Ministerio de Educación el cierre de la institución.

Una vez que las instituciones obtienen la autonomía se pueden someter de manera voluntaria a procesos de acreditación institucional, siendo éste un requisito para que sus alumnos puedan optar a ayudas estudiantiles provenientes del gobierno; este proceso está a cargo de la Comisión Nacional de Acreditación (CNA). Este proceso, además de determinar si una universidad está acreditada o no, entrega la cantidad de años por los que la institución está acreditada (entre 1 y 7) y en qué áreas está acreditada (gestión

institucional, docencia de pregrado, infraestructura y equipamiento, investigación, docencia de postgrado y vinculación con el medio); ambos indicadores entregan información relevante a la hora de decidir en qué institución realizar estudios superiores.

4.3 Rankings

Por otra parte existen distintos rankings privados, tanto nacionales como internacionales, que sirven como fuente de información para los alumnos a la hora de decidir, ya que son interpretables como indicadores de calidad de las distintas instituciones. Los principales ranking disponibles para las universidades del país son el ranking Iberoamericano SIR-SCIImago y los publicados por las revistas *América Economía* y *Qué Pasa*. Si bien estos ranking entregan información relevante sobre la calidad de las distintas universidades y carreras, tienen como principal problema el ser elaborados con datos auto reportados y no auditados, por lo que no es posible tener certeza de la veracidad de la información utilizada. En particular, Reyes y Rosso (2012) plantean que para darle mayor validez a las comparaciones se deberían agrupar las instituciones en tres grupos distintos: (i) universidades con énfasis en la docencia, investigación y programas de doctorado; (ii) universidades con énfasis en la docencia e investigación focalizada, y (iii) universidades con énfasis en la docencia. Argumentan que dentro de cada grupo existe la suficiente variabilidad entre instituciones como para que se pueda realizar un ranking.

De los ranking existentes se puede notar claramente que hay dos universidades por sobre el resto del sistema; la Universidad de Chile y la Pontificia Universidad Católica de Chile. Parece importante notar lo anterior, debido a que se espera que las postulaciones tengan una clara tendencia hacia estas universidades. A continuación se puede notar la existencia de otro grupo, con un mayor número de universidades, que se encuentran en un segundo nivel de calidad; en este se encuentran, entre otras, la Universidad Austral, la Universidad de Concepción, la Universidad de Santiago de Chile y universidades privadas como la Universidad Nacional Andrés Bello y la Universidad Adolfo Ibáñez.

Finalmente hay un grupo final que contiene a todas las demás universidades, con un menor nivel de calidad.

4.4 Matrícula y Cobertura

En los últimos años la cantidad de alumnos matriculados en las distintas instituciones de educación superior ha aumentado sostenidamente, pasando de cerca de 165.000 el año 1983 a más de un millón el 2011. En la Figura 4.1 se puede apreciar cómo ha sido el aumento para los tres distintos tipos de instituciones existentes, constatándose que entre los años 2008 y 2012 los IP han experimentado las mayores tasas de crecimiento con un 16,6% anual en promedio, seguidos de los centros de formación técnica con un 10,1%.

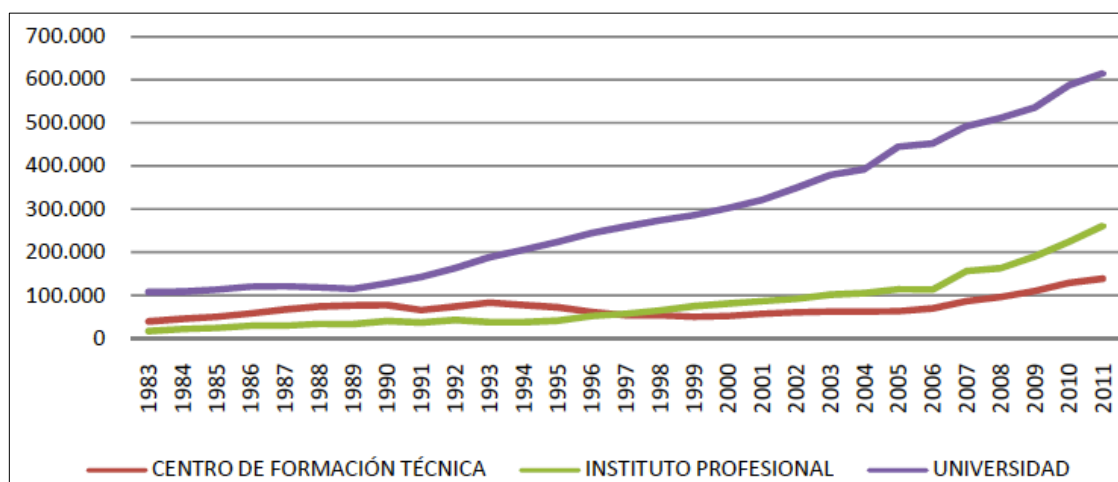


Figura 4.1: Evolución de la matrícula educación superior en Chile

Fuente: SIES (2011)

Al separar las universidades del CRUCH de las privadas, se puede observar que las primeras presentaron un crecimiento absoluto del 5,6% en el período 2008-2012, mientras que las segundas crecieron poco más de un 49% (pasando de una matrícula total de 251.000 alumnos a una de 374.000), siendo así las instituciones que tienen un mayor porcentaje del total de estudiantes matriculados.

Con respecto a la cobertura, el sistema ha llegado a sus más altos niveles, con una cobertura neta del 36,3% el año 2011, mientras el nivel promedio de la OCDE es de un 59% (OCDE, 2011). Por su parte la cobertura bruta aumentó desde un 14,4% en 1990 a un 51,8% en 2011.

4.5 Prueba de Selección Universitaria

El instrumento utilizado para la selección de alumnos por parte de las universidades en la actualidad es la Prueba de Selección Universitaria (PSU). Esta prueba es rendida por todos aquellos estudiantes que se inscriben previamente en el proceso y que cuentan con los requisitos necesarios para rendirla. La validez del puntaje obtenido es de dos años. La prueba evalúa cuatro áreas diferentes: (i) lenguaje y comunicación, (ii) matemáticas, (iii) ciencias y (iv) ciencias sociales, siendo las dos primeras obligatorias y pudiéndose escoger una entre las dos últimas. Los puntajes obtenidos en estas pruebas se complementan con el promedio de notas de la enseñanza media del alumno. A partir del año 2012 también se considera como instrumento de selección el ranking del alumno en su generación.

En la Tabla 4.2 se presentan los puntajes promedio de la prueba para los individuos que egresaron del colegio el 2011; la primera columna presenta el puntaje promedio de todos los individuos que rindieron la prueba, la segunda el puntaje de quienes postulan y la tercera de aquellos que fueron seleccionados en alguna de sus preferencias.

Tabla 4.2: Puntajes promedio PSU²

	Rinden	Postulan	Seleccionados
Total Individuos	162.743	67.880	54.917
Promedio Lenguaje y Comunicaciones	487,9	570,6	588,8
Promedio Matemáticas	487,9	577,1	597,1
Promedio Ciencias Sociales	488,0	568,2	585,5
Promedio Ciencias	491,4	555,2	571,9

Se puede notar claramente que, para quienes rinden la prueba, el puntaje promedio se encuentra cercano a los 490 puntos, mientras que para aquellos que deciden postular (que son poco más del 40% del total) el puntaje promedio sube considerablemente hasta cerca de los 570 puntos. Finalmente, los individuos que son seleccionados son el 81% de quienes postulan y un 33% del total.

4.6 Sistema de Postulaciones

El total de las universidades pertenecientes al CRUCH más un grupo de ocho universidades privadas tienen un sistema de postulación y selección conjunta que está centralizado en el Departamento de Evaluación, Medición y Registro Educacional (DEMRE) de la Universidad de Chile. Este sistema permite que los individuos que rinden la PSU establezcan un ranking de hasta diez preferencias de combinaciones carrera-universidad. Los individuos pueden decidir si quieren ingresar al ranking de postulaciones una combinación o más. Una vez que el DEMRE ha recibido todas las

² Todas las tablas y figuras que no muestran su fuente de manera explícita son de elaboración propia.

postulaciones, realiza la asignación de individuos a las distintas universidades adscritas al sistema.

5. INFORMACIÓN RECOPIADA PARA LA MODELACIÓN

5.1 Datos Disponibles

Para la realización de este trabajo se tuvo disponible un conjunto de datos, a partir de los cuales se construyó el banco de datos utilizado para estimar distintos modelos.

5.1.1 Base de datos Prueba de Selección Universitaria (PSU)

La base de datos del proceso de postulación 2012 de la PSU cuenta con cerca de 380.000 registros correspondientes al total de individuos que rindieron la PSU el año 2011, y consiste de una muestra desagregada a nivel individual, presentando información: (i) del colegio de egreso, (ii) características socioeconómicas y de localización del individuo, (iii) puntajes obtenidos en la prueba y rendimiento escolar, y (iv) postulaciones realizadas y resultados de las mismas. A partir del procesamiento de la información disponible en esta base de datos se extrajo información relativa a las alternativas consideradas en el conjunto de elección del individuo y su elección de primera prioridad.

5.1.2 Base de datos de las universidades (BDU)

Este base de datos es el principal insumo en términos de información descriptiva sobre las alternativas existentes, en particular: (i) costo de la universidad, (ii) si la universidad es o no parte del CRUCH, (iii) áreas y años de acreditación, (iv) localización, (v) amplitud de oferta educacional, y (vi) duración de los programas disponibles. Adicionalmente, se dispone de un conjunto de variables asociadas a la calidad institucional pero que se consideran de poco conocimiento por parte de los individuos, por lo que no fueron consideradas en la modelación. Se cuenta con información para 53 universidades, entre las que se encuentran todas las consideradas en este estudio.

5.1.3 Ranking revista Qué Pasa (RQP)

Se incorporó en la modelación, como un instrumento para medir la calidad universitaria, un ranking publicado por la revista Qué Pasa (Qué Pasa, 2011), construido a partir de la valoración que tiene el mercado laboral de las distintas instituciones universitarias. Se considera relevante dado que es información disponible de manera gratuita en internet y, por lo tanto, de libre acceso al momento en que los individuos realizan su postulación. Este ranking incluye a todas las universidades a las que se puede postular mediante el sistema DEMRE; en particular, considera a todas las universidades presentes en este estudio, algo que otros ranking disponibles no cumplían.

5.2 Banco de Datos de Estimación

Para construir el banco de datos final, utilizado para la estimación de los modelos, fue necesario procesar la información disponible en las tres bases de datos mencionadas anteriormente a fin de obtener una única fuente de datos que tuviera la información dispuesta de manera tal que permitiera la estimación de los modelos.

5.2.1 De carreras a universidades

La información disponible en la base de datos de postulaciones, tiene información de las carreras a que postuló cada individuo codificadas en más de 1.300 alternativas diferentes de carrera-universidad. Para efectos de esta investigación fue necesario identificar, para cada alternativa, su carrera y universidad a fin de clasificar cada opción en base a dos códigos en lugar de uno. Para realizar la codificación, se debió definir un total de 61 carreras diferentes, con el fin de homologar las carreras de distintas universidades bajo un código común, ya que cada universidad tiene la facultad de poner el nombre que

quiera a las carreras que ofrece (lo que claramente hace más difícil el análisis³). Este proceso permitió también agrupar a todas aquellas carreras en que la especialización se elegía al ingresar (como es el caso de ingeniería en la Universidad Técnica Federico Santa María, en que existen más de 10 códigos diferentes), lo que no es de interés para esta investigación. En la Tabla A.1 del Anexo 1 se encuentran los códigos asociados a cada una de las carreras consideradas.

Adicionalmente, dado que se decidió trabajar con las universidades desagregadas a nivel de provincia (para incorporar el efecto de la localización en la elección), fue necesario generar un código universidad-provincia, lo que llevó a un total de 59 opciones diferentes (aunque en la realidad éstas representan solamente a 33 universidades).

Así, luego de haber realizado la codificación, se obtuvo un total de 910 combinaciones de carrera y universidad utilizadas en la estimación de modelos. La reducción con respecto a los más de 1.300 códigos originales se debe principalmente a la agrupación de carreras.

5.2.2 Filtros realizados

Del total de registros presentes en la base PSU, más de 250.000 fueron eliminados ya que no postulaban a ninguna universidad y, por lo tanto, no existía elección observada. Se eliminó también a todos aquellos individuos que no resultaron seleccionados en las opciones a las que postularon.

³ Por ejemplo, la Pontificia Universidad Católica de Chile ofrece la carrera “Arte”, mientras que la Universidad del Desarrollo ofrece la carrera “Arte y Cultura Contemporánea” pero para efectos del estudio son carreras homólogas. Así, bajo la clasificación realizada ambas fueron codificadas como “Arte”.

Se decidió considerar para la estimación solamente a aquellos individuos que egresaron de la enseñanza secundaria el año 2011, a fin de tener una muestra de individuos comparables entre sí (en distintas generaciones pueden haber cambios en los currículos ministeriales o diferencias en las ayudas estudiantiles recibidas). Luego de este filtro, se dispuso en total de poco más de 45.000 observaciones.

Finalmente, dado que en esta investigación se buscaba estudiar los determinantes de la elección de universidad y no la elección vocacional o de carrera, se dejó en el banco de datos solamente a aquellos individuos que postulaban siempre a la misma carrera. Esta decisión se tomó para dejar fuera de la estimación a componentes vocacionales de la decisión universitaria. Luego de haber realizado este último filtro, el banco de datos contuvo un total de 11.065 observaciones para la estimación de modelos.

La muestra de estimación, luego de haber realizado todos los filtros, tiene un sesgo hacia individuos de ingreso alto, con padres con educación universitaria y provenientes de colegios particulares (que son aquellos con más probabilidades de acceder a la educación universitaria) lo que hace que los resultados que se obtengan deban ser utilizados teniendo en cuenta estas características. Aún así, los modelos utilizados permiten obtener valoraciones diferenciadas por grupos de individuos lo que entrega mayor información al momento de utilizar los resultados obtenidos.

Adicionalmente, en términos de las universidades consideradas en el estudio, hay un claro sesgo hacia universidades privadas con un nivel de calidad superior (bajo distintos criterios), dado que éstas son las que son parte del sistema de postulaciones del DEMRE. Así los resultados obtenidos pueden ser aplicables con mayor seguridad a universidades pertenecientes a éste grupo que a universidades no consideradas en el estudio.

5.3 Análisis del Banco de Datos

5.3.1 Variables socioeconómicas de los individuos

a) Características personales

La muestra de estimación está compuesta por 55,6% de hombres y 44,4% de mujeres. Los individuos de sexo masculino se encuentran sobrerrepresentados en comparación con la base de datos previa a la eliminación de individuos que postulaban a más de una carrera diferente, en que los porcentajes eran 49,7% y 50,3% respectivamente. Se puede desprender de aquí, que las mujeres tienen mayor tendencia a incluir más de una carrera diferente en su postulación; lo anterior también puede tener relación con que, en promedio, las mujeres realizan un mayor número de postulaciones que los hombres.

b) Características familiares

La variable ingreso bruto familiar disponible en el banco de datos tiene doce categorías definidas de acuerdo a la Tabla A.2 en el Anexo 1. A partir de esta variable se decidió construir una nueva variable, lineal en el ingreso, a fin de disponer de un ingreso per-cápita cuyo uso es más adecuado para este tipo de modelaciones. Para realizar lo anterior se decidió asignar a cada tramo el valor promedio del intervalo de ingreso de acuerdo a los datos disponibles en la encuesta CASEN (2009), con el fin de obtener un valor representativo de la realidad nacional⁴ para cada intervalo. Finalmente, para calcular el ingreso per-cápita, se dividió este ingreso por el total de integrantes del grupo familiar.

⁴ Para realizar esto se excluyó de la muestra el 2% de ingreso superior por considerar que generaba una distorsión en el ingreso per cápita estimado para el último intervalo.

En la Tabla A.3 del Anexo 1, se presenta los rangos de ingreso per cápita asociados a la pertenencia en cada quintil de ingreso, que son muy relevantes dado que existe una gran cantidad de ayudas estudiantiles directamente relacionadas a éstos. En la Figura A.1 del Anexo 1, se puede observar la composición del banco de datos en términos de quintiles de ingreso, apreciándose que un 23% de los individuos pertenece al quintil de mayor ingreso per cápita (por lo que se puede sospechar que la muestra se encuentra sobre-representada en términos de individuos de ingreso alto). Aun así, el quintil con menor representación en la muestra tiene un 17% de observaciones, por lo que si bien podría existir una sobre-representación de individuos de ingreso alto, el banco de datos se encuentra bastante balanceado en términos del nivel de ingreso individual.

Por simplicidad, en la interpretación de resultados se decidió trabajar sólo con tres categorías de ingreso: bajo, medio y alto. En la Tabla A.4 del Anexo 1 se presentan los rangos de ingreso per cápita asociados a cada categoría, donde se puede apreciar que los límites definidos se encuentran dentro del rango del segundo quintil para la división entre categorías bajo y medio, y en el cuarto para las categorías medio y alto. Al observar la Figura A.2 del Anexo 1, se puede apreciar que la proporción de individuos en cada categoría es relativamente similar.

En la Figura A.3 del Anexo 1 se puede apreciar la composición del banco de datos en términos del nivel educacional de los padres. Este se presenta tanto por separado como el máximo alcanzado por uno de ellos, ya que existe la presunción que el sólo estar en contacto con individuos de un determinado nivel educacional puede generar diferencias en la valoración de ciertas características de las universidades (sin importar si es el padre o la madre). En la figura se aprecia que más del 40% de los individuos cuenta con alguno de los padres con educación universitaria, mientras que en casi el 40% de los casos los padres terminan, como máximo, la educación media. Dado el bajo número de individuos con padres con educación básica o sin educación, se decidió unir éstas a educación media. En la Figura 5.1 se presenta el máximo nivel educacional alcanzado

por los padres, desagregado por el nivel de ingreso per cápita de la persona. Se puede apreciar una clara relación entre ambas variables, donde individuos cuyos padres tienen un nivel educacional bajo son, mayoritariamente, de bajos ingresos, mientras que los individuos cuyos padres tienen educación universitaria son, en su mayoría, de ingreso alto. Algo particular ocurre con los individuos de ingreso medio, donde no se aprecia una tendencia clara sino que se observa que los padres tienen niveles educacionales repartidos en las tres categorías existentes.

c) Características del colegio

Con respecto al tipo de colegio del que egresan los individuos, en la Tabla A.5 del Anexo 1 se puede apreciar que un 53% proviene de colegios subvencionados, mientras que 29% viene de colegios particulares y 17% de colegios municipales. Se nota una clara sobre-representación de individuos de colegios particulares ya que, del total de individuos que rinden la PSU, éstos son solamente el 12%, mientras que en este banco de datos tienen una representación mucho mayor.

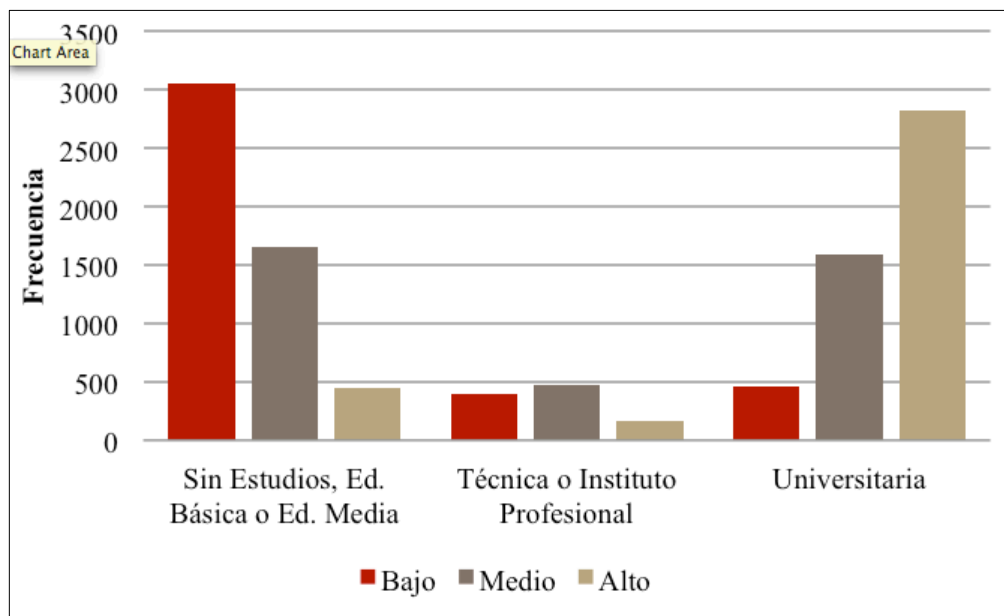


Figura 5.1: Educación de los padres e ingreso

En términos del nivel de ingreso de los individuos de cada tipo de colegio, en la Figura 5.2 se puede apreciar claramente que tanto los alumnos de colegios municipales como subvencionados son principalmente de ingreso bajo, aunque en los últimos existe también una gran presencia de individuos de ingreso medio. Por su parte, un 67% de los individuos de colegios particulares tienen ingreso alto, lo que tiene coherencia con lo esperado ya que, en general, los costos asociados a colegios privados son tan altos que solo pueden ser abordados por individuos de ingreso medio-alto o alto.

d) Rendimiento académico

En términos del rendimiento académico escolar de los individuos, en la Tabla A.6 del Anexo 1 se presenta la estadística descriptiva del *Promedio NEM* y el *Ranking* obtenidos por los individuos al terminar su educación media. El *Ranking* fue incorporado tal que para el mejor alumno tomaras el valor 100% y para el peor 0%. Se aprecia que el valor

medio para el *Promedio NEM* es de 59,5 mientras que el *Ranking* toma un valor promedio de 62%.

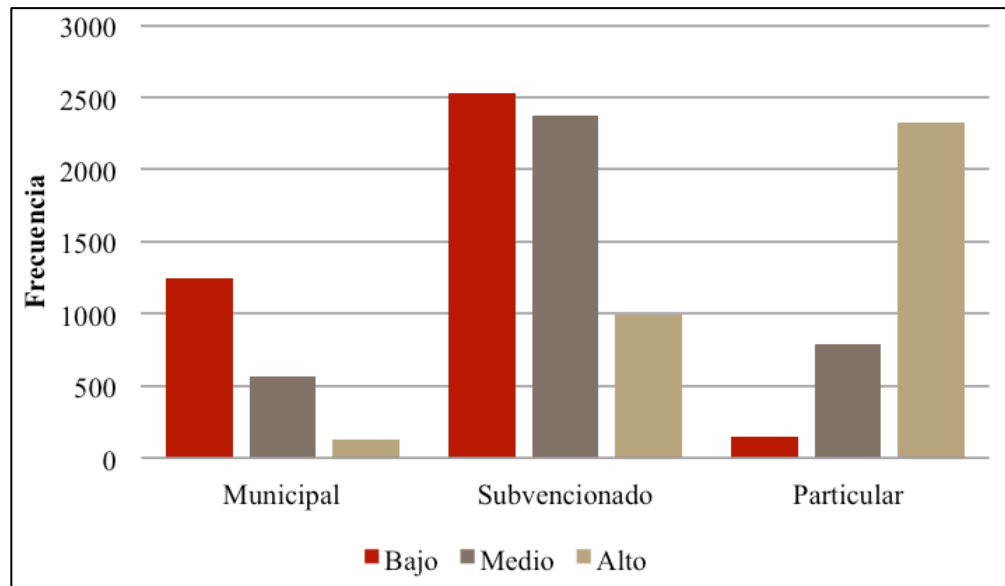


Figura 5.2: Tipo de dependencia escolar e ingreso

En la Figura 5.3 se presenta la relación existente entre el ingreso y las variables de rendimiento escolar consideradas. Se puede notar que a medida que aumenta el ingreso el *Promedio NEM* aumenta, lo que puede deberse a que los alumnos de mayor ingreso asisten a colegios de mejor calidad y, por lo tanto, tienen una mejor formación que los individuos de menor ingreso; alternatively, los sistemas de evaluación existentes en los colegios de ingreso alto pueden tender a evaluar mejor que colegios de menor ingreso. Con respecto al *Ranking*, se aprecia que éste disminuye a medida que aumenta el ingreso. Esto se puede atribuir a que, como la muestra considera solamente alumnos seleccionados por alguna universidad, los individuos de ingreso bajo presentes en ésta son los de mejor rendimiento, mientras que para los individuos de ingreso alto el espectro de rendimiento de los individuos seleccionados es mayor.

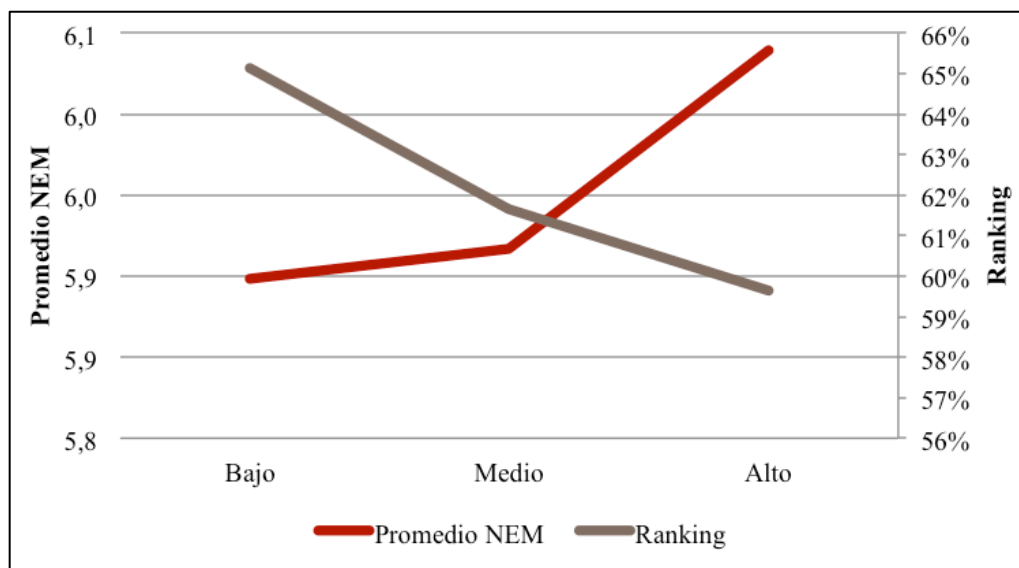


Figura 5.3: Promedio NEM y Ranking según ingreso

Con respecto al rendimiento en la PSU, en la Tabla A.7 del Anexo 1, se presentan los principales estadísticos descriptivos para cada prueba. Al observar la tabla se puede notar que la prueba que presenta los mejores puntajes es la de *Matemáticas*, mientras que la con peor rendimiento es la de *Ciencias Sociales* (aunque esta es la segunda prueba con mayor cantidad de individuos por sobre los 800 puntos).

En la Figura 5.4 se puede observar que sobre los 600 puntos, la proporción del total de puntajes por tramo aumenta con el ingreso, mientras que para puntajes bajo ese límite se produce el efecto contrario, sugiriendo una fuerte correlación entre el rendimiento en la PSU y el nivel de ingreso.

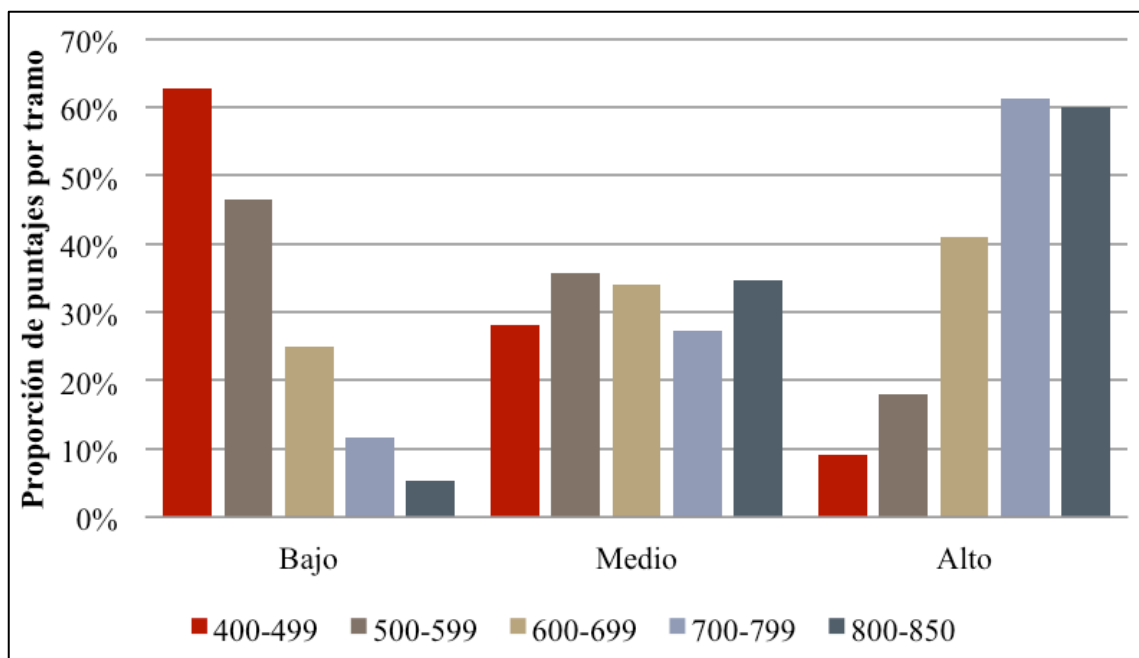


Figura 5.4: Proporción del total de puntajes por tramo e ingreso

En la Figura A.4 del Anexo 1 se muestra la relación entre el *Ranking* del individuo y el *Puntaje PSU*⁵. Se puede ver que los individuos con *Ranking* bajo no logran puntajes altos (por sobre los 700 puntos), mientras que los alumnos con *Ranking* alto obtienen puntajes en todo el espectro. También es interesante notar que la variable *Ranking* parece establecer sólo una cota superior al puntaje, ya que para todos los niveles de esta variable se presentan observaciones con *Puntaje PSU* bajos.

⁵ La variable *Puntaje PSU* se define como el promedio simple entre los puntajes de las pruebas de *Lenguaje y Comunicación* y *Matemáticas*.

e) Expectativas de obtener una beca

En términos de acceso a becas, si bien no se dispone de información sobre su oferta por parte de las distintas universidades, se realizó una estimación del acceso a distintas becas ofrecidas por el gobierno en función de los requisitos para la asignación de las mismas; esta se considera como la expectativa de obtener una determinada beca por parte del alumno, ya que su asignación es posterior al proceso de postulación. Las becas consideradas en el análisis son: (i) Beca de Excelencia Académica (BEA), (ii) Beca Bicentenario (BIC), Beca PSU (PSU) y (iv) Beca Vocación de Profesor (VPR).

En la Tabla 5.1 se puede ver la proporción de individuos que obtendría la beca, bajo la estimación realizada, para cada categoría de ingreso. Se puede notar que tanto la beca PSU como la VPR representan una mínima proporción de los individuos de la muestra, por lo que se consideraron poco relevantes para la modelación.

Tabla 5.1: Expectativas de obtener una beca e ingreso

		BEA	BIC	BPSU	BVPR
Ingreso	Bajo	12,1%	58,9%	0,4%	1,6%
	Medio	8,1%	6,4%	0,6%	1,2%
	Alto	1,1%	0,0%	0,1%	1,1%

La beca BIC es sumamente relevante, con un 58,9% de los individuos de ingreso bajo. Por su parte, la beca BEA que cuenta con un 12,1% de individuos de ingreso en el mismo rango, es también la que presenta la mayor cantidad de individuos de ingreso medio, con un 8,1%. Como era de esperar, los individuos de ingreso alto reciben un muy pequeño porcentaje de las becas.

f) Postulaciones

El número de postulaciones realizado por cada individuo entrega una noción de la cantidad de alternativas que se tienen en cuenta al momento de la elección. En la Tabla 5.2 se puede ver que los individuos de mayor ingreso realizan un mayor número de postulaciones y, además, tienen una menor variabilidad en este número.

Tabla 5.2: Estadísticas de postulaciones por estrato de ingreso

		Desviación		
		Media	Estándar	Mediana
Ingreso	Bajo	3,7	2,2	3
	Medio	3,9	2,1	4
	Alto	3,9	2,0	4

En la Figura 5.5 se presenta el comportamiento de los individuos respecto al puntaje obtenido en la PSU, en términos del número de postulaciones realizadas. Se puede notar que la proporción de individuos que realiza sólo una postulación decrece con el puntaje hasta el intervalo entre 600 y 699 puntos, pero luego sube nuevamente. Esto se puede deber a que cuando las personas obtienen sobre 700 puntos, su certeza de que serán seleccionados en la primera alternativa es muy grande. Lo contrario ocurre con cinco postulaciones o más, donde en el tramo intermedio de puntajes se encuentra la mayor cantidad de individuos. En la figura también se puede apreciar claramente que entre cuatro y seis postulaciones la mayoría de los individuos tiene ingreso alto; mientras que entre una y tres, la mayor cantidad de individuos tiene ingreso bajo o medio.

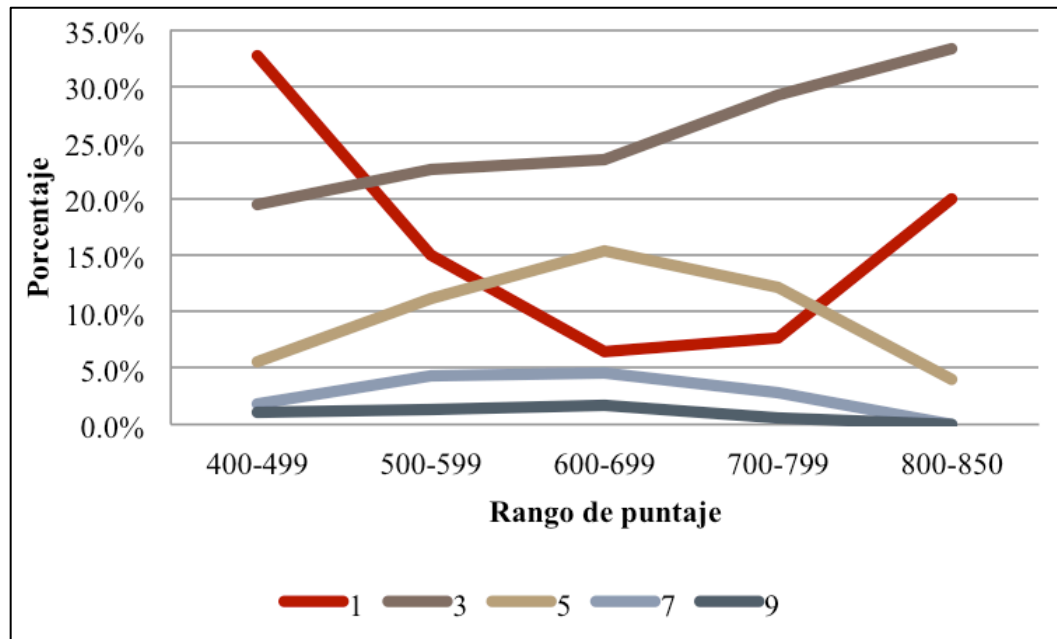


Figura 5.5: Numero de postulaciones según rango de puntaje

5.3.2 Atributos de las alternativas

La estimación de modelos logit se basa en las diferencias existentes entre los atributos de las alternativas disponibles; por esto es relevante presentar la información que se tiene disponible al respecto. Paredes *et al.* (2012) cuentan con una base de datos rica en información descriptiva de las universidades chilenas, la que se utilizó para construir una serie de indicadores agregados de calidad universitaria para el caso chileno. Para este estudio se dispuso de la misma base de datos, que se complementó con información específica de las distintas carreras como el arancel, la duración y el puntaje promedio PSU (habitualmente relacionado a la calidad). Sus principales características se presentan en la Tabla 5.3.

Las variables *Años de Acreditación*, *Ranking Qué Pasa?* y *Promedio PSU*, se consideran como indicadores de calidad universitaria, mientras que *Arancel* y *Duración* tienen relación con los costos asociados a asistir a una universidad determinada.

Tabla 5.3: Atributos de las alternativas

		Media	Desviación Estándar	Mínimo	Máximo
Todas las Universidades	Universidad CRUCH	70,2%	-	-	-
	Años de Acreditación	4,9	1,2	2	7
	Ranking Qué Pasa?	16,4	11,2	1	42
	Promedio PSU	585,9	58,5	478,9	796,08
	Arancel (millones de \$)	2,6	0,9	1,07	5,98
	Duración (semestres)	9,6	1,8	2	15
Universidades Consejo de Rectores	Años de Acreditación	4,9	1,4	2	7
	Ranking Qué Pasa?	17,7	12,4	1	42
	Promedio PSU	589,8	58,9	478,9	796,08
	Arancel (millones de \$)	2,4	0,7	1,07	5,28
	Duración (semestres)	9,7	1,7	2	15
Universidades Privadas	Años de Acreditación	5,0	0,2	5	6
	Ranking Qué Pasa?	13,4	6,8	5	31
	Promedio PSU	576,8	56,5	478,97	789,95
	Arancel (millones de \$)	3,3	0,9	1,61	5,98
	Duración (semestres)	9,4	2,1	2	14

La variable *Años de Acreditación* se puede considerar como una medida de calidad universitaria y está relacionada con el proceso de acreditación al que se someten las universidades chilenas. Al aumentar la cantidad de años de acreditación se presume que mejor es la universidad. En la Tabla 5.3 se puede apreciar que si bien las universidades privadas reciben en promedio más años de acreditación que las pertenecientes al CRUCH, este valor es bastante cercano entre ambos tipos de institución. Así, se espera que el signo estimado para esta variable sea positivo.

La variable *Ranking Qué Pasa?* es un ranking de universidades chilenas que se publica anualmente y tiene relación con la percepción del mercado laboral sobre las distintas instituciones. En la Tabla 5.3 se puede ver que si bien considera 33 universidades diferentes, el valor máximo es 42. Esto se debe a que la variable se incorpora en su escala real y no se ajusta como un ranking exclusivo para las universidades

consideradas, lo que permite que los resultados obtenidos también puedan ser aplicados a universidades con un ranking dentro del rango presentado que no son parte de esta investigación. Se observa, además, que las universidades del CRUCH tienen un ranking promedio de 17,7 claramente inferior al de las universidades privadas (13,4). Esto demuestra que, en promedio, el mercado laboral tiene mejor percepción de las universidades privadas presentes en la muestra que de las universidades pertenecientes al CRUCH. Aún así, es muy importante mencionar que cuatro de las primeras cinco universidades del ranking pertenecen al CRUCH. Se espera que el signo de esta variable sea negativo, ya que mientras menor su valor, mayor debiese ser la utilidad asociada a la alternativa.

La variable *Promedio PSU*, calculada como el valor promedio del puntaje PSU de los individuos que entran a una determinada carrera en una universidad en particular, ha sido ampliamente utilizada para medir la calidad universitaria en términos del nivel académico de los individuos pertenecientes a la institución (Kohn *et al.*, 1976; Montgomery, 2002; Long, 2004). En la Tabla 5.3 se puede observar que el valor promedio que toma la variable para universidades pertenecientes al CRUCH supera por 13 puntos al valor promedio de las universidades privadas, lo que indica que el nivel de los alumnos de las universidades del CRUCH es superior al de las universidades privadas, medido en términos del puntaje obtenido en la PSU. Se espera que el parámetro asociado a este atributo tenga un valor positivo.

En la Figura 5.6 se presentan los puntajes PSU promedio por universidad; se puede apreciar que las universidades privadas se encuentran repartidas en todo el espectro posible de puntajes promedio, contando con universidades tanto entre las tres primeras como entre las tres últimas. Es interesante notar, también, que entre las primeras diez universidades sólo hay una que no esté ubicada en Santiago o Viña del Mar-Valparaíso, y, entre las cinco primeras, solamente hay universidades de Santiago.

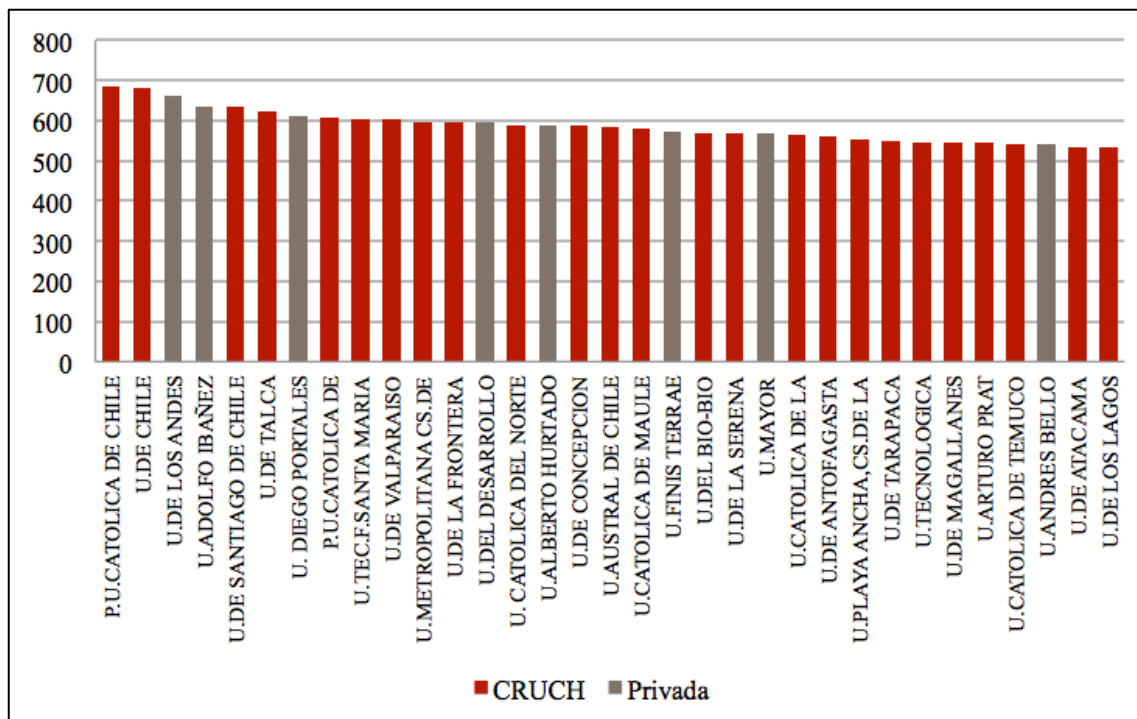


Figura 5.6: Distribución de universidades según promedio PSU

La variable *Arancel* es igual al valor anual, en millones de pesos, de acceder a una determinada carrera en una universidad. Esta variable representa el valor de lista del arancel y no considera posibles descuentos o becas que pueda obtener cada individuo en particular. En la Figura 5.7 se ordenan las universidades de acuerdo al atributo *Arancel*. Se puede notar claramente una concentración de las universidades privadas en el tramo de mayores aranceles. Lo anterior se puede apreciar también al observar la Tabla 5.3, donde se aprecia que, en promedio, el arancel de las universidades privadas es un 37% superior al de las universidades pertenecientes al CRUCH.

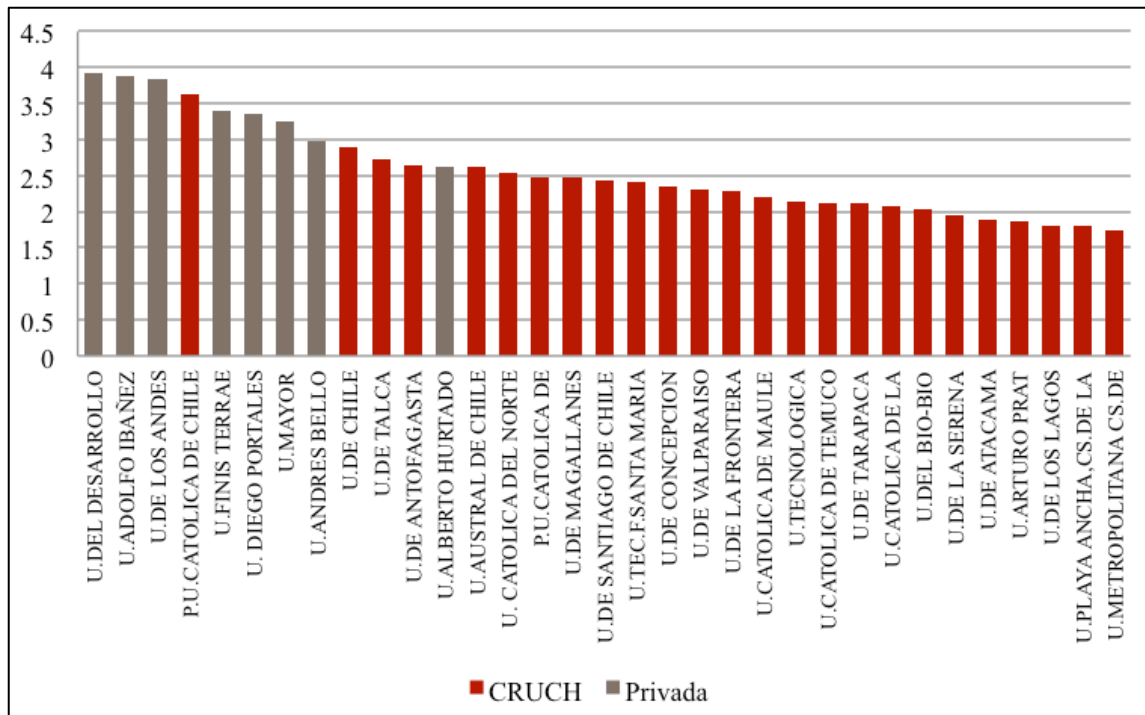


Figura 5.7: Distribución de universidades según arancel

Con respecto a la variable *Duración*, ésta representa el número de semestres considerados en el plan de estudios de cada carrera. Se espera que a medida que aumenta la duración de una carrera, en comparación con la misma en otras universidades, disminuya la preferencia por la institución, por lo que se espera que el signo del parámetro asociado a este atributo sea negativo.

Para incorporar el efecto de la localización en la elección de universidad, se utilizó la variable *MismaProv*, que permite identificar si una universidad se encuentra en la misma provincia que el hogar del individuo. Como se mencionó anteriormente, se decidió trabajar con las alternativas desagregadas al nivel universidad-provincia donde, por ejemplo, las dos sedes de la Pontificia Universidad Católica de Chile (Santiago y Villarrica) pasan a ser dos alternativas separadas, pero los distintos campus de la misma universidad en Santiago se consideran como una única alternativa. *MismaProv* es una

variable dicotómica que toma el valor 1 si la universidad se encuentra en la misma provincia que el hogar del individuo y 0 en otro caso. Al estudiar todas las posibles combinaciones de alternativas y provincias del hogar, se observa que sólo en un 12% de los casos las personas tienen una alternativa en la misma provincia, con un máximo de 27% para los individuos residentes en la provincia de Santiago. Se considera la cercanía entre el hogar y la universidad como algo valioso, por lo que se espera que el signo del parámetro asociado a esta variable sea positivo, aumentando la utilidad de una alternativa si ésta se encuentra en la misma provincia que el hogar del individuo.

A partir de éstos datos, en el próximo capítulo se realizará la estimación de los distintos modelos presentados en el capítulo 3. Se espera que sus resultados sean consistentes tanto con la literatura revisada como con los signos esperados ya discutidos.

6. ESTIMACIÓN DE MODELOS Y RESULTADOS

Para estimar los modelos en esta tesis se utilizó el software de uso libre *BIOGEME* (Bierlaire, 2003). Este cuenta con dos opciones: (i) *Guibiogeme*, que se utiliza para la estimación de modelos más sencillos, como MNL y HL, y (ii) *Pythonbiogeme*, que es más recomendable (por su rapidez) para la estimación de modelos más complejos como el LCL.

6.1 Constantes Específicas

Las constantes específicas se utilizan normalmente en modelación de elecciones discretas para tratar de incorporar elementos que no son capturados por los atributos que se tiene disponibles. Adicionalmente, su utilización permite que, en el caso de modelos MNL, la función estimada reproduzca las proporciones de mercado observadas en la muestra. Dada la gran cantidad de alternativas con que se trabajó en esta investigación, la incorporación de constantes específicas se especificó para grupos de universidades que se consideró similares (a cada grupo se le asignó una constante).

Se definió cuatro grupos de universidades, basándose en un conjunto de rankings discutidos en Paredes *et al.* (2012). En la Tabla B.1 del Anexo 2 se presenta la asignación de cada universidad a las cuatro categorías existentes; como la Universidad Mayor no se encuentra en el ranking, se consideró como parte del grupo de referencia (IV), fijando su valor en cero; por lo tanto, las constantes de los demás grupos representan diferencias con respecto a esta categoría.

En este caso las constantes específicas buscan medir diferencias entre las alternativas en términos de atributos no observables (que no pudieron incluirse en la modelación) tales como selectividad y prestigio entre otros.

6.2 Choice-set

La definición del choice-set o conjunto de alternativas para la elección es clave en la estimación, ya que tiene una incidencia directa sobre los resultados de la misma. Una definición errónea de éste puede inducir a sesgos en los parámetros estimados.

Niu y Tienda (2008) realizan un estudio en que comparan distintos criterios de definición del choice-set para elección universitaria y encuentran que hay diferencias considerables en los resultados de la estimación.

Se consideró como alternativas disponibles a aquellas universidades que los individuos incluían en su ranking de postulaciones. Lo anterior es uno de múltiples criterios posibles en la definición del choice-set y, en caso de que exista individuos que se “autoexcluyen” de algunas alternativas (por obtener un puntaje lo suficientemente bajo como para no tener expectativas de ser seleccionado) podría existir sesgo en algunos de los parámetros estimados, principalmente en aquellos que marcan preferencia por universidades o grupos de éstas (como las constantes específicas utilizadas).

6.3 Modelo Logit Multinomial (MNL)

Se estimó un modelo MNL base, que incorpora solamente atributos de las alternativas sin ninguna interacción y se le consideró como un punto de partida en el proceso de estimación de modelos. En el Anexo 2 se encuentra una descripción detallada del modelo estimado y los resultados obtenidos.

Se estimó también un modelo MNL con variaciones sistemáticas en los gustos (VSG), permitiendo la existencia de valoraciones distintas por parte de grupos de individuos caracterizados en función de las variables socioeconómicas disponibles. Como posteriormente se logró estimar modelos estadísticamente superiores a éste, los resultados de la estimación sólo se muestran en el Anexo 2.

6.4 Modelo Logit Jerárquico (HL)

Para la estimación del modelo HL se usó como partida el modelo MNL con VSG.

6.4.1 Estructura de nidos

Se probó dos estructuras jerárquicas diferentes para la estimación del HL; la primera se basó en el ranking establecido por Paredes *et al.* (2012) y, la segunda en la categorización de universidades chilenas establecida por Reyes y Rosso (2012). En el Anexo 2 se presentan las categorías definidas para ambos criterios y las universidades pertenecientes a cada categoría.

Teniendo en consideración los resultados de la estimación de ambos modelos, se decidió que la categorización generada a partir del ranking presentado por Paredes *et al.* (2012) es más adecuada, ya que incorporaba de mejor manera la estructura de decisión de los individuos de la muestra. En las

Tabla B.9 y Tabla B.10 en el Anexo 2, se pueden ver los resultados de la estimación del HL basado en Reyes y Rosso (2012).

En la Figura 6.1 se presenta la estructura jerárquica definida a partir del ranking de Paredes *et al.* (2012). Se puede apreciar que las universidades pertenecientes al CRUCH son separadas en tres grupos: (i) *Top*, conformado por la Pontificia Universidad Católica de Chile y la Universidad de Chile, (ii) *Alto*, conformado por un grupo de cinco universidades que siguen a las anteriores, y (iii) *Otras*, conformado por todas las demás universidades del CRUCH. Por su parte todas las universidades privadas son agrupadas en un mismo nido.

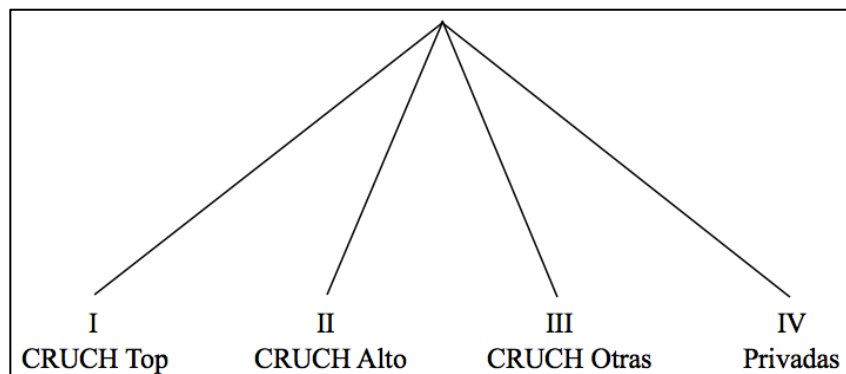


Figura 6.1: Estructura modelo HL basado en Paredes *et al.* (2012)

6.4.2 Modelo estimado

Los resultados de la estimación del modelo HL basado en Paredes *et al.* (2012) se presentan en la Tabla 6.1. Todos los parámetros estimados tienen signos de acuerdo a lo esperado y los parámetros asociados a los nidos son consistentes con la teoría (esto es, estadísticamente distintos de 0 y 1 al 95% de confianza). Se puede ver, también, que las constantes específicas presentan signos y magnitudes de acuerdo a lo esperado, donde a medida que aumenta el puntaje de las universidades de la categoría (ver Tabla B.1 del Anexo 2) aumenta el valor del parámetro asociado. Es interesante notar que la constante de la categoría I es más de cinco veces superior a la de la categoría II, indicando una clara preferencia hacia la Universidad de Chile y Pontificia Universidad Católica de Chile.

Se puede observar que el parámetro asociado a que la universidad se encuentre en la misma provincia que el hogar del individuo es positivo y con gran significancia estadística, lo que sugiere una relevancia alta a la localización de la universidad. Al comparar la magnitud del parámetro con las constantes específicas estimadas, se puede ver que su valor es superior al todas ellas; esto podría interpretarse como que, por ejemplo, un individuo que vive en Talca preferiría, *ceteris paribus*, estudiar en una universidad de la ciudad que hacerlo en alguna de la categoría I. No obstante, hay que

tener cuidado con este tipo de interpretaciones, ya que sabemos – por otro lado – que muchas veces los individuos postulan sólo a las universidades que podrían entrar.

Tabla 6.1: Estimación modelo HL basado en Paredes *et al.* (2012)

Atributo	Parámetro	Test-t
K_I	0,833	(10,3)
K_II	0,160	(2,4)
K_III	0,137	(4,0)
Arancel	-3,820	(-4,7)
Ingreso Medio	1,410	(1,7)
Ingreso Alto	2,530	(2,5)
Beca Bicentenario	3,200	(3,7)
Promedio PSU	0,478	(6,7)
Padres Educación Universitaria	0,129	(1,5)
Ingreso Medio	0,161	(1,8)
Ingreso Alto	0,178	(1,5)
PSU sobre 700	0,601	(5,4)
Colegio Particular	0,563	(6,2)
Duración	-0,023	(-1,8)
Misma Provincia	1,090	(17,6)
Años de Acreditación	6,480	(3,5)
Proporción Mismo NSE	0,481	(3,8)
Ranking Qué Pasa	-0,593	(-2,2)
10% Mejor de su colegio	-1,050	(-2,7)
PSU sobre 700	-2,230	(-3,2)
Padres Educación Universitaria	-1,180	(-2,5)
CRUCH	0,024	(0,4)
Beca Bicentenario	0,401	(3,8)
$\phi_{CRUCH - U y UC}$	0,61	(12,0)
$\phi_{CRUCH - Top}$	0,81	(13,8)
$\phi_{CRUCH - Otras}$	0,65	(17,4)
$\phi_{PRIVADAS}$	0,77	(8,0)
Log-verosimilitud (0)	-11.402,2	
Log-verosimilitud	-9.462,4	
ρ^2	0,170	
$\bar{\rho}^2$	0,168	
# Parámetros	27	

También se puede desprender de los resultados que los individuos parecen preferir universidades a medida que aumenta la proporción de individuos similares a ellos en términos de su nivel socioeconómico. Lo anterior muestra concordancia con la realidad de Chile, en que existen altos niveles de segregación tanto a nivel residencial (Rodríguez y Arriagada, 2004) como escolar (Hsieh y Urquiola, 2003; Elaqua, 2009), por lo que es esperable que las personas prefieran universidades compuestas por individuos similares.

La pertenencia de una universidad al CRUCH tiene un efecto positivo pero con baja significancia estadística. Sin embargo, este efecto se vuelve considerablemente significativo entre aquellos individuos que tienen expectativas de obtener la beca Bicentenario (lo que tiene lógica, ya que esta beca solamente cubre carreras en universidades pertenecientes al CRUCH).

En la Tabla 6.1 también se presentan los principales indicadores de ajuste del modelo HL basado en Paredes *et al.* (2012). Si se hace un test de razón de verosimilitud (LR) respecto al MNL con VSG ($l(\theta) = -9.497,9$), que es su versión restringida, se ve que dada su log-verosimilitud de -9.462,4 se obtiene que $LR = 71$, muy superior al valor de $\chi^2_{gdl=4}$ al 95% (9,49); esto permite rechazar la hipótesis nula que el modelo HL es equivalente al MNL y, por lo tanto, se prefiere al primero.

En la Tabla 6.2 se presentan los parámetros promedio asociados a *Arancel*, *Promedio PSU* y *Ranking Qué Pasa*, que son los atributos en que se incluye la mayor cantidad de variaciones sistemáticas en los gustos, para cada quintil de ingreso.

Tabla 6.2: Parámetros promedio según quintil de ingreso

Quintil	Total	Total BIC	Arancel	Promedio PSU	Ranking Qué Pasa
I	2.285	1.260	-2,055	0,565	-1,169
II	1.973	1.284	-1,487	0,643	-1,298
III	2.135	0	-2,410	0,883	-1,615
IV	2.087	0	-1,950	1,163	-2,240
V	2.585	0	-1,290	1,405	-2,687

En primer lugar se puede notar que el parámetro asociado al *Arancel* decrece en magnitud entre los dos primeros quintiles, para luego crecer en el tercero y volver a decrecer hasta el quinto. El efecto que ocurre en los primeros dos quintiles se explica por la interacción entre *Arancel* y las expectativas de obtener la beca BIC, que tiene como requisito pertenecer a alguno de los dos primeros quintiles y financia hasta el arancel de referencia (por lo que es de esperar que quienes obtienen la beca tengan una sensibilidad menor con respecto al *Arancel* que aquellos que no la obtienen).

Con respecto al *Promedio PSU*, en la Tabla 6.2 se puede apreciar que a medida que aumenta el quintil de ingreso aumenta la valoración de la calidad en términos de los individuos que estudian en cada universidad. Se observa que los individuos del quinto quintil valoran más del doble este atributo que los del primer quintil. Dadas las variaciones sistemáticas en los gustos incorporadas en el modelo, el parámetro aumentará, por una parte, a medida que aumenta el ingreso y, por otra, para aquellos individuos provenientes de colegios particulares que hayan obtenido un muy buen rendimiento en la PSU (sobre 700 puntos) o cuyos padres tengan educación universitaria. Todas las anteriores son variables que presentan una correlación positiva con el ingreso, en especial las dos primeras.

El parámetro asociado a la variable *Ranking Qué Pasa* también presenta una magnitud creciente con el quintil de ingreso, lo que es esperable ya que a mayor ingreso se espera

un mayor conocimiento de las alternativas universitarias y su ranking. En este caso, las variaciones sistemáticas de gustos que resultaron significativas son: (i) si el individuo obtiene más de 700 puntos en la PSU, (ii) si los padres del individuo tienen educación universitaria, y (iii) si el individuo se encuentra entre el 10% superior de su colegio. Como se mencionó anteriormente, las dos primeras se encuentran correlacionadas con el ingreso de manera positiva. Si bien el rendimiento escolar en términos del ranking no tiene relación con el ingreso, se espera que aumente la magnitud del parámetro ya que los individuos con mejor rendimiento escolar tienden a tener mayores expectativas de acceso a la educación superior y, por lo tanto, a estar más informados al momento de elegir entre diferentes alternativas universitarias.

6.5 Modelo Logit de Clases Latentes (LCL)

Para la estimación de LCL se probó con dos, tres y cuatro clases. Finalmente, en esta sección se presentan los mejores modelos de dos y tres clases latentes, ya que los modelos con más clases resultaron inconsistentes.

6.5.1 Modelo con dos clases latentes (2LCL)

Al estimar este modelo no se aplicó restricciones en la disponibilidad de alternativas ni en los parámetros de la función de utilidad por lo que las diferencias existentes entre ambas clases están definidas únicamente por el signo y magnitud de los parámetros estimados. En la Tabla 6.3 se presenta la composición de cada clase latente tras estimar el modelo. En primer lugar es interesante notar que la Clase 2 es más pequeña, con una probabilidad de pertenencia promedio de 39,3% contra un 60,7% de la Clase 1.

Tabla 6.3: Composición de clases modelo 2LCL

		Clase 1	Clase 2
Total		60,7%	39,3%
Ingreso	Bajo	42%	25%
	Medio	35%	32%
	Alto	23%	43%
Beca Bicentenario	-	24%	22%
Sexo	Mujer	45%	43%
Nivel Educativo de los Padres	Escolar	53%	35%
	Técnica o IP	10%	8%
	Universitaria	36%	57%
Tipo de Dependencia Escolar	Municipal	20%	13%
	Subvencionado	60%	42%
	Particular	19%	45%
Rendimiento Académico	10% Mejor del Colegio	19%	20%
	PSU Sobre 700	15%	36%

El perfil de individuo que domina la Clase 2 tiene ingreso alto, proviene de colegio particular (mayoritariamente), obtiene un buen rendimiento en la PSU y sus padres cuentan con educación superior. Se puede suponer también que, dadas las características de cada clase, los individuos de ingreso bajo y medio que pertenecen a la Clase 2 son quienes obtienen un buen resultado en la PSU. Por su parte, la Clase 1 está compuesta principalmente por individuos de ingreso bajo o medio, que egresan de colegios subvencionados y cuyos padres tienen un bajo nivel educacional. En esta clase se concentran los individuos que no obtienen un rendimiento sobresaliente en la PSU.

En términos del ranking de egreso, se puede decir que ambas clases son equivalentes ya que la diferencia en términos de esta variable es prácticamente nula. Lo mismo ocurre para las expectativas de obtener la beca BIC, aunque se puede suponer que, del universo total de individuos que pueden acceder a la beca por las restricciones de ingreso de la

misma, un mayor porcentaje de la Clase 2 la obtiene. Lo anterior se condice con que los individuos de ingreso bajo o medio que pertenecen a esta clase son aquellos con un buen rendimiento académico.

En la Tabla 6.4 se presentan los resultados de la estimación del modelo 2LCL; se puede notar que se han estimado tanto parámetros específicos para cada clase como parámetros comunes a ambas. Además, se puede notar que, para algunos atributos, se ha introducido variaciones sistemáticas en los gustos, aunque en menor medida que las incorporadas en los modelos anteriores. Esto se debe a que en los modelos de clases latentes gran parte de la heterogeneidad existente es capturada por las diferentes clases postuladas.

Tabla 6.4: Estimación modelo 2LCL

	Parámetros Comunes		Clase 1		Clase 2	
K_I	0,709	(7,0)				
K_II-III	0,153	(3,3)				
CRUCH			-0,115	(-1,1)	0,611	(2,8)
Beca Bicentenario	0,266	(1,8)				
Años Acreditación	8,750	(3,7)				
Ranking Qué Pasa			-0,351	(-0,9)	-4,740	(-3,1)
10% Mejor de su colegio	-2,180	(-3,0)				
PSU sobre 700	-2,780	(-2,8)				
Promedio PSU			0,210	(1,9)	4,470	(7,0)
Arancel			-4,250	(-3,7)	-5,140	(-2,6)
Ingreso Medio	1,300	(1,2)				
Ingreso Alto	2,640	(1,9)				
Beca Bicentenario	2,150	(1,6)				
Misma Provincia	1,380	(19,5)				
Proporción Misma CSE	0,821	(5,1)				
Log-verosimilitud (0)			-11.402,2			
Log-verosimilitud			-9.408,5			
ρ^2			0,175			
$\bar{\rho}^2$			0,172			
# Parametros			27			

Al final de la Tabla 6.4 se presentan los principales indicadores de ajuste del modelo 2LCL. Se puede notar que presenta una log-verosimilitud de -9.408,5 por lo que se puede entender que presenta un mejor ajuste que el modelo HL, aunque en la práctica no son comparables estadísticamente.

Respecto a los parámetros estimados, en primer lugar es interesante notar que las constantes específicas II y III han sido agrupadas en una sola. Esto se debe a que al estimar modelos incorporándolas por separado, los valores estimados eran muy similares para ambas. No obstante, la gran diferencia entre la magnitud de la constante I y las demás se mantuvo como en los modelos anteriores.

Al igual que en el modelo HL, el parámetro asociado al atributo *Mismapro* es positivo y de mayor magnitud que las constantes específicas estimadas. Ahora bien, al comparar la razón entre el parámetro y la constante de la categoría I se puede ver que, para este modelo, hay un aumento en la importancia otorgada a que la universidad se encuentre en la misma provincia que el hogar del individuo. La variable *Años de Acreditación*, que se estima como un parámetro común a ambas clases, presenta una significancia estadística alta y su signo es consistente con lo esperado lo que muestra que, en términos generales, los individuos si consideran alguna variable asociada a la calidad al momento de elegir.

En la Tabla 6.5 se presentan los parámetros promedio de cada quintil de ingreso para las variables *Arancel*, *Promedio PSU* y *Ranking Qué Pasa*. Mientras en los modelos MNL y HL la obtención de parámetros para cada individuo es directa, en los modelos LCL es más compleja. Existen clases con parámetros diferentes entre sí y cada individuo tiene probabilidades de “pertenecer” a cada una en función de sus atributos socioeconómicos. Para “recuperar” la valoración de los distintos atributos asignada por el modelo a cada individuo se utilizó la siguiente expresión:

$$\theta_{mq} = \sum_s \theta_{smq} \cdot P_{sq}, \forall m, q \quad (6-1)$$

donde s representa las clases latentes, m el atributo en cuestión y q identifica al individuo. Así la valoración efectiva del atributo m por parte de cada individuo de acuerdo al modelo, es el promedio ponderado del atributo en cada clase por la probabilidad de pertenecer a ésta. Una vez que se tienen los parámetros para cada individuo, se puede calcular el promedio simple por quintil que son los valores presentados finalmente en la Tabla 6.5.

Tabla 6.5: Parámetros promedio según quintil modelo 2LCL

Quintil	Total	Total BIC	Arancel	Promedio PSU	Ranking Qué Pasa
I	2.285	1.260	-3,30	1,35	-2,31
II	1.973	1.284	-2,91	1,59	-2,55
III	2.135	0	-3,25	1,66	-2,74
IV	2.087	0	-2,80	2,13	-3,57
V	2.585	0	-2,10	2,58	-4,33

Al analizar los valores del parámetro correspondiente al *Arancel*, se puede ver que el quintil que entrega mayor relevancia a este atributo es el primero, compuesto por los individuos de menor ingreso, seguido por el tercero. Lo anterior se debe, al igual que en el caso del modelo HL, a la existencia de la beca BIC. En este caso, el efecto que genera esta beca se ve “suavizado” por la heterogeneidad que permite el modelo, acercándose más a la verdadera valoración individual. Al observar los parámetros estimados en la Tabla 6.4, se puede apreciar que las variaciones sistemáticas de gustos introducidas son consistentes con lo esperado, tanto en signo como en magnitud.

Se puede notar que la Clase 2, compuesta por un mayor porcentaje de individuos de ingreso alto, presenta un parámetro del *Arancel* mayor en magnitud que la Clase 1, lo

que *a priori* podría ser considerado un resultado contra-intuitivo. No obstante, no se trata necesariamente de una inconsistencia en la estimación del modelo, ya que se puede deber a que los individuos con mayor probabilidad de pertenecer a esta clase sean aquellos con un mejor rendimiento en la PSU y, por lo tanto, se espera que tengan una mayor sensibilidad respecto al *Arancel* (dado que sus expectativas respecto a la elección de universidad son mayores).

Con respecto al parámetro del atributo *Promedio PSU*, su valor va creciendo significativamente con el ingreso en forma consistente con lo esperado. Al analizar de manera desagregada la valoración de cada clase, se puede notar que para la Clase 1 este atributo tiene una magnitud baja, esto es, en esta clase habría individuos poco sensibles a cambios en el atributo *Promedio PSU*. Por su parte, como la Clase 2 tiene una alta valoración del atributo se puede desprender que ésta estaría compuesta principalmente por quienes otorgan mucha relevancia a la calidad de las alternativas existentes vista en términos del nivel académico de los alumnos pertenecientes a las mismas. Lo anterior tiene concordancia con lo observado para el *Arancel*.

El parámetro asociado al atributo *Ranking Qué Pasa*, aumenta en magnitud con el ingreso. La mayor diferencia se presenta entre los quintiles IV y V, lo que se puede atribuir a que para el último la proporción de individuos con buen rendimiento en la PSU aumenta considerablemente. Adicionalmente, al observar la Tabla 6.4 se puede notar que la significancia estadística del parámetro de la Clase 1 es baja, lo que implica que tiene poca incidencia en la elección. Por el contrario, para los individuos de la Clase 2 tanto la magnitud, como la significancia estadística del parámetro son muy superiores.

El que la universidad pertenezca o no al CRUCH presenta un efecto diferente para cada clase. Por una parte, para la Clase 1 se obtiene un parámetro negativo y poco significativo, lo que se puede interpretar como que los individuos tienden a preferir las universidades privadas (aunque esta no es una variable relevante en la elección); esto se

puede deber al bajo nivel de conocimiento de las alternativas por parte de los individuos pertenecientes a esta clase. Por otra parte, la Clase 2 tiene un parámetro positivo y estadísticamente significativo, lo que se traduce en una preferencia por universidades pertenecientes al CRUCH. Esto se puede explicar por dos motivos: (i) los individuos pertenecientes a esta clase tienen mayores expectativas de obtener becas gubernamentales, las que suelen estar condicionadas a que la institución pertenezca al CRUCH, y (ii) dado que los individuos pertenecientes a esta clase presentan una mayor sensibilidad por el Arancel, el parámetro estimado puede absorber parte del efecto de la diferencia de precios entre universidades del CRUCH y privadas.

6.5.2 Modelo con tres clases latentes (3LCL)

Al igual que en el modelo 2LCL, en este caso no se aplican restricciones ni a la disponibilidad de alternativas ni a los parámetros presentes en la función de utilidad. En la Tabla 6.6 se presenta la composición de las tres clases estimadas. Los individuos de la muestra cuentan, en promedio, con una probabilidad de pertenecer a la Clase 1 del 21,2%, siendo esta la menor de las tres. La Clase 2, por su parte, tiene una probabilidad de pertenencia promedio de 34,2% y la probabilidad de pertenecer a la Clase 3 es, en promedio, de 44,3% siendo la clase que teóricamente debería tener mayor cantidad de individuos en la muestra.

Se puede notar que la Clase 1 está compuesta mayoritariamente por individuos de ingreso alto (57% del total), mientras los individuos de ingreso medio alcanzan solo el 27% y los de ingreso bajo el 16%. Al observar el nivel educacional de los padres, se puede notar claramente que esta clase está compuesta por individuos cuyos padres cuentan con un alto nivel educacional (un 73% tiene educación universitaria). Por otra parte, del total de la clase, un 61% proviene de colegios particulares, un 31% de colegios subvencionados y un 8% de colegios municipales, lo que confirma que la clase está dominada por individuos de ingreso alto. Finalmente, se puede notar que un 37% de los individuos obtiene un puntaje sobresaliente en la PSU.

Tabla 6.6: Composición de clases modelo 3LCL

		Clase 1	Clase 2	Clase 3
Total		21,2%	34,2%	44,6%
Ingreso	Bajo	16%	38%	43%
	Medio	27%	36%	35%
	Alto	57%	26%	23%
Beca Bicentenario	-	13%	32%	21%
Sexo	Mujer	44%	43%	46%
Nivel Educativo de los Padres	Escolar	21%	51%	52%
	Técnica o IP	5%	11%	10%
	Universitaria	73%	36%	36%
Tipo de Dependencia Escolar	Municipal	8%	19%	20%
	Subvencionado	31%	57%	61%
	Particular	61%	23%	19%
Rendimiento Académico	10% Mejor del Colegio	18%	28%	14%
	PSU sobre 700	37%	34%	8%

Tanto la Clase 2 como la Clase 3, presentan una composición similar en cuanto al ingreso, nivel educativo de los padres y tipo de dependencia escolar. De hecho, están compuestas mayoritariamente por individuos de ingreso bajo y medio (cerca del 80%), cuyos padres cuentan con un bajo nivel de educación y que provienen principalmente de colegios subvencionados. La diferencia entre ambas se encuentra al observar el rendimiento académico, notándose claramente que la Clase 2 contiene individuos con un rendimiento superior al promedio (28% de las personas está en el 10% mejor de su colegio, en ranking de notas, en comparación con un 18% y un 14% de las Clases 1 y 3 respectivamente). Existe también una gran diferencia en el rendimiento académico, en términos de la obtención de un puntaje PSU sobresaliente, donde para la Clase 2 es 34% mientras que para la Clase 3 es sólo 8%.

En la Tabla 6.7 se presentan los resultados de estimar el modelo 3LCL. La mayoría de los parámetros comunes son estadísticamente significativos y el único que no es significativo (al 95% de significancia) es la interacción entre el *Arancel* y el nivel de ingreso⁶, aunque es significativo al 90% por lo que se decidió mantenerlo en el modelo.

Tabla 6.7: Estimación del modelo 3LCL

	Parámetros comunes		Clase 1		Clase 2		Clase 3	
K_I	0,575	(5,4)						
K_II-III	0,152	(3,2)						
CRUCH			-0,103	(-0,2)	1,750	(3,5)	-0,369	(-2,5)
Años Acreditación	9,420	(3,8)						
Ranking Qué Pasa			-13,400	(-4,6)	-1,370	(-1,2)	-0,487	(-1,0)
Promedio PSU			5,690	(6,9)	2,620	(6,1)	0,038	(0,3)
Arancel			0,426	(0,1)	-10,100	(-3,9)	-4,13	(-3,1)
Ingreso Medio o Alto	2,480	(1,9)						
Beca Bicentenario	3,400	(2,5)						
Misma Provincia	1,410	(19,3)						
Proporción Misma CSE	0,805	(5,0)						
Log-verosimilitud (0)					-11.402,2			
Log-verosimilitud					-9.374,9			
ρ^2					0,178			
$\bar{\rho}^2$					0,175			
# Parámetros					37			

⁶ Para la interacción se consideró que los individuos con ingreso medio o alto tuvieran un parámetro único. Esto ya que al considerar parámetros independientes para cada categoría de ingreso se obtiene parámetros muy similares en magnitud.

Respecto a las constantes específicas se puede ver que, tal como en el modelo 2LCL, se juntó las categorías II y III. La constante asociada a la categoría I sigue siendo muy superior a las de las demás categorías, aunque la diferencia disminuye al comparar este modelo con los anteriores. Aun así, se confirma una preferencia importante por las dos universidades pertenecientes a esa categoría.

La variable *Mismapro* presenta nuevamente un parámetro con alta significancia estadística. El modelo se estimó primero considerando el atributo como específico a cada clase y luego considerándolo como un parámetro común; posteriormente se realizó un test de razón de verosimilitud para verificar si la diferencia entre ambos era o no significativa. Como el test LR fue aceptado, se eligió el modelo con el atributo común a las tres clases. Se puede ver que la magnitud del parámetro estimado, en comparación con la magnitud de las constantes específicas, aumenta respecto a modelos anteriores, llegando a ser casi el triple que la mayor de las constantes.

El parámetro asociado al atributo *Años de Acreditación* tiene signo positivo, como era de esperar, y sigue siendo un elemento relevante en la elección. Lo anterior es lógico, dado que es un atributo asociado a la calidad de las alternativas y de amplio conocimiento. La proporción de alumnos del mismo nivel socioeconómico que el individuo que tienen las alternativas, se mantiene como un atributo relevante en la elección y, en términos de la magnitud del parámetro, es similar al estimado en el modelo 2LCL.

En la Tabla 6.8 se presentan los parámetros promedio según quintil de ingreso para el modelo 3LCL. Los parámetros individuales se obtuvieron de la misma forma descrita para el modelo 2LCL.

Tabla 6.8: Parámetros promedio según quintil modelo 3LCL

Quintil	Total	Total BIC	Arancel	Promedio PSU	Ranking Qué Pasa
I	2.285	1.260	-4,09	1,55	-2,09
II	1.973	1.284	-3,43	1,78	-2,45
III	2.135	0	-3,30	1,88	-2,78
IV	2.087	0	-2,44	2,45	-4,24
V	2.585	0	-1,53	3,02	-5,75

Al analizar el parámetro *Arancel*, se puede notar que este decrece con el ingreso lo que es esperable *a priori*. Esto representa un cambio respecto a los modelos anteriores, en que había un alza en la valoración del arancel entre los quintiles II y III. En este caso el parámetro promedio de ambos quintiles es muy similar, siendo el segundo levemente menor en magnitud. Además, la disminución de la magnitud del parámetro entre los quintiles I y V es más fuerte en este modelo que en los presentados anteriormente.

La Clase 2, que contiene principalmente a individuos de alto rendimiento académico e ingreso bajo o medio, presenta una sensibilidad alta con respecto al *Arancel*; esto parece razonable ya que, debido a su nivel de ingreso, se trata de individuos con un alto nivel de sensibilidad al costo, lo que puede verse aumentado por las expectativas existentes y el nivel de conocimiento de sus alternativas universitarias. Por su parte, la Clase 3 (que contiene principalmente individuos de ingreso bajo), tiene parámetro menor en magnitud, pero con el mismo signo que la Clase 2. Esto se puede explicar por el bajo nivel de expectativas que tienen los alumnos de bajo rendimiento escolar.

La Clase 1 presenta un parámetro positivo, lo que en principio parece contra-intuitivo; no obstante, se puede deber a que las universidades con mayor *Arancel* son generalmente las que tienen mejor nivel docente y mayor nivel de inversión en infraestructura. Ambos atributos son muy valorados por individuos de ingreso alto como los que componen mayoritariamente la Clase 1.

La valoración del atributo *Promedio PSU* crece a medida que aumenta el nivel de ingreso de la persona. Esto es concordante con lo esperado teóricamente, ya que individuos de ingreso alto tienden a tener un mayor conocimiento de las alternativas y, a su vez, le entregan un mayor valor a los atributos relacionados con la calidad. La valoración implícita correspondiente a los individuos del quinto quintil es casi dos veces mayor que la correspondiente a las personas pertenecientes al primero. Si bien el parámetro del atributo tiene signo correcto en todas las clases, al observarlas por separado se puede apreciar claramente que tanto la Clase 1 como la Clase 2 presentan un parámetro estadísticamente significativo, mientras que la Clase 3 no. La magnitud del parámetro para la primera clase es más del doble que el de la segunda, lo que muestra que los individuos con mayor nivel de ingreso tienen mayor sensibilidad respecto al atributo que los de ingreso bajo y buen rendimiento académico. El parámetro estimado para la tercera clase es menos del 2% que el de la Clase 2, lo que, en conjunto con la baja significancia estadística, implica que este atributo sería irrelevante en la elección de los miembros de la clase.

El parámetro asociado al *Ranking Qué Pasa* es estadísticamente significativo sólo para los individuos de la Clase 1; esto parece razonable ya que el acceso a la información entregada por este atributo no se encuentra igualmente disponible para todos los individuos, sino que es de mucho más fácil acceso para las personas de ingreso alto. La magnitud del atributo para la segunda clase es cerca del 10% de la primera, y la de la tercera clase es aún menor, con poco más del 3%.

En relación a la pertenencia de las universidades al CRUCH, la Clase 2 es la única que presenta beneficio, lo que tiene relación directa con que esta clase es la que cuenta con el mayor porcentaje de individuos beneficiados por la beca BIC. Las otras dos clases presentan un parámetro negativo, aunque para la Clase 1 éste no es estadísticamente significativo (por lo que se puede considerar como irrelevante en la elección). En el caso de la Clase 3, el parámetro es estadísticamente significativo sugiriendo una preferencia

hacia las universidades privadas. Lo anterior se puede deber a que los individuos de esta clase (bajos ingresos y bajo rendimiento académico) tienden a preferir universidades privadas donde tienen mejores opciones de conseguir ayuda estudiantil, que a través de becas gubernamentales (que generalmente tienen como requisito que la universidad pertenezca al CRUCH).

6.6 Comparación de Modelos

6.6.1 Disposición al pago por atributos

En esta sección se calculan las disposiciones al pago por los atributos asociados a la calidad: (i) *Años de Acreditación*, (ii) *Promedio PSU*, y (iii) *Ranking Qué Pasa*. En particular, en la Tabla 4.1 se presentan las disposiciones al pago por *Años de Acreditación* para cada uno de los tres modelos seleccionados. En términos generales parece razonable que aumentar en un año la acreditación de una universidad tenga una disposición al pago promedio entre \$235.300 y \$629.000 para el primer (más bajo) y quinto (mayor) quintil de ingreso respectivamente.

Tabla 6.9: Disposición al pago (miles de \$/año) por años de acreditación según ingreso

Quintil de Ingreso	Modelo HL	Modelo 2LCL	Modelo 3LCL
I	315,3	265,0	235,3
II	435,9	300,9	281,0
III	268,9	269,1	291,6
IV	332,4	312,6	394,7
V	502,3	415,7	629,0

Las magnitudes obtenidas parecen adecuadas considerando que, tal como se muestra en la Tabla 6.10, la diferencia de arancel para distintos niveles de acreditación es en promedio \$253.000. Se puede notar que, dadas las disposiciones al pago encontradas, la

gran mayoría de los individuos estaría dispuesto a pagar la diferencia existente entre las universidades con dos años de acreditación y las con cinco pertenecientes al CRUCH.

Tabla 6.10: Arancel promedio (millones de \$/año) según años de acreditación

Años de Acreditación	CRUCH	Privada
2	1,966	-
3	2,015	-
4	2,210	-
5	2,247	3,255
6	2,462	3,870
7	3,230	-

Por su parte, para las universidades del CRUCH, la diferencia de arancel existente entre seis y siete años de acreditación no es suficiente para generar un cambio basado solamente en términos del aumento en el nivel de acreditación. Algo similar ocurre en las universidades privadas, donde solamente los individuos del quinto quintil de ingreso estarían dispuestos a pagar la diferencia existente en el arancel, y sólo al considerar la disposición al pago entregada por el modelo 3LCL.

Con respecto al *Promedio PSU*, en la Tabla 6.11 se presentan las disposiciones a pagar por aumentar en un punto el puntaje promedio de la universidad. Se puede notar que para todos los quintiles de ingreso, la disposición a pagar encontrada es alta en comparación con la disposición a pagar por *Años de Acreditación*, lo que muestra que se le da un mayor valor a la variable *Promedio PSU* al momento de elegir que estudiar.

Al observar las diferencias entre distintos quintiles, se puede notar que los de menor ingreso presentan – relativamente – una baja disposición al pago por aumentar el *Promedio PSU* de la universidad (con un valor promedio, entre los tres modelos, de \$35.400), mientras que los del quintil más alto tienen una disposición a pagar promedio

de \$142.900 por año, lo que parece un valor elevado y refleja una baja sensibilidad respecto al costo.

Tabla 6.11: Disposiciones al pago (miles de \$/año) por promedio PSU por ingreso

Quintil de Ingreso	Modelo HL	Modelo 2LCL	Modelo 3LCL
I	27,5	40,8	37,8
II	43,3	54,7	51,8
III	36,6	50,9	57,0
IV	59,7	76,0	100,1
V	108,9	122,5	197,3

La Tabla 6.12 presenta los valores medios de la variable *Promedio PSU* para universidades con distintos *Años de Acreditación*. Se puede ver que, dadas las disposiciones al pago presentadas y la diferencia de puntajes entre universidades, la gran mayoría de los individuos estaría dispuesto a pagar más que la diferencia de *Arancel* existente por acceder a una mejor universidad. En este punto, es importante mencionar que la oferta de vacantes es limitada y, por lo tanto, no es posible que todos los individuos realicen el cambio como si fuera un mercado sin restricciones de capacidad.

Al observar la diferencia de *Promedio PSU* existente entre universidades del CRUCH y privadas, se puede notar que en las universidades con cinco años de acreditación, la disposición a pagar por pasar de una universidad del CRUCH a una privada es negativa, por lo que aquellos individuos estudiando en universidades privadas debiesen preferir, *ceteris paribus*, cambiarse a universidades tradicionales.

Tabla 6.12: Promedio PSU según años de acreditación

Años de Acreditación	CRUCH	Privada
2	545,2	-
3	566,9	-
4	565,2	-
5	576,4	574,5
6	597,9	635,3
7	682,4	-

En la Tabla 6.13 se presentan las disposiciones al pago por el atributo *Ranking Qué Pasa*. Se puede notar que esta es negativa, ya que al aumentar el valor de la variable disminuye la calidad y, por lo tanto, las disposiciones al pago podrían ser consideradas como una disposición a aceptar compensación por disminuir la calidad.

Tabla 6.13: Disposiciones al pago (miles de \$/año) por ranking Qué Pasa según ingreso

Quintil	Modelo HL	Modelo 2LCL	Modelo 3LCL
I	-56,9	-70,1	-51,1
II	-87,3	-87,6	-71,6
III	-67,0	-84,3	-84,4
IV	-114,9	-127,4	-173,7
V	-208,3	-205,9	-375,2

Se puede ver que, como era de esperar, a medida que aumenta el nivel de ingreso aumenta en términos absolutos la disposición al pago por el atributo. Individuos del primer quintil de ingreso estarían dispuestos a aceptar \$59.400 al año – como promedio de los tres modelos – por una disminución en el *Ranking Qué Pasa*. Por su parte, individuos del quintil de ingreso superior requerirían una compensación cercana a los \$260.000 si disminuyera el ranking en la misma magnitud.

Finalmente, En la Tabla 6.14 se presenta el *Ranking Qué Pasa* promedio según *Años de Acreditación*, separando universidades del CRUCH y privadas. Dadas las disposiciones a pagar por el ranking presentadas en la Tabla 6.13, se puede ver que los individuos del cuarto y quinto quintil no estarían dispuestos a aceptar una disminución del ranking dadas las diferencias existentes en el *Arancel*, sin importar los *Años de Acreditación* que tenga la universidad en que se encuentran. Por su parte, individuos del primer quintil estarían dispuestos a aceptar una disminución del *Ranking Qué Pasa*, para pasar de universidades con seis a universidades con cinco *Años de Acreditación*.

Tabla 6.14: Ranking Qué Pasa según años de acreditación

Años de Acreditación	CRUCH	Privada
2	33,9	-
3	25,7	-
4	29,3	-
5	19,9	13,7
6	8,3	5,0
7	1,5	-

6.6.2 Capacidad predictiva de los modelos

Para analizar la capacidad predictiva de los modelos estimados, se utilizó una muestra de validación. Esta está compuesta por 3.569 individuos que rindieron la PSU el año 2011, esto es, el mismo año que los individuos de la muestra de estimación, pero que egresaron de la enseñanza media el año 2010. En la Tabla 6.15 se presenta una comparación de las características socioeconómicas de los individuos de ambas muestras, con el objetivo de analizar las diferencias existentes. Se puede apreciar que en términos del ingreso, el banco de datos de validación cuenta con una mayor representación de individuos de ingreso bajo, con un 42% contra un 35% en la muestra de estimación. Dada esta tendencia, se espera que en la muestra de validación la elección de los individuos sea más sensible al arancel y menos sensible a atributos asociados a la calidad.

Tabla 6.15: Características socioeconómicas muestras de estimación y validación

		Estimación	Validación
Total		11.065	3.569
Ingreso	Bajo	35%	42%
	Medio	34%	32%
	Alto	31%	26%
Beca Bicentenario	-	23%	22%
Sexo	Mujer	44%	46%
Nivel Educativo de los Padres	Escolar	46%	48%
	Técnica o IP	9%	10%
	Universitaria	44%	40%
Tipo de Dependencia Escolar	Municipal	17%	22%
	Subvencionado	53%	53%
	Particular	29%	25%
Rendimiento Académico	10% Mejor del Colegio	19%	15%
	PSU Sobre 700	23%	17%

Al analizar las diferencias existentes en términos del tipo de colegio del que provienen los individuos, se puede notar que la muestra de validación tiene un mayor porcentaje de individuos de colegios municipales (22%) que la muestra de estimación (17%), mientras que la segunda tiene un mayor porcentaje proveniente de colegios particulares (29% contra 25%). En términos del rendimiento académico, se puede notar que en la base de estimación hay un mayor número de individuos de alto rendimiento (19% en el 10% superior de sus colegios y 23% con puntajes PSU por sobre 700 vs. la base de validación con 15% y 17% respectivamente), lo que representa una diferencia superior al 20% en el rendimiento académico de los individuos. No obstante, aun considerando las diferencias existentes entre ambos bancos de datos, se puede notar que son similares, por lo que se esperaría obtener buenos resultados al utilizar los modelos en modalidad predictiva.

La capacidad predictiva de un modelo se puede analizar de dos maneras: (i) observando la elección predicha para cada individuo, o, más correctamente, (ii) estimando las proporciones de mercado de cada alternativa según cada modelo. La diferencia radica en

que bajo la primera, se debe realizar una asignación, consistente en considerar que cada persona elegiría la opción de mayor utilidad; pero en un mercado con muchas opciones esto conduce a resultados sesgados. Bajo la segunda forma se computan las proporciones de mercado de cada alternativa con un método de agregación adecuado. En este caso se estimaron las proporciones de mercado agregadas utilizando el método de enumeración muestral (Ortúzar y Willumsen, 2011, Cap 9).

En la Tabla 6.16 se presentan las proporciones de mercado de las diez universidades con mayor demanda observada en la muestra de validación. Además, se presentan las proporciones de mercado calculadas utilizando cada uno de los modelos estimados. Se puede ver que el modelo 3LCL presenta un mejor desempeño que los restantes, ya que sus predicciones son más certeras (presenta el menor error de los tres). En segundo lugar se encuentra el modelo 2LCL y finalmente el modelo HL.

Tabla 6.16: Proporciones de mercado observada y modeladas

Alternativa	Observada	Modelo HL	Modelo 2LCL	Modelo 3LCL
111	13,3%	11,0%	11,9%	13,3%
121	10,6%	8,3%	9,1%	9,8%
131	7,3%	7,3%	7,5%	7,4%
391	6,0%	5,1%	5,0%	4,4%
141	5,4%	5,7%	5,7%	5,6%
411	5,2%	4,9%	4,9%	5,0%
381	5,0%	5,6%	5,4%	5,2%
161	3,6%	5,7%	5,3%	5,1%
152	3,3%	3,1%	2,9%	2,9%
191	2,6%	2,8%	2,7%	2,7%
Error⁷	-	0,22%	0,13%	0,09%

⁷ El error que se presenta se calcula como la suma de las diferencias al cuadrado entre los porcentajes predichos y observados.

Como complemento, en la Tabla 6.17 se presenta el porcentaje correctamente predicho por cada modelo (de acuerdo a una asignación de mayor utilidad); es decir, del total de elecciones predichas a nivel individual en qué porcentaje la alternativa predicha es igual a la elección observada. También se presenta el test χ^2_{FPR} para ver si los modelos aplicados son consistentes con los datos (Ortúzar y Willumsen, 2011, Cap. 8).

Tabla 6.17: Capacidad predictiva y test chi-cuadrado FPR

	Modelo HL	Modelo 2LCL	Modelo 3LCL
% Predicción Correcta	57,0%	57,4%	57,6%
χ^2_{FPR} ⁸	23,78	4,63	4,68

Se puede notar que los tres modelos tienen porcentajes de recuperación de la primera preferencia (FPR) muy similares (del orden de 57%). El modelo con mayor valor es el 3LCL. Al analizar los valores del test χ^2_{FPR} , se puede ver que este se rechazaría para el modelo HL, pero sería aceptado para los modelos de clases latentes al 85% de significancia. Así, estos últimos dos modelos parecen comportarse mejor para el banco de datos de validación utilizado.

En la Tabla 6.18 se presenta el test de razón de verosimilitud directa⁹ para comparar los tres modelos estimados. Se puede ver que al comparar el modelo HL con el 2LCL el

⁸ El test χ^2_{FPR} permite indicar si un modelo estimado es consistente con los datos en su fase predictiva. Este test se presenta en detalle en Ortúzar (2000), página 147. En este caso las bandas de probabilidad son de 0,1 por lo que el valor crítico de χ^2_{m-1} es 4,82 al 85% de significancia.

⁹ El test de razón de verosimilitud directa permite comparar cuál, entre dos modelos, se ajusta mejor a los datos de validación. Este test tiene la ventaja de que se puede aplicar aun cuando los modelos

segundo tiene un mejor ajuste a los datos. Lo mismo ocurre al comparar el HL con el 3LCL, donde el segundo es notoriamente superior al primero. Finalmente, al comparar el modelo 2LCL con el 3LCL, se puede ver que el segundo tiene un mejor ajuste a los datos, resultando el mejor modelo bajo todos los criterios de análisis probados.

Tabla 6.18: Test de razón de verosimilitud directa

	Razón de Verosimilitud Directa
HL vs 2LCL	1,17E-28
HL vs 3LCL	1,94E-45
2LCL vs 3LCL	1,66E-17

6.7 Aplicación Práctica

Para ejemplificar la utilidad práctica que pueden tener los modelos estimados, se decidió realizar el ejercicio de dar gratuidad en la educación universitaria en dos casos: (i) cobertura para los primeros cuatro quintiles de ingreso y (ii) gratuidad a la totalidad de la población. La aplicación práctica se hace utilizando la base de datos de validación presentada en el punto anterior.

Al momento de estimar los modelos se consideró la elección entre las alternativas a las que los individuos efectivamente postulan, es decir, considerando restricciones de disponibilidad de alternativas. Se espera que la implementación de una política como la gratuidad, con coberturas como las consideradas, genere cambios en la estructura de postulaciones, por lo que se prefiere comparar los distintos casos utilizando las

comparados no sean anidados. Una descripción completa del test se puede encontrar en Ortúzar (2000), páginas 148 y 149.

proporciones de mercado de la elección sin considerar las restricciones de disponibilidad.

Tabla 6.19: Universidades con mayor proporción de mercado modelada.

	Actualidad	80% Gratuidad	100% Gratuidad
P .U. CATOLICA DE CHILE - Santiago	23,4%	28,0%	28,8%
U. DE CHILE - Santiago	16,1%	15,8%	15,7%
U. DE SANTIAGO DE CHILE - Santiago	4,2%	3,9%	3,8%
U. DE CONCEPCION - Concepción	2,4%	2,1%	2,1%
P. U. CATOLICA DE VALPARAISO - Valparaíso	2,1%	2,0%	2,0%
U. TEC. F. SANTA MARIA - Santiago	2,0%	1,6%	1,5%
P. U. CATOLICA DE CHILE - Villarrica	1,7%	1,3%	1,2%
U. ADOLFO IBAÑEZ - Santiago	1,6%	1,6%	1,7%
U. ALBERTO HURTADO - Santiago	1,5%	1,4%	1,4%
U. DE LOS ANDES - Santiago	1,5%	1,7%	1,7%

En la Tabla 6.19 se presentan las proporciones de mercado de las diez universidades con mayor preferencia modelada¹⁰ en la actualidad. Se puede ver que, al considerar la elección sin las restricciones de disponibilidad la universidad con una mayor preferencia sería la Pontificia Universidad Católica de Chile, seguida por la Universidad de Chile. Ambas absorben cerca del 40% del total de primeras preferencias mientras que las ocho universidades restantes entre las diez mejores sólo representan el 16,9%.

Se puede ver que tanto en una situación con un 80% de gratuidad como con un 100% de gratuidad, de las diez universidades sólo la Pontificia Universidad Católica de Chile, en

¹⁰ Se utiliza el término “preferencia modelada” para hacer referencia a las preferencias sin considerar las restricciones de disponibilidad.

sus sedes de Santiago, la Universidad Adolfo Ibáñez, en su sede de Santiago, y la Universidad de los Andes ven aumentada su preferencia. La primera experimenta un aumento considerable de un 19% lo que se puede asociar a que es una universidad con un alto nivel de calidad, pero con aranceles bastante elevados, lo que hace que una cantidad importante de individuos con un nivel de ingreso entre los primeros cuatro quintiles no la consideren como su opción preferida, lo que cambia al momento en que los alumnos se ven enfrentados a una situación de gratuidad.

Por otra parte, instituciones como la Universidad de Santiago, la Universidad de Concepción y la Universidad Técnica Federico Santa María ven una disminución considerable en la preferencia de los alumnos, lo que puede tener relación con que sean universidades con aranceles bajos en comparación con otras universidades, pero con un nivel de calidad inferior a universidades como la Pontificia Universidad Católica de Chile, en términos de las variables incorporadas en los modelos.

7. CONCLUSIONES Y RECOMENDACIONES

Los resultados obtenidos para cada modelo estimado muestran que la elección de universidad es consistente con la teoría económica y estaría efectivamente basada en los atributos de las distintas alternativas. Variables asociadas a la calidad influyen positivamente en la probabilidad de elección, mientras que variables asociadas al costo lo hacen de forma negativa. A partir del modelo que permite capturar heterogeneidad entre individuos, encontramos que los atributos socioeconómicos incorporados en la modelación también tienen el impacto esperado, pero lo hacen en forma diferenciada. A medida que aumenta el ingreso individual la sensibilidad por la calidad aumenta y la del costo disminuye, lo que puede tener que ver con la mayor información y expectativas respecto a la educación superior. Además, individuos con alto nivel de rendimiento académico en el colegio otorgan mayor valor a la calidad y también parecen tener mayor sensibilidad con respecto al costo.

Las expectativas de obtener una beca – que se entregan con criterios objetivos, por lo que son bastante certeras – hacen que los individuos favorecidos sean relativamente insensibles respecto al valor del *Arancel*. Debido a lo anterior, estos individuos eligen su universidad sin considerar el costo, por lo que su elección pierde consistencia con los objetivos de, por ejemplo, contener costos para el Estado (obviamente que es racional). Si el sistema de ayudas estudiantiles tuviese mayor participación de créditos en lugar de becas, este problema podría ser incorporado por los individuos.

Los resultados de la estimación son consistentes en los tres modelos estimados. Las valoraciones promedio por quintil, la significancia estadística de los distintos atributos y las disposiciones a pagar por aumentar la calidad, son similares para los tres modelos, tanto en signo como en magnitud.

Las categorías de universidades existentes son también consistentes con las constantes específicas estimadas y evidencian una clara – y muy superior – preferencia por las dos universidades más prestigiosas: Pontificia Universidad Católica de Chile y Universidad de Chile. Esto tiene concordancia con que estas son las universidades que reciben a los mejores alumnos cada año, reciben más aporte estatal y generan más investigación, sólo por nombrar algunas de las dimensiones en que superan claramente al resto del sistema de educación superior.

Al estudiar el efecto que tendría una política como la gratuidad en la educación superior sobre las preferencias individuales mediante los modelos, es posible ver que aumenta la proporción de individuos que prefieren universidades con mejores indicadores de calidad y que actualmente tienen aranceles muy por sobre el promedio, tales como la Pontificia Universidad Católica de Chile, la Universidad Adolfo Ibáñez y la Universidad de los Andes.

Los correctos resultados de los dos modelos de clases latentes estimados, permiten concluir que la utilización de este tipo de herramientas econométricas es adecuada para modelar la elección de universidad. Ambos modelos entregan resultados en línea con lo esperado, tanto en términos de sus parámetros como de las disposiciones al pago por sus distintos atributos. Además, permiten realizar un análisis de posibles agrupaciones de individuos con comportamientos similares y de acuerdo a las principales características socioeconómicas que los caracterizan. En particular, del modelo con tres clases latentes se puede concluir que los tres grupos existentes son: (i) individuos de ingreso alto, (ii) individuos de ingreso medio-bajo con un alto rendimiento académico, e (iii) individuos de ingreso medio-bajo con un bajo rendimiento académico. Lo anterior puede ser una herramienta útil para orientar políticas hacia alguno de los grupos, conociendo su comportamiento y pudiendo prever los efectos que tendría la misma.

Como una de las posibles líneas de investigación futura se propone la flexibilización del conjunto de elección, permitiendo incorporar de mejor manera el efecto de la “autoexclusión” de los individuos de aquellas universidades a las que no les alcance el puntaje, de manera tal de eliminar posibles sesgos en la estimación. Para esto se propone la utilización de modelos con choice-set probabilísticos definidos a partir de la diferencia existente entre el puntaje obtenido por los individuos y los puntajes de corte de las distintas carreras (en las distintas universidades consideradas).

Se propone también como una futura investigación, la incorporación con un mayor nivel de detalle de créditos y becas ofrecidos tanto por el gobierno como por las distintas instituciones universitarias, de manera tal de poder obtener el efecto qué estos tienen sobre la elección y cuales son las diferencias existentes entre diferentes instituciones. Para desarrollar esta línea de investigación es necesario un extenso proceso de recolección de información sobre las ayudas estudiantiles de las distintas universidades.

Finalmente, se propone desarrollar un modelo conjunto de elección de carrera y universidad de manera de realizar una modelación más real de la elección a la que se ven enfrentados los individuos, lo que permitiría además analizar las diferencias existentes entre las distintas carreras. Para esto se propone utilizar una combinación de modelos de clases latentes y modelos de nidos cruzados, lo que permite dar flexibilidad a la modelación (ya que no hay certeza de si el orden de la elección es primero carrera y luego universidad o lo contrario). Además es necesario recopilar información relativa a las diferentes carreras (ingreso esperado, campo laboral, etc.).

BIBLIOGRAFÍA

Arellana, J., Daly, A.J., Hess, S., Ortuzar, J. de D. & Rizzi, L. (2012). Developing surveys for the study of departure time choice: a two-stage efficient design approach. *Transportation Research Record*, 2303, pp. 9-18.

Avery, C. & Hoxby, C.M. (2004). Do and should financial aid packages affect students' college choices? En: Hoxby, C.M. (ed.), *College Choices: The Economics of Where to Go, When to Go, and How to Pay for It*, University of Chicago Press, Chicago, pp. 239-299.

Becker, G. (1964). *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*. Columbia University Press, Nueva York.

Becker, W. (1990). The demand for higher education. En: Hoenack, S.A. & Collins, E.L. (eds.), *The Economics of American Universities*. State University of New York Press, Albany, pp. 155-188.

Beggs, S., Cardell, S. & Haussman, J. (1981). Assessing the potential demand for electric cars. *Journal of Econometrics*, 17(1), pp. 1-19.

Bezman, T. & Depken, C.A. (1998). School characteristics and the demand for college. *Economics of Education Review*, 17(2), pp. 205-210.

Bierlaire, M. (2003). BIOGEME: A free package for the estimation of discrete choice models. *Proceedings of the 3rd Swiss Transportation Research Conference*, Ascona, Switzerland.

Cabrera, A.F. & La Nasa, S.M. (2000). Understanding the College-Choice Process. *New Directions for Institutional Research*, 2000(107), pp. 5-22.

Carrasco, J.A. & Ortúzar, J. de D. (2002). Review and assessment of the nested logit model. *Transport Reviews*, 22(2), pp. 197-218.

Carter, R.E. & Curry, D.J. (2011). Using student-choice behaviour to estimate tuition elasticity in higher education. *Journal of Marketing Management*, 27(11-12), pp. 1186-1207.

Chapman, D.W. (1981). A model of student college choice. *Journal of Higher Education*, 52(5), pp. 490-505.

Chapman, R.G. & Staelin, R. (1982). Exploding rank-ordered choice set data within the stochastic utility model. *Journal of Marketing Research*, 19(3), pp. 288-301.

Christensen, S., Melder, J. & Weisbrod, B.A. (1975). Factors affecting college attendance. *Journal of Human Resources*, 10(2), pp. 174-188.

Comisión de Financiamiento Estudiantil para la Educación Superior. (2012). *Análisis y Recomendaciones para el Sistema de Financiamiento Estudiantil*. Recuperado de: <http://200.6.99.248/~bru487cl/files/ComParedes.pdf>

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

Drewes, T. & Michael, C. (2006). How do students choose a university?: An analysis of applications to universities in Ontario, Canada. *Research in Higher Education*, 47(7), 781-800.

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

Elacqua, G. (2009). The impact of school choice and public policy on segregation: Evidence from Chile. *Documento de Trabajo CPCE N°10*, Centro de Políticas Públicas Comparadas de Educación.

Ellwood, D. & Kane, T. (2000). Who is getting a college education? Family background and the growing gaps in enrolment. En: Danziger, S. & Waldfogel, J. (eds.) *Securing the Future*, Russell Sage, Nueva York, pp. 283-324.

Freeman, R. (1986). Demand for education. En: Ashenfelter, O. & Layard, R. (eds.), *Handbook of Labour Economics*, Elsevier, Amsterdam.

Fuller, W.C., Manski, C.F. & Wise, D.A. (1982). New evidence on the economic determinants of postsecondary schooling choices. *Journal of Human Resources*, 17 (4), pp. 477-498.

Goodman, L.A. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61(2), pp. 215-231.

Gopinath, D.A. (1995). *Modelling Heterogeneity in Discrete Choice Processes: Application to Travel Demand*. Ph.D. Thesis, Department of Civil and Environmental Engineering, Massachusetts Institute of Technology.

Greene, W.H. & Hensher, D.A. (2003). A latent class model for discrete choice analysis: contrasts with mixed logit. *Transportation Research Part B*, 37(8), pp. 681-698.

Heckman, J.J. & Singer, B. (1984). Econometric duration analysis. *Journal of Econometrics*, 24(1-2), pp. 63-132.

Heller, D.E. (1997). Student price response in higher education: an update to Leslie and Brinkman. *Journal of Higher Education*, 68(6), pp. 624-659.

Hess, S., Daly, A.J., Rohr, C. & Hyman, G. (2007). On the development of time period and mode choice models for use in large scale modelling forecasting systems. *Transportation Research Part A*, 41(9), pp. 802-826.

Hsieh, C. & Urquiola, M. (2003). When schools compete, how do they compete? An assessment of the Chile's nationwide school voucher program. *NBER Working Paper 10008*, The National Bureau of Economics Research.

Horowitz, J.L. & Louviere, J.J. (1995). What is the role of consideration sets in choice modelling? *International Journal of Research in Marketing*, 12(1), pp. 39-54.

Hurtubia, R., Nguyen, M.H., Glerum, A. & Bierlaire, M. (2013). *Estimation of mode choice models with latent classes and psychometric indicators*. XVI Congreso Chileno de Ingeniería de Transporte, Santiago, Chile.

Jackson, G.A. & Weathersby, G.B. (1975). Individual demand for higher education: a review and analysis of recent empirical studies. *Journal of Higher Education*, 46(6), pp. 623-652.

Jensen, A.F., Cherchi, E. & Ortúzar, J. de D. (2012). A “short panel” SC survey to elicit preferences for electric vehicles and charging infrastructure. *XVII Congreso Panamericano de Ingeniería de Transito, Transporte y Logística*, Universidad de los Andes, Santiago, Septiembre 24-27, 2012.

Kohn, M.G., Manski, C.F. & Mundel, D.S. (1976). An empirical investigation of factors influencing college-going behaviour. *Annals of Economic and Social Measurement*, Vol. 5(4), pp. 391-419.

Lehr, D.K. & Newton, J.M. (1978). Time series and cross-sectional investigations of the demand for higher education. *Economic Inquiry*, 16(3), pp. 411-422.

Leslie, L. & Brinkman, P. (1987). Student price response: the student demand studies. *Journal of Higher Education*, 58(2), pp. 181-204.

Lockshin, L., Jarvis, W., d’Hauteville, F. & Perrouty, J-P. (2006). Using simulations from discrete choice experiments to measure consumer sensitivity to brand, region, price, and awards in wine choice. *Food Quality and Preference*, 17(3-4), pp. 166-178.

Long, B.T. (2004a). How have college decisions changed over time? An application of the conditional logistic choice model. *Journal of Econometrics*, 121(1-2), pp. 271-296.

Long, B.T. (2004b). Does the format of a financial aid program matter? The effect of state in-kind tuition subsidies. *The Review of Economics and Statistics*, 86(3), pp. 767-782.

Louviere, J.A., Hensher, D.A. & Swait, J.D. (2000). *Stated Choice Methods: Analysis and Applications*. Cambridge University Press, Cambridge.

Mark, D.R., Lusk, J.L. & Daniel, M.S. (2004). Recruiting agricultural economics graduate students: student demand for program attributes. *American Journal of Agricultural Economics*, 86(1), pp. 175-184.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behaviour. En: Zarembka, P. (ed.), *Frontiers in Econometrics*, Academic Press, Nueva York, pp. 105-142.

McFadden, D. (1974). The measurement of urban travel demand. *Journal of Public Economics*, 3(1), pp. 303-328.

Mincer, J. (1974). *Schooling, Experience, and Earnings*. Columbia University Press, Nueva York.

Ministerio de Desarrollo Social. (2009). *Encuesta de Caracterización Socioeconómica Nacional (CASEN)*. Recuperado de Ministerio de Desarrollo Social: http://observatorio.ministeriodesarrollosocial.gob.cl/casen_usuarios.php

Ministerio de Educación. (2011). *Tabla de quintiles de ingreso del Ministerio de Educación para el año 2011*. Recuperado de Ministerio de Educación: http://www.mineduc.cl/index2.php?id_contenido=18386&id_portal=74&id_seccion=4033

Montgomery, M. (2002). A nested logit model of the choice of a graduate business school. *Economics of Education Review*, 21(5), pp. 471-480.

Niu, S.X., Tienda, M. & Cortes, K. (2006). College selectivity and the Texas top 10% law. *Economics of Education Review*, 25(3), pp. 259-272.

Niu, S.X. & Tienda, M. (2008). Choosing colleges: identifying and modelling choice sets. *Social Science Research*, 37(2), pp. 416-433.

OCDE (2011). *Education at a Glance: OECD Indicators*. OECD Publishing. Recuperado de Organisation for Economic Co-operation and Development: <http://www.oecd.org/education/skills-beyond-school/48631582.pdf>

Ortúzar, J. de D. & Garrido, R.A. (1994). A practical assessment of stated preferences methods. *Transportation*, 21(3), pp. 289-305.

Ortúzar J. de D. (2000) *Modelos Econométricos de Elección Discreta*. Ediciones Universidad Católica de Chile, Santiago.

Ortúzar, J. de D. & Willumsen, L.G. (2011). *Modelling Transport*. Cuarta Edición, John Wiley and Sons, Chichester.

- Oosterbeek, H., Groot, W., & Hartog, J. (1992). An empirical analysis of university choice and earnings. *De Economist*, 140(3), pp. 293-309.
- Palma, D., Ortúzar, J. de D., Casaubon, G., Rizzi, L.I. & Agosin, E. (2013). Measuring Consumer Preferences Using Hybrid Discrete Choice Models. *American Asociation of Wine Economists (AAWE) Working Paper* N°137.
- Paredes, R.D., Valdivia, C. & Barrios, A. (2013). University quality: rankings, categories and public policy in a dynamic context. *Higher Education Review* (under review).
- Punj, G. & Staelin, R. (1978). The Choice Process for Graduate Business Schools. *Journal of Marketing Research*, 15(Nov), pp. 588-598.
- Qué Pasa. (2011). *Ranking Universidades 2011*. Revista Qué Pasa, Edición 16 de Diciembre de 2011. Recuperado de Revista Qué Pasa: <http://papeldigital.info/qp/index.html?2011121602#>
- Radner, R. & Miller, L.S. (1970). Demand and supply in higher education: a progress report. *American Economic Review*, 60(2), pp. 326-334.
- Reyes, C. & Rosso, P.P. (2012). *Una nueva clasificación de las universidades chilenas*. Documento de Trabajo, Red Universitaria Cruz del Sur.
- Rodriguez, J. & Arriagada, C. (2004). Segregación residencial en la ciudad latinoamericana. *EURE*, 30(89), pp. 5-24.
- Sá, C., Florax, R.J.G.M. & Rietveld, P. (2012). Living arrangement and university choice of Dutch prospective students. *Regional Studies*, 46(5), pp. 651-667.

Shen, J. (2009). Latent class model or mixed logit model? A comparison by transport mode choice data. *Applied Economics*, 42(22), pp. 2915-2924.

Sistema Nacional de Información de la Educación Superior (SIES). (2010). *Evolución Matrícula Educación Superior de Chile, Período 1990-2009*. Recuperado de Mi Futuro: http://www.mifuturo.cl/images/Estudios/Estudios_SIES_DIVESUP/estudio_evolucion_matricula_historica_1990_2009.pdf

Spence, M. (1973). Job market signalling. *Quarterly Journal of Economics*, 87(3), pp. 355-374.

Swait, J.D. & Ben-Akiva, M. (1986). An analysis of the effects of captivity on travel time and cost elasticities. En: *Annals of the 1985 International Conference on Travel Behavior*, 16-19 Abril 1985, Noordwijk, Holanda, pp. 113-128.

Train, K.E. (2009). *Discrete Choice Methods with Simulation*. Segunda Edición, Cambridge University Press, Cambridge.

Varela, N. (1996). *Un Modelo de Preferencias para la Elección de una Carrera Universitaria*. Tesis de Magister, Departamento de Ingeniería de Sistemas, Pontificia Universidad Católica de Chile.

Walker, J.L. & Li, J. (2007). Latent lifestyle preferences and household location decisions. *Journal of Geographical Systems*, 9(1), pp. 77-101.

Wetzel, J., O'Toole, D. & Peterson, S. (1998). An analysis of student enrolment demand. *Economics of Education Review*, 17(1), pp. 47-54.

Williams, H.C.W.L. (1977). On the formation of travel demand models and economic evaluation measures of user benefit. *Environment and Planning A*, 9(3), pp. 285-344.

ANEXOS

A. ANEXO 1: Análisis de Datos

Tabla A.1: Listado de carreras

Código Carrera	Nombre Carrera	Código Carrera	Nombre Carrera
1	Educación Física	32	Contador Auditor
2	Pedagogía	33	Psicología
3	Educación de Párvulos	34	Sociología
4	Educación Diferencial	35	Periodismo
5	Pedagogía en Inglés (Idiomas)	36	Trabajo Social
6	Educación Artes y Música	37	Ciencia Política
7	Educación Otros	38	Geografía
8	Ingeniería Comercial	39	Geología
9	Derecho	40	Historia
10	Medicina	41	Letras y Literatura
11	Odontología	42	Publicidad
12	Enfermería	43	Antropología
13	Kinesiología	44	Filosofía
14	Tecnología Médica	45	Bibliotecología
15	Nutrición y Dietética	46	Administración
16	Obstetricia	47	Administración Pública
17	Terapia Ocupacional	48	Teología
18	Fonoaudiología	49	Piloto Comercial
19	Química	50	Literatura Inglesa y Traducción
20	Biología	51	Turismo
21	Física y Astronomía	52	Arquitectura
22	Matemática y Estadística	53	Diseño
23	Bioquímica	54	Arte
24	Agronomía e Ingeniería Forestal	55	Música
25	Veterinaria	56	Teatro
26	Ingeniería Ambiental	57	Cine y Audiovisual
27	Ingeniería en Recursos Naturales	58	Bachillerato/College - Ciencias
28	Construcción Civil	59	Bachillerato/College - Ciencias Sociales
29	Ingeniería	60	Bachillerato/College - Arte
30	Ingeniería - Ejecución	61	Bachillerato/College - General
31	Ingeniería - Técnico		

Tabla A.2: Categorías de ingreso e ingreso promedio CASEN.

Categoría	Mínimo	Máximo	Promedio CASEN 2009
1	\$0	\$144.000	\$88.332
2	\$144.001	\$288.000	\$220.345
3	\$288.001	\$432.000	\$359.109
4	\$432.001	\$576.000	\$498.979
5	\$576.001	\$720.000	\$643.705
6	\$720.001	\$864.000	\$787.814
7	\$864.001	\$1.008.000	\$931.231
8	\$1.008.001	\$1.152.000	\$1.075.965
9	\$1.152.001	\$1.296.000	\$1.221.394
10	\$1.296.001	\$1.440.000	\$1.364.074
11	\$1.440.001	\$1.584.000	\$1.507.474
12	\$1.584.001	más	\$1.980.101

Fuente: Elaboración propia en base a CASEN (2009).

Tabla A.3: Definición quintiles de ingreso per cápita.

Quintil	Inferior	Superior
I	\$0	\$62.090
II	\$62.091	\$106.214
III	\$106.215	\$168.366
IV	\$168.367	\$301.741
V	\$301.742	más

Fuente: Elaboración propia en base a Ministerio de Educación (2012).

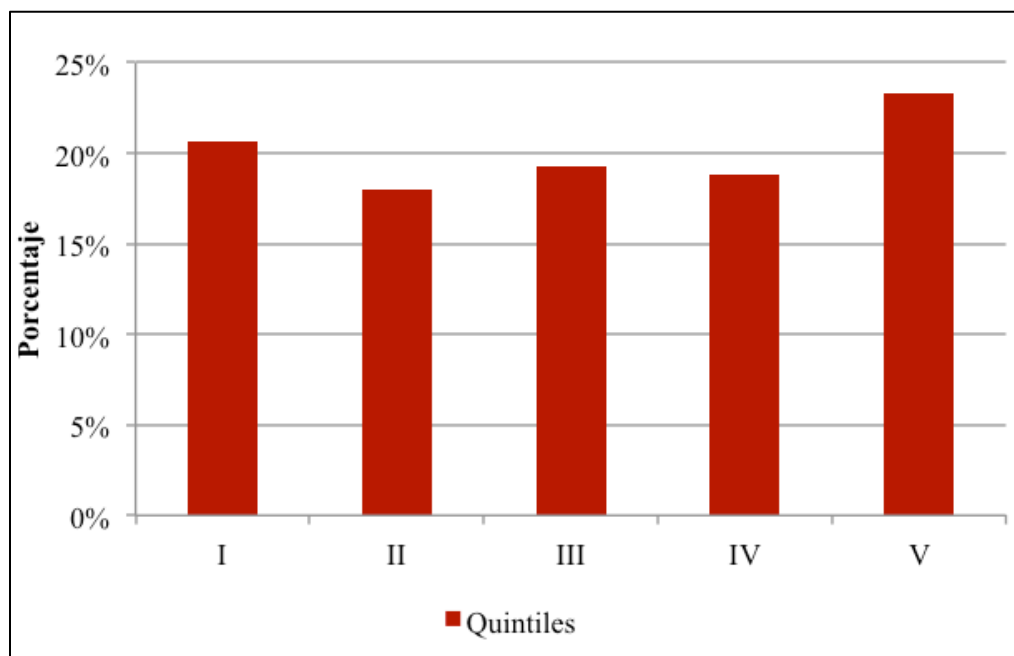


Figura A.1: Proporción de individuos según quintil de ingreso.

Tabla A.4: Definición categorías de ingreso.

Ingreso	Inferior	Superior
Bajo	\$0	\$89.777
Medio	\$89.778	\$232.807
Alto	\$232.808	más

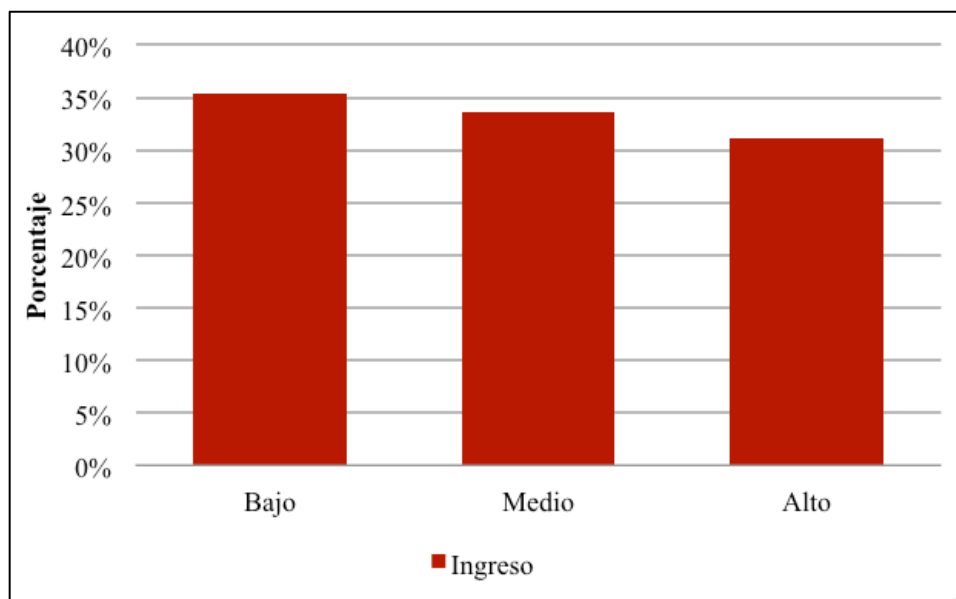


Figura A.2: Composición de la muestra en términos de categorías de ingreso utilizadas.

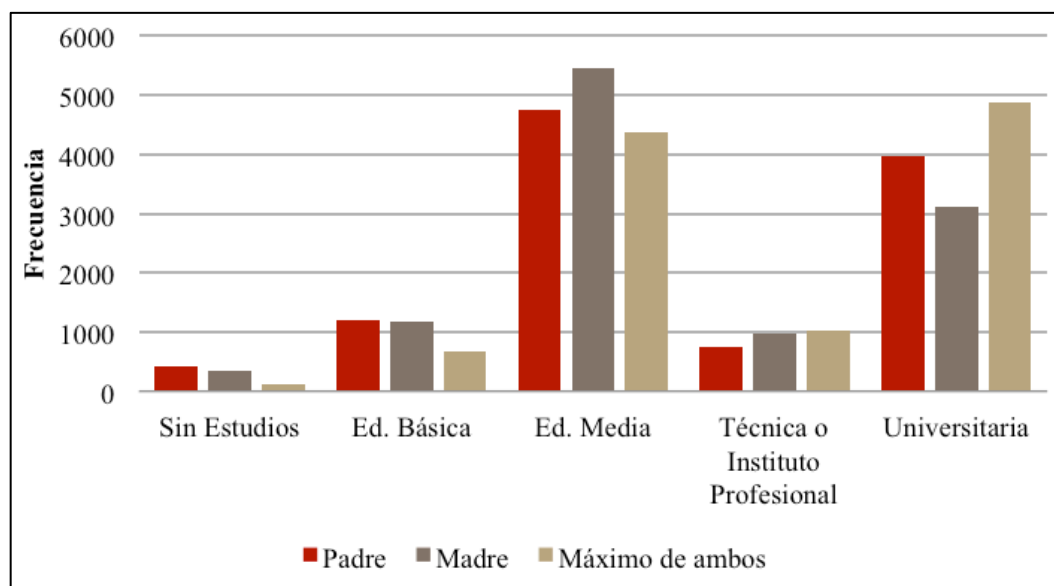


Figura A.3: Nivel educacional de los padres.

Tabla A.5: Individuos de acuerdo a Tipo de Dependencia Escolar.

	Frecuencia	Porcentaje
Municipal	1.934	18%
Subvencionado	5.891	53%
Particular	3.255	29%

Tabla A.6: Estadísticas descriptivas Promedio NEM y Ranking.

	Media	Desviación Estándar	Mínimo	Máximo
Promedio NEM	59.5	4.6	44	70
Ranking	62%	27%	0%	100%

Tabla A.7: Estadísticas descriptivas Puntajes PSU.

	N° Obs.	Media	Desviación Estándar	Mínimo	Máximo
Puntaje NEM	11.065	609,5	94,4	290	826
Lenguaje y Comunicación	11.065	597,4	82,0	363	850
Matemáticas	11.065	610,4	85,1	383	850
Ciencias	6.093	591,5	92,1	286	850
Ciencias Sociales	7.391	585,7	89,7	195	850

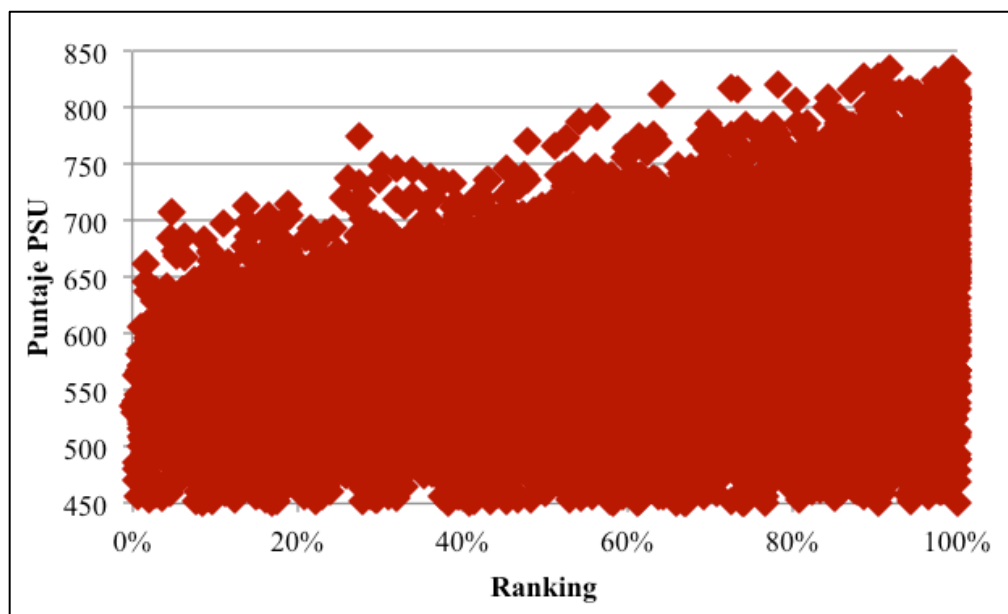


Figura A.4: Dispersión Ranking y Puntaje PSU.

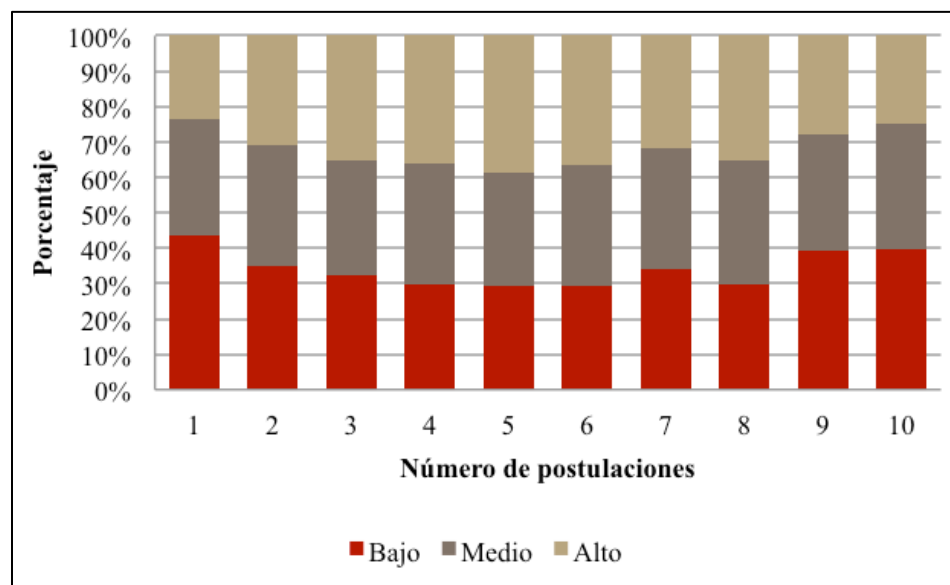


Figura A.5: Proporción de categorías de ingreso según número de postulaciones.

B. ANEXO 2: Modelación

a) Tests Estadísticos Asociados

Se presentan en esta sección los principales tests estadísticos e indicadores de ajuste de los modelos que serán utilizados para el análisis y comparación de los modelos estimados.

- Test t de significancia estadística de parámetros

Permite rechazar la hipótesis nula de que un parámetro estimado es estadísticamente igual a un valor determinado, comúnmente se suele comparar con cero. Este test es de gran utilidad ya que permite conocer qué parámetros del modelo son relevantes y cuales son estadísticamente iguales a cero. El estadígrafo se define como:

$$t^* = \frac{\hat{\theta}_k - X}{\sqrt{\sigma_k^2}} \quad (B-1)$$

Donde $\hat{\theta}_k$ es el valor del parámetro estimado, X es el valor con que se quiere comparar y σ_k^2 es la varianza del parámetro estimado. Una vez calculado el estadígrafo se rechaza la hipótesis nula de que el parámetro es igual al valor X si su valor es mayor que un valor t que se obtiene a partir de la tabla de la distribución t de Student, que para muestras grandes tiende a una Normal. Las significancias estadísticas utilizadas comúnmente son de un 1%, 5% y 10% cuyos valores de t son respectivamente 2,57, 1,96 y 1,65.

- Test t de significancia conjunta de parámetros

Este test permite ver si la suma de dos parámetros es estadísticamente significativa o no. El test se aplica de igual manera que el test t de significancia estadística de parámetros, pero el estadígrafo es calculado de una manera diferente:

$$t^* = \frac{\hat{\theta}_k + \hat{\theta}_l - X}{\sqrt{\sigma_k^2 - 2\rho_{kl}\sigma_k\sigma_l + \sigma_l^2}} \quad (\text{B-2})$$

Donde $\hat{\theta}_k$ y $\hat{\theta}_l$ son los parámetros estimados, X es el valor con que se quiere comparar – normalmente cero –, σ_k^2 y σ_l^2 son las varianzas de los parámetros estimados y el término $\rho_{kl}\sigma_k\sigma_l$ es la covarianza entre ambos. Los valores de comparación para la aplicación del test

- Test de razón de verosimilitud

El test de razón de verosimilitud permite la comparación estadística de dos modelos anidados, esto requiere que un modelo sea la versión restringida del otro. El test se realiza en base a la comparación de las log-verosimilitudes de ambos, y se construye el siguiente estadígrafo:

$$LR = 2 * \{l(\theta_r) - l(\theta_g)\} \quad (\text{B-3})$$

Donde $l(\theta_r)$ es la log-verosimilitud del modelo restringido y $l(\theta_g)$ es la log-verosimilitud del modelo general. El estadígrafo LR distribuye asintóticamente χ^2 con r grados de libertad y nivel de confianza $(1 - \alpha)$, donde r es el número de restricciones aplicadas al modelo general para obtener el modelo restringido. La hipótesis nula de que ambos modelos son equivalentes se rechaza si el valor del estadígrafo es mayor que un valor $\chi_{r,1-\alpha}^2$ de tabla indicando que el modelo general es preferido al restringido.

- Índice ρ^2 y ρ^2 ajustado

Los índices ρ^2 y ρ^2 ajustado (o $\bar{\rho}^2$) son indicadores del ajuste general del modelo, y pueden ser calculados para cada modelo estimado. El primero mide la bondad de ajuste del modelo en comparación con un modelo equiprobable, es decir, que asigna a cada

alternativa la misma probabilidad de elección. El segundo hace la comparación con un modelo solo constantes, es decir, considerando únicamente la presencia de constantes modales en la estimación. Este es un mejor indicador del ajuste del modelo porque el modelo solo constantes permite reproducir las proporciones de mercado lo que lo convierte en un referente mucho mejor para realizar la comparación. Los indicadores se definen como:

$$\rho^2 = 1 - \frac{l(\theta)}{l_i} \quad (\text{B-4})$$

Donde $l(\theta)$ es la log-verosimilitud del modelo estimado y l_i es la log-verosimilitud del modelo equiprobable para el caso de ρ^2 y del modelo de solo constantes en el caso de $\bar{\rho}^2$.

b) Constantes Modales

En la Tabla B.1 se presenta la clasificación realizada de las universidades para incorporar las constantes específicas en los distintos modelos estimados. Se presenta además el puntaje obtenido por cada institución en los ranking realizados por Paredes *et al.* (2012) en base a los cuales fue realizada la categorización.

Tabla B.1: Universidades y Constantes Modales.

Grupo	Universidad	Puntaje
I	U.DE CHILE	0,092
I	P.U.CATOLICA DE CHILE	0,080
II	U.DE CONCEPCION	0,053
III	P.U.CATOLICA DE VALPARAISO	0,030
III	U.TEC.F.SANTA MARIA	0.023
III	U.DE SANTIAGO DE CHILE	0.035
III	U.AUSTRAL DE CHILE	0.036
III	U. CATOLICA DEL NORTE	0.025
III	U.DE VALPARAISO	0.020
III	U.METROPOLITANA CS.DE LA EDUCACION	0.018
III	U.TECNOLOGICA METROPOLITANA	0.016
III	U.DE TARAPACA	0.022
III	U.DE ANTOFAGASTA	0.020
III	U.DE LA SERENA	0.017
III	U.PLAYA ANCHA,CS.DE LA EDUCACION	0.015
III	U.DEL BIO-BIO	0.021
III	U.DE LA FRONTERA	0.024
III	U.DE MAGALLANES	0.015
III	U.DE TALCA	0.029
III	U.CATOLICA DE MAULE	0.016
III	U.DIEGO PORTALES	0.018
III	U.ANDRES BELLO	0.022
III	U.ADOLFO IBAÑEZ	0.020
III	U.DE LOS ANDES	0.019
III	U.ALBERTO HURTADO	0.016
IV	U.ARTURO PRAT	0.014
IV	U.DE ATACAMA	0.014
IV	U.DE LOS LAGOS	0.014
IV	U.CATOLICA DE LA SANTISIMA CONCEPCION	0.014
IV	U.CATOLICA DE TEMUCO	0.014
IV	U.MAYOR	S/I
IV	U.FINIS TERRAE	0.013
IV	U.DEL DESARROLLO	0.014

c) Modelo MNL Base

Se considera como modelo base a un modelo MNL en que se incorporan todos los atributos disponibles, pero no se incorpora interacciones entre los mismos ni variaciones sistemáticas de los gustos. La forma funcional de su función de utilidad es la siguiente:

$$\begin{aligned}
 V_{i,q} = & \theta_K + \theta_{ARANC}ARANC + \theta_{PPSU}PPSU + \theta_{DURAC}DURAC \\
 & + \theta_{CRUCH}CRUCH + \theta_{RKQP}RKQP + \theta_{ANAC}ANAC \\
 & + \theta_{MISMAPROV}MISMAPROV
 \end{aligned} \tag{B-5}$$

Luego de estimar el modelo, se obtiene los parámetros presentados en la Tabla B.2, donde se puede apreciar que si bien el signo del parámetro asociado al *Arancel* es correcto, presenta una baja significancia estadística, aún cuando el costo de una alternativa se considera como relevante. Además se puede observar que las constantes modales estimadas siguen la lógica esperada, en que la constante de la categoría I es mayor que la de la II y así sucesivamente, mostrando una preferencia por las universidades con mayor “puntaje” en el ranking utilizado para establecer las categorías.

Tabla B.2: Modelo MNL Base.

	MNL	
K_I	1.05	(13.06)
K_II	0.281	(3.78)
K_III	0.215	(5.03)
Arancel	-0.282	(-0.7)
Promedio PSU	0.973	(17.31)
Duración	-0.04	(-2.76)
Misma Provincia	1.15	(19.31)
Años de Acreditación	3.6	(1.68)
Ranking Qué Pasa	-1.28	(-4.76)
CRUCH (<i>Dummy</i>)	0.23	(4.15)

Es interesante mencionar que, del total de parámetros estimados, 8 son estadísticamente significativos al 95% y uno al 90%, presentando todos el signo esperado. Lo anterior muestra que los atributos considerados son relevantes en la elección de los individuos.

En la Tabla B.3 se presentan los principales indicadores de ajuste del modelo, donde se puede notar que cuenta con una log-verosimilitud de -9.699,265 y un ρ^2 de 0,149 de lo que se puede desprender que el modelo estimado es superior al modelo equiprobable.

Tabla B.3: Indicadores de Ajuste Modelo MNL Base.

Log-verosimilitud	-9.699,265
ρ^2	0,149
$\frac{\rho^2}{\rho^2}$	0,148
# Parámetros	10

d) Modelo MNL con Variaciones Sistemáticas en los Gustos

Se presenta a continuación el modelo MNL estimado incorporando variaciones sistemáticas en los gustos.

En la Tabla B.4 se presentan los resultados de la estimación del modelo, donde se puede ver claramente que los signos de los parámetros estimados son los correctos, tanto para variables relacionadas con el costo como el *Arancel*, la *Duración* y *MismaProv* (La universidad se encuentra en la misma provincia que la del hogar del individuo), como para los atributos asociados a la calidad de la universidad (*Promedio PSU*, *Años de Acreditación* y *Ranking Qué Pasa*). Además el atributo *Proporción Mismo NSE* presenta el signo esperado.

Tabla B.4: Modelo MNL con Variaciones Sistemáticas en los Gustos.

	MNL Interacciones	
K_I	0.708	(8.33)
K_II	0.21	(2.81)
K_III	0.18	(4.16)
Arancel	-0.427	(-4.45)
Ingreso Medio	0.149	(1.54)
Ingreso Alto	0.32	(2.74)
Beca Bicentenario	0.373	(3.61)
Promedio PSU	0.000586	(6.96)
Padres Educación Universitaria	0.000141	(1.39)
Ingreso Medio	0.000208	(1.98)
Ingreso Alto	0.000222	(1.63)
PSU sobre 700	0.000638	(5.15)
Colegio Particular	0.000622	(6.04)
Duración	-0.0319	(-2.16)
Misma Provincia	1.3	(20.93)
Años de Acreditación	0.0787	(3.63)
Proporción Mismo NSE	0.551	(3.73)
Ranking Qué Pasa	-0.00844	(-2.61)
10% Mejor de su colegio	-0.0169	(-3.08)
PSU sobre 700	-0.0264	(-3.31)
Padres Educación Universitaria	-0.0129	(-2.8)
CRUCH	0.0444	(0.74)
Beca Bicentenario	0.435	(3.79)

Al observar los indicadores de ajuste del modelo se puede notar que el modelo es claramente mejor que el modelo equiprobable y que el modelo sólo constantes. Presentando un valor de log-verosimilitud de -9.497,9 y un ρ^2 de 0,167.

Tabla B.5: Indicadores de ajuste modelo MNL con Variaciones sistemáticas en los gustos.

Log-verosimilitud $l(\theta)$	-9.497,9
ρ^2	0,167
$\overline{\rho^2}$	0,165
# Parámetros	23

e) Modelo HL

Se presentan en la Tabla B.6 a continuación las categorías definidas en base al ranking presentado en Paredes *et al.* (2012) que son consideradas al momento de establecer los nidos del modelo HL.

Tabla B.6: Categorías Universidades basada en Paredes et al. (2012).

I	CRUCH - Top
II	CRUCH - Alto
III	CRUCH - Otras
IV	Privadas

Se presentan en la Tabla B.7 a continuación las categorías definidas a partir de la clasificación de universidades presentada en Reyes y Rosso (2012) que son consideradas para la estimación del modelo HL.

Tabla B.7: Categorías Universidades basada en Reyes y Rosso (2012).

I	CRUCH - Investigación y Doctorado
II	CRUCH - Investigación y Doctorado Selectivas
III	CRUCH - Otras
IV	Privadas - Investigación y Doctorado Selectivas
V	Privadas - Otras

Tabla B.8: Universidades y categorías modelo HL.

Clasificación Paredes <i>et al.</i> (2012)	Clasificación Reyes y Rosso (2012)	Universidad
I	I	U.DE CHILE
I	I	P.U.CATOLICA DE CHILE
II	I	U.DE CONCEPCION
II	I	P.U.CATOLICA DE VALPARAISO
II	I	U.DE SANTIAGO DE CHILE
II	I	U.AUSTRAL DE CHILE
II	II	U.DE TALCA
III	II	U.TEC.F.SANTA MARIA
III	II	U. CATOLICA DEL NORTE
III	II	U.DE VALPARAISO
III	II	U.DE TARAPACA
III	II	U.DE ANTOFAGASTA
III	II	U.DE LA SERENA
III	II	U.DE LA FRONTERA
III	III	U.METROPOLITANA CS.DE LA EDUCACION
III	III	U.TECNOLOGICA METROPOLITANA
III	III	U.DEL BIO-BIO
III	III	U.DE MAGALLANES
III	III	U.CATOLICA DE MAULE
III	III	U.ARTURO PRAT
III	III	U.DE LOS LAGOS
III	III	U.CATOLICA DE LA SANTISIMA CONCEPCION
III	III	U.CATOLICA DE TEMUCO
III	III	U.PLAYA ANCHA,CS.DE LA EDUCACION
III	III	U.DE ATACAMA
IV	IV	U.ANDRES BELLO
IV	IV	U.ADOLFO IBAÑEZ
IV	IV	U.DE LOS ANDES
IV	IV	U.ALBERTO HURTADO
IV	V	U. DIEGO PORTALES
IV	V	U.DEL DESARROLLO
IV	V	U.FINIS TERRAE
IV	V	U.MAYOR

- Estimación Modelo HL basado en Reyes y Rosso (2012)

En la Figura B.1 se puede ver de manera gráfica la estructura jerárquica utilizada en el modelo utilizando la estructura jerárquica definida a partir de la propuesta de Reyes y Rosso (2012). Por su parte, en la

Tabla B.9 se presentan los resultados de la estimación del mismo. Se puede notar que los parámetros estimados son similares a los del modelo MNL con Variaciones Sistemáticas en los Gustos y el modelo HL basado Paredes *et al.* (2012). Comparándolo con este último la principal diferencia se da en la valoración de la variable *CRUCH*, la que aumenta un 325%, aumentando su significancia estadística de un 32% a un 92%. Al observar los parámetros asociados a los distintos nidos se puede notar que los cinco nidos considerados presentan parámetros entre 0 y 1, siendo todos estadísticamente distintos de ambos valores.

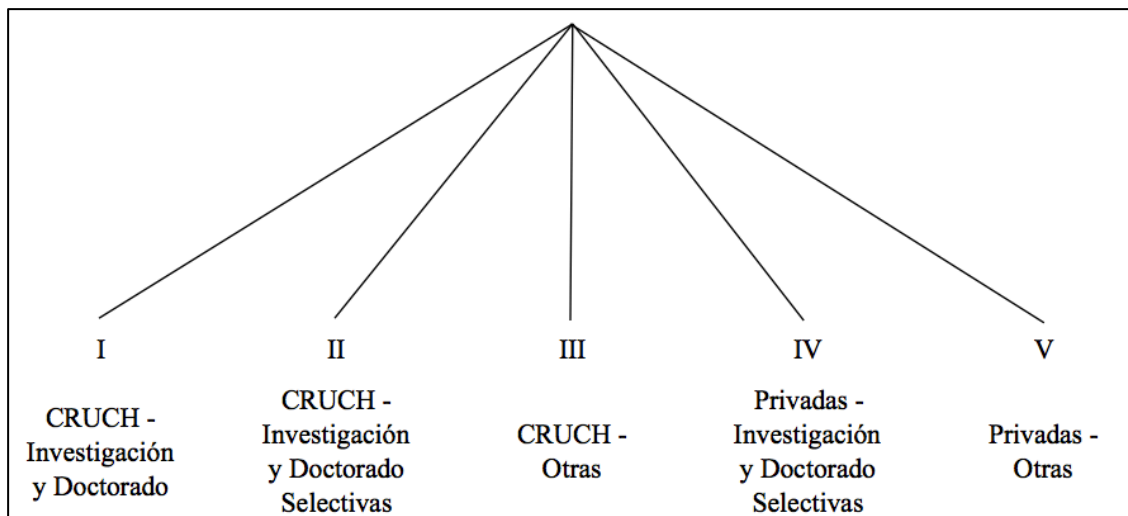


Figura B.1: Estructura modelo HL basado en Reyes y Rosso (2012).

Tabla B.9: Estimación modelo HL basado en Reyes y Rosso (2012).

	NL Interacciones	
K_I	0.653	(7.8)
K_II	0.176	(2.45)
K_III	0.143	(3.12)
Arancel	-3.96	(-4.45)
Ingreso Medio	1.42	(1.57)
Ingreso Alto	2.82	(2.59)
Beca Bicentenario	3.44	(3.66)
Promedio PSU	0.519	(6.57)
Padres Educación Universitaria	0.139	(1.49)
Ingreso Medio	0.203	(2.08)
Ingreso Alto	0.198	(1.56)
PSU sobre 700	0.539	(4.67)
Colegio Particular	0.599	(6.24)
Duración	-0.026	(-1.91)
Misma Provincia	1.17	(19.1)
Años de Acreditación	9.39	(4.44)
Proporción Mismo NSE	0.505	(3.69)
Ranking Qué Pasa	-0.659	(-2.18)
10% Mejor de su colegio	-1.54	(-2.98)
PSU sobre 700	-2.66	(-3.56)
Padres Educación Universitaria	-1.16	(-2.69)
CRUCH	0.102	(1.74)
Beca Bicentenario	0.427	(3.88)
$\phi_{CRUCH-Inv.y Doct.}$	0.81	
$\phi_{CRUCH-Inv.y Doct. Sel.}$	0.67	
$\phi_{CRUCH-Otras}$	0.83	
$\phi_{Privadas-Inv.y Doct. Sel.}$	0.85	
$\phi_{CRUCH-Otras.}$	0.79	

En la Tabla B.10 se presentan los principales indicadores de ajuste del modelo. Se puede apreciar que, en comparación con el modelo HL basado en Paredes *et al.* (2012), la log-verosimilitud del modelo es inferior y la cantidad de parámetros estimados es mayor.

Tabla B.10: Indicadores de ajuste modelo HL basado en Reyes y Rosso (2012).

Log-verosimilitud $l(\theta)$	-9.475,310
ρ^2	0,169
$\frac{\rho^2}{\rho^2}$	0,167
# Parámetros	28

f) Modelo HL sin Proporción Mismo NSE

Se presenta a continuación los parámetros obtenidos de la estimación del modelo HL basado en Paredes *et al.* (2012) pero excluyendo el atributo *Proporción Mismo NSE*. Se muestra en la quinta columna de la Tabla B.11 los resultados de la realización de un test de comparación de parámetros. Se puede ver que solamente el parámetro asociado a la interacción entre el *Arancel* y el ingreso alto es diferente. Pero al realizar un test de significancia estadística conjunta de este parámetro y el del *Arancel*, no se puede rechazar la hipótesis nula de que ambos parámetros sean conjuntamente iguales a cero, lo que ocurre para ambos modelos.

Tabla B.11: Comparación parámetros de modelos sin y con Proporción Mismo NSE.

	Sin Proporción Mismo NSE		Con Proporción Mismo NSE		Test-t de Comparación de Parámetros
	Parámetro	Test-t	Parámetro	Test-t	
K_I	0,811	10,0	0,833	10,3	-0,3
K_II	0,163	2,4	0,160	2,4	0,0
K_III	0,142	4,1	0,137	4,0	0,1
Arancel	-4,750	-6,0	-3,820	-4,7	-1,2
Ingreso Medio	3,510	4,1	3,200	3,7	0,4
Ingreso Alto	4,630	5,3	2,530	2,5	2,4
Beca	2,230	2,8	1,410	1,7	1,0
Bicentenario					
Promedio PSU	0,402	5,7	0,478	6,6	-1,1
Padres					
Educación	0,558	6,1	0,563	6,2	-0,1
Universitaria					
Ingreso Medio	0,129	1,5	0,129	1,5	0,0
Ingreso Alto	0,320	2,8	0,178	1,5	1,3
PSU sobre 700	0,257	2,9	0,161	1,8	1,1
Colegio					
Particular	0,596	5,3	0,601	5,4	0,0
Duración	-0,021	-1,7	-0,023	-1,8	0,2
Misma Provincia	-		0,481	3,8	-
Años de Acreditación	1,090	17,7	1,090	17,6	0,0
Proporción Mismo NSE	6,710	3,6	6,480	3,5	0,1
Ranking Qué Pasa	0,002	0,0	0,024	0,4	-0,4
10% Mejor de su colegio	0,476	4,6	0,401	3,8	0,7
PSU sobre 700	-0,547	-2,0	-0,593	-2,2	0,2
Padres					
Educación	-1,080	-2,8	-1,050	-2,7	-0,1
Universitaria					
CRUCH	-2,220	-3,2	-2,230	-3,2	0,0
Beca					
Bicentenario	-1,200	-2,6	-1,180	-2,5	0,0

g) Significancia Conjunta de Parámetros

Se presentan en esta sección las significancias conjuntas de los parámetros y sus interacciones, de manera tal de poder analizar con mayor detalle el efecto que éstos tienen.

En la Tabla B.12 se presentan los tests para el modelo HL, donde se puede ver que la significancia estadística conjunta de los parámetros estimados y sus interacciones permite concluir que, para los atributos asociados al ingreso, con un nivel de significancia del 95% se puede decir que el efecto conjunto es igual a cero. Para todos los demás parámetros e interacciones se puede afirmar que el efecto conjunto es significativamente distinto de cero.

Tabla B.12: Test de significancia estadística conjunta de parámetros del modelo HL.

Atributo	Interacción con	Test-t conjunto
Arancel	Ingreso Medio	-1,5
Arancel	Ingreso Alto	-0,7
Arancel	Beca BIC	-0,4
Promedio PSU	Padres Educación Universitaria	5,0
Promedio PSU	Ingreso Medio	4,4
Promedio PSU	Ingreso Alto	4,0
Promedio PSU	PSU sobre 700	8,0
Promedio PSU	Colegio Particular	9,1
Ranking Qué Pasa	10% Mejor de su colegio	-3,0
Ranking Qué Pasa	PSU sobre 700	-3,7
Ranking Qué Pasa	Padres Educación Universitaria	-2,9
CRUCH	Beca BIC	3,1

En la Tabla B.13 se puede ver que para el modelo 2LCL el efecto es similar al existente en el modelo HL, ya que las interacciones asociadas al parámetro *Arancel* no son estadísticamente significativas (con respecto a cero). Pero para todos los demás atributos que incorporan interacciones, el efecto conjunto entre el parámetro del atributo y la interacción es significativamente distinto de cero. La única diferencia se produce al observar la significancia estadística conjunta del atributo *CRUCH* y la beca BIC para la clase 1, donde la interacción parece significativa, pero de manera conjunta con el parámetro del atributo no lo es.

Tabla B.13: Test de significancia estadística conjunta de parámetros del modelo 2LCL.

Atributo	Interacción	Clase 1	Clase 2
		Test-t conjunto	Test-t conjunto
Arancel	Ingreso Medio	-1,4	-1,5
Arancel	Ingreso Alto	-0,7	-0,9
Arancel	Beca BIC	-0,9	-1,1
Ranking Qué Pasa	10% Mejor de su colegio	-2,8	-4,2
Ranking Qué Pasa	PSU sobre 700	-3,0	-4,0
CRUCH	Beca BIC	0,7	3,1

En la Tabla B.14 se puede notar que, para las clases 1 y 3 del modelo 3LCL la significancia conjunta de los atributos y las interacciones incorporadas es baja, mientras que para la clase 2, en ambas interacciones la significancia conjunta es alta, superando el test con un 95% de significancia.

Tabla B.14: Test de significancia estadística conjunta de parámetros del modelo 3LCL.

Atributo	Interacción	Clase 1	Clase 2	Clase 3
		Test-t conjunto	Test-t conjunto	Test-t conjunto
Arancel	Ingreso Medio/Alto	0.7	-2.2	-0.7
Arancel	Beca BIC	0.9	-2.0	-0.3