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

PREDICCIÓN DE FUGA DE CLIENTES DESDE UN ENFOQUE DE COMPETENCIA

TESIS PARA OPTAR AL GRADO DE MAGISTER EN GESTIÓN DE OPERACIONES

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERO CIVIL INDUSTRIAL

LEONARDO ANDRÉS LÓPEZ HERRERA

PROFESOR GUÍA:
RICHARD WEBER HAAS

MIEMBROS DE LA COMISIÓN:
NICOLÁS FIGUEROA GONZÁLEZ
CRISTIÁN BRAVO ROMÁN
GASTÓN L'HUILLIER CHAPARRO

SANTIAGO DE CHILE
JUNIO 2013

Resumen de la tesis para optar al título de Ingeniero Civil
Industrial y grado de Magíster en Gestión de Operaciones
Alumno: Leonardo López Herrera
Prof. Guía: Richard Weber Haas
Fecha: 04/06/2013

PREDICCIÓN DE FUGA DE CLIENTES DESDE UN ENFOQUE DE COMPETENCIA

Las firmas requieren contar con herramientas que les permitan estimar probabilidades de fuga para su cartera de clientes y así decidir sobre qué clientes concentrar sus esfuerzos de retención. Este problema ha sido estudiado profundamente a través de métodos estadísticos tradicionales, como por ejemplo la Regresión Logística.

En esta tesis se propone estimar las probabilidades de fuga de clientes a partir de la forma funcional utilizada en modelos de Regresión Logística, pero imponiendo que los parámetros del modelo cumplan con las condiciones de equilibrio de un modelo de Teoría de Juegos, donde las firmas utilizan sus herramientas competitivas para atraer a los clientes y los clientes escogen fugarse o no según estas mismas probabilidades estimadas endógenamente. Para esto se plantea un problema de optimización no lineal cuya función objetivo corresponde a maximizar la función de verosimilitud del modelo tradicional de Regresión Logística y cuyas restricciones corresponden al sistema de ecuaciones que determinan las condiciones de equilibrio del juego. Se plantean tres modelos de Teoría de Juegos para abordar la estimación a través del esquema propuesto: el primero, donde las firmas compiten solamente en precios; el segundo, donde las firmas compiten en precios y descuentos; y finalmente, uno donde las firmas compiten en precios, descuentos y esfuerzos en marketing.

Se propone un método de solución numérica para la metodología planteada y luego es aplicada a un problema con datos reales. Los modelos propuestos modifican la importancia relativa de las variables permitiendo mejorar la interpretación de los resultados y los niveles de significancia de las variables que son consideradas endógenas en los respectivos juegos. Los resultados muestran que el modelo tradicional siempre supera a los modelos propuestos en las bases de entrenamiento y de prueba. Esto sucede porque estos últimos imponen condiciones sobre los parámetros estimados buscando que estos además satisfagan las condiciones de equilibrio del juego subyacente. Sin embargo, se observa que los modelos propuestos superan los resultados del enfoque tradicional en particular en la base de validación. Lo anterior es relevante ya que esta base corresponde a una muestra fuera del periodo temporal utilizado en la determinación de los modelos, lo que indica que los modelos propuestos se adaptan mejor a cambios en las condiciones de mercado con respecto a la situación bajo la que se estiman sus parámetros.

En definitiva los modelos propuestos sacrifican precisión en la estimación de sus parámetros para mejorar su capacidad de entender el comportamiento de compra de los clientes, recogiendo efectos de largo plazo de la base de datos y siendo consistentes con la racionalidad de los jugadores. Como línea de trabajo futuro se recomienda la aplicación de la metodología propuesta a cualquier otro problema de Clasificación Binaria donde los agentes se comporten de forma estratégica y donde se pueda caracterizar un juego de competencia subyacente al problema estadístico estudiado.

A mi familia y a mi pareja, por todos estos años de amor y apoyo incondicional

Agradecimientos

Quiero agradecer a todos quienes han sido parte y me han ayudado, de muy diferentes formas, durante esta importante etapa de mi vida. En particular agradecer a quienes me apoyaron en el desarrollo de este trabajo que lamentablemente postergué por mucho tiempo.

En primer lugar quiero agradecer de forma muy especial a Richard Weber por motivar esta tesis, por la constante preocupación por mis avances, por los consejos y excelentes observaciones a este trabajo. En particular agradezco haberme dado la oportunidad de dictar la cátedra del curso de Optimización durante el semestre Primavera 2008, la que para mi fue una experiencia académica extraordinaria. Agradezco también a Nicolás Figueroa por aceptar ayudarme en esta tesis sobretodo en la parte inicial, por entender los objetivos y alcances del trabajo y por las valiosas observaciones entregadas. Mi agradecimiento también para Cristián Bravo y Gastón L'Huillier por aceptar ser miembros de la comisión de esta tesis, por la excelente disposición, por la detallada revisión que realizaron de este trabajo y por los valiosos comentarios de ambos.

Le doy las gracias a todos mis amigos, dentro y fuera de la Universidad, quienes me han acompañado y han sido fundamentales en esta etapa de mi vida.

Agradezco infinitamente a mi familia. Sin ustedes esto no sería posible. A mis padres Juan y Edith, gracias por ser un ejemplo de vida, por el esfuerzo y entrega de todos estos años. Les agradezco haberme transmitido sus valores, por motivarme desde siempre a estudiar, por confiar en mis capacidades y por todo el amor que me entregan. A mis hermanos les agradezco por la compañía, la lealtad y el cariño. Por sus talentos y valores. Por ser como son.

Finalmente quiero agradecer mi pareja Ma. Eugenia por todo el amor y por enseñarme a descubrir mi faceta más sensible. Durante estos años has sido fundamental para conseguir este objetivo, siempre tuviste la palabra y el abrazo que necesite. Gracias por la paciencia, los sacrificios y por tu apoyo incondicional.

Tabla de Contenido

I Presentación de la Tesis **19**

1. Descripción del Problema **21**
1.1. Introducción . . . 21
1.2. Retención de Clientes y Competencia de Mercado . . . 22
1.3. Predicción de Fuga de Clientes . . . 26
1.4. Justificación de la Tesis . . . 26
1.5. Objetivos . . . 28
1.5.1. Objetivo General . . . 28
1.5.2. Objetivos Específicos . . . 28
1.6. Motivación: Data Mining y Teoría de Juegos . . . 29
1.7. Estructura del trabajo . . . 29

II Marco Teórico **31**

2. Conceptos de Teoría de Juegos **33**
2.1. Introducción . . . 33
2.2. Comportamiento Racional y la Función de Utilidad . . . 34
2.3. Representación de un Juego . . . 35
2.3.1. Juego en su Forma Estratégica o Normal . . . 35
2.3.2. Juego en su Forma Extensiva . . . 35
2.3.3. Juegos con Estrategias Mixtas . . . 36
2.4. Solución de un Juego y el Concepto de Equilibrio . . . 37
2.4.1. Definición de Equilibrio . . . 37
2.4.2. Equilibrio en Estrategias Dominantes . . . 38
2.4.3. Equilibrio en Nash . . . 38
2.5. Juegos Dinámicos con Información Completa . . . 39
2.5.1. Juegos Dinámicos, Subjuegos y Equilibrio Perfecto en el Subjuego . . . 39
2.5.2. Inducción Hacia Atrás y Perfección en el Subjuego . . . 40

3. Retención de Clientes y Competencia **41**
3.1. Introducción . . . 41
3.2. Competencia en Precios . . . 42
3.2.1. Competencia en Precios de Hotelling . . . 42

3.2.2. Competencia con Discriminación en Precios y Costos de Cambio . . . 43
3.2.3. Competencia con Discriminación en Precios y Preferencias de los Consumidores . . 44
3.2.4. Pagar para Cambiar o Pagar para Retener . . . 45
3.2.5. Competencia con Promociones Uno a Uno . . . 46
3.3. Competencia en Esfuerzos de Marketing . . . 47
3.4. Competencia en Precios y Esfuerzos en Marketing . . . 48
3.4.1. Formulación . . . 48
3.4.2. Resolución . . . 50
3.4.3. Resultados y Conclusiones . . . 51

4. Clasificación Binaria, Modelos de Elección Discreta y de Regresión Logística 53
4.1. Introducción . . . 53
4.2. Clasificación Binaria . . . 54
4.3. Modelos de Elección Discreta . . . 55
4.3.1. Definiciones . . . 55
4.3.2. Probabilidades de Elección . . . 55
4.3.3. Modelo de Elección Discreta Logit Multinomial . . . 57
4.3.4. Propiedades de las Probabilidades Logit . . . 57
4.4. Regresión Logística . . . 61
4.4.1. Definición . . . 61
4.4.2. Regresión Logística como Modelo de Elección Discreta Binaria . . . 62
4.4.3. Estimación del Modelo de Regresión Logística . . . 63
4.4.4. Solución Numérica y Método de Newton-Raphson . . . 65
4.4.5. Implementación Computacional del Modelo de Regresión Logística . . . 67

III Modelos Propuestos y Resultados 69

5. Predicción de Fuga de Clientes con Restricciones de Competencia 71
5.1. Fuga de Clientes utilizando Regresión Logística . . . 71
5.2. Regresión Logistica con Restricciones de Competencia . . . 72
5.2.1. Definición del modelo . . . 72
5.2.2. Solución del modelo . . . 74
5.3. Predicción de Fuga de Clientes con Restricciones de Competencia en Precios . . . 76
5.3.1. Definición del Problema . . . 76
5.3.2. Resolución del Problema . . . 78
5.4. Predicción de Fuga de Clientes con Restricciones de Competencia en Precios y Descuentos en Precios . . . 79
5.4.1. Definición del Problema . . . 79
5.4.2. Resolución del Problema . . . 81
5.5. Predicción de Fuga de Clientes con Restricciones de Competencia en Precios, Descuentos en Precios y Esfuerzos en Marketing . . . 81
5.5.1. Definición del Problema . . . 81
5.5.2. Resolución del Problema . . . 84
5.6. Implementación Computacional de las Metodologías Propuestas . . . 84

6. Predicción de Fuga de Clientes con Restricciones de Competencia: Caso Aplicado **85**
6.1. Base de datos . . . 85
6.1.1. Descripción y variables disponibles . . . 85
6.1.2. Análisis del mercado a estudiar . . . 86
6.2. Aplicación de metodologías . . . 91
6.2.1. Descripción . . . 91
6.2.2. Base de datos de trabajo . . . 93
6.2.3. Selección de atributos . . . 96
6.3. Resultados . . . 98
6.3.1. Resultados Generales . . . 98
6.3.2. Análisis Resultados Estadísticos . . . 99
6.3.3. Interpretación de Variables . . . 103

IV Conclusiones **109**

7. Conclusiones **111**
7.1. Conclusiones Generales . . . 111
7.2. Conclusiones Específicas . . . 112
7.2.1. Modelamiento del Proceso de Fuga de Clientes . . . 112
7.2.2. Probabilidad de Fuga de Clientes desde el Enfoque Tradicional . . . 113
7.2.3. Probabilidades de Fuga de Clientes según Metodología Propuesta . . . 113
7.2.4. Aplicación de las Metodologías Propuestas a un Caso Real . . . 114
7.2.5. Implementación de las Metodologías Propuestas . . . 115
7.3. Desafíos Futuros . . . 116

Bibliografía **119**

Anexos **124**

A. Conceptos de Teoría de Juegos **125**
A.1. El Dilema del Prisionero . . . 125
A.1.1. Definición del Juego . . . 125
A.1.2. Representación en Forma Estratégica o Normal . . . 126
A.1.3. Representación en Forma Extensiva . . . 126
A.1.4. Equilibrios en el Juego . . . 126

B. Retención de Clientes y Competencia **129**
B.1. Competencia con Discriminación en Precios y Costos de Cambio . . . 129
B.2. Competencia con Discriminación en Precios y Preferencias de los Consumidores . . . 131
B.3. Pagar para Cambiar o Pagar para Retener . . . 132
B.4. Competencia con Promociones Uno a Uno . . . 133
B.5. Competencia en Esfuerzos de Marketing . . . 135

C. Modelo de Elección Discreta Logit Multinomial: Estimación de las Probabilidades Elección **141**

D. Criterios de Evaluación de los Modelos **145**
D.1. Bondad de Ajuste del Modelo . . . 145
D.1.1. Razón de Verosimilitud . . . 145
D.1.2. Significancia del Modelo y de los Atributos del Modelo . . . 146
D.2. Estadístico de Kolmogórov-Smirnov (KS) . . . 148
D.3. Curva ROC . . . 149
D.3.1. Área Bajo la Curva ROC . . . 149
D.3.2. Estadístico J de Youden . . . 151
D.4. Error Cuadrático Medio (MSE) . . . 152

E. Modelos de Fuga de Clientes con Restricciones de Competencia: Detalle de Cálculos **153**
E.1. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios . . . 153
E.1.1. Cálculo del Gradiente $\nabla_b \tilde{\ell}(b)$ . . . 153
E.1.2. Cálculo del Hessiano $\mathbf{H}_{\tilde{\ell}}(b)$ . . . 154
E.2. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios y Descuentos en Precios . . . 156
E.2.1. Cálculo del Gradiente $\nabla_b \tilde{\ell}(b)$ . . . 156
E.2.2. Cálculo del Hessiano $\mathbf{H}_{\tilde{\ell}}(b)$ . . . 157
E.3. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios, Descuentos en Precios y Esfuerzos en Marketing . . . 160
E.3.1. Cálculo del Gradiente $\nabla_b \tilde{\ell}(b)$ . . . 160
E.3.2. Cálculo del Hessiano $\mathbf{H}_{\tilde{\ell}}(b)$ . . . 161

F. Implementación Computacional de las Metodologías **165**
F.1. Modelo de Regresión Logística . . . 165
F.1.1. Cálculo del Gradiente $\nabla_b \ell(b/x)$ . . . 165
F.1.2. Cálculo del Hessiano $\mathbf{H}_{\ell}(b/x)$ . . . 166
F.1.3. Método de Newton-Raphson . . . 169
F.2. Modelo de Regresión Logística con Restricciones de Competencia . . . 170
F.2.1. Método de Newton-Raphson . . . 170
F.2.2. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios . . . 172
F.2.3. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios y Descuentos en Precios . . . 172
F.2.4. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios, Descuentos en Precios y Esfuerzos en Marketing . . . 173

G. Implementación Computacional de las Metodologías: Códigos de Fuente **175**
G.1. Regresión Logística . . . 175
G.2. Modelos Propuestos . . . 177

H. Diccionario de variables **191**
H.1. Variables asociadas a elementos competitivos . . . 191
H.2. Variables demográficas . . . 192
H.3. Variables construidas . . . 195

I. Análisis descriptivo de variables **197**

I.1. Variables asociadas a elementos competitivos . . . 197
I.1.1. Variables continuas . . . 197
I.1.2. Variables categóricas . . . 199
I.2. Variables construidas . . . 201
I.2.1. Variables continuas . . . 201
I.2.2. Variables categóricas . . . 203
I.3. Variables demográficas . . . 205
I.3.1. Variables continuas . . . 205
I.3.2. Variables categóricas . . . 207

J. Resultados **217**
J.1. Selección de atributos . . . 218
J.2. Conjunto de atributos A . . . 220
J.2.1. Modelo Regresión Logística (*Benchmark*) . . . 220
J.2.2. Modelo propuesto (1) . . . 221
J.2.3. Modelo propuesto (2) . . . 223
J.2.4. Modelo propuesto (3) . . . 224
J.3. Conjunto de atributos B . . . 225
J.3.1. Modelo Regresión Logística (*Benchmark*) . . . 225
J.3.2. Modelo propuesto (1) . . . 229
J.3.3. Modelo propuesto (2) . . . 232
J.3.4. Modelo propuesto (3) . . . 235

Índice de Tablas

4.1. Tabla de confusión e identificación de aciertos y errores 55

6.1. Compras de Ketchups en formato 32 oz. por marca y año. Fuente: Elaboración Propia. . . . 87
6.2. Hogares y compras por clasificación de frecuencia de compra. Fuente: Elaboración Propia. . 88
6.3. Hogares expuestos a comerciales de TV de cada marca. Fuente: Elaboración Propia. 89
6.4. Matriz de transición entre marcas, en número de casos. Fuente: Elaboración Propia. 91
6.5. Matriz de transición entre marcas, en porcentaje de casos. Fuente: Elaboración Propia. . . . 91
6.6. Matriz de transición entre marcas y según clasificación de frecuencia de compra. Fuente: Elaboración Propia. 92
6.7. Promedio de días entre compras, según frecuencia de compra. Fuente: Elaboración Propia. . 92
6.8. Asociación entre variables de los modelos y variables de la base de datos. 94
6.9. Número de compras según base de datos. Fuente: Elaboración Propia. 95
6.10. Distribución variable dependiente por base de datos. Fuente: Elaboración Propia. 96
6.11. Parámetros utilizados en la estimación de los modelos propuestos. Fuente: Elaboración Propia. 99
6.12. Resultados estadístico Kolmogórov-Smirnov (KS) por modelo y base de datos. Fuente: Elaboración Propia. 100
6.13. Resultados área bajo la curva (AUC) por modelo y base de datos. Fuente: Elaboración Propia. 101
6.14. Resultados error cuadrático medio (MSE) por modelo y base de datos. Fuente: Elaboración Propia. 102
6.15. Coeficientes para modelos con atributos *A* y cambio porcentual de su valor absoluto respecto del *Benchmark*. Fuente: Elaboración Propia. 105
6.16. Valor absoluto de estadístico *Z* para modelos con atributos *A* y cambio porcentual de su valor absoluto respecto del *Benchmark*. Fuente: Elaboración Propia. 106

A.1. Juego *Dilema del Prisionero* en su forma estratégica o normal 126

H.1. Definición de variables asociadas a elementos competitivos 191
H.2. Definición de variables demográficas 195
H.3. Definición de variables construidas I 196
H.4. Definición de variables construidas II 196

I.1. Análisis descriptivo 197
I.2. Análisis de Varianza (ANOVA) 198
I.3. Análisis descriptivo 199
I.4. Pearson Chi-Square tests 200

I.5. Análisis descriptivo . . . 201
I.6. Análisis de Varianza (ANOVA) . . . 202
I.7. Análisis descriptivo . . . 203
I.8. Pearson Chi-Square tests . . . 204
I.9. Análisis descriptivo . . . 205
I.10. Análisis de Varianza (ANOVA) . . . 206
I.11. Análisis descriptivo . . . 212
I.12. Pearson Chi-Square tests . . . 215

J.1. Resultado selección de atributos. S: variables seleccionadas por algoritmo. M: variables agregadas manualmente. . . . 219
J.2. Set atributos A. Resultado Modelo de Regresión Logística (*Benchmark*). . . . 220
J.3. Set atributos A. Modelo *Benchmark*, Estadísticos de Training. . . . 220
J.4. Set atributos A. Modelo *Benchmark*, Estadísticos de Eficiencia por Base de datos. . . . 221
J.5. Set atributos A. Resultado Modelo (1): Competencia en precios. . . . 221
J.6. Set atributos A. Modelo (1), Estadísticos de Training. . . . 222
J.7. Set atributos A. Modelo (1), Estadísticos de Eficiencia por Base de datos. . . . 222
J.8. Set atributos A. Resultado Modelo (2): Competencia en precios y descuentos. . . . 223
J.9. Set atributos A. Modelo (2), Estadísticos de Training. . . . 223
J.10. Set atributos A. Modelo (2), Estadísticos de Eficiencia por Base de datos. . . . 224
J.11. Set atributos A. Resultado Modelo (3): Competencia en precios, descuentos y marketing. . . . 224
J.12. Set atributos B. Modelo (3), Estadísticos de Training. . . . 225
J.13. Set atributos B. Modelo (3), Estadísticos de Eficiencia por Base de datos. . . . 225
J.14. Set atributos A. Resultado Modelo de Regresión Logística (*Benchmark*). . . . 228
J.15. Set atributos B. Modelo *Benchmark*, Estadísticos de Training. . . . 228
J.16. Set atributos B. Modelo *Benchmark*, Estadísticos de Eficiencia por Base de datos. . . . 228
J.17. Set atributos B. Resultado Modelo (1): Competencia en precios. . . . 231
J.18. Set atributos B. Modelo (1), Estadísticos de Training. . . . 231
J.19. Set atributos B. Modelo (1), Estadísticos de Eficiencia por Base de datos. . . . 232
J.20. Set atributos B. Resultado Modelo (2): Competencia en precios y descuentos. . . . 234
J.21. Set atributos B. Modelo (2), Estadísticos de Training. . . . 235
J.22. Set atributos B. Modelo (2), Estadísticos de Eficiencia por Base de datos. . . . 235
J.23. Set atributos B. Resultado Modelo (3): Competencia en precios, descuentos y marketing. . . . 238
J.24. Set atributos B. Modelo (3), Estadísticos de Training. . . . 238
J.25. Set atributos B. Modelo (3), Estadísticos de Eficiencia por Base de datos. . . . 238

Índice de Figuras

1.1. Aumentos de utilidad por aumentos de un 5 % en la tasa de retención. Fuente: Elaboración propia, con datos de Reichheld y Sasser (2000) 23
1.2. Esquema de ganancias por cliente en el tiempo. 23
1.3. Esquema de interacción entre agentes involucrados. 28

3.1. Modelo de Competencia de Hotelling 43
3.2. Demandas de cada firma 49

4.1. Forma funcional *Sigmoidal* de la probabilidad Logit 58

6.1. Clasificación de hogares según número de compras. Fuente: Elaboración Propia. 88
6.2. Número de hogares expuestos a comerciales de TV y exposiciones promedio. Fuente: Elaboración Propia. 89
6.3. Distribución de hogares según número de exposiciones a comerciales de TV. Fuente: Elaboración Propia. 90
6.4. Distribución de compras según promedio de días entre compras. Fuente: Elaboración Propia. 93
6.5. Esquema de bases de datos: Entrenamiento, Prueba y Validación 95

A.1. Juego *Dilema del Prisionero* en su forma extensiva 127

D.1. Ejemplo estadístico Kolmogórov-Smirnov (KS) 149
D.2. Ejemplo estadístico área bajo la curva ROC (AUC) 150
D.3. Ejemplo estadístico J de Younden 151

Parte I

Presentación de la Tesis

Capítulo 1

Descripción del Problema

1.1. Introducción

La realidad competitiva que en estos días enfrentan las firmas ha provocado que éstas ya no concentren sus esfuerzos de marketing exclusivamente en estrategias de captación de clientes, sino también en estrategias de retención y fidelización. En una estrategia de captación las firmas buscan aumentar su cartera de clientes incentivando a clientes de la competencia y a nuevos consumidores que aún no participan del mercado para que la elijan. Una estrategia de retención busca aumentar la lealtad y fidelidad de la cartera de clientes, generando preferencia hacia la firma y evitando que decidan cambiarse a una firma competidora. Esto permite a la firma profundizar y rentabilizar la relación con el cliente a lo largo de su vida, traduciéndola en mayores utilidades. Distintos estudios destacan la importancia de una estrategia eficiente de retención por sobre una estrategia de captación, y su impacto positivo sobre las utilidades de las firmas que las ejecutan eficientemente.

En determinados mercados la cartera de clientes resulta ser uno de los principales activos de las firmas y por lo tanto defenderla con eficacia resulta ser una tarea estratégica. En este contexto los modelos cuantitativos para apoyar la retención de clientes son un componente fundamental de las estrategias de marketing que ejecutan las firmas en mercados competitivos. Los resultados entregados por este tipo de modelos permite a las firmas concentrar los esfuerzos de retención en los clientes más valiosos y con mayor probabilidad de abandonar la firma. Evitar la fuga de sus mejores clientes permite a las firmas aumentar la rentabilidad de su cartera de clientes en el largo plazo.

Para abordar el problema de retención de clientes, las firmas utilizan enfoques basados en técnicas tradicionales de la estadística y el Data Mining. Para construir estos modelos, se utilizan datos históricos sobre el comportamiento y características de la cartera de clientes. Este problema también puede ser abordado desde un enfoque microeconómico, modelando la competencia de la industria y entendiendo la interacción y decisiones de los agentes participantes. Las firmas del mercado compiten en precios, descuentos y esfuerzos de marketing para atraer y retener a los consumidores. Los consumidores analizan las ofertas que ofrecen las firmas del mercado y escogen la más conveniente según los costos de cada opción y sus preferencias. La literatura que aborda el problema a través de este enfoque utiliza principalmente modelos de Teoría de

Juegos. Éstos tienden a ser altamente teóricos al intentar conceptualizar y comprender el fenómeno económico subyacente. Este tipo de modelos todavía no han podido ser llevados a la práctica en aplicaciones reales debido al nivel de complejidad y la necesidad de que estas soluciones permitan a las firmas que las utilizan traducirlas en decisiones de retención a nivel de cada cliente de la cartera.

El propósito del presente trabajo es abordar el problema de la retención de clientes intentando incorporar el enfoque de competencia en las técnicas tradicionalmente utilizadas por las firmas para implementar sus estrategias de retención. Más particularmente se busca abordar el problema de predicción de las probabilidades de fuga de clientes en un contexto de mercados competitivos.

1.2. Retención de Clientes y Competencia de Mercado

Las firmas desean determinar cuáles son los esfuerzos que deben realizar en captación y retención de clientes para maximizar sus utilidades, dado que poseen recursos limitados. Si bien durante muchos años las firmas concentraron la mayor parte de sus esfuerzos en estrategias de captación de clientes, en la actualidad cada vez tienen mayor relevancia las estrategias de retención, lo que se puede observar principalmente en mercados maduros y/o fuertemente competitivos. Varios estudios señalan la importancia de las estrategias de retención de clientes. Una de las razones para concentrar esfuerzos en retención, es el costo de adquirir nuevos clientes en comparación con el costo de retener clientes existentes. Grand (2002) plantea que puede ser por lo menos cinco veces más costoso adquirir un nuevo cliente que retener uno existente. Por otro lado, Reichheld y Sasser (2000) concluyen que aumentos de un 5 % en la tasa de retención de clientes puede generar incrementos de utilidades del orden de un 25 % a un 85 % promedio, dependiendo del tipo de negocio o industria, como se observa en la figura 1.1. Según Reichheld y Sasser (2000), a medida que aumenta la permanencia de los clientes en el tiempo, aumentan los beneficios que estos entregan a la empresa. Esto se explica por aumentos en los niveles de transacciones con los clientes, posibilidad de aplicar estrategias de *cross-selling* y *up-selling*, reducciones en los costos operacionales, referencias que atraen nuevos clientes, entre otros factores. En definitiva, resulta fundamental para el éxito de una empresa su capacidad de mantener por largo tiempo a sus mejores clientes para forjar relaciones más duraderas que se traduzcan en en mayores niveles de utilidad, tal como lo esquematiza la figura 1.2.

El mayor costo de adquirir un nuevo cliente, con respecto a retener uno existente, se explica porque el consumidor debe tomar la decisión de abandonar una firma competidora a la que ya se encuentra afiliado y esta decisión tiene costos importantes para el cliente. Un cliente que decide cambiarse de firma debe estar convencido que el beneficio que obtendrá en la nueva firma será superior al costo asociado al cambio más los beneficios que obtendría manteniéndose en la misma firma.

Se define el *switching cost* como el costo que debe enfrentar un cliente cuando decide abandonar una firma y escoger una firma rival. Klemperer (1995) clasifica los *switching costs* en distintos tipos:

- **Costos por compatibilidad:** Al cambiar de proveedor de algún producto y/o servicio el cliente debe evaluar si el nuevo proveedor entregará productos compatibles con otros que el consumidor posee en otros ámbitos. Por ejemplo si un cliente decide cambiar algún Software que utilice en su computador,

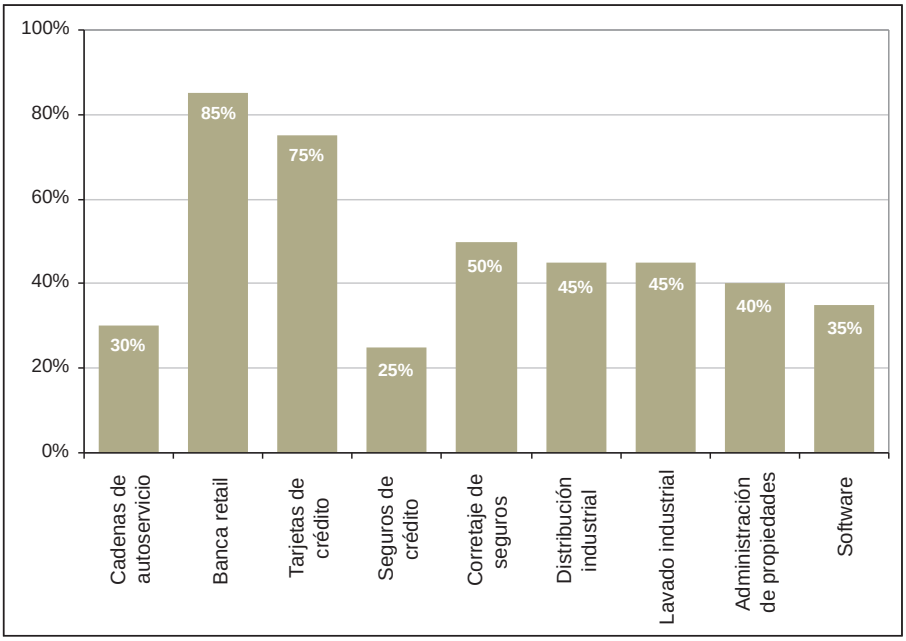

Figura 1.1: Aumentos de utilidad por aumentos de un 5 % en la tasa de retención. Fuente: Elaboración propia, con datos de Reichheld y Sasser (2000)

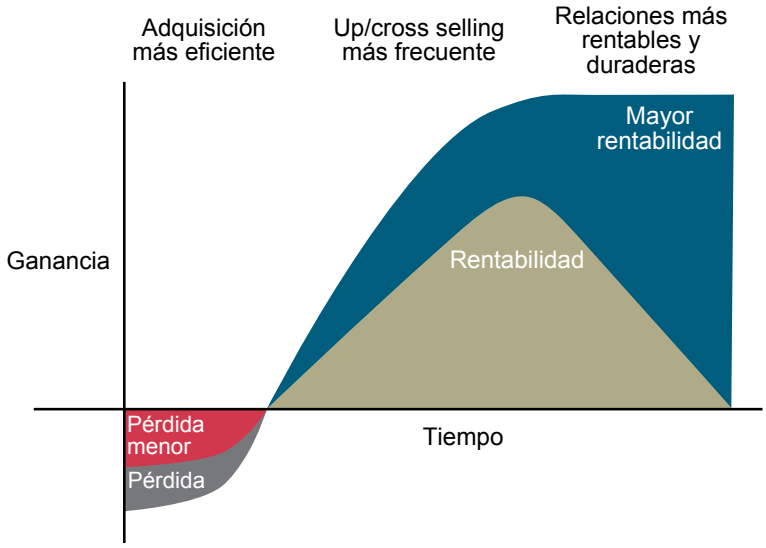

Figura 1.2: Esquema de ganancias por cliente en el tiempo.

debe evaluar si el nuevo Software es compatible con su Sistema Operativo, si es capaz de comunicarse con los otros Softwares que utiliza, si es compatible con la impresora que posee, etc.

- **Costos de transacción por cambio de proveedor:** Existen costos asociados a la transacción, al trámite y al cumplimiento de cláusulas contractuales asociadas a la decisión de dejar de ser cliente de una firma y afiliarse a una nueva. Estos costos pueden ser por ejemplo: el tiempo invertido por el cliente en la transacción, el pago de multas por dar término por adelantado a un contrato y los costos notariales de firmar un nuevo contrato. Algunos de este tipo de costos se incurren por ejemplo cuando un cliente decide cerrar una cuenta corriente en un banco y abrir una nueva en otro.

- **Costos de aprendizaje:** Hay un costo asociado a tener que aprender a utilizar un nuevo producto y/o servicio. Si un cliente se cambia de banco, debe aprender a utilizar el nuevo sitio web del banco, conocer los nuevos teléfonos de atención al cliente, conocer las nuevas políticas del banco, etc. Otro ejemplo es el usuario que decide cambiar su sistema operativo de Windows a Linux, quien debe aprender un nuevo lenguaje, funciones y *layout* del sistema.

- **Incertidumbre acerca de la calidad de productos no probados:** Cuando un consumidor escoge un producto y/o servicio que satisface sus necesidades es reacio a cambiarse a un competidor por el riesgo de que éste no cumpla sus expectativas. Por ejemplo las personas que utilizan un medicamento de un laboratorio determinado no quieren tomar el mismo medicamento comercializado por otro laboratorio, ya que tienen dudas con respecto a su calidad y a que no tenga la misma efectividad del que ya conoce.

- **Pérdidas de programas de beneficios:** Las firmas premian a sus clientes frecuentes con beneficios a través de sus programas de lealtad. Si un consumidor decide cambiar de firma perdería sus beneficios acumulados. Por ejemplo, los programas de viajero frecuente de las aerolíneas incentivan a sus pasajeros a seguir prefieriendola para seguir acumulando kilómetros o millas que luego podrán canjear por pasajes en futuros viajes.

- **Costos psicológicos:** Referidos a la lealtad no económica que las firmas generan sobre los consumidores, principalmente a través del marketing. Las marcas a través del marketing generan una personalidad y se asocian con determinadas causas con la que los consumidores se identifican. Esto genera una preferencia a priori del consumidor hacia determinada firma. Un consumidor frente a dos productos equivalentes escogerá el que ofrece la firma de su preferencia.

A través de esfuerzos de marketing las firmas pueden crear *switching costs* sobre su cartera actual de clientes. Sobre los clientes que tienen intensiones de abandonar la firma, se pueden aplicar estrategias de retención para crear o aumentar estos costos asociados al cambio. El objetivo es hacer más difícil la decisión del cliente de abandonar la firma. Los programas de fidelidad que las firmas ofrecen buscan que los clientes obtengan beneficios por preferirla recurrentemente. Si una firma identifica que un cliente tiene intenciones de abandonar la firma, ésta puede ofrecer por ejemplo descuentos en el precios o algún beneficio especial por mantenerse en la firma.

Los *switching costs* relacionados a costos económicos son fácilmente compensables por las otras firmas que buscan atraer a un cliente. El competidor que busca captar a un cliente de una firma rival puede asumir y pagar los *switching costs* que enfrenta el consumidor. Un competidor fácilmente puede igualar el descuento ofrecido por otra firma como oferta de retención. Si el cliente enfrenta un costo de aprendizaje, la firma

competidora puede ofrecer las facilidades para que el nuevo cliente conozca como utilizar su producto y/o servicio. En caso que exista incertidumbre con respecto a la calidad de producto ofrecido por la competencia, estas pueden hacer demostraciones y dar un periodo de prueba para que el consumidor se entere de las bondades de sus productos. Por esta razón es que las firmas no pueden escoger estrategias de retención que se basen únicamente en crear *switching costs* asociados a costos económicos. La literatura especializada pone énfasis en que las estrategias de retención que utilicen las firmas tienen que buscar crear la lealtad de los clientes además de poner barreras económicas para evitar su fuga (Peppers y Rogers, 2011; Buttle, 2008; DeBonis y otros, 2003; Thompson, 2004).

Como se plantea en Peppers y Rogers (2011) un cliente retenido no es necesariamente un cliente leal a la firma. Las firmas deben invertir en construir relaciones individuales más fuertes para potenciar los niveles de lealtad de sus clientes. Para esto es necesario que las firmas pongan énfasis en la calidad de sus productos y servicios y demostrar un interés real en construir una relación de largo plazo con cada uno de sus clientes. La lealtad de los clientes está asociada a la calidad de la relación que mantiene la firma con el cliente y al nivel de satisfacción que siente cada cliente a lo largo del tiempo. La lealtad de un cliente entonces es un intangible que resulta del nivel de identificación del cliente con la firma y de los niveles de satisfacción que el cliente perciba como resultado de una relación de largo plazo. Por esta razón, la lealtad de los clientes representa un *switching costs* que las firmas rivales difícilmente pueden compensar al momento de intentar atraer un cliente. Las firmas con altos niveles de satisfacción en sus clientes, tienen clientes más leales, mayores tasas de retención y en consecuencia obtienen mayores utilidades a lo largo del tiempo.

Existen varios trabajos que concluyen que la presencia de *switching costs* en un mercado lo vuelve menos competitivo, dando mayor poder de mercado a las firmas y permitiéndoles decidir precios más altos (Klemperer, 1987b,a; Chen, 1997; Fudenberg y Tirole, 2000). En otros trabajos se concluye que las firmas se benefician incluso si ellas mismas crean y determinan endógenamente los *switching costs*, por ejemplo creando programas de beneficios (Kim y otros, 2001; Caminal y Matutes, 1990). Esto quiere decir que cuando existen *switching costs* las firmas obtienen mayores utilidades a través de precios más altos. Los clientes están dispuestos a pagar estos mayores precios pues el costo de cambiarse de firma puede ser muy alto. Estos trabajos abordan el problema de competencia a través de enfoques microeconómicos utilizando modelos de Teoría de Juegos.

Cuando una firma construye los *switching costs* a partir de servicios de calidad y con altos niveles de satisfacción, los consumidores la prefieren de forma natural. Sin embargo existen *switching costs* que en algunos mercados no son deseados y en algunos casos los gobiernos incluso los regulan para mejorar el nivel de competitividad del mercado. Un ejemplo son los mercados de telefonía móvil y la portabilidad numérica. La portabilidad númerica permite que los consumidores sean los dueños de sus propios números de teléfonos. Cuando la portabilidad no existe los dueños de los números son las compañías, por lo que si un consumidor evalúa cambiar de proveedor debe estar dispuesto a perder su número telefónico. Esto representa un importante *switching cost* que distorsiona el mercado y le entrega poder adicional a los proveedores de telefonía móvil. Las compañías pueden cobrar precios mayores ya que saben que los consumidores prefieren pagar un precio más alto que perder sus números telefónicos. A través de regulaciones que exijan la portabilidad numérica, un gobierno puede eliminar este *switching cost* e introducir mayores niveles de competencia al mercado. En los trabajos Shi y otros (2006), Reiko y John (2010) y Lee y otros (2006) se analizan los efectos de introducir la portabilidad numérica en mercados de telefonía móvil. En los dos primeros trabajos los autores abordan el problema planteando modelos de Teoría de Juegos.

1.3. Predicción de Fuga de Clientes

Hasta el momento se ha discutido la importancia de las estrategias de retención de clientes y el objetivo de construir *switching costs* que hagan más difícil su decisión de cambio. Sin embargo, no se ha planteado la forma en que las firmas deciden cómo distribuir sus esfuerzos de retención a lo largo de su cartera de clientes en un contexto de recursos limitados. Para esto las firmas utilizan estimaciones de la probabilidad de fuga de sus clientes. La fuga se define como el proceso de abandono de un cliente hacia la firma. Este abandono puede estar asociado a la inactividad de parte de un cliente respecto de la compra de los productos y/o servicios ofrecidos por la firma o simplemente a la elección del cliente por utilizar los productos y/o servicios de las firmas rivales.

Para poder contar con estimaciones de las probabilidades de fuga de sus carteras de clientes, las firmas utilizan metodologías que tratan de identificar patrones comunes entre los clientes que han abandonado alguna vez la firma y los que aún la prefieren. Para identificar estos patrones las firmas utilizan principalmente técnicas de Data Mining, específicamente técnicas de clasificación binaria, que les permite construir un clasificador que pueden aplicar sobre la cartera actual de clientes para predecir el perfil de cada uno de sus clientes y determinar su probabilidad de fuga de la firma. Dentro de las técnicas de clasificación binaria utilizadas para estos fines podemos encontrar aplicaciones de Regresión Logística, Support Vector Machines, Redes Neuronales, Árboles de Decisión, Cadenas de Markov y Redes Bayesianas, entre otras (Haddena y otros, 2007; Neslin y otros, 2006; Van˜den Poel y Larivière, 2004; Hsieh, 2004; Chiang y otros, 2003; Ngai y otros, 2009; Miranda y otros, 2005; Hung y otros, 2006; Zhao y otros, 2005; Coussement y den Poel, 2008).

En conjunto con estimaciones del valor presente de las utilidades que generará cada cliente para la firma, es posible concentrar recursos limitados en esfuerzos de retención sobre los mejores clientes con alta probabilidad de fuga. A estos clientes la firma aplicará acciones de marketing con el objetivo de aumentar sus *switching costs* haciendo más difícil su decisión de abandonar la firma. Existen diversos trabajos que plantean metodologías para valorar la lealtad de un cliente o estimar el valor de la relación con cada cliente. La mayoría abordan diferentes métodos para estimar el *Customer Lifetime value*, que representa el valor presente de los flujos de caja futuros que se espera que existan en la relación con un cliente (Bolton, 1998; Rosset y otros, 2003; Dwyer, 1997; Panda, 2003; Lewis, 2005; Ryals y Knox, 2005; Glady y otros, 2009; Gupta y otros, 2006; Glady y otros, 2006; Haenlein y otros, 2006). Otros trabajos proponen diferentes metodologías para tomar algunas decisiones propias de una estrategia de marketing y para la asignación de recursos (Gupta y Steenburgh, 2005; Prinzie y den Poel, 2006; Lemon y otros, 2002; Jonker y otros, 2004; Bitran y Mondschein, 1996; Mizuno y otros, 2008; Haenlein y otros, 2006).

1.4. Justificación de la Tesis

Una tarea fundamental dentro de una estrategia de retención, es la predicción de fuga de clientes. El problema de detección de fuga de clientes se ha abordado en diversos trabajos, principalmente utilizando técnicas de Data Mining. Haddena y otros (2007) revisan las principales técnicas de Data Mining aplicadas para predecir fuga de clientes que han tenido buenos resultados. Sin embargo, los autores discuten la necesidad de concentrar esfuerzos de investigación en crear sistemas de predicción de fuga de clientes más poderosos,

exactos y robustos, que permitan entender mejor el comportamiento de los clientes y las causas que provocan las fugas. Este tipo de modelos considera un único tomador de decisiones, la firma que intenta retener a sus clientes, sin embargo el proceso de fuga es más complejo y considera la interacción de la firma, su cartera de clientes y su competencia. En este contexto, resulta interesante seguir desarrollando modelos de predicción de fuga de clientes para utilizar sus resultados como parte de una estrategia de retención de clientes más completo.

Boulding y otros (2005) plantean proposiciones frente a lo que es conocido y aceptado en la literatura del *Customer Relationship Management* (CRM) y los desafíos y obstáculos que las firmas enfrentan en la implementación de una estrategia de CRM. Dentro de sus proposiciones, los autores plantean que para una implementación exitosa de una estrategia de CRM se requiere que las firmas incorporen su conocimiento acerca de la competencia y las reacciones competitivas de las firmas rivales en sus procesos y decisiones asociadas al CRM. Plantean con sorpresa que la literatura asociada al CRM ha omitido largamente los tópicos asociados a la competencia y que las firmas al ignorar este efecto ponen en serio peligro su capacidad de innovación y de mantenerse en el tiempo de manera sustentable. Boulding y otros (2005) creen que la investigación futura en este campo debe concentrarse en cómo integrar la competencia en los procesos que las firmas ejecutan en sus estrategias de CRM y en estudiar bajo qué condiciones el CRM da una ventaja sostenible frente a las reacciones de la competencia. En estas futuras investigaciones se debe reconocer que las firmas no escogen las actividades de CRM que realizan en el abstracto, por el contrario, las escogen sobre base de la respuesta del mercado a estas actividades junto con otros factores, tales como las habilidades y capacidades particulares de la firma.

En este contexto es necesario desarrollar un modelo de retención de clientes que incorpore el carácter estratégico de las decisiones que toma una firma, sus clientes y sus competidores, pero que también aproveche el potencial de los datos que posee la firma interesada en utilizar el modelo. En la figura 1.3 es posible observar la interacción que existe entre los distintos agentes involucrados en el problema de retención de clientes. La decisión que toma un cliente de cambiarse o no de una firma a otra, no depende únicamente de las decisiones que tome una de las firmas sino también de las decisiones que tome su competencia y del entorno asociado al mercado. Cada cliente observa en el mercado los precios que escogen la firma y su competencia. Cada cliente también está influenciado por los esfuerzos de marketing que la firma y su competencia realizan. En base a estas decisiones de la firma y su competencia, cada cliente decide mantenerse como cliente de la firma o fugarse haciendose cliente de la firma competidora. Las decisiones de todos los agentes determinan las utilidades que obtienen la firma y sus competidores y la satisfacción de los clientes.

En este trabajo se desea abordar el problema de predicción de fuga de clientes a través de un enfoque que incorpore la realidad competitiva que enfrentan las firmas. Las técnicas de clasificación binaria del Data Mining permiten abordar el problema de predecir las probabilidades de fuga de los clientes de una firma. Por otro lado, los modelos de Teoría de Juegos permiten estudiar el comportamiento estratégico y las decisiones de las firmas pertenecientes a un mercado y sus clientes en un ambiente competitivo.

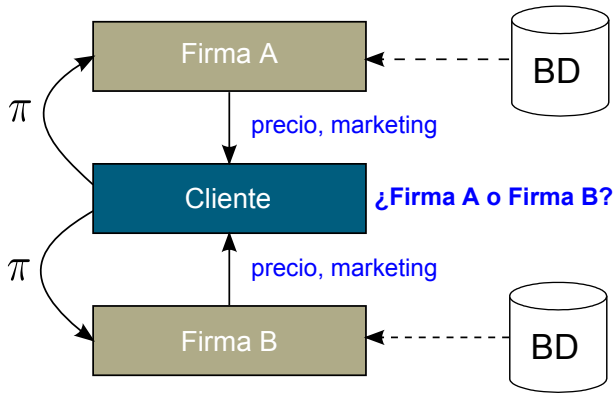

Figura 1.3: Esquema de interacción entre agentes involucrados.

1.5. Objetivos

1.5.1. Objetivo General

El presente trabajo tiene por objetivo abordar el problema de predecir las probabilidades de fuga de los clientes de una firma, incorporando los aspectos de competencia del mercado en que participan. Para esto se busca plantear un esquema cuantitativo que permita aprovechar la sinergia de utilizar el Data Mining, como herramienta estadística, y la Teoría de Juegos, para modelar el problema microeconómico.

1.5.2. Objetivos Específicos

- Modelar el proceso de fuga de clientes como un juego entre una firma, sus competidores y sus clientes, incorporando el carácter estratégico de las decisiones de estos agentes. En particular se abordará el caso de mercados con dos firmas. Se busca estudiar las decisiones de precios y esfuerzos en marketing que realizan las firmas en un ambiente de competencia y analizar cómo estas afectan la decisión de los consumidores con respecto a la firma que escogen.
- Utilizar técnicas tradicionales de Data Mining para abordar el problema de predicción de fuga de clientes de una firma. En particular, se escoge utilizar de modelos de Regresión Logística para construir un clasificador binario que sirva como *benchmark* para las metodologías propuestas.
- Plantear un modelo cuantitativo que permita abordar el problema de predicción de fuga de clientes de una firma, incorporando de forma endógena el modelo de competencia que enfrentan los agentes. Para esto se busca la construcción de un clasificador binario a partir de un modelo de Regresión Logística modificado para incorporar elementos de competencia del mercado en que participa la firma.
- Validar el modelo desarrollado a través de su aplicación en un caso de complejidad real. Concluir sobre los beneficios de utilizar un modelo de predicción de fuga de clientes que incorpore un enfoque de competencia y proponer una linea de investigación futura en este ámbito.

1.6. Motivación: Data Mining y Teoría de Juegos

Las distintas técnicas de Data Mining utilizadas para abordar el problema de detección de fuga de clientes tienen en común el supuesto de un sólo tomador de decisiones e ignoran el comportamiento estratégico del resto de los agentes involucrados. Consideremos el caso de una firma que baja el precio de sus productos esperando aumentos en su demanda. Si la firma no observa aumentos en la demanda podría concluir que el precio no es una variable relevante en la decisión de compra de sus clientes. Sin embargo, esta situación se podría deber a que sus competidores han disminuido aún más sus precios y la decisión de muchos de sus potenciales clientes ha sido preferir los precios más bajos. Al considerar las acciones de todos los agentes se llegaría a la conclusión que el precio es una variable relevante en la decisión de los clientes. Si bien las técnicas tradicionales del Data Mining entregan información muy valiosa con respecto a los patrones de comportamiento de los agentes, en determinados problemas no es suficiente para modelar acabadamente la situación.

Por otro lado, tenemos que la Teoría de Juegos ha sido utilizada con mucho éxito para caracterizar el comportamiento de agentes económicos que interactúan motivados por la maximización de sus utilidades, conscientes cada uno que los otros jugadores actúan también de forma racional. Este enfoque permite analizar el comportamiento estratégico de los jugadores. Sin embargo, la Teoría de Juegos no aprovecha todo el potencial de información que se encuentra en las grandes cantidades de datos que hoy las empresas almacenan y que el mercado tiene disponible.

La motivación de este trabajo radica en poder complementar técnicas de Data Mining, en particular Regresión Logística, y modelos de Teoría de Juegos para elaborar un modelo de predicción de fuga de clientes más robusto que los tradicionalmente utilizados por las firmas en la práctica.

1.7. Estructura del trabajo

En el capítulo 2 se introducen los principales conceptos de Teoría de Juegos que se utilizan a lo largo de este trabajo. En el capítulo 3 se hace una revisión de la literatura sobre retención de clientes bajo un enfoque de competencia. Los trabajos revisados plantean modelos de Teoría de Juegos para introducir los aspectos de competencia entre las firmas de un mercado.

Para caracterizar las decisiones de los clientes, en el capítulo 4 se define el problema de Clasificación Binaria, se presentan los fundamentos de los Modelos de Elección Discreta bajo un enfoque de utilidad aleatoria y en particular el problema de elección binaria a través de Regresión Logística, metodología que es intensamente utilizada en los modelos propuestos en este trabajo.

En el capítulo 5 se propone una metodología para predecir las probabilidades de fuga de clientes a través de un esquema de Regresión Logística, el cual es modificado para incorporar elementos de competencia de acuerdo al mercado en que participan las firmas. Se propone el esquema teórico del modelo y una metodología para encontrar la solución numérica del problema. En la práctica, se plantean tres modelos bajo este esquema de acuerdo a las alternativas de decisiones que endógenamente pueden tomar las firmas.

En el capítulo 6 se aplican las tres alternativas de modelos planteadas en el capítulo anterior a un caso aplicado. Estas alternativas son comparadas con los resultados de un modelo tradicional de Regresión Logística y se concluye sobre las bondades del esquema propuesto.

Finalmente en el capítulo 7 se plantean las conclusiones generales y específicas de esta tesis, además de plantear nuevos desafíos asociados a este trabajo y potenciales líneas de investigación a futuro.

Parte II

Marco Teórico

Capítulo 2

Conceptos de Teoría de Juegos

2.1. Introducción

La Teoría de Juegos es un conjunto de herramientas analíticas que permiten estudiar y entender los problemas de decisión de multiples tomadores de decisiones o jugadores, caracterizando sus interacciones y comportamiento estratégico. El supuesto básico de la Teoría de Juegos es que los jugadores persiguen objetivos exógenos bien definidos (los jugadores son racionales) y que toman en cuenta su conocimiento o expectativas respecto del comportamiento de los otros jugadores (los jugadores razonan de forma estratégica porque saben que los otros jugadores también son racionales).

Un *juego* es la descripción de la interacción estratégica que llevan a cabo los jugadores, especificando las decisiones que los jugadores pueden tomar y sus preferencias. En un juego la entidad principal es el jugador. El *jugador* puede ser interpretado como el individuo o grupo de individuos que toman decisiones en el ambiente definido por el juego.

El objetivo de este capítulo es presentar los fundamentos básicos sobre Teoría de Juegos que son utilizados en este trabajo. La sección 2.2 introduce el concepto de racionalidad en la toma de decisiones, supuesto fundamental en Teoría de Juegos. En la sección 2.3 se definen las dos formas de representación de un juego: forma normal y forma extensiva; y los tipos de estrategias que pueden seguir los jugadores: puras y mixtas. La sección 2.4 introduce el concepto de equilibrio definiendo en particular el equilibrio bajo estrategias dominantes y el equilibrio de Nash. La sección 2.5 presenta los Juegos Dinámicos, el concepto subjuego y la perfección en los subjuegos. Se deja fuera de este capitulo la presentación de los juegos con información imperfecta, ya que escapan a las definiciones utilizadas en este trabajo que siempre suponen información perfecta.

Para la construcción de este capítulo se han tomado como fuentes bibliográfica a Fudenberg y Tirole (1991), Osborne y Rubinstein (1994), Gibbons (1992) y Fischer (2006).

2.2. Comportamiento Racional y la Función de Utilidad

El supuesto de racionalidad permite establecer las preferencias de los jugadores y caracterizar su proceso de toma de decisiones. Un tomador de decisiones racional es consiente de sus alternativas de decisión, se forma expectativas sobre cualquier elemento de incertidumbre, tiene preferencias claras y escoge sus acciones de forma deliberada luego de algún proceso de optimización.

Cuando no existe incertidumbre, un modelo de elección racional está compuesto por los siguientes elementos.

- Un conjunto A de *acciones* sobre las que el tomador de decisiones debe escoger.
- Un conjunto C de posibles *consecuencias* de estas acciones.
- Una *función de consecuencias* $g : A \to C$, que asocie a cada una de las acciones posibles una consecuencia.
- Una *relación de preferencias* $\succsim$ en el conjunto de consecuencias C. Ésta debe ser una relación binaria completa, transitiva y reflexiva.

Generalmente las preferencias de los tomadores de decisiones son especificadas a través de una función de utilidad $U : C \to \mathbb{R}$, la que transforma la relación de preferencias $\succsim$ equivalentemente en una relación en el conjunto de los números reales: $x \succsim y$ si y sólo si $U(x) \geq U(y)$.

Si el conjunto de acciones factibles que tiene un tomador de decisión está dado por $\mathcal{A} \subseteq A$, entonces un tomador de decisiones racional escogerá una acción $a^* \in \mathcal{A}$ factible y óptima en el sentido de que $g(a^*) \succsim g(a)$ para todo $a \in \mathcal{A}$. Alternativamente se puede definir que el tomador de decisiones racional escogerá la acción a^* que resuelve el problema de optimización $\max_{a \in \mathcal{A}} U(g(a))$.

Para ambientes bajo incertidumbre, se encuentra ampliamente aceptada en la literatura la teoría de decisiones racionales planteada por Von˜Neumann y Morgenstern (1944). Si la función de consecuencias $g(\cdot)$ es estocástica pero conocida por el tomador de decisiones entonces éste se comporta de forma racional si maximiza el valor esperado de la función de utilidad de Von Neumann-Morgenstern. La función de consecuencias es estocástica si para cada acción a, su consecuencia $g(a)$ es una distribución de probabilidad en el conjunto de consecuencias C (para cada acción se asocia una probabilidad de ocurrencia de cada una de las consecuencias). El tomador de decisiones tiene en mente un espacio de estados de la naturaleza Ω; una medida de probabilidad sobre Ω; una función $g : S \times \Omega \to C$, que asocia a cada par acción/estado de la naturaleza (s, ω) una consecuencia; y una función de utilidad $U : C \to \mathbb{R}$ que permite al tomador de decisiones valorar cada una de las consecuencias. El tomador de decisiones racional escogerá la acción a^* que maximice el valor esperado de $U(g(a, \omega))$ con respecto a la medida de probabilidad sobre Ω.

En adelante, se denotará la función de utilidad de Von Neumann-Morgenstern como una función $u(a, \omega) = U(g(a, \omega))$, que asocia directamente al par acción/estado de la naturaleza (a, ω) una valoración de la consecuencia para el tomador de decisiones. De esta forma, si el conjunto de estrategias factibles está dado por

$\mathcal{A} \subseteq A$ y si la medida de probabilidad sobre Ω está dada por $P : \Omega \to [0,1]$ entonces el tomador de decisión racional escogerá la acción a^* que resuelve el siguiente problema de optimización.

$$\text{máx}_{a \in \mathcal{A}} \sum_{\omega} u(a, \omega) P(\omega) \tag{2.1}$$

2.3. Representación de un Juego

2.3.1. Juego en su Forma Estratégica o Normal

Un juego en su forma *estratégica* o *normal*, es un juego en el cual cada jugador decide su estrategia de una vez, por todo el juego y de forma simultánea. Este juego se modela a través de la definición de los siguientes elementos.

- Un conjunto de I jugadores racionales.
- Para cada jugador $i \in I$, un conjunto no vació S_i de estrategias.
- Una función de pagos u_i que entrega a cada jugador $i \in I$ la utilidad $u_i(s)$ de Von Neumann-Morgenstern para cada perfil de estrategias $s = (s_1, \ldots, s_i, \ldots, s_I)$. El perfil de estrategias es el vector con las decisiones de todos los jugadores. Se define el conjunto de perfiles de estrategias como el producto cruz de los conjuntos de estrategias de cada jugador $S = \times_i S_i$. De esta forma se tiene que $s \in S$.

Las utilidades de cada jugador se definen sobre el perfil de estrategias s y no sobre su estrategia s_i particular. Esto porque los jugadores tienen preferencias sobre las consecuencias del juego, las que se explican no sólo por sus decisiones sino también por las decisiones del resto de los jugadores.

Esta definición de juego supone que todos los jugadores conocen la estructura del juego, que saben que sus oponentes también la conocen, que saben que sus oponentes saben que ellos la conocen, y así al infinito.

En adelante se abusará de la notación para referirse a los jugadores diferentes del jugador i (al resto de los jugadores) bajo el índice $-i$. De esta forma, así como s_i se define como la estrategia que decide el jugador i; se define $s_{-i} = (s_1, \ldots, s_{i-1}, s_{i+1}, \ldots, s_I)$ como el vector de estrategias de todos los jugadores exceptuando al jugador i. Así, el perfil de estrategias s se puede escribir como $s = (s_i, s_{-i})$.

2.3.2. Juego en su Forma Extensiva

Un juego en su forma extensiva es una descripción detallada de la estructura secuencial de los problemas de decisión que enfrentan los jugadores en una situación estratégica. Un juego en su forma *extensiva* está compuesto por los siguientes elementos.

- Un conjunto de I jugadores racionales.
- Un *árbol* que representa el juego y que está compuesto de nodos y ramas. Cada uno de los nodos está asociado a alguno de los jugadores. Cada una de las ramas representa las acciones que dispone el jugador en cada uno de los nodos.
- La *información* que se encuentra a disposición de un jugador en cada nodo en que le corresponde una decisión. La información se describe mediante conjuntos de información, que son conjuntos de nodos que el jugador no puede distinguir entre sí.
- Las *estrategias* $s_i \in S_i$ de los jugadores. Las estrategias son un plan de acción completo que le dice al jugador que acción elegir cuando llega a uno de los conjuntos de información. Las estrategias son funciones desde los conjuntos de información del jugador hacia las acciones que tiene en cada conjunto de información.
- Los pagos u_i que recibe cada jugador $i \in I$ en cada una de las hojas o nodos terminales del árbol que representa el juego.

Cada secuencia de nodos y ramas lleva a una hoja del árbol. Por lo tanto, la utilidad que recibe cada jugador en cada hoja del árbol corresponde a la utilidad que entrega la historia o secuencia de decisiones de todos los jugadores. De esta forma, u_i se puede representar como una función $u_i : S \to \mathbb{R}$ que para cada combinación de decisiones de los jugadores $s \in S$ asigna el pago correspondiente para cada jugador.

Notemos que en la forma extensiva las acciones no son lo mismo que las estrategias. Las acciones son las alternativas de decisión que tienen los jugadores en cada conjunto de información. Las estrategias son un libro de instrucciones que le indica al jugador que acción tomar en cada uno de los conjuntos de información del juego.

La forma extensiva del juego permite dar una representación gráfica de los juegos, facilitando su interpretación. Sin embargo, requiere de formalizar una estructura que en ocasiones puede ser compleja. La forma extensiva de los juegos se utiliza para representar Juegos Dinámicos (sección 2.5).

2.3.3. Juegos con Estrategias Mixtas

Supongamos que un jugador i cuenta con un conjunto S_i de potenciales estrategias a elegir.

Las dos formas de representar un juego definidas en las secciones anteriores suponen que los jugadores utilizan estrategias puras. Una *estrategia pura* corresponde a una estrategia $s_i \in S_i$ que decide el jugador como respuesta competitiva a las elecciones que realizarán los otros jugadores. Se le llama estrategia pura porque el jugador debe escoger con probabilidad uno, sólo uno de los elementos de S_i (en el juego o en alguna de las etapas del juego).

Se define una *estrategia mixta* σ_i como una distribución de probabilidad sobre las estrategias puras que decide el jugador i. De esta forma, $\sigma_i(s_i)$ corresponde a la probabilidad que el jugador i le asigna a la estrategia s_i. Las distribuciones de probabilidad de las estrategias mixtas de cada jugador son independientes de las de

sus oponentes. Cuando un jugador escoge una estrategia mixta no elige una acción a seguir en particular del conjunto S_i, sino decide una probabilidad de utilizar cada una de las estrategias puras del conjunto S_i. Si el juego se jugara infinitas veces entonces con una frecuencia $\sigma_i(s_i)$ el jugador escogerá la estrategia pura s_i.

El espacio de estrategias mixtas del jugador i está dado por Σ_i y el conjunto de los perfiles de estrategias mixtas se define como $\Sigma = \times_i \Sigma_i$, cuyos elementos son los perfiles de estrategias mixtas $\sigma = (\sigma_i, \sigma_{-i})$.

La utilidad de Von Neumann-Morgenstern del jugador i en estrategias mixtas se define a través de la función $u_i(\sigma)$. Esta utilidad queda determinada a partir del perfil de estrategias mixtas σ que decidan simultáneamente todos lo jugadores y corresponde al valor esperado de las correspondientes utilidades $u_i(s)$ en estrategias puras.

$$u_i(\sigma) = u_i(\sigma_i, \sigma_{-i}) = \sum_{s \in S} \left(\prod_{j=1}^{I} \sigma_j(s_j) \right) u_i(s) \tag{2.2}$$

Las dos formas de representar los juegos se pueden definir equivalentemente para estrategias mixtas reemplazando s_i, S_i, s, S, $u(s)$ por σ_i, Σ_i, σ, Σ y $u(\sigma)$ respectivamente.

2.4. Solución de un Juego y el Concepto de Equilibrio

2.4.1. Definición de Equilibrio

Las definiciones hasta ahora presentadas permiten modelar a través de un juego las situaciones de competencia que enfrentan los tomadores de decisiones que participan de ellas. Sin embargo, hasta ahora no se ha caracterizado el concepto de solución de un juego. Para resolver un juego se requiere determinar cuáles son las decisiones que escogerán cada uno de los jugadores. De acuerdo al supuesto de racionalidad, la solución del juego tiene que ver con la maximización conjunta que realizan los jugadores de sus propias utilidades.

El requisito fundamental para que una estrategia sea la solución de un juego es que esta estrategia represente una situación de equilibrio. En Teoría de Juegos, un *equilibrio* corresponde a una situación tal que ninguno de los jugadores quiere cambiar su estrategia dada la combinación de estrategias del resto de los jugadores. Un jugador tiene incentivos de cambiar su estrategia si así puede aumentar sus utilidades. Cuando un jugador cambia su estrategia podría disminuir las utilidades del resto de los jugadores. Si esto sucede, el resto de los jugadores acomodará sus decisiones buscando aumentar sus utilidades en respuesta a la acción del primer jugador. Esta situación se repetirá hasta que todos los jugadores no tengan una mejor alternativa y no quieran cambiar sus decisiones, alcanzándose la situación de equilibrio.

2.4.2. Equilibrio en Estrategias Dominantes

Una estrategia σ_i^* del jugador *i* es *mejor respuesta* a σ_{-i} (las estrategias de los demás jugadores) si se cumple la siguiente condición.

$$u_i(\sigma_i^*, \sigma_{-i}) \geq u_i(\sigma_i, \sigma_{-i}) \qquad \forall \sigma_i \in \Sigma_i \tag{2.3}$$

Esto quiere decir que si una estrategia de un jugador es mejor respuesta entonces es lo mejor que el jugador puede hacer ante una determinada elección de los demás jugadores.

Una estrategia σ_i^* del jugador *i* es *dominante* si es la mejor respuesta a todas las estrategias de los demás jugadores, y por lo tanto se satisface la siguiente desigualdad, cumpliéndose que esta desigualdad sea estricta para al menos una de las estrategias σ_i.

$$u_i(\sigma_i^*, \sigma_{-i}) \geq u_i(\sigma_i, \sigma_{-i}) \qquad \forall \sigma_i \in \Sigma_i, \forall \sigma_{-i} \in \Sigma_{-i} \tag{2.4}$$

De esta forma, una estrategia es dominante si es la mejor respuesta que tiene un jugador ante todas las estrategias de los demás jugadores. Cuando un jugador posee una estrategia dominante entonces siempre la utilizará, porque es lo mejor que puede hacer, independientemente de lo que hagan los demás jugadores.

Se dice que una combinación de estrategias $\sigma^* = (\sigma_i^*)_{i=1}^I$ es un *equilibrio en estrategias dominantes* si cada σ_i^* es dominante.

Este tipo de equilibrio se alcanza cuando todos los jugadores poseen una estrategia dominante, y por lo tanto siempre la escogen sin importar lo que decidan el resto de los jugadores. Cuando existe este tipo de equilibrio, este es único y ningún jugador tiene mejores alternativas desde el punto de vista individual. Esta es una situación que es difícil de encontrar en los juegos, por lo que son pocos los juegos que poseen este tipo de equilibrios.

2.4.3. Equilibrio en Nash

El equilibrio de Nash es el concepto de solución más utilizado en Teoría de Juegos. Este concepto captura el estado estacionario del juego en el cual cada jugador tiene las expectativas correctas acerca del comportamiento de los otros jugadores y actúa racionalmente en consecuencia.

Un *equilibrio de Nash* es una combinación de estrategias $\sigma^* = (\sigma_i^*)_{i=1}^I$ tal que se satisface la siguiente condición.

$$u_i(\sigma_i^*, \sigma_{-i}^*) \geq u_i(\sigma_i, \sigma_{-i}^*) \qquad \forall i \in I, \forall \sigma_i \in \Sigma_i \tag{2.5}$$

En un equilibrio de Nash la estrategia que utiliza cada jugador es la mejor respuesta que tiene ante las estrategias de los otros jugadores. Si σ^* es un equilibrio de Nash entonces la estrategia σ_i^* de cada jugador debe resolver el siguiente problema de optimización: $\max_{\sigma_i \in \Sigma_i} u_i(\sigma_i, \sigma_{-i}^*)$.

Supongamos que para un juego se tiene una solución candidata que indica las estrategias que seguirá cada jugador. Para que la solución candidata efectivamente sea una solución al juego en cuestión, es necesario que cada jugador este dispuesto a seguir la estrategia indicada. Para que cada jugador siga esta estrategia, ésta debe ser la mejor respuesta competitiva que tenga el jugador en contra las estrategias que se propone que utilizarán el resto de los jugadores. De lo contrario, si alguno de los jugadores tiene al menos una alternativa de decisión que le entregue una mayor utilidad, entonces se desviará hacia esta nueva estrategia. Por lo tanto, esta solución candidata debe representar necesariamente un equilibrio de Nash.

El equilibrio de Nash no entrega información respecto de la calidad de la estrategia del jugador frente a otras estrategias de los demás jugadores, sólo indica que es la mejor respuesta frente a las estrategias que el resto utiliza en el mismo equilibrio. En un equilibrio de Nash las estrategias de cada jugador son la mejor respuesta a las mejores respuestas del resto de los jugadores.

Si se tiene un equilibrio en estrategias dominantes, todos los jugadores siguen una estrategia que es dominante y que es mejor respuesta a todas las estrategias de los demás jugadores, en particular a las estrategias dominantes de los demás jugadores. Por lo tanto, un equilibrio en estrategias dominantes es también un equilibrio de Nash.

Si bien la existencia de equilibrios de Nash en estrategias puras no se puede asegurar, en estrategias mixtas es posible asegurar que siempre existen equilibrios de Nash. La demostración de estos resultados escapa a los objetivos de este trabajo pero tanto Fudenberg y Tirole (1991) como Osborne y Rubinstein (1994) los presentan de forma detallada.

2.5. Juegos Dinámicos con Información Completa

2.5.1. Juegos Dinámicos, Subjuegos y Equilibrio Perfecto en el Subjuego

Hasta el momento se han planteado juegos en los cuales los jugadores toman decisiones de forma simultánea. En Juegos Dinámicos, existe de forma explícita una secuencia que describe el momento en que cada jugador debe tomar sus decisiones. Para representar este tipo de juegos se utiliza la forma extensiva (sección 2.3.2), que permite esquematizar de forma natural la secuencia de decisiones de los jugadores en este tipo de juegos. Este tipo de juegos permiten modelar por ejemplo situaciones donde existen etapas temporales o cuando un mismo juego se repite un número determinado de veces.

Para entender como los jugadores toman decisiones se puede suponer que estos se comportarán de forma racional en cada una de las etapas del juego. Esto permite ir estudiando las decisiones de los jugadores en cada etapa y luego concluir sobre la estrategia que utilizarán en el juego completo. Para esto es necesario introducir los conceptos de subárbol, subjuego y equilibrio perfecto en el subjuego.

Se define un *subárbol* de un juego como un subconjunto de nodos y acciones de un juego que se origina en un conjunto de información que es un singleton. Un conjunto de información es un singleton cuando este está compuesto por un único nodo.

Al considerar un subárbol del juego, se puede definir un *subjuego* del juego original, que corresponde al juego restringido al subárbol. Este subjuego comienza en un único nodo de decisión y contiene todos los nodos que siguen a su nodo inicial. Si un subjuego contiene cualquier punto de un conjunto de información, entonces contiene todo el conjunto de información.

Un equilibrio de Nash es un *equilibrio perfecto en el subjuego* (EPS) si al considerar cada subárbol, la combinación de estrategias restringidas al subárbol es un equilibrio de Nash del juego restringido al subárbol.

En el concepto del EPS se utiliza la racionalidad secuencial de los jugadores, quienes juegan en forma óptima en cada nodo del juego. El EPS es un equilibrio de Nash al que se le exige que cada jugador optimice en cada nodo, llegue al nodo o no en el juego, dadas las estrategias que deciden el resto de los jugadores. Una estrategia óptima para un jugador debe maximizar sus pagos esperados, en cada conjunto de información en el que al jugador le toque decidir. Es decir, la estrategia óptima de cada jugador debe especificar una acción óptima en cada conjunto de información, aunque el jugador piense ex ante que dicho conjunto de información no será alcanzado en el juego.

2.5.2. Inducción Hacia Atrás y Perfección en el Subjuego

La inducción hacia atrás o *backward induction* es un método que permite analizar un juego dinámico desde el final, la última etapa, hacia el principio: desde los conjuntos de información en el final del árbol hasta los conjuntos de información del principio.

Para encontrar los EPS en juegos de información perfecta se puede utilizar el método de inducción hacia atrás. Se parte desde los penúltimos nodos (ya que los últimos corresponden a los pagos) y se elige en cada nodo la mejor acción. Como en cada nodo sólo hay un jugador entonces la mejor acción se determina a partir de la solución del problema de decisión de un único tomador de decisiones. Luego, se debe reemplazar el juego original por uno en que se eliminan los nodos terminales del juego y los nodos penúltimos se transforman en los nodos terminales de un juego simplificado, cuyos pagos corresponden al pago de la mejor estrategia a usar en cada nodo penúltimo. Esta iteración hacía atrás se realiza hasta terminar el juego. Este procedimiento lleva a una solución única, salvo que los pagos a los jugadores sean los mismos en nodos terminales distintos, que corresponde al EPS del juego.

Se puede demostrar que todo juego dinámico finito con información perfecta tiene al menos un EPS. Se dice que un juego dinámico es finito cuando el juego posee un número finito de etapas y un número finito de acciones factibles en cada etapa.

Capítulo 3

Retención de Clientes y Competencia

3.1. Introducción

Si bien la consideración de aspectos competitivos en la literatura existente relativa al CRM es considerada por algunos autores como insuficiente (Boulding y otros, 2005), y en particular a lo referido al desarrollo de políticas de retención de clientes (Haddena y otros, 2007), existe una serie de trabajos que utilizando Teoría de Juegos estudian los fenómenos de competencia que enfrentan firmas y consumidores. Estos entregan algunos resultados interesantes que permiten entender cómo las firmas configuran sus estrategias para atraer y retener a los consumidores, y las consecuencias sobre los niveles de competencia en el mercado.

Las firmas manejan diferentes herramientas para competir en el mercado. A través de precios y descuentos adecuados pueden atraer a los clientes de la competencia e intentar retener a sus propios clientes. Sin embargo, las decisiones que toman las firmas no son triviales ya que los consumidores sufren costos por cambiar de firma y tienen preferencias por determinadas firmas del mercado. La situación se puede volver más compleja al considerar que las propias firmas son las que construyen las lealtades que sus clientes tienen por ellas. Estos elementos introducen elementos que distorsionan la posición competitiva de las firmas en el mercado y que es necesario entender en detalle.

A continuación se revisarán algunos trabajos que entregan interesantes resultados relativos al tema de retención de clientes, estudiados desde un punto de vista competitivo. Varios de los modelos que se presentan suponen la existencia de costos para los clientes que deciden cambiarse de firma de un periodo a otro (*switching cost*), o suponen que los consumidores están caracterizados por preferencias intrínsecas hacia las firmas del mercado, las que representan medidas de lealtad hacia cada firma.

3.2. Competencia en Precios

3.2.1. Competencia en Precios de Hotelling

El modelo Hotelling (1929) es uno de los más simple de los modelos de competencia en precios. Según Fudenberg y Tirole (1991) el modelo de competencia de Hotelling consiste en una ciudad lineal de largo uno, donde existe un continuo de consumidores con demanda unitaria distribuidos uniformemente a lo largo de la ciudad. Existen dos firmas, cada una localizada en un extremo de la ciudad, las que ofrecen un mismo producto con costo marginal $c \geq 0$. Si la posición dentro de la ciudad se representa a través de la variable x, entonces la firma A está ubicada en $x = 0$ y la firma B en $x = 1$. Si un cliente se está ubicado en la posición x dentro de la ciudad, entonces se encuentra a x unidades de distancia de la firma A y a $(1-x)$ unidades de distancia de la firma B. Los consumidores incurren en un costo t por cada unidad de distancia que deben recorrer para llegar hasta alguna de las firmas. La demanda por cada firma corresponde a la proporción de consumidores que encuentra más conveniente comprar en cada una de las firmas. En la figura 3.1 se observa el esquema de competencia en el modelo de Hotelling (1929). El consumidor ubicado en la posición x, comprará en la firma A si $p_A + tx \leq p_B + t(1-x)$. Es decir, todos los consumidores ubicados hasta $x \leq (p_B - p_A + t)/2t$ comprarán en la firma A. El resto escogerá la firma B, por lo que las demandas de cada firma están dadas por las siguientes expresiones.

$$D_A(p_A, p_B) = \frac{p_B - p_A + t}{2t} \tag{3.1}$$

$$D_B(p_A, p_B) = 1 - D_A(p_A, p_B) \tag{3.2}$$

El equilibrio de Nash de este juego está dado por $p_A^* = p_B^* = c + t$. Es decir, las firmas en equilibrio deciden no diferenciarse en precios y por lo tanto la demanda de ambas firmas resulta ser la misma y está dada por $D_A(p_A, p_B) = D_B(p_A, p_B) = 1/2$.

Se puede interpretar la ciudad de hotelling como una escala de lealtad de los consumidores hacia cada firma. Un consumidor con mayor lealtad por la firma A (más cercano a la firma A) puede escoger comprar en A aunque la firma B le ofrezca un menor precio. Esto se debe al costo que enfrenta el consumidor al elegir la firma que no prefiere (un mayor costo de transporte en el modelo de Hotelling), el que representa un *switching cost* para el consumidor. Las firmas saben que tienen una fracción de consumidores muy leales a ellas y que por lo tanto tiene un porcentaje de demanda relativamente segura. También saben que aunque bajen mucho sus precios, hay otra fracción de consumidores que de todas formas elegirá a la firma rival. Por lo tanto, si cada firma debe cobrar un único precio para todos los consumidores, éstas deciden no intensificar la competencia cobrando el mismo precio y obteniendo rentas de sus mercados cautivos a través de precios por sobre el costo marginal c del producto. La presencia de *switching costs* hace que el mercado se vuelva menos competitivo. Cuando el *switching costs* tiende a cero, es decir $t \to 0$, entonces el mercado se vuelve completamente competitivo haciendo que las firmas cobren el costo marginal del producto: $p_A^* = p_B^* \to c$.

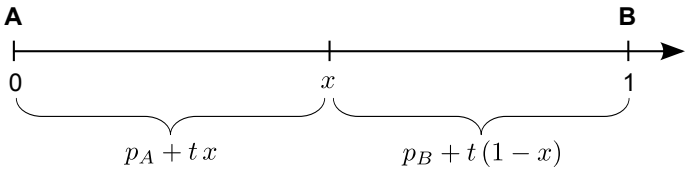

Figura 3.1: Modelo de Competencia de Hotelling

3.2.2. Competencia con Discriminación en Precios y Costos de Cambio

En muchos mercados los consumidores enfrentan costos de cambio cuando deciden cambiar el proveedor de sus producto y/o servicios, tales como costos de transacción o costos por aprender a utilizar el nuevo producto. Si bien los efectos de los costos de cambio han sido estudiados en la literatura previa a Chen (1997), típicamente se ha asumido que las firmas cargan siempre un mismo precio para todos sus clientes. El trabajo de Klemperer (1995) resume muy bien estos conceptos y plantea de forma muy clara los conceptos del *switching cost*.

Chen (1997) estudia la práctica de ofrecer a algunos clientes descuentos en el precio. A esta práctica se le conoce como discriminación en precios, ya que la firma no cobra un mismo precio a todos los consumidores a pesar de ofrecerles el mismo producto. En el trabajo de Chen (1997) se comparan las estrategias de discriminación en precios con un regimen donde las firmas sólo pueden ofrecer un único precio.

Chen (1997) plantea un juego entre dos firmas. En el Anexo B.1 se detallan algunos antecedentes de este trabajo. En el primer periodo, las firmas compiten en precio. En el segundo periodo, las firmas conocen en que firma compró cada consumidor en el primer periodo. Si un consumidor se cambia de la firma donde compro en el primer periodo, incurrirá en un costo de cambio s, conocido sólo por él al inicio del segundo periodo. La variable s de cada consumidor se considera un realización independiente de una variable aleatoria S que se distribuye uniformemente. Posteriormente, las firmas compiten nuevamente en precios en el segundo periodo. Bajo un regimen de discriminación en precios, cada firma puede proveer un pago monetario o descuento para que un consumidor se cambie desde la firma rival. Bajo un regimen de precios uniformes, este pago no está permitido, y por lo tanto las firmas cobran un único precio.

Chen (1997) demuestra la existencia de un único equilibrio de Nash en el segundo periodo bajo un régimen de discriminación en precios. Estos resultados son consistentes con lo que se obtiene en la literatura (Hotelling, 1929; Klemperer, 1995): si los costos de cambio se aproximan a cero el precio de equilibrio del mercado se aproxima al costo marginal, que corresponde al precio competitivo, y el descuento para los nuevos clientes se aproxima a cero.

Chen (1997) también concluye que bajo un régimen sin discriminación en precios, la firma que obtiene la mayor participación de mercado en el primer periodo, en el segundo periodo decide un precio mayor que bajo un régimen con discriminación. El excedente de los consumidores es menor en un regimen de precios uniformes que cuando las firmas pueden discriminar en precios. De la misma forma, las utilidades de equilibrio del segundo periodo para ambas firmas son menores en un régimen con discriminación en precios que en un régimen de precios uniformes.

Chen (1997) demuestra que cuando está permitida la discriminación en precios en el segundo periodo, entonces las firmas escogen para el primer periodo un mismo precio que se encuentra debajo del costo marginal del producto. Esto quiere decir que ambas firmas deciden intensificar la competencia en el primer periodo para asegurar una buena participación de mercado, ya que la única forma de aumentar participación de mercado en el segundo periodo es a través del pago de descuento a los clientes que en el primer periodo escogen la firma rival.

Por otro lado, cuando en el segundo periodo se cobran precios uniformes, Chen (1997) concluye que las firmas deciden precios sobre el costo marginal en el primer y segundo periodo. En el segundo periodo las firmas saben que tienen una fracción del mercado asegurada y otra fracción a la que no podrán seducir, tal como sucede en el modelo de competencia de Hotelling (1929). En efecto, dado que en el primer periodo las firmas deciden el mismo precio ambas consiguen la misma participación de mercado.

Se puede concluir que la suma descontada de las utilidades totales de cada firma son mayores bajo un régimen de precios uniformes que en un régimen de discriminación en precios. Cuando a las firmas les está permitido discriminar en precio una de las decisiones que pueden tomar es no discriminar, aplicando una estrategia de precios uniformes. Sin embargo, los resultados encontrados por Chen (1997) indican que las firmas, empujadas por la competencia, deciden discriminar en precios a pesar de que sin discriminación podrían estar mejor obteniendo mayores utilidades. Es decir, es posible observar en el comportamiento de las firmas el fenómeno del dilema del prisionero. En el Anexo A.1 se describe en detalle el juego *el dilema del prisionero*.

3.2.3. Competencia con Discriminación en Precios y Preferencias de los Consumidores

Fudenberg y Tirole (2000) proponen modelo de dos firmas con la misma estructura que el planteado por Chen (1997). En el Anexo B.2 se detallan algunos antecedentes de este trabajo. En el primer periodo las firmas compiten con un único precio para todos los consumidores y en el segundo periodo las firmas pueden discriminar en precios para atraer a los consumidores que escogieron a la firma rival. La diferencia entre estos dos trabajos está en la definición de preferencias. Chen (1997) plantea que los consumidores enfrentan costos de cambios y en Fudenberg y Tirole (2000) se definen preferencias relativas por cada firma asociadas a las utilidades de los consumidores.

Fudenberg y Tirole (2000) determinan que cuando está permitida la discriminación en precios, se obtiene un precio menor que el precio obtenido cuando la discriminación está prohibida. Esta situación se encuentra en línea con los resultados que se obtienen en Chen (1997). Los resultados de Fudenberg y Tirole (2000) difieren a los de Chen (1997) en el caso del primer periodo, donde los precios de equilibrio son mayores al obtenido cuando la discriminación en precios no está permitida. Estos resultados son contrarios a los obtenidos por Chen (1997) pues, bajo un régimen discriminación en precios, se obtiene un precio que es menor al caso de precios uniformes y que es incluso es menor que el costo marginal del producto.

En Chen (1997) la competencia en el primer periodo es mucho más fuerte que en Fudenberg y Tirole (2000). Esto es consecuencia de que en Chen (1997) el *switching cost* de un consumidor en el segundo periodo es

el mismo independiente de la firma que haya escogido en el primer periodo, por lo que las firmas luchan descarnadamente en el primer periodo por capturar la mayor parte del mercado. En cambio en Fudenberg y Tirole (2000), dada la definición de preferencias intrínsecas por cada firma, hay consumidores que a priori son fuertemente leales a una firma, por lo que su *switching cost* es muy alto para la firma rival. Las firmas no tienen incentivos a bajar mucho los precios pues tienen asegurada una fracción del mercado. En definitiva, las firmas competirán por la fracción de clientes con lealtades más débiles, por lo que los precios de equilibrio serán más altos comparados con los obtenidos por Chen (1997).

3.2.4. Pagar para Cambiar o Pagar para Retener

Si bien tanto en Chen (1997) como en Fudenberg y Tirole (2000) se concluye que es óptimo cobrar menores precios a los consumidores que son clientes de la competencia con el objetivo de atraerlos, hay evidencia que existen firmas que cobran menores precios a sus propios clientes para retenerlos. Shaffer y Zhang (2000) estudian bajo que condiciones para una firma es óptimo cobrar menos a los clientes de la competencia (estrategia *pagar para cambiar* de firma) y bajo que condiciones es óptimo cobrar menos a sus propios clientes (*pagar para retener*).

En el modelo planteado por Shaffer y Zhang (2000) existen dos firmas. En el Anexo B.3 se detallan algunos antecedentes de este trabajo. El mercado está divido en dos: el grupo de consumidores que comprarían en la firma A si los precios de ambas firmas fuesen los mismos y el grupo con el resto de los consumidores que comprarían en la firma B si los precios de ambas firmas fuesen los mismos. Las firmas deben decidir los precios que cobran cada grupo. Los costos de cambios son representados a través de una medida de lealtad, definida como la mínima diferencia de precio necesaria para inducir a un consumidor para cambiar de firma. Cada consumidor está caracterizado por su lealtad distribuida uniformemente.

Shaffer y Zhang (2000) concluyen que si sólo una de las firmas puede discriminar en precios, o si la demanda es simétrica, siempre será óptima la estrategia *pagar para cambiar*. Si ambas firmas pueden discriminar en precios y si la demanda es asimétrica, puede ser óptimo para ambas firmas seguir una estrategia *pagar para cambiar*, o puede ser óptimo para sólo una de las firmas seguir una estrategia *pagar para cambiar* (en este caso la otra firma debe seguir una estrategia *pagar para retener*). Nunca es óptimo para ambas firmas seguir una estrategia *pagar para retener*.

Con respecto a los precios, dependiendo de los parámetros, se pueden observar precios más bajos para todos los consumidores con respecto al caso sin discriminación en precios. Puede suceder que los precios para algún segmento de los consumidores baje y para otro suba. En este caso, al segmento que una firma en particular le cobra el mayor precio, la firma competidora le cobrará su menor precio y viceversa, tal como sucede en Chen (1997); Fudenberg y Tirole (2000). La otra alternativa que se puede dar es que la firma más grande baje sus dos precios mientras que la más pequeña suba uno de sus precios y baje el otro. Con respecto a las utilidades, si sólo una firma puede realizar discriminación en precios, está firma siempre aumenta sus utilidades a expensas de la otra firma. Si ambas firmas pueden realizar discriminación en precio y las demandas son simétricas, se observa como resultado el fenómeno del dilema del prisionero, como sucede en Chen (1997); Fudenberg y Tirole (2000). Pero si la demanda es asimétrica entonces dependiendo de los parámetros, ambas firmas pueden estar aún peor que como estarían sin algún tipo de discriminación, o sólo

una de las firmas podría estar mejor.

El aporte del trabajo de Shaffer y Zhang (2000) está en la generalización del concepto de discriminación en precios, permitiendo que las firmas puedan ofrecer descuentos en precio no sólo a los consumidores que escogen la firma rival, sino también hacer ofertas de retención a sus propios clientes. Shaffer y Zhang (2000) también plantean un mercado de lealtades asimétricas, reconociendo que no todas las firmas tienen la misma capacidad de generar lealtades de parte de los consumidores, o que una de las firmas domine las lealtades de los consumidores. Esto permite verificar las condiciones sobre los parámetros para obtener equilibrios donde una de las firmas domina el mercado.

3.2.5. Competencia con Promociones Uno a Uno

En los trabajos citados en las secciones anteriores, se proponen modelos de discriminación de precios, la que se realiza a nivel de segmentos de consumidores. En trabajos como Chen y otros (2001); Shaffer y Zhang (2002); Fruchter y Zhang (2004) se plantean modelos donde las firmas buscan precios particulares para cada consumidor de manera de maximizar sus utilidades, de acuerdo a las características que conoce de estos.

Shaffer y Zhang (2002) propone un juego de dos firmas donde cada consumidor tiene una demanda unitaria por el producto. En el Anexo B.4 se detallan algunos antecedentes de este trabajo. Los consumidores se diferencian en cuanto están dispuestos a pagar por el producto de su firma menos preferida. La lealtad de los consumidores se definen de forma asimétrica como en Shaffer y Zhang (2000). Se asume que las firmas conocen las lealtades de todos los consumidores. Las firmas participan de un juego de dos periodos, en donde en el primer periodo escogen precios uniformes para todos los consumidores y luego en el segundo se escogen las estrategias de promociones. En el segundo periodo, las firmas deben decidir qué segmentos de consumidores serán objeto de promociones y el valor de los descuentos que se les ofrecerán. Las firmas enfrentan un costo de segmentar cada consumidor independiente si este escoge la firma o no.

Según Shaffer y Zhang (2002), una estrategia de *promoción uno a uno* permite a las firmas la flexibilidad de generar ventas adicionales ofreciendo sólo los descuentos estrictamente necesarios. Si el mercado utiliza una estrategia de *promoción uno a uno* entonces la competencia se intensifica porque todos los consumidores, incluyendo a los consumidores de lealtades más fuertes, son potencialmente accesibles para todas las firmas. Cuando se asume simetría en las condiciones de las firmas, una estrategia de *promoción uno a uno* inevitablemente lleva al mercado a un dilema del prisionero, donde todas las firmas se encuentran en una situación peor. Por otro lado, cuando firmas son asimétricas, existirá un efecto en las participaciones de mercado. Shaffer y Zhang (2002) demuestran que las firmas que poseen la mayor fidelidad de parte de los consumidores son las que están mejor posicionadas para ganar una mayor participación de mercado. Una estrategia de *promoción uno a uno* entrega a estas firmas la flexibilidad que necesitan para sacar provecho de la lealtad de sus clientes. De este modo, una estrategia de *promoción uno a uno* representa una gran ayuda para las firmas más grandes, ya que les permite competir en los diferentes nichos de mercados que cada vez son más fragmentados.

Shaffer y Zhang (2002) también concluye que las firmas deben entender la relevancia estratégica que tiene la construcción de lealtades en los consumidores hacia sus marcas. Esto porque en mercados con niveles de competencia cada vez más agresivos, todos los consumidores están potencialmente expuestos a las ofertas

de las firmas rivales y la lealtad de estos consumidores representa la principal línea de defensa que tienen las firmas para retener a sus clientes.

3.3. Competencia en Esfuerzos de Marketing

Musalem y Joshi (2007) plantean un modelo donde los márgenes de ganancia que obtienen las firmas son parámetros conocidos y por lo tanto las firmas compiten realizando esfuerzos por atraer o retener a un único consumidor. En el Anexo B.5 se detallan algunos antecedentes de este trabajo. El consumidor tiene lealtades intrínsecas hacia las firmas, pero además se caracteriza por una capacidad de recepción de los esfuerzos que hace cada firma por retenerlo. Los esfuerzos en marketing que hacen las firmas pueden o no tener éxito, por lo que los costos que involucran estos esfuerzos se deben asumir incluso si la firma no consigue adquirir o retener al consumidor por el que compiten. Esto es diferente a cuando la competencia entre las firmas es en precios y las firmas ofrecen descuentos a los consumidores: los descuentos son pagados por la firma sólo cuando el consumidor la elige.

Cada firma decide el esfuerzo que realiza por adquirir al consumidor en el primer periodo. En el segundo periodo, la firma que captó al consumidor en el primer periodo compite por retenerlo y la otra firma por robarlo a su competidor. El consumidor está caracterizado por su lealtad a cada firma y escoge a la firma que le entrega la mayor utilidad. La utilidad que obtiene el consumidor se modela a través de una función de utilidad aleatoria.

$$U_{i1} \quad = \quad V_{i1} + \varepsilon_{i1} \quad = \quad \ln(l_i + \theta_i a_{i1}) + \varepsilon_{i1} \tag{3.3}$$

La expresión V_{i1} recoge la utilidad que le entrega al consumidor su grado de lealtad hacia la firma y el esfuerzo de marketing que realiza la firma por atraerlo, mientras que ε_{i1} captura los factores no observables. De esta forma, los autores determinan las probabilidades de elección del consumidor a través de un modelo de elección discreta.

Musalem y Joshi (2007) concluyen que las firmas deben concentrar sus esfuerzos de retención en los clientes con sensibilidades moderadas hacia sus esfuerzos de retención. Los clientes altamente sensibles no requieren esfuerzos altos para ser retenidos, mientras que los esfuerzos se pierden en los clientes menos sensibles. Los esfuerzos de atracción deben ser más agresivos para los clientes que son moderadamente rentables para la firma competidora. Los clientes más rentables del competidor resultan muy difíciles de atraer, porque el competidor invierte mucho para retenerlos, y los clientes menos rentables para el competidor no requieren mayores esfuerzos.

El hecho que las firmas puedan mejorar su relación con un cliente de un periodo a otro aumentan los incentivos de las firmas por atraer a este cliente. Esta mejora en la relación se puede dar aumentando el margen de ganancia de la firma, mejorando la sensibilidad del cliente a los esfuerzos de marketing de la firma y aumentando su lealtad intrínseca del cliente hacia la firma. Sin embargo, la competencia por un cliente con

estos potenciales de mejora a lo largo de los periodos, en ocasiones también puede conducir a la disminución de los beneficios para las empresas. Musalem y Joshi (2007) concluyen que si bien un cliente se puede volver más rentable en el futuro, esta situación no siempre es deseable para las firmas. Un incremento en los esfuerzos de atracción de una de las firmas puede ser compensado y potencialmente neutralizado por la firma rival, llevando a un equilibrio con menos utilidades para ambas firmas. En este sentido, para las firmas resulta fundamental conocer y entender el potencial de desarrollo de sus clientes. Una firma que no logra identificar y tener en cuenta su entorno competitivo puede evaluar incorrectamente los beneficios del desarrollo potencial de un cliente y, por lo tanto, tomar decisiones de gestión equivocadas.

3.4. Competencia en Precios y Esfuerzos en Marketing

Como resultado de la revisión bibliográfica presentada en este capítulo, el autor de esta tesis ha propuesto un modelo de competencia en precios y esfuerzos en marketing (López y otros, 2008). Este trabajo fue presentado en una conferencia internacional y a continuación se presenta en el contexto de modelos de competencia que caracterizan el fenómeno de retención de clientes.

Un factor no considerado en la literatura revisada es que el marketing permite construir lealtades, modificando las preferencias de los consumidores. Estas lealtades luego se traducen en *switching costs*, los que son aprovechados por las firmas al momento de tomar sus decisiones de precios. Por lo tanto, no basta considerar que las lealtades son exógenas, como en Chen (1997); Shaffer y Zhang (2002), ni modelos en que las firmas sólo compiten en marketing para construir lealtades, como en Musalem y Joshi (2007). En la práctica, las firmas manipulan dos herramientas estratégicas: precios y esfuerzos en marketing.

En López y otros (2008) se proponen dos modelos de competencia. En el primero, que es el que se presenta en esta sección, las firmas inicialmente establecen endógenamente las preferencias de los consumidores a través de inversiones en marketing. Luego de establecer estas preferencias, las firmas compiten en precios para maximizar sus utilidades. En el segundo modelo, las firmas compiten igual que en el modelo anterior pero luego existe un nuevo periodo donde las firmas vuelven a competir en precios para retener a sus clientes y para atraer los clientes de la firma rival. Para los fines de esta tesis, basta con presentar el primero de estos modelos.

3.4.1. Formulación

Se considera un mercado con dos firmas A y B, que ofrecen un único producto de costo marginal $c \geq 0$, donde los consumidores son heterogéneos sólo por su lealtad l a cada firma. Se supone que la lealtad l de cada consumidor es una variable aleatoria distribuida uniformemente entre $-l_B$ y l_A: $l \sim U(-l_B, l_A)$, con $l_A, l_B > 0$.

Se define la lealtad de cada consumidor por una firma como la mínima diferencia de precio, ponderada por un parámetro β, necesaria para inducirlo ha escoger la firma rival. El consumidor con lealtad $l \in [-l_B, l_A]$ escoge la firma A si y solo si $\beta(p_A - p_B) \leq l$, donde p_i es el precio de la firma i y β es el parámetro que

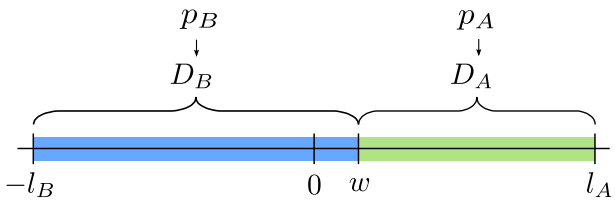

Figura 3.2: Demandas de cada firma

transforma las unidades de precio en unidades de lealtad. El consumidor escoge A si y sólo si $p_A \leq p_B + l/\beta$. Esta definición de lealtades es la propuesta en Shaffer y Zhang (2002) cuando se utiliza $\beta = 1$. La lealtad l de un consumidor se puede interpretar como lealtades hacia la firma A: si $l > 0$ el consumidor prefiere a priori la firma A y si $l < 0$ prefiere a priori la firma B. Si el consumidor tiene lealtad $l = 0$ entonces estará indiferente entre ambas firmas. Si $l_A = l_B$, como la distribución de consumidores es uniforme, entonces ambas firmas se encuentran en igualdad de condiciones, pues las preferencias son simétricas e igualmente repartidas entre las firmas. Si por ejemplo $l_A > l_B$ la distribución pierde su simetría y la firma A tendrá proporcionalmente un mayor número de consumidores que la prefiere, representando esto una ventaja competitiva.

El modelo consiste en un juego con un periodo de decisión inicial, o periodo cero, donde cada firma $i \in \{A,B\}$ decide su inversión en marketing x_i. Estas inversiones afectan la distribución de las lealtades de los consumidores, es decir, se asume que $l_i = l_i(x_A, x_B)$ con $i \in \{A,B\}$. Se supone que $l_i(x_i, x_{-i})$ es creciente en x_i y decreciente en x_{-i}. Si la firma i realiza una fuerte inversión en marketing entonces conseguirá hacer crecer l_i, haciendo que la distribución de lealtades se cargue hacia su lado y construyendo una ventaja competitiva. La firma $-i$ puede perjudicar el crecimiento de l_i aumentando su propia inversión x_{-i}, lo cual también favorecerá la construcción del parámetro l_{-i}. Esta definición asume que la distribución de lealtades está determinada por factores exógenos y por los esfuerzos en marketing que endógenamente deciden las firmas en el periodo cero. Posteriormente, en el primer periodo, cada firma $i \in \{A,B\}$ decide el precio p_i con el que competirá contra la firma rival observando las lealtades de cada consumidor, ya establecidas en el periodo anterior.

Como el consumidor de lealtad l escoge A si y sólo si $\beta(p_A - p_B) \leq l$, el consumidor que queda indiferente entre las firmas es el con lealtad $w = \beta(p_A - p_B)$. De esta forma, la demanda D_A por la firma A está dada por la fracción de mercado con lealtades mayores a w. La demanda D_B por la firma B está dada por la fracción de mercado con lealtades menores a w. La figura 3.2 resume estos resultados.

Las demandas por cada firma están dadas según las siguientes expresiones.

$$D_A = \frac{l_A - w}{l_A + l_B} = \frac{l_A - \beta(p_A - p_B)}{l_A + l_B} \tag{3.4}$$

$$D_B = \frac{w + l_B}{l_A + l_B} = \frac{\beta(p_A - p_B) + l_B}{l_A + l_B} \tag{3.5}$$

A partir de las demandas se pueden construir las funciones de utilidad de cada firma. Por cada unidad de demanda, cada firma recibe el precio del producto pero debe cancelar su costo marginal. Adicionalmente,

las firmas deben financiar la inversión en marketing que realizan en el periodo inicial. La función de utilidad resultante es la siguiente para cada firma.

$$\pi_i = (p_i - c)D_i - x_i \quad \text{con } i \in \{A, B\} \tag{3.6}$$

Desarrollando la expresión anterior a partir de las demandas de cada firma se llega a las siguientes funciones de utilidad.

$$\pi_A = (p_A - c)\frac{l_A - \beta(p_A - p_B)}{l_A + l_B} - x_A \tag{3.7}$$

$$\pi_B = (p_B - c)\frac{\beta(p_A - p_B) + l_B}{l_A + l_B} - x_B \tag{3.8}$$

3.4.2. Resolución

Para resolver este juego, se utiliza inducción hacía atrás según se plantea en la sección 2.5. Para el primer periodo se supone que los límites de la distribución de lealtades l_A y l_B son conocidas por las firmas. Estos límites son decididos por las firmas en el periodo inicial a través de su inversiones en marketing. Conocidos l_A y l_B, la firma A decide el precio p_A que maximiza su utilidad π_A y la firma B decide el precio p_B que maximiza su utilidad π_B. A partir de las condiciones de primer orden de ambos problemas es posible despejar los precios de equilibrio que deciden las firmas para el primer periodo, los que están dados por la siguiente expresión.

$$p_A^* = c + \frac{2}{3}\frac{l_A}{\beta} + \frac{1}{3}\frac{l_B}{\beta} \tag{3.9}$$

$$p_B^* = c + \frac{1}{3}\frac{l_A}{\beta} + \frac{2}{3}\frac{l_B}{\beta} \tag{3.10}$$

Para continuar con el procedimiento de inducción hacia atrás es necesario que las firmas maximicen sus utilidades π_i pero ahora con respecto a sus inversiones en marketing x_i. Como $l_i = l_i(x_A, x_B)$, para resolver el problema es necesario caracterizar esta dependencia funcional. Se supone que los límites de la distribución de lealtad tienen la siguiente expresión.

$$l_i(x_A, x_B) \quad = \quad \theta + \frac{x_i}{x_A + x_B} \cdot \mu - \frac{x_{-i}}{x_A + x_B} \cdot \phi \quad \forall i \in \{A, B\} \tag{3.11}$$

Esto quiere decir que cada límite de la distribución de lealtades tiene un valor base entregado por el parámetro θ. Luego, por cada punto porcentual que represente la inversión en marketing que realice la firma i respecto de la inversión total de la industria, el parámetro l_i crecerá en u. Sin embargo, l_i disminuirá en ϕ por cada

punto porcentual que represente la inversión en marketing que realice la firma $-i$ respecto a la inversión total de la industria.

Planteando las condiciones de primer orden es posible determinar que si se cumple que $\phi < \frac{12\theta+5\mu}{7}$, entonces en el periodo cero las inversiones en marketing de equilibrio de cada firma están dadas por la siguiente expresión.

$$x_A^* = x_B^* = \frac{1}{12}\frac{\mu+\phi}{\beta} \tag{3.12}$$

Es decir, las inversiones en marketing que realizan las firmas son iguales pues de lo contrario abren la puerta para que la firma rival establezca una ventaja competitiva y les quite participación de mercado. Esto se da porque el modelo hace que las firmas partan de condiciones simétricas dadas por el parámetro θ, dejando a las propias firmas la responsabilidad de construir las asimetrías a partir de su competencia primero en esfuerzos de marketing y luego en precios.

3.4.3. Resultados y Conclusiones

Es importante notar que los esfuerzos en marketing dependen de los parámetros μ y ϕ, que representan la eficiencia marginal en construir lealtades propias y destruir lealtades del rival. Adicionalmente, depende del parámetro $1/\beta$ que representa el parámetro que transforma las unidades de lealtad en unidades de monetarias o de utilidad. A partir de este resultado, es posible construir el resto de las variables que componen el equilibrio del juego.

$$\begin{aligned}
l_i^* &= \theta + \frac{1}{2}(\mu-\phi) \quad \forall i \in \{A,B\} && (3.13)\\
w^* &= 0 && (3.14)\\
D_i^* &= \frac{1}{2} \quad \forall i \in \{A,B\} && (3.15)\\
p_i^* &= c + \frac{\theta}{\beta} + \frac{\mu-\phi}{2\beta} \quad \forall i \in \{A,B\} && (3.16)\\
\pi_i^* &= \frac{\theta}{2\beta} + \frac{\mu-2\phi}{6\beta} \quad \forall i \in \{A,B\} && (3.17)
\end{aligned}$$

Se puede observar que la distribución de lealtades construidas a partir de la competencia es simétrica. Como los esfuerzos en marketing son positivos, las firmas invierten en marketing para mantener su posición competitiva y no entregar participación de mercado a la firma rival. Esto implica que el consumidor que se mantiene indiferente entre ambas firmas es el que se encuentra al centro de la distribución de clientes ($w^* = 0$) y por lo tanto el mercado se reparte en partes iguales para ambas firmas ($D_i^* = 1/2$).

Los precios y utilidades son estrictamente crecientes en θ y μ y estrictamente decrecientes en β y ϕ.

Si $\mu = \phi = 0$ las lealtades son exógenas y no hay competencia en esfuerzos de marketing. Si además $\beta = 1$, se recuperan los resultados de la literatura Chen (1997); Shaffer y Zhang (2000, 2002).

Cuando $\mu > \phi$, la competencia en esfuerzos de marketing provoca un alza en los precios respecto del caso de lealtades exógenas debido a los mayores *switching costs* que estos generan. En este caso las firmas se benefician de poder invertir en marketing.

Si $\mu > 2\phi$, la competencia en esfuerzos de marketing provoca un alza en las utilidades de las firmas con respecto al caso de lealtades exógenas, debido a los mayores precios que los consumidores están dispuestos a pagar, consecuencia de las lealtades construidas por las firmas a partir de los esfuerzos en marketing realizados en el periodo incial. Si por el contrario $\mu < 2\phi$, las firmas igualmente invierten en marketing pero obtienen menores utilidades con respecto al caso de lealtades exógenas. Esto se da porque las firmas son más eficientes destruyendo lealtades de los consumidores hacia la firma rival que construyendo lealtades hacia ellos mismos. Esto se asemeja a una competencia de destrucción de imagen de los rivales, en lugar de promover las bondades de la propia marca. En un mercado con estas caracteristicas, las firmas preferirían no contar con la facultad de invertir en lealtades a través de esfuerzos en marketing y estarían mejor en un mercado con lealtades exógenas. En este caso las firmas caen en un dilema de los prisioneros e igualmente compiten en marketing a pesar que saben que terminarán peor.

Cuando se cumple que $2\phi > 3\theta + \mu$, entonces las firmas son tan buenas destruyendo la lealtad de los consumidores hacia la firma rival que ambas firmas terminan con utilidades negativas y sufriendo un dilema de los prisioneros. Cuando $\phi > 2\theta + \mu$ en equilibrio existe tal nivel de competencia que las firmas están dispuestas a cobrar un precio bajo el costo marginal del producto.

El aporte del modelo propuesto en López y otros (2008), radica en considerar simultáneamente y de forma endógena las decisiones de precios y marketing que realizan las firmas, cuando las inversiones en marketing determinan el nivel de los *switching costs* que enfrentan los consumidores.

Capítulo 4

Clasificación Binaria, Modelos de Elección Discreta y de Regresión Logística

4.1. Introducción

El objetivo de este capítulo es presentar las metodológicas estadísticas tradicionales que son utilizadas a lo largo del desarrollo de este trabajo. Éstas se utilizan para caracterizar el problema de predicción de la probabilidad de fuga de los clientes.

En la sección 4.2 se define el concepto de Clasificación Binaria. El problema de Clasificación es utilizado en estadística y Data Mining cuando se debe predecir a qué grupo o clase pertenece determinado objeto. Cuando existen sólo dos alternativas de clasificación entonces es un problema de Clasificación Binaria. En la sección 4.3 se presenta el Modelo de Elección Discreta Logit Multinomial, el cuál es ampliamente utilizado para estimar las probabilidades de elección de las alternativas de decisión que enfrentan agentes económicos. En el capítulo 4.4 se presenta en detalle la metodología de Regresión Logística, que propone una forma funcional para estimar las probabilidades de elección cuando existen sólo dos alternativas. Las probabilidades propuestas por la regresión Logística se pueden utilizar para la construcción de un Clasificador Binario. Adicionalmente, en el Anexo D se plantean diferentes criterios estadísticos para evaluar los resultados de los modelos de estimación de probabilidades de elección y de Clasificación Binaria.

Para la construcción de este capítulo se han tomado como fuentes bibliográficas a Train (2003), Larose (2006), Hosmer y Lemeshow (2004), Vapnik (2000) y Venables y Ripley (2002).

4.2. Clasificación Binaria

Es de interés para este trabajo estudiar el problema de clasificación binaria. Dado un número predefinido de clases y un conjunto de objetos, en un problema de clasificación se desea asignar cada objeto (cuya clase es desconocida) como perteneciente a alguna de estas clases. Para esto es necesaria la construcción de una función de clasificación que asigne a cada objeto alguna de las clases. La construcción de la función de clasificación se realiza en base a un conjunto de observaciones de entrenamiento. Cada observación de entrenamiento está compuesta por un objeto y por la clase a la que pertenece. De esta forma, se busca una función que de acuerdo a las características de cada objeto asigne la clase a la que pertenece sin equivocarse. Cuando existen sólo dos clases predefinidas, se habla de un problema de clasificación binaria.

Supongamos que se cuenta con un conjunto de entrenamiento con I observaciones. Los objetos de cada observación se representan a través de un vector de características con n atributos, incluyendo una constante. Sea x_i el vector de atributos del objeto i, con $i \in \{1,...,I\}$ y $x_i \in \mathbb{R}^n$. El primer componente del vector x_i es igual 1 para todos los objetos, ya que corresponde al atributo asociado a la constante. En un problema de clasificación binaria cada observación puede pertenecer a alguna de las dos clases, la cual se representa a través de la variables $y_i \in \{0,1\}$ con $i \in \{1,...,I\}$. Se define la variable binaria y_i de la siguiente forma:

$$y_i = \begin{cases} 1 & \text{Si el objeto } i \text{ pertenece a la clase Positiva o clase base.} \\ 0 & \text{Si el objeto } i \text{ pertenece a la clase Negativa.} \end{cases} \tag{4.1}$$

Se dirá que una observación es positiva si $y_i = 1$ y una observación es negativa si $y_i = 0$.

De esta forma, el conjunto de entrenamiento queda compuesto por cada una de las observaciones (x_1, y_1), (x_2, y_2), ..., (x_I, y_I). Cada observación corresponde al par ordenado dado por las características del objeto y la variable que identifica su clase.

Un clasificador binario se trata de una función $f : \mathbb{R}^n \to \{0,1\}$ que permite asignar un objeto $x_i \in \mathbb{R}^n$, cuya clase es desconocida, a una predicción de su clase $\widehat{y_i} = f(x_i) \in \{0,1\}$.

Esta predicción puede ser correcta o incorrecta con respecto a la clase real $y_i \in \{0,1\}$ del objeto, que es desconocida al momento de realizar la predicción. Se distinguen cuatro escenarios posibles que pueden suceder con la predicción $\widehat{y_i}$:

- La predicción es Verdadera Positiva (TP, *True Positive*) cuando el clasificador indica que la observación es Positiva ($\widehat{y_i} = 1$) y la predicción es acertada porque la observación realmente es Positiva ($y_i = 1$).
- Una estimación es Falsa Positiva (FP, *False Positive*) cuando el clasificador indica que la observación es Positiva ($\widehat{y_i} = 1$) y la predicción es incorrecta porque la observación en realidad es Negativa ($y_i = 0$).
- Una estimación es Verdadera Negativa (TN, *True Negative*) cuando el clasificador indica que la observación es Negativa ($\widehat{y_i} = 0$) y la predicción es acertada porque la observación realmente es Negativa ($y_i = 0$).

- La predicción es Falsa Negativa (FN, *False Negative*) cuando el clasificador indica que la observación es Negativa ($\widehat{y}_i = 0$) y la predicción es incorrecta porque la observación en realidad es Positiva ($y_i = 1$).

La tabla 4.1 muestra la decisión del clasificador en relación a la clase real de las observaciones.

	$y_i = 1$	$y_i = 0$
$\widehat{y}_i = 1$	Verdadero Positivo (TP)	Falso Positivo (FP)
$\widehat{y}_i = 0$	Falso Negativo (FN)	Verdadero Negativo (TN)

Tabla 4.1: Tabla de confusión e identificación de aciertos y errores

Diferentes clasificadores pueden tener diferentes niveles de eficiencia en términos de las tasas de Verdaderos Positivos, Verdaderos Negativos, Falsos Positivos o Falsos Negativos. Para cada problema de clasificación se pueden definir diferentes criterios que se busca optimizar, por ejemplo, asociando utilidades o costos a cada tipo de acierto o error del clasificador. Por ejemplo, en algunos problemas es más importante minimizar la tasa de error de Falsos Negativos por sobre la de Falsos Positivos. Según los niveles de eficiencia entonces se pueden evaluar los diferentes clasificadores y luego escoger el mejor clasificador para el objetivo definido.

4.3. Modelos de Elección Discreta

4.3.1. Definiciones

Se entiende por elección al proceso de toma de decisión de un individuo o grupo de personas, en adelante agente, al tener que seleccionar una entre varias alternativas. Los modelos de elección discreta ofrecen una representación matemática de los problemas de elección donde el número de alternativas es finito, las alternativas son mutuamente excluyentes y el conjunto de elección es conocido por el tomador de decisiones (Manski, 1977). En lo que sigue se caracterizará el proceso de elección a través del enfoque utilidad aleatoria propuesto por McFadden (1976), que realiza una representación econométrica del comportamiento de maximización de utilidades de los agentes, bajo el supuesto de que sus comportamientos son racionales.

Un agente tomador de decisiones debe escoger entre J alternativas. En un modelo de elección discreta se supone que cada agente tomador de decisión busca maximizar su utilidad. La utilidad que obtiene el agente i por elegir la alternativa j es U_{ij}, donde $i \in \{1,..,I\}$ y $j \in \{1,...,J\}$. Esta utilidad es conocida por el agente pero no por el investigador. El agente escoge la alternativa k si y sólo si $U_{ik} > U_{ij}\ \forall j \neq k$.

4.3.2. Probabilidades de Elección

El investigador no observa las utilidades de los agentes. El investigador observa algunos atributos que el agente también observa sobre las alternativas y algunos atributos propios del agente. Supongamos que el

agente i observa n atributos relativos a cada alternativa j y que son caracterizados a través del vector $x_{ij} \in \mathbb{R}^n$. De esta forma el investigador puede especificar una función para representar la utilidad de cada agente frente a cada alternativa. Se define esta función como $V_{ij} = V(b, x_{ij})\ \forall i, j$. La función $V(\cdot)$ relaciona cada atributo con niveles de utilidad a través de parámetros $b \in \mathbb{R}^n$ que son desconocidos y que el investigador intentará estimar estadísticamente.

Como existen aspectos de la utilidad que el investigador no puede observar entonces $U_{ij} \neq V_{ij}$. La utilidad entonces se puede descomponer como $U_{ij} = V_{ij} + \varepsilon_{ij}$, donde ε_{ij} captura los factores que afectan la utilidad pero que no están incluidas en V_{ij}. Hasta el momento esta descomposición es general, ya que ε_{ij} se define como la diferencia entre la utilidad real U_{ij} y la parte de la utilidad capturada por el investigador V_{ij}.

El investigador no conoce ε_{ij} y por tanto trata este término como un elemento aleatorio. Se define la densidad de probabilidad conjunta del vector $\varepsilon_i = \langle \varepsilon_{i1}, ..., \varepsilon_{iJ} \rangle$ como $f(\varepsilon_i)$. Con esto, el investigador puede afirmar que la probabilidad de que el agente i escoja la alternativa k esta dada por la siguiente formula.

$$\begin{aligned} P_{ik} &= Prob(U_{ik} > U_{ij} \quad \forall j \neq k) \\ &= Prob(V_{ik} + \varepsilon_{ik} > V_{ij} + \varepsilon_{ij} \quad \forall j \neq k) \\ &= Prob(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij} \quad \forall j \neq k) \end{aligned} \tag{4.2}$$

Esta probabilidad representa la distribución acumulada asociada a que cada término aleatorio $\varepsilon_{ij} - \varepsilon_{ik}$ sea menor que el diferencial de utilidad de las alternativas $V_{ik} - V_{ij}$. Utilizando la distribución de densidad de probabilidad $f(\varepsilon_i)$, esta probabilidad acumulada se puede escribir según la siguiente expresión.

$$\begin{aligned} P_{ik} &= Prob(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij} \quad \forall j \neq k) \\ &= \int_{\varepsilon} I(\varepsilon_{ij} - \varepsilon_{ik} < V_{ik} - V_{ij} \quad \forall j \neq k) f(\varepsilon_i) d\varepsilon_i \end{aligned} \tag{4.3}$$

Donde $I(\cdot)$ es la función indicatriz, que toma el valor 1 cuando la expresión del paréntesis es verdadera y 0 cuando no. Esta es una integral multidimensional sobre la densidad de probabilidad de la porción de utilidad no observable $f(\varepsilon_i)$. Se pueden obtener diferentes modelos de elección discreta con diferentes especificaciones de esta función de densidad, lo que implica diferentes supuestos sobre la distribución de probabilidad de la porción de utilidad no observable. La solución de esta integral tiene formulas cerradas sólo para algunas especificaciones de $f(\cdot)$. El modelo de elección discreta Logit, que es explicado a continuación, es derivado a partir de supuestos que llegan a una formula cerrada para esta integral. Con otras especificaciones, como el Probit o el Mixed Logit, no se llegan a formulas cerradas pero si a expresiones que pueden ser evaluadas numéricamente a través de técnicas de simulación.

Es importante destacar que los valores absolutos de utilidad son irrelevantes tanto para los agentes tomadores de decisión como para el investigador. Como se mostró recién, la decisión depende de las diferencias de utilidad y no de los valores absolutos. Esto quiere decir que los únicos parámetros que pueden ser estimados por el investigador tienen que ver con los que capturan las diferencias entre las alternativas.

4.3.3. Modelo de Elección Discreta Logit Multinomial

El modelo de elección discreta Logit Multinomial, Logit en adelante, es una de las especificaciones más utilizadas en la literatura porque sus supuestos permiten llegar a una formula cerrada para la integral que define la probabilidad de que un agente elija determinada alternativa.

El modelo Logit se obtiene de asumir que cada termino εij es independiente e idénticamente distribuido a través de una distribución Gumbel. La función de densidad y la función de probabilidad acumulada de esta distribución son las siguientes.

$$f(\varepsilon_{ij}) = e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}} \tag{4.4}$$

$$F(\varepsilon_{ij}) = e^{-e^{-\varepsilon_{ij}}} \tag{4.5}$$

Utilizando estas funciones se puede derivar el modelo Logit encontrando una expresión para la probabilidad expresada a través de la integral de la formula (4.3). En Train (2003) se detalla el tratamiento algebraico para llegar a la siguiente formula cerrada.

$$P_{ik} = \frac{e^{V_{ik}}}{\sum_j e^{V_{ij}}} \tag{4.6}$$

Esta expresión representa la probabilidad de elección que plantea un modelo de elección discreta Logit Multinomial. La probabilidad de elección de la alternativa k corresponde a la importancia relativa del término $e^{V_{ik}}$ sobre la suma de todos los términos $e^{V_{ij}}$ de todas las alternativas j (incluyendo la alternativa k). La utilidad observada por el investigador usualmente se expresa como una función lineal sobre el vector de atributos observados: $V_{ik} = b^T x_{ik}$. Donde x_{ik} es el vector de atributos observados y b es el vector de parámetros asociados a cada atributo.

4.3.4. Propiedades de las Probabilidades Logit

Propiedades Generales

Las probabilidades Logit tienen varias propiedades deseables. Las probabilidades P_{ik} se encuentran necesariamente entre cero y uno, como se requiere para una expresión de probabilidad. Cuando la utilidad observable de una alternativa V_{ik} crece y el resto de se mantiene constante (V_{ij} se mantiene constante $\forall j \neq k$), entonces P_{ik} tiende a uno. Por otro lado, cuando V_{ik} decrece y el resto de las utilidades observables se mantienen constantes, P_{ik} tiende a cero ya que la exponencial del numerador de la formula (4.6) tiende a cero cuando V_{ik} tiende a $-\infty$. La probabilidad Logit asociada a una alternativa no puede ser cero y la única forma que sea igual a uno es cuando existe sólo una alternativa.

Forma Funcional *Sigmoidal*

La forma funcional de la probabilidad Logit asociada a una alternativa con respecto al nivel de utilidad observable de la misma, sigue una forma funcional *Sigmoidal* tal como se observa en la figura 4.1.

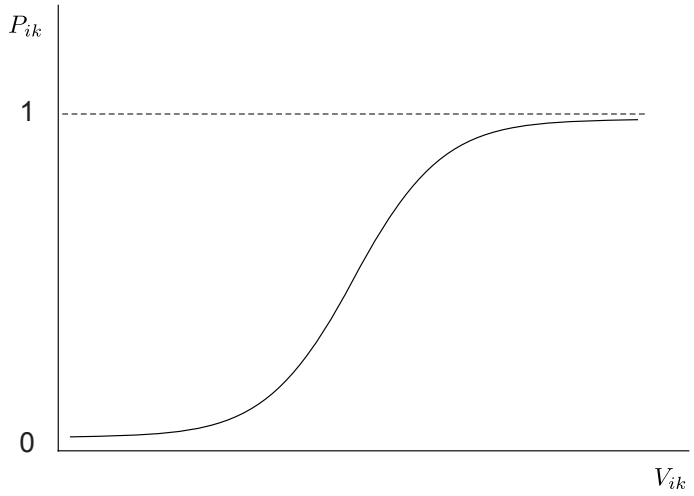

Figura 4.1: Forma funcional *Sigmoidal* de la probabilidad Logit

Esta forma funcional implica que la sensibilidad de la probabilidad asociada a una alternativa con respecto al nivel de utilidad observable de la misma no es lineal y depende del nivel de esta utilidad. En los extremos, cuando la utilidad observable de una alternativa es muy superior o muy inferior al resto, cambios en los niveles de su utilidad (suponiendo que no hay cambios en las utilidades de las otras alternativas) se traducen en pequeños cambios en la probabilidad respectiva. En el caso de un nivel bajo de utilidad para una alternativa, el que ésta baje aún más no hace que la probabilidad de elección siga disminuyendo: ya que las otras alternativas ya dominan la elección y en el nuevo escenario continúan dominando. En el caso de un nivel alto de utilidad para una alternativa, el que ésta aumente aún más no hace que la probabilidad de elección siga aumentando: ya que esta alternativa ya domina la elección y en el nuevo escenario continua dominando. Cuando la probabilidad se encuentra en torno al 50 %, cualquier cambio en los niveles de utilidad de una alternativa se traduce en una respuesta casi lineal en términos de probabilidad.

Sensibilidad de la Probabilidad Logit ante Cambios en los Atributos

Las conclusiones sobre la sensibilidad de las probabilidades Logit a los niveles de utilidad observables se pueden extrapolar para la sensibilidad que tiene la probabilidad de una alternativa con respecto a cambios en el valor de algún atributo asociado a la misma o alguna otra alternativa.

La sensibilidad de la probabilidad de una alternativa ante cambios en los atributos de la misma se puede representar a través de la derivada parcial de P_{ik} con respecto a cambios en alguno de los atributos z_{ik} asociado a la misma alternativa k.

$$\frac{\partial P_{ik}}{\partial z_{ik}} = \frac{e^{V_{ik}}}{\sum_j e^{V_{ij}}} \frac{\partial V_{ik}}{\partial z_{ik}} - \frac{e^{V_{ik}}}{\left(\sum_j e^{V_{ij}}\right)^2} e^{V_{ik}} \frac{\partial V_{ik}}{\partial z_{ik}} = \frac{\partial V_{ik}}{\partial z_{ik}} P_{ik}(1 - P_{ik}) \tag{4.7}$$

Si V_{ik} se expresa como una función lineal de los atributos y el parámetro que acompaña al atributo z_{ik} es b_z entonces la derivada anterior queda dada la siguiente expresión.

$$\frac{\partial P_{ik}}{\partial z_{ik}} = b_z P_{ik}(1 - P_{ik}) \tag{4.8}$$

Esta derivada toma el valor más alto cuando $P_{ik} = 1 - P_{ik}$ =0.5. La derivada toma los valores más bajos cuando P_{ik} se acerca a cero o a uno. Como se discutía anteriormente, esto es consecuencia y consistente con la forma funcional *Sigmoidal*. Cuando $b_z > 0$, aumentos en el atributo z_{ik} se traducen en un aumento de la probabilidad de elección de la alternativa k y se interpreta que la variable es deseable pues a mayor valor de la variable mejora la probabilidad de elección de la alternativa que la posee.

La sensibilidad de la probabilidad de una alternativa ante cambios en los atributos de otra alternativa, se puede representar a través de la derivada parcial de P_{ik} con respecto a cambios en alguno de los atributos z_{ij} asociado a otra alternativa j.

$$\frac{\partial P_{ik}}{\partial z_{ij}} = -\frac{e^{V_{ik}}}{(\sum_h e^{V_{ih}})^2} e^{V_{ij}} \frac{\partial V_{ij}}{\partial z_{ij}} = -\frac{\partial V_{ij}}{\partial z_{ij}} P_{ik} P_{ij} \qquad \forall j \neq k \tag{4.9}$$

Cuando V_{ij} se expresa como una función lineal de los atributos y el parámetro que acompaña al atributo z_{ij} es b_z entonces la derivada anterior queda dada la expresión que sigue.

$$\frac{\partial P_{ik}}{\partial z_{ij}} = -b_z P_{ik} P_{ij} \qquad \forall j \neq k \tag{4.10}$$

Cuando $b_z > 0$, aumentos en el atributo z_{ij} (asociado a la alternativa j) se traducen en una disminución de la probabilidad de elección de cada una de las otras alternativas (cada alternativa $k \neq j$). Esto sucede porque un aumento en z_{ij} mejora la probabilidad de elección de la alternativa j, ya que la variable z es una variable deseable cuando $b_z > 0$. Como mejora la valoración de la alternativa j, las otras alternativas se vuelven menos deseables y por lo tanto disminuyen su probabilidad de elección. El cambio en la probabilidad de las otras alternativas (disminución cuando $b_z > 0$) es proporcional al valor de la probabilidad justo antes del cambio que se observa en z_{ij}.

Sustitución Proporcional de las Alternativas

La probabilidad Logit cumple que la suma de las probabilidades de todas las alternativas suman uno. Esta es una propiedad básica para una distribución de probabilidad y aplicado al caso de elección discreta se traduce

en que siempre al menos una de las alternativas debe ser escogida.

$$\sum_k P_{ik} = \sum_k \frac{e^{V_{ik}}}{\sum_j e^{V_{ij}}} = \frac{\sum_k e^{V_{ik}}}{\sum_j e^{V_{ij}}} = 1 \tag{4.11}$$

Esto implica que cuando se tienen J alternativas basta con determinar la probabilidad asociada a $J-1$ de ellas, pues la faltante se obtiene a partir de la ecuación anterior (uno menos la suma de las $J-1$ probabilidades conocidas). Otra consecuencia de lo anterior es que el modelo Logit implica sustituciones proporcionales a través de las alternativas. Esto quiere decir en términos prácticos que cualquier mejora en probabilidad de alguna alternativa tiene que necesariamente provocar la disminución de alguna de las probabilidades de las otras alternativas. Las probabilidades Logit cumplen entonces con que el efecto de sustitución total entre las alternativas suman cero, lo cual queda demostrado a través de las siguientes expresiones.

$$\begin{aligned}
\sum_{k=1}^{J} \frac{\partial P_{ik}}{\partial z_{ij}} &= \frac{\partial V_{ij}}{\partial z_{ij}} P_{ij}(1-P_{ij}) + \sum_{k \neq j} \left(-\frac{\partial V_{ij}}{\partial z_{ij}} \right) P_{ik} P_{ij} \\
&= \frac{\partial V_{ij}}{\partial z_{ij}} P_{ij} \left((1-P_{ij}) - \sum_{k \neq j} P_{ik} \right) \\
&= \frac{\partial V_{ij}}{\partial z_{ij}} P_{ij} \left((1-P_{ij}) - (1-P_{ij}) \right) \\
&= 0
\end{aligned} \tag{4.12}$$

Esto es consistente con la interpretación de las derivadas parciales de las formulas (4.8) y (4.10). Cuando un atributo deseable aumenta su valor en la alternativa j entonces, de acuerdo a la expresión (4.8), la probabilidad de elección de la misma alternativa aumenta. Por otro lado, de acuerdo a la expresión (4.10), la probabilidad de elección del resto de las alternativas disminuye. La magnitud del aumento de la probabilidad de elección de la alternativa j entonces depende de cuanto disminuyan las probabilidades de elección de las otras alternativas, ya que la suma de todas las probabilidades se mantiene constante e igual a uno.

Independencia a las Alternativas Irrelevantes

Otra característica de las probabilidades Logit es la independencia con respecto a las alternativas irrelevantes (IIA por sus siglas en inglés). Esta propiedad se refiere a que la relación que existe entre las probabilidades de elección de cualquiera de dos alternativas, sólo se explica por las características asociadas a ambas alternativas (específicamente las utilidades que generan), sin importar lo que suceda con el resto de las alternativas. Consideremos el ratio de las probabilidades Logit asociadas a las alternativas k y j.

$$\frac{P_{ik}}{P_{ij}} = \frac{e^{V_{ik}} / \sum_h e^{V_{ih}}}{e^{V_{ij}} / \sum_h e^{V_{ih}}} = \frac{e^{V_{ik}}}{e^{V_{ij}}} = e^{V_{ik} - V_{ij}} \tag{4.13}$$

En particular el ratio entre las probabilidades de elección de cualquiera de dos alternativas, sólo se explica por las diferencias de utilidad que existe entre ambas alternativas. Las utilidades que generan el resto de las alternativas son irrelevantes al comparar dos alternativas especificas.

4.4. Regresión Logística

4.4.1. Definición

La Regresión Logística permite abordar el problema de Clasificación Binaria. Ésta plantea una forma funcional Logística $\sigma(\cdot)$ que a partir de las características de un objeto x_i y un vector de parámetros b estima la probabilidad de que el objeto pertenezca a la clase positiva o clase base.

$$\begin{aligned} \sigma: \quad & \mathbb{R}^n \longrightarrow (0,1) \\ & x_i \longmapsto \sigma(b, x_i) \end{aligned} \tag{4.14}$$

Se define $\sigma_i = \Pr(y_i = 1) = \sigma(b, x_i)$ como la probabilidad de que el objeto i pertenezca a la clase positiva estimada a partir del modelo de Regresión Logística. La función $\sigma(\cdot)$ propuesta por la regresión logística es la siguiente.

$$\sigma_i = \sigma(b, x_i) = \frac{1}{1 + e^{-V(b, x_i)}} \tag{4.15}$$

Donde $V(\cdot)$ es una función lineal sobre las características del cliente $V(b, x_i) = b^T x_i$ y $b \in \mathbb{R}^n$ es el vector de parámetros asociado a cada atributo.

$$V(b, x_i) = b^T x_i = b_1 + b_2 x_{i2} + b_3 x_{i3} + \ldots + b_n x_{in} \qquad \forall i \in \{1, \ldots, I\} \tag{4.16}$$

Se debe notar que b_1 corresponde a la constante del problema y por lo tanto el primer componente del vector de características de cada objeto se define como: $x_{i1} = 1 \quad \forall i$.

Dado el conjunto de atributos que se incluyen en el modelo de Regresión Logística, el modelo queda definido cuando se determinan los parámetros b que utiliza la función $\sigma(\cdot)$. Existen tantas alternativas de modelos como posibles parámetros b. El objetivo al resolver el modelo de Regresión Logística es encontrar los parámetros b que relacionen de la mejor forma las características de las observaciones x_i con la clase y_i a la que pertenece.

4.4.2. Regresión Logística como Modelo de Elección Discreta Binaria

Si se supone un modelo de elección discreta Logit Multinomial de acuerdo a lo expuesto en la sección 4.3, las probabilidades de elección están dadas por la siguiente expresión.

$$P_{ik} = \frac{e^{V_{ik}}}{\sum_j e^{V_{ij}}} \tag{4.17}$$

Esta expresión corresponde a la probabilidad de que el agente i elija la alternativa k. Si se asume un problema de elección discreta binaria, cuando existen sólo dos alternativas de elección, entonces las probabilidades de elección para el cliente i de cada alternativa son las siguientes.

$$P_{i0} = \frac{e^{V_{i0}}}{e^{V_{i0}} + e^{V_{i1}}} \tag{4.18}$$

$$P_{i1} = \frac{e^{V_{i1}}}{e^{V_{i0}} + e^{V_{i1}}} \tag{4.19}$$

De acuerdo a las propiedades de los modelos Logit presentadas en la sección 4.3.4, las probabilidades de elección Logit se caracterizan porque son explicadas sólo por las diferencias de utilidad entre las alternativas relevantes. En el caso binario, como sólo existen dos alternativas las probabilidades de elección pueden explicarse sólo con las diferencias de utilidad de una alternativa respecto de la otra. Esto quiere decir que es posible definir arbitrariamente el nivel de utilidad que entrega una de las dos alternativas, y que por lo tanto el modelo se explique por la utilidad que entrega la otra alternativa. Definamos entonces que la alternativa $k = 0$ tiene una utilidad base de $V_{i0} = 0$. De esta forma, como $e^0 = 1$ entonces las probabilidades de elección quedan de la siguiente forma.

$$P_{i0} = \frac{1}{1 + e^{V_{i1}}} \tag{4.20}$$

$$P_{i1} = \frac{e^{V_{i1}}}{1 + e^{V_{i1}}} \tag{4.21}$$

En este caso la función V_{i1} representa la utilidad que entrega la alternativa $k = 1$ al agente respecto al nivel base definido para la alternativa $k = 0$. Si además se supone que $V_{i1} = -V(b, x_i)$, con $V(b, x_i) = b^T x_i$, entonces el modelo de elección discreta Logit binario es equivalente al modelo de Regresión Logística, donde la alternativa $k = 0$ del modelo Logit corresponde a la clase positiva o clase base del modelo de Regresión Logística.

$$P_{i0} = \sigma_i = \frac{1}{1 + e^{-V(b, x_i)}} \tag{4.22}$$

De esta forma, la alternativa $k = 1$ del modelo Logit corresponde a la clase negativa del modelo de Regresión Logística y su probabilidad de elección basta con definirla como el complemento de la probabilidad de elección de la clase base.

4.4.3. Estimación del Modelo de Regresión Logística

Considerando el conjunto de observaciones $\{(x_i, y_i)\}_{i=1}^{I}$ disponibles, el modelo de Regresión Logística estima una probabilidad σ_i de que el objeto i pertenezca a la clase positiva y una probabilidad $1 - \sigma_i$ de que el objeto i pertenezca a la otra clase o clase negativa. De esta forma, la probabilidad de que el objeto i pertenezca a la clase observada, independiente de cuál clase sea, esta dada por la siguiente expresión.

$$\sigma_i^{y_i}(1-\sigma_i)^{1-y_i} \tag{4.23}$$

Cuando el objeto i pertenece a la clase positiva entonces $y_i = 1$ y por lo tanto la probabilidad anterior es igual a $\sigma_i^1(1-\sigma_i)^0 = \sigma_i$. En el otro caso, cuando la observación pertenece a la clase negativa entonces $y_i = 0$ y por lo tanto la probabilidad anterior es igual a $\sigma_i^0(1-\sigma_i)^1 = 1 - \sigma_i$.

Si se asume que la asignación de los objetos a las dos clases posibles es un proceso de probabilidad independiente, entonces la probabilidad conjunta de que todos los objetos pertenezcan a las clases que se observan está dada por la multiplicación de las probabilidades definidas en (4.23) para cada objeto.

$$\mathcal{L}(b/x) = \prod_{i=1}^{I} \sigma_i^{y_i}(1-\sigma_i)^{1-y_i} \tag{4.24}$$

La probabilidad conjunta anterior corresponde a la función de verosimilitud del modelo de Regresión Logística, al utilizar los parámetros b dadas las características de cada objeto. Los parámetros b que mejor explican la realidad observada son los que maximizan la función de verosimilitud y que por lo tanto maximizan la probabilidad conjunta estimada por el modelo. Los parámetros de máxima verosimilitud son los que necesariamente estiman en promedio las probabilidades más cercanas a las clases de los objetos. Esto porque para maximizar la función $\mathcal{L}(b/x)$ idealmente para cada objeto el término $\sigma_i^{y_i}(1-\sigma_i)^{1-y_i}$ tiene que ser lo más grande posible. Esto se consigue si para cada objeto donde $y_i = 1$ el modelo estima probabilidades σ_i que son lo más grande posible con respecto a $1 - \sigma_i$; y viceversa cuando $y_i = 0$.

La maximización de la función de verosimilitud es equivalente a la maximización del logaritmo de la función de verosimilitud, dado que el logaritmo es una función monótona creciente. Al aplicar el logaritmo natural a la función de verosimilitud se obtiene una función que es más fácil de tratar algebraicamente y por lo tanto de resolver computacionalmente. Si se define $\ell(b/x) = \ln(\mathcal{L}(b/x))$ entonces los parámetros de máxima verosimilitud $\hat{b}$ se obtienen equivalentemente a través de la solución del siguiente problema de optimización.

$$\hat{b} = \underset{b \in \mathbb{R}^n}{\arg\max} \{\ell(b/x)\} \tag{4.25}$$

McFadden (1974) demostró que el logaritmo de la función de verosimilitud $\ell(b/x)$ es globalmente cóncavo con respecto a los parámetros b. Esto permite asegurar que el problema de optimización (4.25) tiene una única solución $\hat{b}$ que maximiza $\ell(b/x)$. De esta forma las probabilidades estimadas por el modelo de Regresión Logística corresponden a las siguientes.

$$\hat{\sigma}_i = \sigma(\hat{b}, x_i) = \frac{1}{1+\exp\left(-\left[\hat{b}_1 + \sum_{j=2}^{n} \hat{b}_j x_{ij}\right]\right)} \tag{4.26}$$

Donde x_{ij} es el valor del atributo j del vector $x_i = \langle 1, x_{i2}, ..., x_{in} \rangle$ que caracteriza al cliente i, con $x_{i1} = 1$. $\hat{b}_j$ es el j-ésimo componente del vector $\hat{b} = \langle \hat{b}_1, \hat{b}_2, ..., \hat{b}_n \rangle$ y es también el parámetro estimado que pondera al atributo j en la función lineal $V(b, x_i)$.

A partir de las propiedades de la función $\ln(\cdot)$ es posible representar $\ell(b/x)$ transformando las multiplicaciones que forman la función $\mathcal{L}(b/x)$ en sumas.

$$\begin{aligned} \ell(b/x) &= \sum_{i=1}^{I} \ln\left(\sigma_i^{y_i}(1-\sigma_i)^{1-y_i}\right) \\ &= \sum_{i=1}^{I} \left(y_i \ln(\sigma_i) + (1-y_i)\ln(1-\sigma_i)\right) \end{aligned} \tag{4.27}$$

La solución $\hat{b}$ del problema de optimización (4.25) debe satisfacer las condiciones de primer orden del problema, las que se obtienen igualando a cero el gradiente de $\ell(b/x)$.

$$\nabla_b \ell(b/x) = 0 \quad \Leftrightarrow \quad \frac{\partial \ell(b/x)}{\partial b_k} = 0 \quad \forall k \in \{1, ..., n\} \tag{4.28}$$

El gradiente de $\ell(b/x)$ corresponde al vector con las derivadas parciales de la función $\ell(b/x)$ con respecto a cada componente del vector b. La derivada parcial con respecto al k-ésimo componente del vector b queda determinado por la siguiente expresión.

$$\begin{aligned} \frac{\partial \ell(b/x)}{\partial b_k} &= \sum_{i=1}^{I} \left(y_i \frac{1}{\sigma_i} \frac{\partial \sigma_i}{\partial b_k} - (1-y_i) \frac{1}{1-\sigma_i} \frac{\partial \sigma_i}{\partial b_k} \right) \\ &= \sum_{i=1}^{I} \left(y_i \frac{1}{\sigma_i} \sigma_i(1-\sigma_i) x_{ik} - (1-y_i) \frac{1}{1-\sigma_i} \sigma_i(1-\sigma_i) x_{ik} \right) \\ &= \sum_{i=1}^{I} \left(y_i(1-\sigma_i) - (1-y_i)\sigma_i \right) x_{ik} \\ &= \sum_{i=1}^{I} \left(y_i - \sigma_i \right) x_{ik} \quad \forall k \in \{1, ..., n\} \end{aligned} \tag{4.29}$$

De esta forma, imponiendo que las derivadas parciales sean iguales a cero se obtienen las condiciones de primer orden del Modelo de Regresión Logística. Estas condiciones establecen un sistema de n ecuaciones que deben satisfacer los n parámetros de vector $\hat{b}$ y que permite resolver el problema (4.25).

Si se reordenan los términos y se divide por el número de agentes I en cada lado, entonces se obtiene el siguiente sistema de ecuaciones.

$$\frac{\partial \ell(b/x)}{\partial b_k} = 0 \quad \Leftrightarrow \quad \frac{1}{I}\sum_{i=1}^{I} y_i x_{ik} = \frac{1}{I}\sum_{i=1}^{I} \sigma_i x_{ik} \quad \forall k \in \{1,...,n\} \tag{4.30}$$

Las condiciones de primer orden del Modelo de Regresión Logística imponen que el estimador de máxima verosimilitud $\hat{b}$ sea el vector de parámetros que hace que para los objetos pertenecientes a la clase positiva el valor promedio estimado por el modelo para cada variable explicativa sea igual al promedio efectivamente observado en la muestra utilizada para su estimación. Se puede verificar que estas condiciones de primer orden también hacen que el promedio estimado por el modelo y el promedio observado para los objetos pertenecientes a la clase negativa sean iguales. De esta forma las condiciones de primer orden del problema inducen al Modelo de Regresión Logística a reproducir las características promedio de los objetos observadas en la muestra para ambas clase.

Para verificar que la interpretación anterior también aplica para los objetos de la clase negativa, basta con multiplicar por menos uno la expresión (4.30) y luego sumar $\frac{1}{I}\sum x_{ik}$ en ambos lados de la ecuación, obteniendo el siguiente sistema de ecuaciones equivalente al anterior.

$$\frac{1}{I}\sum_{i=1}^{I}(1-y_i)x_{ik} = \frac{1}{I}\sum_{i=1}^{I}(1-\sigma_i)x_{ik} \quad \forall k \in \{1,...,n\} \tag{4.31}$$

4.4.4. Solución Numérica y Método de Newton-Raphson

Si se suponen conocidos los vectores de características x_i y las clases y_i de los objetos entonces las únicas incógnitas del sistema de ecuaciones (4.30) son los n parámetros del vector b que determinan las probabilidades $\sigma_i = \sigma(b, x_i)$. A partir de las n ecuaciones que plantean las condiciones de primer orden (4.30) es posible obtener una solución $\hat{b}$ del problema de optimización planteado en (4.25).

Dado que el sistema de ecuaciones es no lineal, para poder resolver este tipo problemas es necesario recurrir a métodos numéricos que convergen a la solución $\hat{b}$ de forma iterativa. Se utiliza el método de Newton-Raphson para resolver este problema de forma numérica. Este método se utiliza para resolver sistemas de ecuaciones no lineales. Como un problema de optimización sin restricciones se puede traducir en un sistema de ecuaciones a partir de las condiciones de optimalidad de primer orden, el método de Newton-Raphson se puede adaptar para resolver problemas de optimización no lineales sin restricciones.

Sea $f(X)$ la función no lineal que se desea maximizar o minimizar y sea $X \in \mathbb{R}^n$ la variable de decisión del problema. En la iteración q se tiene que la solución candidata es $X^{(q)}$. Una condición necesaria para

que una solución $X^{(q)}$ sea la solución óptima del problema es que sea un punto estacionario, es decir, que se cumpla que $\nabla f(X^{(q)}) = 0$. El método de Newton-Raphson en cada iteración avanza hacia una solución que se acerque cada vez más al punto estacionario más cercano. De esta forma, dado $X^{(q)}$, la nueva solución candidata $X^{(q+1)}$ se define a partir de la siguiente recurrencia.

$$X^{(q+1)} = X^{(q)} - \mathbf{H}_f^{-1}(X^{(q)})\nabla f(X^{(q)}) \tag{4.32}$$

Este método requiere una solución inicial $X^{(0)}$ para comenzar las iteraciones y de un criterio de parada que indique que el punto estacionario ha sido alcanzado. Existen varios criterios de parada comúnmente utilizados:

- Si $|\nabla f(X^{(q)})| < \varepsilon$, donde $\varepsilon > 0$ corresponde a un número muy pequeño, entonces el gradiente se acerca mucho a cero y por lo tanto se ha encontrado un punto estacionario.
- Si $|f(X^{(q+1)}) - f(X^{(q)})| < \varepsilon$ entonces la nueva iteración ya no mejora el valor de la función objetivo y por lo tanto se ha encontrado un punto estacionario. Notemos que si se cumple esta condición entonces también se cumplirá la anterior ya que $\nabla f(X^{(q+1)}) \approx \frac{f(X^{(q+1)}) - f(X^{(q)})}{X^{(q+1)} - X^{(q)}}$.
- Si $\|X^{(q+1)} - X^{(q)}\|_1 < \varepsilon$ entonces la nueva iteración no cambia la solución candidata. Si la solución candidata no cambia entonces la función objetivo tampoco cambiará por lo que esta condición es equivalente a las condiciones anteriores. Esta condición tiene la ventaja de ser la más simple de verificar en términos computacionales.

Si se cumple la condición de parada en la Q-ésima iteración entonces $\nabla f(X^{(Q)}) \approx 0$, siendo $X^{(Q)}$ un punto estacionario y cumpliendo con la condición necesaria para ser la solución óptima del problema. Esta condición necesaria además es suficiente cuando $f(\cdot)$ es Convexa o Cóncava, ya que existe un único punto estacionario que corresponde a la solución óptima del problema. Si $f(\cdot)$ es Convexa entonces basta con que $X^{(Q)}$ sea un punto estacionario para asegurar que es el mínimo global de la función objetivo. Si $f(\cdot)$ es Cóncava entonces basta con que $X^{(Q)}$ sea un punto estacionario para asegurar que es el máximo global de la función objetivo. En caso contrario, el punto estacionario encontrado podría corresponder sólo a un óptimo local del problema.

Para utilizar el método de Newton-Raphson para resolver el problema de optimización (4.25), que estima los parámetros del modelo de Regresión Logística, se debe considerar que la función objetivo corresponde al logaritmo de la función de verosimilitud $\ell(b/x)$ y la variable decisión es el vector de parámetros b. De esta forma, a partir de una solución inicial $b^{(0)}$ es posible construir la recurrencia de Newton-Raphson.

$$b^{(q+1)} = b^{(q)} - \mathbf{H}_\ell^{-1}(b^{(q)})\nabla_b \ell(b^{(q)}) \tag{4.33}$$

Cuando se cumple la condición de parada en la Q-ésima iteración entonces el valor $b^{(Q)}$ corresponde a la solución óptima del problema (4.25) y determina el estimador de los parámetros del modelo de Regresión Logística: $\hat{b} = b^{(Q)}$.

Notemos que para resolver el problema se requiere conocer el gradiente $\nabla_b \ell(b/x)$ y el hessiano $\mathbf{H}_\ell(b/x)$. Cada componente del gradiente $\nabla_b \ell(b/x)$ fue definido en la expresión (4.29). El hessiano $\mathbf{H}_\ell(b/x)$ se define como la matriz de segundas derivadas parciales de $\ell(b/x)$ con respecto a los componentes del vector b.

$$\mathbf{H}_\ell(b/x)_{kj} = \frac{\partial^2 \ell(b/x)}{\partial b_j \partial b_k} \quad \forall k, j \in \{1, ..., n\} \tag{4.34}$$

De esta forma, cada componente de la matriz hessiana $\mathbf{H}_\ell(b/x)$ está dado por la siguiente expresión.

$$\begin{aligned} \frac{\partial^2 \ell(b/x)}{\partial b_j \partial b_k} &= \sum_{i=1}^{I} -\frac{\partial \sigma_i}{\partial b_j} x_{ik} \\ &= -\sum_{i=1}^{I} \sigma_i (1 - \sigma_i) x_{ij} x_{ik} \quad \forall k, j \in \{1, ..., n\} \end{aligned} \tag{4.35}$$

4.4.5. Implementación Computacional del Modelo de Regresión Logística

Para resolver el modelo de Regresión Logística descrito en este capítulo, se realiza una implementación propia utilizando el lenguaje del paquete estadístico R (R Development Core Team, 2008). En el Anexo F se presenta la estrategia de implementación de las diferentes metodologías utilizadas en esta Tesis a través del lenguaje R. En particular en el Anexo F.1 se describe la implementación de la metodología de solución del modelo de Regresión Logística a través del método Newton-Raphson, como se propone en este capítulo.

Parte III

Modelos Propuestos y Resultados

Capítulo 5

Predicción de Fuga de Clientes con Restricciones de Competencia

5.1. Fuga de Clientes utilizando Regresión Logística

El problema que se desea abordar en este trabajo es la estimación de la probabilidad de fuga de clientes de una firma. Este problema puede ser abordado como un problema de clasificación binaria. El problema de clasificación binaria se representa a través de las observaciones $\{(x_i, y_i)\}_{i=1}^{I}$, donde $x_i \in \mathbb{R}^n$ es el vector de características de la observación i e $y_i \in \{0,1\}$ identifica la clase a la que pertenece la observación i. En el problema de clasificación binaria se intenta calcular $\Pr(y_i = 1)$ que representa la probabilidad de que una observación i pertenezca a la clase base $y_i = 1$. Para aplicar el problema de clasificación binaria a la estimación de probabilidad de fuga de clientes se tiene que considerar que cada una de las observaciones representa el caso de un cliente y que el vector x_i representa las características del cliente i.

Dentro de la muestra de clientes se deben considerar clientes de dos tipos: clientes que se han fugado de la firma y clientes que no. Para identificar estas dos clases de clientes se utiliza la variable y_i, la cual toma el valor $y_i = 1$ para los clientes que siguen prefiriendo la firma y toma el valor $y_i = 0$ para los clientes que se fugan de la firma. Se supone que en el mercado sólo existen dos firmas y que los clientes siempre escogen una de las dos. De esta forma, se busca estimar $\Pr(y_i = 1)$ que representa entonces la probabilidad de que el cliente i se mantenga en la firma. La probabilidad de fuga, y de que escoja la firma rival, entonces está dada por $\Pr(y_i = 0) = 1 - \Pr(y_i = 1)$. Al modelar el problema de predecir las probabilidades de fuga de clientes como un problema de clasificación binaria, éste se puede resolver a través de un modelo de Regresión Logística: sea $\sigma_i = \Pr(y_i = 1)$, donde $\sigma_i = \sigma(b, x_i)$ es función del vector de características x_i.

$$\sigma_i = \sigma(b, x_i) = \frac{1}{1 + e^{-b^T x_i}} \tag{5.1}$$

Donde $b \in \mathbb{R}^n$ son los parámetros del modelo de Regresión Logística que se busca estimar.

5.2. Regresión Logistica con Restricciones de Competencia

5.2.1. Definición del modelo

En este trabajo se busca proponer un modelo de estimación basado en Regresión Logística pero introduciendo restricciones de acuerdo al equilibrio del juego de competencia subyacente en el proceso de clasificación binaria.

De acuerdo al capítulo 4 en la expresión (4.25), el parámetro estimado $\hat{b}$ para el modelo de Regresión Logística se obtiene a partir de la maximización del logaritmo de la función de verosimilitud $\ell(b/x)$, sin imponer restricciones a los valores que pueda tomar b. Por otra parte y de acuerdo a lo discutido en los capítulos anteriores, en los diferentes problemas que se pueden modelar a través de clasificación binaria existe un juego de competencia subyacente al problema estadístico de estimar la función de clasificación. Por ejemplo, en el problema de predicción de la probabilidad de fuga de clientes, cada cliente debe tomar la decisión de cambiarse o no de firma y cada una de las firmas del mercado debe decidir precios y acciones de marketing para atraer a los clientes que demandan sus productos y/o servicios. Esta situación de competencia se puede modelar a través de Teoría de Juegos y como resultado de este juego se pueden determinar las condiciones de equilibrio que rigen el comportamiento de los agentes.

Supongamos que se busca abordar un problema de clasificación binaria a través de un modelo de Regresión Logística y por lo tanto se desea estimar el parámetro b de la función de probabilidad. Supongamos que es posible modelar el juego de competencia asociado al problema de clasificación. El equilibrio del juego debe depender necesariamente del valor de la probabilidad $\sigma_i = \Pr(y_i = 1)$ asociada a cada observación i, pues en función de estas probabilidades los agentes tomarán sus decisiones. Esto implica que el equilibrio del juego necesariamente se traduce en un conjunto de condiciones que dependen del valor de b y de x_i pues a partir de este parámetro y de las características de los clientes se determina el valor de las probabilidades σ_i. Supongamos entonces que las condiciones de equilibrio del juego se traducen en un conjunto $\mathcal{A}(b,x) \subseteq \mathbb{R}^n$ que restringe los posibles valores que puede tomar b. Se propone resolver el problema de maximización de la función de verosimilitud del modelo de Regresión Logistica abordado en la expresión (4.25), pero acotando ahora los valores de b al subconjunto que representa las condiciones de equilibrio del juego de competencia. De esta forma, el b estimado bajo este enfoque modificado está dado por la siguiente expresión.

$$\tilde{b} = \underset{b \in \mathcal{A}(b,x)}{\arg\max} \{\ell(b/x)\} \tag{5.2}$$

Supongamos adicionalmente que el conjunto $\mathcal{A}(b,x) \subseteq \mathbb{R}^n$ se puede traducir en un número determinado de ecuaciones que reproducen este conjunto. Si se supone que los agentes que participan del juego de competencia interactúan a nivel de cada observación, entonces existirán restricciones a nivel de cada observación $i \in \{1,...,I\}$. Sea $g_{im}(b,x) = 0$ la m-ésima condición de equilibrio sobre la observación i, con $m \in \{1,...,M\}$. De esta forma, la solución de $\tilde{b}$ obtenida de acuerdo a la fórmula (5.2) se puede obtener de forma equivalente a través de resolver el siguiente problema de optimización.

$$
\begin{array}{lll}
(P) & \text{máx} & \ell(b/x) \\
 & s.a. & g_{im}(b,x) = 0 \quad \forall i \in \{1,...,I\} \; \forall m \in \{1,...,M\} \\
 & & b \in \mathbb{R}^n
\end{array}
\tag{5.3}
$$

Este problema tiene solución si se satisfacen las condiciones de equilibrio sobre cada una de las observaciones de la muestra estudiada. En una aplicación a un caso real, esta situación puede ser muy exigente y por lo tanto el problema de optimización se puede volver infactible. Dado que ningún método estadístico pretende encontrar una función de probabilidad que reproduzca las clases de cada observación sin error, tampoco parece adecuado plantear un modelo que no permita que exista cierto nivel de error o de desviación con respecto al equilibrio del juego. Por otro lado, parece natural permitir desviaciones con respecto al equilibrio del juego porque es muy difícil encontrar una situación real donde todas las observaciones se adecúen de forma perfecta a las condiciones de equilibrio. Una alternativa para permitir desviaciones, y buscar una solución que las disminuya lo más posible, es a través de incorporar penalizaciones en la función objetivo del problema de optimización (P) para cada una de las restricciones del mismo. Por ejemplo si a la función objetivo del problema (P) se le resta $(g_{im}(b,x))^2$ entonces para máximizar esta nueva función objetivo necesariamente se tendrá que buscar un b que además de hacer crecer $\ell(b/x)$ haga que $(g_{im}(b,x))^2$ se acerque lo más posible a cero, lo cual haría cumplir una de las condiciones de equilibrio.

De esta forma, el estimador $\tilde{b}$ propuesto por el modelo planteado es el que resuelve el problema de maximizar el logaritmo de la función de verosimilitud, de una distribución de probabilidad logística, penalizada por las condiciones de equilibrio del juego de competencia subyacente en el problema de clasificación binaria estudiado. Expresando lo anterior en términos formales, en el modelo propuesto se plantea utilizar el siguiente estimador de b.

$$
\tilde{b} = \underset{b \in \mathbb{R}^n}{\arg\max} \left\{ \ell(b/x) + \mu \sum_{i=1}^{I} \sum_{m=1}^{M} (g_{im}(b,x))^2 \right\} \tag{5.4}
$$

Donde $\mu < 0$ es un parámetro que pondera el peso relativo del cumplimiento de las condiciones de equilibrio con respecto a la maximización de la función de verosimilitud.

Es importante destacar que el estimador $\hat{b}$ −asociado a la estimación clásica de la regresión logística− domina al estimador $\tilde{b}$ −asociado al modelo propuesto en este trabajo− en términos de eficiencia en probabilidad en la muestra de calibración $\{(x_i, y_i)\}_{i=1}^{I}$. Esto pues en el segundo caso, la existencia de restricciones hace que se busque los b's candidatos sobre un subconjunto de las alternativas de solución que existen en el primer caso, por lo que en el mejor de los casos este estimador igualará al primero en eficiencia en probabilidad. Dado lo anterior, necesariamente se debe cumplir que $\ell(\hat{b}/x) \geq \ell(\tilde{b}/x)$.

La motivación del modelo propuesto no tiene que ver con la búsqueda de un estimador de b que tenga un menor error dentro de la muestra de calibración sino que se espera que, al incorporar las condiciones de competencia del juego subyacente, pueda tener un mejor comportamiento fuera de la muestra de calibración. Esto es de suma importancia cuando este tipo de modelos son utilizados para predecir sobre nuevas observaciones con una clase y_i todavia desconocida. Por ejemplo, las firmas que estiman este tipo de modelos para calcular probabilidades de fuga de clientes, estiman un b a partir de una muestra histórica de clientes para

los cuales ya conocen su desenlace: saben si se fugaron o no. Luego, el modelo obtenido es aplicado sobre las carteras de clientes vigentes para estimar las probabilidades de fuga y tomar decisiones de marketing tempranas sobre cada uno de ellos. Si el modelo obtenido está sobreajustado a la muestra de calibración, los resultados en la aplicación del modelo sobre las carteras vigentes no será satisfactorio. Con el tipo de modelo planteado en este capítulo se espera que, al considerar las condiciones de equilibrio de juego subyacente, se consiga un modelo que pueda adaptarse mejor a los cambios de comportamiento promedio de las nuevas observaciones que no fueron parte de la estimación del modelo.

5.2.2. Solución del modelo

El estimador $\tilde{b}$ se obtiene a partir de la solución del problema de optimización no lineal sin restricciones propuesto en (5.4). Este tipo de problemas se pueden resolver a través de la solución del sistema de ecuaciones no lineales que proponen las condiciones de primer orden. Las condiciones de primer orden de un problema de optimización se obtienen cuando se igualan a cero las derivadas de la función objetivo con respecto a cada una de las variables de decisión. En este caso las derivadas de la función $\tilde{\ell}(b/x)$ con respecto a cada componente b_k del vector b con $k \in \{1,...,n\}$, donde $\tilde{\ell}(b/x)$ se define como la función objetivo penalizada propuesta a continuación.

$$\tilde{\ell}(b/x) = \sum_{i=1}^{I} \left(y_i \ln(\sigma_i) + (1-y_i)\ln(1-\sigma_i)\right) + \mu \sum_{i=1}^{I}\sum_{m=1}^{M} \left(g_{im}(b,x)\right)^2 \tag{5.5}$$

De esta forma, se puede resolver el problema propuesto en (5.4) encontrando el parametro b que resuelve el siguiente sistema de n ecuaciones no lineales.

$$\nabla_b \tilde{\ell}(b/x) = 0 \quad \Leftrightarrow \quad \frac{\partial \tilde{\ell}(b/x)}{\partial b_k} = 0 \quad \forall k \tag{5.6}$$

A partir de la definición de $\tilde{\ell}(b/x)$ el sistema de ecuaciones anterior se puede reescribir de la siguiente forma.

$$\sum_{i=1}^{I}(y_i - \sigma_i)x_{ik} + 2\mu \sum_{i=1}^{I}\sum_{m=1}^{M}\left(g_{im}(b,x)\frac{\partial g_{im}(b,x)}{\partial b_k}\right) = 0 \quad \forall k \tag{5.7}$$

Si se supone que los vectores de caracteristicas de cada observación x_i son conocidos, entonces las incognicas sólo son cada uno de los n parámetros del vector b. A partir de las n ecuaciones anteriores entonces es posible obtener una solución $\tilde{b}$ del sistema de ecuaciones no lineales y por lo tanto del problema de optimización planteado en (5.4).

Dado que el sistema de ecuaciones es no lineal, en la práctica para poder resolver este tipo problemas es necesario recurrir a métodos numéricos que convergen a una solución $\tilde{b}$.

Tal como se hace para resolver el modelo de Regresión Logística en el capítulo 4, se utilizará el método de Newton-Raphson para resolver este problema de forma iterativa. La función objetivo en este caso es la función $\tilde{\ell}(b/x)$. La variable de decisión en este caso es el vector b. De esta forma, a partir de una solución inicial $b^{(0)}$ es posible construir la recurrencia de Newton-Raphson.

$$b^{(q+1)} = b^{(q)} - \mathbf{H}_{\tilde{\ell}}^{-1}(b^{(q)})\nabla_b \tilde{\ell}(b^{(q)}) \tag{5.8}$$

Cuando se cumple la condición de parada en la Q-ésima iteración entonces el valor $b^{(Q)}$ corresponde a la solución óptima del problema (5.4) y determina el estimador de los parámetros del modelo de Regresión Logística: $\tilde{b} = b^{(Q)}$.

Notemos que para resolver el problema se requiere conocer el gradiente $\nabla_b \tilde{\ell}(b)$ y el hessiano $\mathbf{H}_{\tilde{\ell}}(b)$. El gradiente en este caso es el vector de dimensión n, compuesto por cada una de las derivadas parciales de la función $\tilde{\ell}(b/x)$ con respecto a cada componente del vector b. La expresión para cada componente del gradiente ya fue mostrada en (5.7) como parte de las condiciones de primer orden del problema.

$$\frac{\partial \tilde{\ell}(b/x)}{\partial b_k} = \sum_{i=1}^{I}(y_i - \sigma_i)x_{ik} + 2\mu \sum_{i=1}^{I}\sum_{k=1}^{M}\left(g_{im}(b,x)\,\frac{\partial g_{im}(b,x)}{\partial b_k}\right) \quad \forall k \tag{5.9}$$

El hessiano de $\tilde{\ell}(b/x)$ es una matriz de dimensión $n \times n$ compuesta por las segundas derivadas parciales de la función con respecto a los componentes del vector b. El componente (j,k) de esta matriz está dado por la siguiente expresión.

$$\mathbf{H}_{\tilde{\ell}}(b/x)_{kj} = \frac{\partial^2 \tilde{\ell}(b/x)}{\partial b_j \partial b_k} \quad \forall j,k \in \{1,...,n\} \tag{5.10}$$

A partir de la definición de $\tilde{\ell}(b/x)$ podemos reescribir estos términos de la siguiente forma.

$$\frac{\partial^2 \tilde{\ell}(b/x)}{\partial b_j \partial b_k} = -\sum_{i=1}^{I}\sigma_i(1-\sigma_i)\,x_{ik}x_{ij} + 2\mu \sum_{i=1}^{I}\sum_{k=1}^{M}\left(\frac{\partial g_{im}}{\partial b_j}\,\frac{\partial g_{im}}{\partial b_k} + g_{im}\,\frac{\partial^2 g_{im}}{\partial b_j \partial b_k}\right) \quad \forall j,k \tag{5.11}$$

El modelo planteado hasta ahora considera las restricciones de equilibrio de un juego genérico. La aplicación de la metodología propuesta requiere la definición de un juego específico para conocer la definición particular de sus condiciones de equilibrio. Para poder resolver el modelo se requiere entonces poder caracterizar el gradiente de la función objetivo y su hessiano. En las siguientes secciones se aborda el problema de estimación de la probabilidad de fuga de clientes bajo el esquema propuesto en esta sección pero a través de tres aproximaciones diferentes de juegos de competencia. En la primera se considera que las firmas compiten e intentan atraer a los clientes sólo a través del precio como herramienta competitiva. En el segundo caso se consdiera que además del precio, las firmas pueden realizar descuentos en el precio a cada cliente. En el tercer caso se considera que además del precio y los descuentos, las firmas puede realizar acciones de marketing

específicas sobre los clientes para atraerlos. Para cada aproximación se define el juego de competencia, estableciendo las condiciones de equilibrio del juego, construyendo la función $\tilde{\ell}(b/x)$ y luego caracterizando su gradiente y hessiano para poder posteriormente utilizar el método Newton-Raphson para su resolución.

5.3. Predicción de Fuga de Clientes con Restricciones de Competencia en Precios

5.3.1. Definición del Problema

En esta sección se presenta un modelo bajo el esquema propuesto en la sección 5.2, para estimar la probabilidad de fuga de clientes donde se considera que las restricciones de equilibrio provienen de un juego donde las firmas compiten sólo en precio para intentar atraer y retener a los clientes del mercado. El objetivo es encontrar un estimador de b a partir del enfoque de regresión logística con restricciones de equilibrio de un juego de competencia.

En la interacción entre cada cliente y las firmas del mercado existe un juego de competencia. Cada firma del mercado intenta atraer o retener a los clientes, buscando un mayor nivel de participación de mercado y por lo tanto mayores utilidades. Los clientes escogerán la firma que le entregue mayores utilidades considerando los niveles de atracción o de lealtad que sienten hacia cada una de ellas.

Supongamos un mercado de dos firmas $j \in \{A,B\}$. Supongamos que la única herramienta competitiva con la que cuentan las firmas es el precio. Cada firma $j \in \{A,B\}$ decide el precio p_{ji} que cobrará a cada cliente $i \in \{1,...,I\}$. Supongamos que los clientes se comportan de acuerdo a un modelo de elección discreta con dos alternativas de elección: firma A o firma B. Esto permite caracterizar la probabilidad de elección de acuerdo a la forma funcional de un modelo de Regresión Lógística, como se plantea en la sección 4.4.2. Se define que la clase base $y_i = 1$ corresponde a la firma A y la clase $y_i = 0$ a la firma B. De esta forma $\sigma_i = 1/(1+\exp(-V(b,x_i))$ corresponde a la probabilidad de elección de la firma A y $1-\sigma_i$ la probabilidad de elección de la firma B. Para los clientes pertenecientes a la firma A, la probabilidad de fuga está dada por $1-\sigma_i$; y para los clientes pertenecientes a la firma B, la probabilidad de fuga está dada por σ_i. Si la base de observaciones $\{(x_i,y_i)\}_{i=1}^{I}$ contiene sólo clientes que incialmente pertenecen a la firma A entonces el esquema anterior es equivalente al planteado en la sección 5.1 para estimar las probabilidades de fuga de los clientes de una firma (en este caso la firma A) utilizando Regresión Logística.

Se supone que la función $V(b,x_i)$ es lineal y que el vector x_i con los atributos que caracterizan la decisión de cada cliente incluyen una constante y los precios que debe pagar en ambas firmas.

$$V(b,x_i) = \alpha + \beta_A\, p_{Ai} + \beta_B\, p_{Bi} + \gamma^T\, \bar{x}_i \tag{5.12}$$

El vector de atributos del cliente i entonces corresponde a un vector con los componentes anteriores: $x_i = \langle 1, p_{Ai}, p_{Bi}, \bar{x}_i \rangle$, donde $\bar{x}_i \in \mathbb{R}^{n-3}$ es un vector que contiene los atributos adicionales a la constante y los

precios. De acuerdo a esta definición el vector de parámetros b tiene los siguientes componentes: $b = \langle \alpha, \beta_A, \beta_B, \gamma \rangle$, donde $\gamma \in \mathbb{R}^{n-3}$ es el vector de parámetros relacionados a los atributos $\bar{x}_i$. Se utilizará la notación x_{ik} para referirse al k-ésimo atributo del cliente i, con $k \in \{1, ..., n\}$. Notar que $x_{i1} = 1$, $x_{i2} = p_{Ai}$ y $x_{i3} = p_{Bi}$.

A partir de los precios que deciden las firmas y los otros atributos de los clientes, se determinan las probabilidades de elección de los clientes. Supongamos que los precios ya tienen descontados los costos marginales del producto y/o servicio. De esta forma las utilidades esperadas que reciben las firmas de acuerdo a las probabilidades de elección son las siguientes.

$$\Pi_A(b,x) = \sum_{i=1}^{I} p_{Ai}\,\sigma_i(b,x_i) \tag{5.13}$$

$$\Pi_B(b,x) = \sum_{i=1}^{I} p_{Bi}\,(1-\sigma_i(b,x_i)) \tag{5.14}$$

Las firmas simultáneamente deciden los precios que maximizan sus utilidades. Las condiciones de primer orden sobre las utilidades de ambas firmas son las siguientes.

$$\frac{\partial \Pi_A(b,x)}{\partial p_{Ai}} = \sigma_i + \beta_A\, p_{Ai}\,\sigma_i\,(1-\sigma_i) = 0 \quad \forall i \tag{5.15}$$

$$\frac{\partial \Pi_B(b,x)}{\partial p_{Bi}} = (1-\sigma_i) - \beta_B\, p_{Bi}\,\sigma_i\,(1-\sigma_i) = 0 \quad \forall i \tag{5.16}$$

Donde $\sigma_i = \sigma_i(b,x)$. Para llegar a las expresiones anteriores se utiliza el siguiente resultado.

$$\frac{\partial \sigma_i(b,x)}{\partial x_{ik}} = b_k\,\sigma\,(1-\sigma) \tag{5.17}$$

Esto significa entonces que $\frac{\partial \sigma_i}{\partial p_{Ai}} = \beta_A\,\sigma\,(1-\sigma)$ y $\frac{\partial (1-\sigma_i)}{\partial p_{Bi}} = -\beta_B\,\sigma\,(1-\sigma)$.

Las condiciones de primer orden en (5.15) y (5.16) representan las condiciones de equilibrio del juego de competencia en precios recién planteado. Supongamos que se conocen los atributos x_i que caracterizan la decisión del cliente i (vector que incluye los precios p_{Ai} y p_{Bi}). Para encontrar entonces un estimador de b bajo el enfoque de regresión logística con restricciones de equilibrio propuesto en la sección anterior, es necesario plantear el siguiente modelo de optimización que busca el vector b que maximiza la función de verosimilitud $\ell(b/x)$ y que satisface las condiciones de equilibrio recién planteadas.

$$
\begin{array}{lll}
(P_1) & \text{máx} & \ell(b/x) = \sum_{i=1}^{I} \left(y_i \ln(\sigma_i) + (1-y_i)\ln(1-\sigma_i)\right) \\
& s.a. & \sigma_i + \beta_A\, p_{Ai}\, \sigma_i\,(1-\sigma_i) = 0 \quad \forall i \\
& & (1-\sigma_i) + \beta_B\, p_{Bi}\, \sigma_i\,(1-\sigma_i) = 0 \quad \forall i \\
& & b \in \mathbb{R}^n
\end{array}
\tag{5.18}
$$

Para resolver este problema se modifica la función objetivo incorporando las restricciones del problema como penalizaciones, de acuerdo a lo propuesto en la expresión (5.4). Se busca el $\tilde{b}$ que resuelve el siguiente problema de optimización sin restricciones.

$$
\tilde{b} = \underset{b \in \mathbb{R}^n}{\arg\max} \left\{ \tilde{\ell}(b/x) \right\} \tag{5.19}
$$

La función $\tilde{\ell}(b/x)$ corresponde a la función objetivo del problema (P_1) penalizada por sus restricciones.

$$
\begin{array}{rcl}
\tilde{\ell}(b/x) & = & \sum_{i=1}^{I} \left(y_i \ln(\sigma_i) + (1-y_i)\ln(1-\sigma_i)\right) + \mu \sum_{i=1}^{I} \Big\{ \left[\sigma_i + \beta_A\, p_{Ai}\, \sigma_i\,(1-\sigma_i)\right]^2 \\
& & + \left[(1-\sigma_i) + \beta_B\, p_{Bi}\, \sigma_i\,(1-\sigma_i)\right]^2 \Big\}
\end{array}
\tag{5.20}
$$

Donde $\mu < 0$ es el parámetro de penalización. El primer término de la expresión anterior corresponde a la función de verosimilitud $\ell(b/x)$ y los siguientes son las expresiones de cada restricción del problema (P_1) al cuadrado. En el problema no se busca el cumplimiento cabal de todas las restricciones de equilibrio, sino asegurar a través de las penalizaciones que la mayor cantidad de condiciones de equilibrio se satisfagan.

5.3.2. Resolución del Problema

De acuerdo al esquema de solución planteado en la sección 5.2.2 es posible encontrar el estimador $\tilde{b}$, solución del problema (5.19), a través del método iterativo Newton-Raphson. Para poder aplicar este método es necesario caracterizar el gradiente y el hessiano de la función objetivo del problema en cuestión $\tilde{\ell}(b/x)$. Con esto es posible utilizar el método de Newton-Rapshon para resolver el problema a través de la recurrencia ya definida anteriormente.

$$
b^{(q+1)} = b^{(q)} - \mathbf{H}_{\tilde{\ell}}^{-1}(b^{(q)}) \nabla_b \tilde{\ell}(b^{(q)}) \tag{5.21}
$$

En el Anexo E.1 se detallan los cálculo que se deben realizar para llegar a una expresión del gradiente y el hessiano de la función $\tilde{\ell}(b/x)$ asociada a este modelo. En el Anexo E.1.1 se detallan las expresiones para determinar este gradiente y en el Anexo E.1.2 para determinar este hessiano.

5.4. Predicción de Fuga de Clientes con Restricciones de Competencia en Precios y Descuentos en Precios

5.4.1. Definición del Problema

Consideremos el problema de competencia en precios planteado en la sección 5.3, pero supongamos que las firmas ya no sólo compiten en precio, sino que también pueden ofrecer a cada uno de los clientes del mercado un descuento en el precio. En este caso se está incorporando el descuento en precio como una nueva herramienta competitiva para las firmas.

Bajo la misma nomenclatura planteada en la sección 5.3, supongamos que la firma $j \in \{A,B\}$ decide el precio p_{ji} que cobrará a cada cliente $i \in \{1,...,I\}$, pero ahora también decide el descuento d_{ji} que ofrecerá como estrategia de atracción o retención. Supongamos que los clientes se comportan de acuerdo a un modelo de elección discreta con dos alternativas de elección: firma A o firma B; con probabilidades de elección σ_i y $1-\sigma_i$ respectivamente. Al igual que en la sección 5.3, estas probabilidades se pueden estimar a partir de un modelo de Regresión Logística utilizando la forma funcional de la expresión (4.26).

Se supone que la función $V(b,x_i)$ es lineal y que el vector x_i con los atributos que caracterizan la decisión de cada cliente incluyen una constante, los precios y los descuentos que las firmas ofrecen.

$$V(b,x_i) = \alpha + \beta_A\, p_{Ai} + \beta_B\, p_{Bi} + \theta_A\, d_{Ai} + \theta_B\, d_{Bi} + \gamma^T\, \bar{x}_i \tag{5.22}$$

El vector de atributos del cliente i entonces corresponde a un vector con los componentes anteriores: $x_i = \langle 1, p_{Ai}, p_{Bi}, d_{Ai}, d_{Bi}, \bar{x}_i \rangle$, donde $\bar{x}_i \in \mathbb{R}^{n-5}$ es un vector que contiene los atributos adicionales a la constante, los precios y los descuentos. De acuerdo a esta definición el vector de parámetros b tiene los siguientes componentes: $b = \langle \alpha, \beta_A, \beta_B, \theta_A, \theta_B, \gamma \rangle$, donde $\gamma \in \mathbb{R}^{n-5}$ es el vector de parámetros relacionados a los atributos $\bar{x}_i$.

A partir de los precios y descuentos que deciden las firmas y los otros atributos de los clientes, se determinan las probabilidades de elección de los clientes. Supongamos que los precios ya tienen descontados los costos marginales del producto y/o servicio. De esta forma las utilidades esperadas que reciben las firmas de acuerdo a las probabilidades de elección son las siguientes.

$$\Pi_A(b,x) = \sum_{i=1}^{I} (p_{Ai} - d_{Ai})\, \sigma_i(b,x_i) \tag{5.23}$$

$$\Pi_B(b,x) = \sum_{i=1}^{I} (p_{Bi} - d_{Bi})\, (1 - \sigma_i(b,x_i)) \tag{5.24}$$

Las firmas simultáneamente deciden los precios, descuentos que maximizan sus utilidades. Las condiciones de primer orden sobre las utilidades de ambas firmas son las siguientes.

$$\frac{\partial \Pi_A(b,x)}{\partial p_{Ai}} = \sigma_i + \beta_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \tag{5.25}$$

$$\frac{\partial \Pi_A(b,x)}{\partial d_{Ai}} = -\sigma_i + \theta_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \tag{5.26}$$

$$\frac{\partial \Pi_B(b,x)}{\partial p_{Bi}} = (1-\sigma_i) - \beta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \tag{5.27}$$

$$\frac{\partial \Pi_B(b,x)}{\partial p_{Bi}} = (1-\sigma_i) + \theta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \tag{5.28}$$

Las condiciones de primer orden planteadas en las ecuaciones de la (5.25) hasta la (5.28), representan las condiciones de equilibrio del juego de competencia en precios y descuentos recién planteado. Para encontrar el estimador de b bajo el enfoque de regresión logística con restricciones de equilibrio propuesto, es necesario plantear el siguiente modelo de optimización que busca el vector b que maximiza la función de verosimilitud $\ell(b/x)$ y que satisface las condiciones de equilibrio recién planteadas.

$$\begin{array}{lll} (P_2) & \text{máx} & \ell(b/x) = \sum_{i=1}^{I} (y_i \ln(\sigma_i) + (1-y_i)\ln(1-\sigma_i)) \\ & s.a. & \sigma_i + \beta_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \\ & & -\sigma_i + \theta_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \\ & & (1-\sigma_i) - \beta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \\ & & (1-\sigma_i) + \theta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \\ & & b \in \mathbb{R}^n \end{array} \tag{5.29}$$

Para resolver este problema se modifica la función objetivo incorporando las restricciones del problema como penalizaciones, de acuerdo a lo propuesto en la expresión (5.4). Se busca el $\tilde{b}$ que resuelve el siguiente problema de optimización sin restricciones.

$$\tilde{b} = \underset{b \in \mathbb{R}^n}{\arg\max} \left\{ \tilde{\ell}(b/x) \right\} \tag{5.30}$$

La función $\tilde{\ell}(b/x)$ corresponde a la función objetivo del problema (P_2) penalizada por sus restricciones.

$$\begin{aligned} \tilde{\ell}(b/x) = & \sum_{i=1}^{I} (y_i \ln(\sigma_i) + (1-y_i)\ln(1-\sigma_i)) + \mu \sum_{i=1}^{I} \Big\{ [\sigma_i + \beta_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i)]^2 \\ & + [(1-\sigma_i) - \beta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i)]^2 + [-\sigma_i + \theta_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i)]^2 \\ & + [(1-\sigma_i) + \theta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i)]^2 \Big\} \end{aligned} \tag{5.31}$$

Donde $\mu < 0$ es el parámetro de penalización. El primer término de la expresión anterior corresponde a la función de verosimilitud $\ell(b/x)$ y los siguientes son las expresiones de cada restricción del problema (P_2) al cuadrado. En el problema no se busca el cumplimiento cabal de todas las restricciones de equilibrio, sino asegurar a través de las penalizaciones que la mayor cantidad de condiciones de equilibrio se satisfagan.

5.4.2. Resolución del Problema

De acuerdo al esquema de solución planteado en la sección 5.2.2 es posible encontrar el estimador $\tilde{b}$, solución del problema 5.30, a través del método iterativo Newton-Raphson. Para poder aplicar este método es necesario caracterizar el gradiente y el hessiano de la función objetivo del problema en cuestión $\tilde{\ell}(b/x)$. Con esto es posible utilizar el método de Newton-Rapshon para resolver el problema a través de la recurrencia ya definida anteriormente.

$$b^{(q+1)} = b^{(q)} - \mathbf{H}_{\tilde{\ell}}^{-1}(b^{(q)})\nabla_b \tilde{\ell}(b^{(q)}) \tag{5.32}$$

En el Anexo E.2 se detallan los cálculo que se deben realizar para llegar a una expresión del gradiente y el hessiano de la función $\tilde{\ell}(b/x)$ asociada a este modelo. En el Anexo E.2.1 se detallan las expresiones para determinar este gradiente y en el Anexo E.2.2 para determinar este hessiano.

5.5. Predicción de Fuga de Clientes con Restricciones de Competencia en Precios, Descuentos en Precios y Esfuerzos en Marketing

5.5.1. Definición del Problema

Se busca abordar el problema de estimar la probabilidad de fuga de clientes. Consideremos el problema de competencia planteado en la sección 5.4, pero supongamos que las firmas ya no sólo compiten en precios y descuentos, sino que también pueden realizar esfuerzos de marketing sobre cada uno de los clientes del mercado. En este caso se está incorporando el esfuerzo en marketing como una nueva herramienta competitiva para las firmas.

Bajo la misma nomenclatura planteada en la sección 5.4, supongamos que la firma $j \in \{A, B\}$ decide el precio p_{ji} y el descuento d_{ji} para cada cliente $i \in \{1, ..., I\}$, pero ahora también decide el esfuerzo en marketing x_{ji} que realizará sobre cada cliente como estrategia de atracción o retención. Supongamos que los clientes se comportan de acuerdo a un modelo de elección discreta con dos alternativas de elección: firma A o firma B; con probabilidades de elección σ_i y $1 - \sigma_i$ respectivamente. Al igual que en las secciones 5.3 y 5.4, estas probabilidades se pueden estimar a partir de un modelo de Regresión Logística utilizando la forma funcional de la expresión (4.26).

Se supone que la función $V(b,x_i)$ es lineal y que el vector x_i con los atributos que caracterizan la decisión de cada cliente incluyen una constante, los precios, los descuentos y los esfuerzos en marketing que las firmas deciden.

$$V(b,x_i) = \alpha + \beta_A\, p_{Ai} + \beta_B\, p_{Bi} + \theta_A\, d_{Ai} + \theta_B\, d_{Bi} + \phi_A\, m_{Ai} + \phi_B\, m_{Bi} + \gamma^T\, \bar{x}_i \tag{5.33}$$

El vector de atributos del cliente i entonces corresponde a un vector con los componentes anteriores: $x_i = \langle 1, p_{Ai}, p_{Bi}, d_{Ai}, d_{Bi}, m_{Ai}, m_{Bi}, \bar{x}_i \rangle$, donde $\bar{x}_i \in \mathbb{R}^{n-7}$ es un vector que contiene los atributos adicionales a la constante, los precios, los descuentos y los esfuerzos en marketing. De acuerdo a esta definición el vector de parámetros b tiene los siguientes componentes: $b = \langle \alpha, \beta_A, \beta_B, \theta_A, \theta_B, \phi_A, \phi_B, \gamma \rangle$, donde $\gamma \in \mathbb{R}^{n-7}$ es el vector de parámetros relacionados a los atributos $\bar{x}_i$.

A partir de los precios, descuentos y esfuerzos en marketing que deciden las firmas y los otros atributos de los clientes, se determinan las probabilidades de elección de los clientes. Supongamos que los precios ya tienen descontados los costos marginales del producto y/o servicio. Se supone además que el costo marginal de los esfuerzos en marketing están dados por la constante e para ambas firmas. Por ejemplo si una firma realiza llamadas por teléfono a sus clientes, la constante e representa el costo que tiene para las firmas cada llamada que se realiza. De esta forma las utilidades esperadas que reciben las firmas de acuerdo a las probabilidades de elección son las siguientes.

$$\Pi_A(b,x) = \sum_{i=1}^{I} \left[(p_{Ai} - d_{Ai})\, \sigma_i(b,x_i) - e\, m_{Ai} \right] \tag{5.34}$$

$$\Pi_B(b,x) = \sum_{i=1}^{I} \left[(p_{Bi} - d_{Bi})\, (1 - \sigma_i(b,x_i)) - e\, m_{Bi} \right] \tag{5.35}$$

Las firmas simultáneamente deciden los precios, descuentos y esfuerzos en marketing que maximizan sus utilidades. Las condiciones de primer orden sobre las utilidades de ambas firmas son las siguientes.

$$\frac{\partial \Pi_A(b,x)}{\partial p_{Ai}} = \sigma_i + \beta_A\, (p_{Ai} - d_{Ai})\, \sigma_i\, (1 - \sigma_i) = 0 \quad \forall i \tag{5.36}$$

$$\frac{\partial \Pi_A(b,x)}{\partial d_{Ai}} = -\sigma_i + \theta_A\, (p_{Ai} - d_{Ai})\, \sigma_i\, (1 - \sigma_i) = 0 \quad \forall i \tag{5.37}$$

$$\frac{\partial \Pi_A(b,x)}{\partial m_{Ai}} = \phi_A\, (p_{Ai} - d_{Ai})\, \sigma_i\, (1 - \sigma_i) - e = 0 \quad \forall i \tag{5.38}$$

$$\frac{\partial \Pi_B(b,x)}{\partial p_{Bi}} = (1-\sigma_i) - \beta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \tag{5.39}$$

$$\frac{\partial \Pi_B(b,x)}{\partial p_{Bi}} = (1-\sigma_i) + \theta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \tag{5.40}$$

$$\frac{\partial \Pi_A(b,x)}{\partial m_{Ai}} = \phi_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i) + e = 0 \quad \forall i \tag{5.41}$$

Las condiciones de primer orden planteadas en las ecuaciones de la (5.36) hasta la (5.41), representan las condiciones de equilibrio del juego de competencia en precios, descuentos y marketing recién planteado. Para encontrar el estimador de b bajo el enfoque de regresión logística con restricciones de equilibrio propuesto, es necesario plantear el siguiente modelo de optimización que busca el vector b que maximiza la función de verosimilitud $\ell(b/x)$ y que satisface las condiciones de equilibrio recién planteadas.

$$\begin{array}{lll} (P_3) & \text{máx} & \ell(b/x) = \sum_{i=1}^{I} (y_i \ln(\sigma_i) + (1-y_i)\ln(1-\sigma_i)) \\ & s.a. & \sigma_i + \beta_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \\ & & -\sigma_i + \theta_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \\ & & \phi_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i) - e = 0 \quad \forall i \\ & & (1-\sigma_i) - \beta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \\ & & (1-\sigma_i) + \theta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i) = 0 \quad \forall i \\ & & \phi_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i) + e = 0 \quad \forall i \\ & & b \in \mathbb{R}^n \end{array} \tag{5.42}$$

Para resolver este problema se modifica la función objetivo incorporando las restricciones del problema como penalizaciones, de acuerdo a lo propuesto en la expresión (5.4). Se busca el $\tilde{b}$ que resuelve el siguiente problema de optimización sin restricciones.

$$\tilde{b} = \underset{b \in \mathbb{R}^n}{\arg\max} \left\{ \tilde{\ell}(b/x) \right\} \tag{5.43}$$

La función $\tilde{\ell}(b/x)$ corresponde a la función objetivo del problema (P_3) penalizada por sus restricciones.

$$\begin{aligned} \tilde{\ell}(b/x) = & \sum_{i=1}^{I} (y_i \ln(\sigma_i) + (1-y_i)\ln(1-\sigma_i)) + \mu \sum_{i=1}^{I} \Big\{ [\sigma_i + \beta_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i)]^2 \\ & + [(1-\sigma_i) - \beta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i)]^2 + [-\sigma_i + \theta_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i)]^2 \\ & + [(1-\sigma_i) + \theta_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i)]^2 + [\phi_A (p_{Ai} - d_{Ai}) \sigma_i (1-\sigma_i) - e]^2 \\ & + [\phi_B (p_{Bi} - d_{Bi}) \sigma_i (1-\sigma_i) + e]^2 \Big\} \end{aligned} \tag{5.44}$$

Donde $\mu < 0$ es el parámetro de penalización. El primer término de la expresión anterior corresponde a la función de verosimilitud $\ell(b/x)$ y los siguientes son las expresiones de cada restricción del problema (P_3)

al cuadrado. En el problema no se busca el cumplimiento cabal de todas las restricciones de equilibrio, sino asegurar a través de las penalizaciones que la mayor cantidad de condiciones de equilibrio se satisfagan.

5.5.2. Resolución del Problema

De acuerdo al esquema de solución planteado en la sección 5.2.2 es posible encontrar el estimador $\tilde{b}$, solución del problema (5.43), a través del método iterativo Newton-Raphson. Para poder aplicar este método es necesario caracterizar el gradiente y el hessiano de la función objetivo del problema en cuestión $\tilde{\ell}(b/x)$. Con esto es posible utilizar el método de Newton-Rapshon para resolver el problema a través de la recurrencia ya definida anteriormente.

$$b^{(q+1)} = b^{(q)} - \mathbf{H}_{\tilde{\ell}}^{-1}(b^{(q)})\nabla_b \tilde{\ell}(b^{(q)}) \tag{5.45}$$

En el Anexo E.3 se detallan los cálculo que se deben realizar para llegar a una expresión del gradiente y el hessiano de la función $\tilde{\ell}(b/x)$ asociada a este modelo. En el Anexo E.3.1 se detallan las expresiones para determinar este gradiente y en el Anexo E.3.2 para determinar este hessiano.

5.6. Implementación Computacional de las Metodologías Propuestas

Para resolver los modelos de Regresión Logística con Restricciones de Competencia propuestos en este capítulo, se realiza una implementación propia utilizando el lenguaje del paquete estadístico R (R Development Core Team, 2008). En el Anexo F se presenta la estrategia de implementación de las diferentes metodologías utilizadas en esta Tesis a través del lenguaje R. En el Anexo F.2.2 se describe la implementación del modelo propuesto con restricciones de competencia en precios. En el Anexo F.2.3 se describe la implementación del modelo propuesto con restricciones de competencia en precios y descuentos en precios. En el Anexo F.2.4 se describe la implementación del modelo propuesto con restricciones de competencia en precios, descuentos en precios y esfuerzos en marketing.

Capítulo 6

Predicción de Fuga de Clientes con Restricciones de Competencia: Caso Aplicado

6.1. Base de datos

6.1.1. Descripción y variables disponibles

En este capítulo se presentan los resultados de la implementación en un caso aplicado de los modelos propuestos en el capítulo 5. El objetivo es poder concluir sobre la bondad de estos modelos y contrastarlos con los resultados obtenidos al utilizar la Regresión Logística tradicional.

En esta aplicación se utilizan datos de escáner de un panel de consumidores elaborado por la empresa A. C. Nielsen y que ha sido publicado[1] por el *Kilts Center for Marketing* de la *University of Chicago Booth School of Business*. Estos datos detallan las compras del panel de consumidores en un mercado de ketchups. El panel de consumidores está cubierto por hogares de dos ciudades de Estados Unidos: Sioux Falls, de Dakota del Sur; y Springfield, de Misouri. Estas ciudades fueron escogidas por la empresa porque eran buenas representantes la realidad demográfica de Estados Unidos al momento del estudio.

Los datos consideran una historia desde Enero de 1986 hasta Agosto de 1988. En la base de datos se detallan cada una de las compras registradas durante el periodo de estudio. Para cada compra se identifica la fecha de compra, la marca del Ketchup comprado y el hogar que realiza la compra.

Para cada compra se cuenta con una serie de atributos asociados a elementos competitivos al momento de la transacción: el precio de compra, las unidades compradas, el valor de los descuentos asociados a la compra

[1] http://research.chicagobooth.edu/marketing/databases/erim/index.aspx [Consulta: Octubre, 2012]

(descuentos de la tienda y del fabricante), variables que describen la situación de exhibición del producto dentro de la tienda (*display*), variables que indican si existe alguna campaña publicitaria asociada al producto (*advertising*), variables que indican el tipo de cupones de descuentos que existen y una variable que indica si el precio del producto responde a alguna situación especial. La descripción de estas variables se detalla en el Anexo H.1.

Adicionalmente se cuenta con más de cuarenta atributos de tipo demográfico que permiten caracterizar a los diferentes hogares que son parte de la base de datos. Entre las variables disponibles se encuentran por ejemplo: el tipo de vivienda, si es propia o arrendada y el tiempo que la familia lleva habitándola; los ingresos y el número de integrantes del hogar; el número de gatos, perros y televisores en el hogar; el estado laboral, la ocupación, el número promedio de horas de trabajo y el nivel educacional tanto para el jefe como para la jefa de hogar. También existe un grupo de atributos que indican la presencia de una serie de artículos en el hogar: maquina de lavar, refrigerador, tostador, microondas, cafetera, entre otros. El detalle de estas variables y sus posibles valores se encuentran en el Anexo H.2.

A partir de los datos anteriores es posible identificar los precios que en cada compra los hogares pagan por el producto escogido. Sin embargo, no se encuentran disponibles explícitamente los precios que el resto de las marcas ofrecían al momento de cada compra. Esta información es relevante para un análisis de competencia entre las marcas, por lo que se sigue el siguiente esquema para construir un muy buen aproximado de estos precios alternativos. Como existe el detalle de las compras de los hogares es posible clasificar estas transacciones según su fecha, el producto involucrado y la ciudad de la compra (Sioux Falls o Springfield). A partir de este análisis se puede obtener el precio promedio de cada día, para cada producto y en cada ciudad. Se supone que estos son los precios que los consumidores observaron para las otras marcas al momento de realizar su compra según la ciudad en que se encuentran. De esta forma, para cada compra se puede identificar el precio del producto comprado a través del precio real de esa transacción; y el precio del resto de las marcas se puede construir según el esquema anterior de acuerdo a la ciudad donde se realiza la compra. En caso que para un producto no existan transacciones en un día en particular, se utiliza el precio promedio de la semana de la compra. Y en caso de que no existan transacciones en una semana en particular, se utiliza el último promedio semanal disponible.

En la base de datos también se cuenta con información de las veces que un hogar ha sido expuesto a comerciales de televisión (TV) de una determinada marca de Ketchups. Estos datos se recolectan identificando los hogares que sintonizan en sus televisores cada comercial de TV emitido por cada marca. A parir de estos registros se pueden construir variables que indiquen para cada compra si un hogar se ha visto expuesto previamente a algún comercial de TV de cada una de las marcas participantes del mercado.

6.1.2. Análisis del mercado a estudiar

Mercado objetivo

Los modelos propuestos en el capítulo 5 fueron construidos para analizar mercados de dos firmas competidoras, por lo que es necesario acotar el mercado objetivo a estudiar. Para esto, se identifica el formato de Ketchups que explica el mayor porcentaje de demanda y las dos marcas que explican el mayor porcentaje de

las compras en este mercado.

Los ketchups se comercializan en varios formatos, donde el formato de 32 oz. es que concentra la mayor parte del mercado. Un 62 % de las compras registradas son en este formato. Existen cuatro marcas que ofrecen ketchups el formato de 32 oz: Heinz, Hunt's, Del Monte y Brooks. El resto de las compras en el formato de 32 oz. corresponden a varias marcas genéricas que se agregan bajo el nombre de House Brand. En la tabla 6.1 se observa la distribución de las compras de Ketchups en el formato de 32 oz. a través de las cinco marcas descritas y a través de los años descritos. Es necesario recordar que para el año 1988, el estudio considera datos sólo hasta el mes de Agosto.

Marca	**1986**	**1987**	**1988**	**Total**	**Total (%)**
HEINZ	4,691	3,058	2,085	9,834	51.5 %
HUNTS	1,736	1,315	1,382	4,433	23.2 %
HOUSE BRAND	1,154	1,001	538	2,693	14.1 %
DEL MONTE	931	690	371	1,992	10.4 %
BROOKS	56	29	48	133	0.7 %
Total	**8,568**	**6,093**	**4,424**	**19,085**	**100.0 %**

Tabla 6.1: Compras de Ketchups en formato 32 oz. por marca y año. Fuente: Elaboración Propia.

Las marcas Heinz y Hunt's concentran el 75 % de las compras de Ketchups en el formato de 32 oz. Para fines de este trabajo entonces se estudiará la competencia de estas dos marcas en este formato. Considerando las compras asociadas a estas dos marcas en el formato en cuestión, se cubre un 47 % de las compras totales del mercado de los Ketchups: 62 % del total corresponden al formato escogido; y 75 % de este 62 % a las dos marcas seleccionadas.

Uno de los objetivos de este estudio es entender la dinámica de elección de marca en este mercado. Para evitar distorsionar las conclusiones a través de compradores ocasionales, se eliminan del estudio todos los hogares que participan del panel de datos por un periodo menor a seis meses. Un 6 % de las compras asociadas al formato y marcas escogidas son hogares que compran de forma ocasional. Esto implica que las compras estudiadas disminuyen de 14,267 a 13,382 bajando el porcentaje de compras estudiadas de un 47 % a un 44 % del total.

Frecuencia de compra

Para fines de este estudio, los hogares se han clasificado de acuerdo a su frecuencia de compra (número de compras por hogar). Los hogares con frecuencia de compra baja son los que se encuentran por debajo del percentil 50 % de la distribución de la variable número de compras. Los de frecuencia normal tienen números de compras entre el percentil 50 % al 75 % de la variable y los de frecuencia alta son los que se encuentran sobre el percentil 75 %. En la tabla H.4 del Anexo H.3 se definen las variables construidas para registrar la frecuencia de compra de los hogares. En la figura 6.1 se puede observar la distribución de hogares de acuerdo al número de compras realizadas. En la tabla 6.2 se observa la cantidad de hogares y número de compras de

acuerdo a esta clasificación. Los hogares de frecuencia de compra baja representan el 52 % de los hogares y sólo el 24 % de las compras. Los de frecuencia de compra normal representan el 25 % de los hogares y el 25 % de las compras. Los de frecuencia de compra alta representan el 23 % de los hogares y el 51 % de las compras.

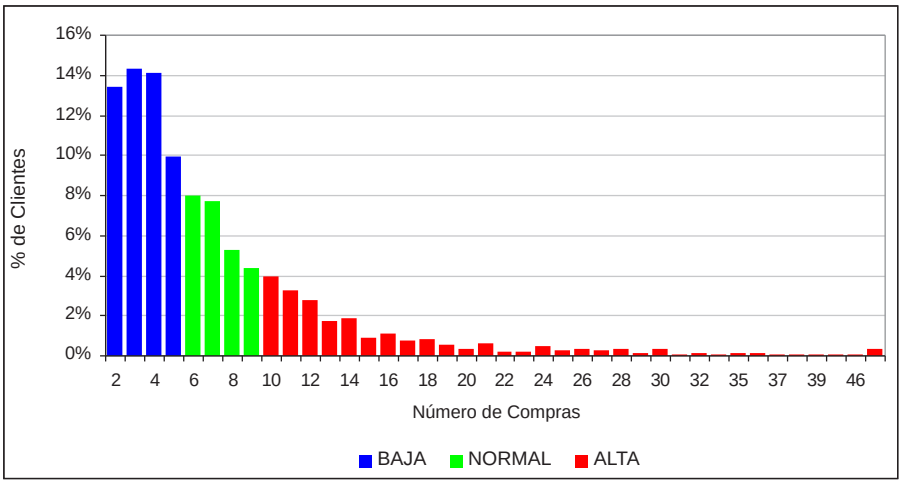

Figura 6.1: Clasificación de hogares según número de compras. Fuente: Elaboración Propia.

Frecuencia	**Hogares**	**Compras**	**Hogares (%)**	**Compras (%)**
BAJA	954	3,241	52 %	24 %
NORMAL	467	3,381	25 %	25 %
ALTA	417	6,760	23 %	51 %
Total	**1,838**	**13,382**	**100 %**	**100 %**

Tabla 6.2: Hogares y compras por clasificación de frecuencia de compra. Fuente: Elaboración Propia.

Esfuerzos de marketing

En la base de datos también se cuenta con información de las veces que un hogar ha sido expuesto a comerciales de televisión (TV) de una determinada marca de Ketchups. En la tabla 6.3 se presentan el número de hogares que son expuestos a los comerciales de TV, el porcentaje que estos hogares representan con respecto al total de clientes del estudio y el número de veces que han sido expuesto a estos comerciales. Los comerciales de Heinz fueron visto por el 53 % de los hogares incluidos en el estudio mientras que los de Hunt's fueron vistos sólo por el 35 %. Las campañas de Heinz son mucho más agresivas que las de Hunt's lo que explica que la primera registre 30,781 exposiciones sobre los hogares mientras que la segunda sólo alrededor de 2,846. El total de hogares expuestos no corresponde a la suma de los hogares expuestos por cada marca, ya que un mismo hogar puede verse expuesto a comerciales de ambas marcas. En efecto, se puede concluir que un porcentaje importante de hogares que son expuestos a los comerciales de Heinz no fueron expuestos

a los comerciales de Hunt's. En cambio prácticamente todos los hogares que son expuestos a comerciales de Hunt's también fueron expuestos a comerciales de Heinz.

Marca	Hogares	Hogares (%)	N° Exposiciones
HEINZ	978	53 %	30,781
HUNTS	642	35 %	2,846
Total	983	53 %	33,627

Tabla 6.3: Hogares expuestos a comerciales de TV de cada marca. Fuente: Elaboración Propia.

La figura 6.2 muestra la cantidad de hogares que se exponen a los comerciales de TV de cada marca a lo largo de los meses y la cantidad de exposiciones promedio a la que se enfrentaron los hogares de cada marca. Se observa que las campañas de marketing de Heinz son constantes en el tiempo mientras que las Hunt's sólo cubren algunos meses del estudio. De la misma forma, las campañas de Heinz consigue tener un número promedio de exposiciones por hogar más alto que las campañas de Hunt's. Lo anterior también puede confirmarse en la figura 6.3 donde se puede ver la distribución de hogares según el número de exposiciones por cada marca. La mayor parte de los hogares que ven comerciales de Hunt's fueron expuestos un número bajo de veces, mientras que los hogares que ven comerciales de Heinz se distribuyen en un nivel alto de exposiciones.

De acuerdo al análisis anterior, se construyen dos variables a partir de la información de exposiciones de los hogares a los comerciales de TV: *HUNTS_Dummy_CE* y *HEINZ_Dummy_CE*, tal como se detalla en la tabla H.4 del Anexo H.3. Estas variables indican en cada compra si el hogar que realiza la transacción ha sido expuesto en algún periodo anterior a algún comercial de TV de la respectiva marca.

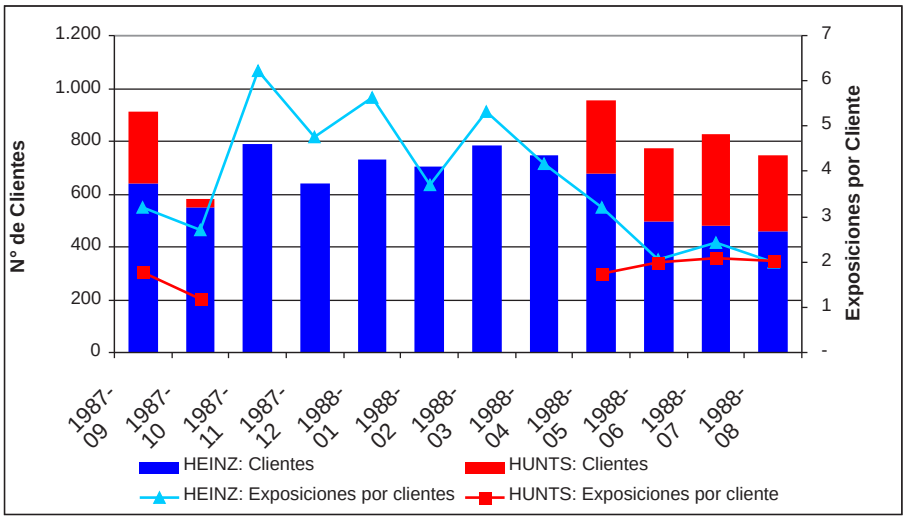

Figura 6.2: Número de hogares expuestos a comerciales de TV y exposiciones promedio. Fuente: Elaboración Propia.

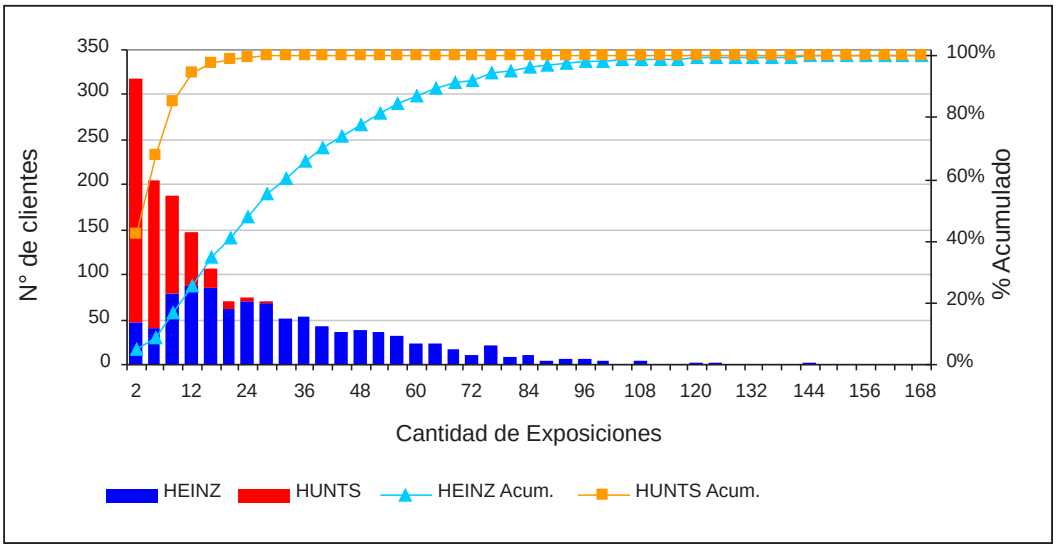

Figura 6.3: Distribución de hogares según número de exposiciones a comerciales de TV. Fuente: Elaboración Propia.

Transición entre marcas

Del total de 13,382 compras que se registran en el periodo de estudio, el 14 % corresponde a las primeras compras de cada uno de los hogares en el periodo de estudio. El 63 % corresponde a compras donde el consumidor escogió la misma marca que en la compra anterior y el 23 % corresponde a compras donde el consumidor escogió una marca diferente que en la compra anterior. Las siguientes tablas representan matrices de transición entre las marcas que escogen los hogares. La tabla 6.4 representa el número de transiciones entre las marcas y la tabla 6.5 presenta las mismas transiciones en términos de porcentaje de casos. En las columnas se encuentran las marcas escogidas por los hogares en cada compra (marca de destino) y en las filas se encuentran las marcas escogidas por los hogares en la compra anterior (marca de origen). Como para la primera compra de cada hogar no se conoce la marca escogida en la compra anterior (ya que no es parte de la base de datos), se asigna una marca de origen "S/I" (sin información) y como destino a la marca escogida. De esta forma, en la tabla 6.4 se observa el origen y destino de cada una de las compras que son parte de este análisis. En la tabla 6.5 se puede observar que en las primeras compras, el 67 % escoge la marca Heinz y un 33 % la marca Hunt's. Se puede concluir también sobre el alto nivel de lealtad de los consumidores sobre la marca Heinz: en el 81 % de los casos los hogares que la escogen la vuelven a elegir en una nueva compra y sólo en un 19 % de los casos los hogares se cambian a la marca Hunt's. La marca Hunt's en cambio sólo puede retener en un 56 % de los casos a sus clientes y en el 44 % de los casos se cambian a la marca Heinz.

La tabla 6.6 representa las transiciones entre las marcas de acuerdo a la clasificación de frecuencia de compra de los hogares. Se observa que la lealtad de los hogares que escogen Heinz crece a mayor frecuencia de compra, mientras que disminuye para los que escogen Hunt's.

En la tabla 6.7 se pueden observar la cantidad de días promedio entre una compra y otra de los hogares de acuerdo a su clasificación según su frecuencia de compra. Tal como se espera, la cantidad de días promedio

Origen\Destino	HEINZ	HUNTS	Total
S/I	1,240	598	1,838
HEINZ	6,524	1,577	8,101
HUNTS	1,503	1,940	3,443
Total	9,267	4,115	13,382

Tabla 6.4: Matriz de transición entre marcas, en número de casos. Fuente: Elaboración Propia.

Origen\Destino	HEINZ	HUNTS
S/I	67 %	33 %
HEINZ	81 %	19 %
HUNTS	44 %	56 %
Total	**69 %**	**31 %**

Tabla 6.5: Matriz de transición entre marcas, en porcentaje de casos. Fuente: Elaboración Propia.

entre compras decrece a mayor frecuencia de compra, lo que se confirma en la distribución mostrada en el gráfico de la figura 6.4: los hogares que compran con mayor frecuencia registran el mayor número de compras y los menores tiempos entre compras.

6.2. Aplicación de metodologías

6.2.1. Descripción

Los modelos propuestos en el capítulo 5 utilizan como base el esquema de Regresión Logística y lo modifican incorporando en su resolución restricciones que nacen del equilibrio de un juego de competencia entre dos firmas. Para poder evaluar estos modelos alternativos, concluir sobre su bondad de ajuste y contrastarlos con la utilización de la Regresión Logística tradicional, se requiere implementar y resolver cada uno de estos modelos.

Los modelos que se desean resolver entonces son los siguientes:

- **Regresión Logística (0):** Corresponde al modelo tradicional de Regresión Logística y sus resultados serán utilizados como base de comparación (*benchmark*) para los modelos propuestos en el capítulo 5.
- **Regresión Logística con restricciones de competencia en precios (1):** Corresponde al primero de los modelos propuestos, que considera restricciones del equilibrio de un juego de competencia en

Frecuencia	Origen\Destino	HEINZ	HUNTS
BAJA	S/I	68 %	32 %
	HEINZ	76 %	24 %
	HUNTS	41 %	59 %
NORMAL	S/I	66 %	34 %
	HEINZ	77 %	23 %
	HUNTS	43 %	57 %
ALTA	S/I	67 %	33 %
	HEINZ	84 %	16 %
	HUNTS	45 %	55 %

Tabla 6.6: Matriz de transición entre marcas y según clasificación de frecuencia de compra. Fuente: Elaboración Propia.

Frecuencia	Origen\Destino	HEINZ	HUNTS	Total
BAJA	HEINZ	217	270	230
	HUNTS	230	221	225
NORMAL	HEINZ	116	147	124
	HUNTS	101	116	110
ALTA	HEINZ	53	63	55
	HUNTS	54	52	52
Total		98	121	105

Tabla 6.7: Promedio de días entre compras, según frecuencia de compra. Fuente: Elaboración Propia.

precios entre dos firmas y que se presenta en la sección 5.3. Este modelo requiere que se consideren como atributos los precios de ambas marcas.

- **Regresión Logística con restricciones de competencia en precios y descuentos (2):** Corresponde al segundo de los modelos propuestos, que considera restricciones del equilibrio de un juego de competencia en precios y descuentos entre dos firmas y que se presenta en la sección 5.4. Este modelo requiere que se consideren como atributos los precios y descuentos de ambas marcas.

- **Regresión Logística con restricciones de competencia en precios, descuentos y esfuerzos en marketing (3):** Corresponde al tercero de los modelos propuestos, que considera restricciones del equilibrio de un juego de competencia en precios, descuentos y esfuerzos en marketing entre dos firmas y que se presenta en la sección 5.5. Este modelo requiere que se consideren como atributos los precios, descuentos y esfuerzos en marketing de ambas marcas.

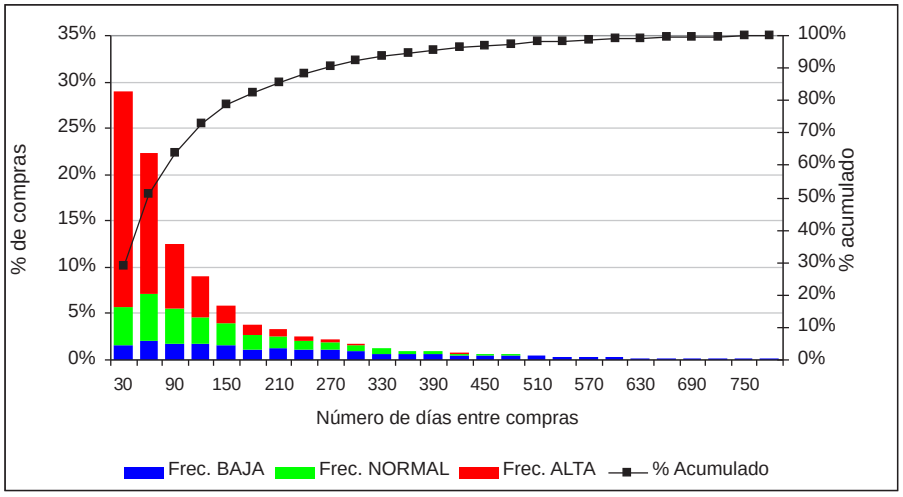

Figura 6.4: Distribución de compras según promedio de días entre compras. Fuente: Elaboración Propia.

Estos modelos se utilizarán para determinar las probabilidades de que un hogar compre en cada una de las marcas estudiadas. Para su implementación se debe utiliza una base de datos con observaciones $\{(x_i, y_i)\}_{i=1}^{I}$, donde $x_i \in \mathbb{R}^n$ es el vector de características de la observación i e $y_i \in \{0,1\}$ identifica la clase a la que pertenece la observación i. En este caso aplicado, las observaciones corresponden a las compras que realizan los hogares y las clases a las marcas escogidas en cada transacción. Un vector x_i está compuesto por cada realización de los atributos asociados a la compra i: por ejemplo el precio de las marcas, atributos de promoción de la marca al momento de la compra, variables demográficas del hogar, etc. Se define que la clase base $y_i = 1$ corresponde a las compras de la marca Hunt's y que la clase $y_i = 0$ a las compras de la marca Heinz. De esta forma, a través de estos modelos se estimará $\Pr(y_i = 1)$ que representa la probabilidad de que el hogar en la compra i prefiera la marca Hunt's.

6.2.2. Base de datos de trabajo

La base de datos de trabajo está constituida por los siguientes elementos:

- **Variables asociadas a elementos competitivos:** A partir de las variables descritas en el Anexo H.1, se construyeron las variables: *Display*, si el producto tiene un tipo de exhibición especial en el local al momento de la compra; *Advertising*, si existe algún tipo de publicidad al instante de la compra; *Type*, si cuando existe algún tipo de publicidad ésta corresponde a algún tipo especial; *Coupon*, si existe algún cupón de descuento asociado a la compra; y *Special_Price*, si el producto comprado presenta algún tipo de precio especial. El detalle de la definición utilizada para construir cada una de estas variables se encuentra en la tabla H.3 del Anexo H.3.
- **Variables de precio:** Se identifican a partir de las variables *P_HUNTS_Unit_Extended_Price* y *P_HEINZ_Unit_Extended_Price* los precios unitarios de cada marca antes de la aplicación de cupones

de descuentos. Si una compra está asociada a la marca Hunt's entonces la variable *P_HUNTS_Unit_Extended_Price* contiene el precio real de la transacción y la variable *P_HEINZ_Unit_Extended_Price* contiene el precio promedio del día de las compras de Heinz en la misma ciudad de la compra. En el caso que la compra está asociada a la marca Heinz entonces la situación es inversa. Se incorporan también dos variables que representan el diferencial de precios entre ambas marcas. La variable *Dif_Unit_Extended_Price* representa la resta entre el precio de la marca comprada menos el precio de la otra marca. Esta variable toma como referencia la marca que se escoge en cada compra. También se construye la variable *Dif2_Unit_Extended_Price* que representa la resta entre el precio de Hunt's menos el precio de Heinz, independiente de la marca escogida en la respectiva compra. Por último, también se utilizan los valores unitarios de los cupones de descuento de cada marca. Estos valores se representan a través de las variables *P_HUNTS_Unit_TV_Coupons* y *P_HEINZ_Unit_TV_Coupons*. El detalle de la definición utilizada para construir cada una de estas variables se encuentra en la tabla H.4 del Anexo H.3.

- **Frecuencias de compra:** Se incorporan al estudio las variables que identifican la frecuencia de compra de cada hogar de acuerdo a lo definido en la sección 6.1.2. Para esto se utilizan las variables indicadoras *Frecuencia_Baja*, *Frecuencia_Normal* y *Frecuencia_Alta*, según si el hogar corresponde a uno de frecuencia baja, normal o alta respectivamente. El detalle de la definición utilizada para construir cada una de estas variables se encuentra en la tabla H.4 del Anexo H.3.

- **Esfuerzos de marketing:** Se utilizan también dos variables que permiten identificar si previo al momento de la compra, el hogar se ha visto expuesto a algún comercial de TV de cada marca. Para esto se utilizan las variables indicadoras *HUNTS_Dummy_CE* y *HEINZ_Dummy_CE*. El detalle de la definición utilizada para construir cada una de estas variables se encuentra en la tabla H.4 del Anexo H.3.

- **Variables Demográficas:** Finalmente las variables demográficas descritas en el Anexo H.2 también son utilizadas en la base de datos de trabajo.

Los modelos (1), (2) y (3) plantean diferentes esquemas de competencia por lo que requieren de la utilización de variables de precio, descuento y esfuerzos en marketing; a diferencia del modelo *benchmark* que a priori no requiere de variables particulares para poder ser estimado. En la tabla 6.8 se detallan las variables que son requeridas por los modelos (1), (2) y (3); se identifica el nombre de estas variables en la base de datos de trabajo; y se especifica para cada variable que modelo las exige para ser estimado.

Variable Modelo	**Variable Base de Datos**	**Modelos que las requieren**
p_{Ai}	*P_HUNTS_Unit_Extended_Price*	Modelo (1), Modelo (2), Modelo (3)
p_{Bi}	*P_HEINZ_Unit_Extended_Price*	Modelo (1), Modelo (2), Modelo (3)
d_{Ai}	*P_HUNTS_Unit_TV_Coupons*	Modelo (2), Modelo (3)
d_{Bi}	*P_HEINZ_Unit_TV_Coupons*	Modelo (2), Modelo (3)
m_{Ai}	*HUNTS_Dummy_CE*	Modelo (3)
m_{Bi}	*HEINZ_Dummy_CE*	Modelo (3)

Tabla 6.8: Asociación entre variables de los modelos y variables de la base de datos.

Como se describe anteriormente, la base de datos contiene un historial de compras desde Enero de 1986 hasta Agosto de 1988. A lo largo de este periodo se distribuyen las compras que son de interés para este estudio: las compras asociadas a las marcas Hunt's y Heinz en el formato de 32 oz., excluyendo las compras ocasionales según se define en la sección 6.1.2. Para la resolución de los modelos antes descritos la base de datos será separada en tres secciones: una para realizar la calibración o entrenamiento de los modelos, una de prueba y una de validación.

La base de entrenamiento (*Training*) está compuesta por una muestra aleatoria del 70 % de las compras realizadas entre Enero de 1986 y Diciembre de 1987. El 30 % restante compone la base de prueba (*Test*), cuyo objetivo es asegurar que los modelos obtenidos con la base de entrenamiento no se encuentran sobreajustados y que pueden explicar las decisiones de compra de las transacciones no consideradas en la calibración de los modelos. La base de validación (*Validation*) está compuesta por el total de compras entre Enero de 1988 y Agosto 1988, que es el periodo no considerado en la base de entrenamiento y en la base de prueba. El objetivo de la base de validación es evaluar la bondad de los modelos en un periodo temporal diferente al utilizado en su calibración y en la base de prueba. Este esquema para segmentar las bases de datos se describe en la figura 6.5 y el número de compras por base de datos se pueden observar en la tabla 6.9.

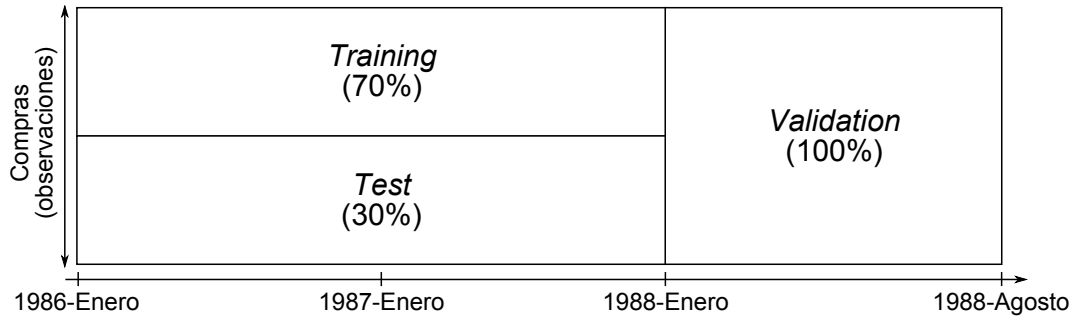

Figura 6.5: Esquema de bases de datos: Entrenamiento, Prueba y Validación

Base de datos	**Total**	**Total (%)**
Training	7,117	53 %
Test	3,008	22 %
Validation	3,257	24 %
Total	**13,382**	**100 %**

Tabla 6.9: Número de compras según base de datos. Fuente: Elaboración Propia.

Los modelos de forma natural presentan mejores indicadores de calidad en las bases de entrenamiento, ya que sus parámetros son determinados para ajustarse a la realidad de las compras de esta base. Sin embargo, un modelo que funciona de forma perfecta en la base de entrenamiento y no es capaz de explicar el comportamiento de compra de las bases de prueba y validación, es un modelo que no puede ser utilizado en la práctica ya que no es útil para estudiar el comportamiento futuro del fenómeno de análisis. En este sentido, se considera que los modelos más potentes son aquellos que presentan además un buen comportamiento en las bases de prueba y validación. Es más, la base de validación tiene una importancia particular debido a

que está fuera del periodo temporal de calibración y por lo tanto recoge potenciales cambios temporales en los patrones de comportamiento de los hogares, permitiendo evaluar la capacidad del modelo obtenido para responder ante estas situaciones. En efecto, de la tabla 6.10 se puede observar que la base de entrenamiento y la base de prueba tienen la misma distribución de casos para la variable dependiente y. En efecto, los casos $y = 1$ en la base de entrenamiento y la base de prueba corresponden a un 28 % y 27 % de los casos. Esto tiene sentido ya que ambas bases se construyen a partir de un muestreo aleatorio (70 % de los casos en una y el restante 30 % en la otra). De la misma tabla también se puede ver que esta proporción cambia de forma importante en la base de validación: los casos $y = 1$ son un 39 % de los casos. Este aumento responde a un aumento en la participación de mercado de la marca Hunt's (marca asociada al estado $y = 1$), por lo que en este caso aplicado la base de validación servirá realmente para evaluar la capacidad de los modelos obtenidos para adaptarse a cambios en las condiciones de mercado.

Base de datos	**y=0**	**y=0 (%)**	**y=1**	**y=1 (%)**	**Total**
Training	5,096	72 %	2,021	28 %	7,117
Test	2,196	73 %	812	27 %	3,008
Validation	1,975	61 %	1,282	39 %	3,257
Total	**9,267**	**69 %**	**4,115**	**31 %**	**13,382**

Tabla 6.10: Distribución variable dependiente por base de datos. Fuente: Elaboración Propia.

6.2.3. Selección de atributos

Las variables independientes candidatas a ser partes del vector de características x_i corresponden a las pertenecientes a la base de datos de trabajo descrita en la sección 6.2.2. A partir de estas variables se deben seleccionar las que serán parte de los modelos a construir.

Para escoger las variables asociadas a los cuatro modelos anteriores se utiliza un algoritmo de selección secuencial de atributos aplicado sobre el modelo tradicional de Regresión Logística, denominado en la sección anterior como modelo (0) o *benchmark*. El subconjunto de atributos que resulta es el que será utilizado para resolver los modelos (1), (2) y (3). Los modelos (1), (2) y (3) requieren la utilización obligatoria de algunos atributos (precios, descuentos y esfuerzos en marketing según corresponda), por lo que en caso de que el subconjunto de atributos seleccionado no incluya alguna de estas variables obligatorias para alguno de los modelos, para este caso las variables faltantes serán agregadas manualmente.

El algoritmo de selección de atributos utilizado es de tipo *stepwise*. Las técnicas de selección de atributos tienen como objetivo encontrar el conjunto de predictores que sea más eficiente en predecir la variable dependiente y que a la vez tenga un alto nivel de parsimonia. Las técnicas más utilizadas para seleccionar secuencialmente los atributos son tres: selección *forward*, eliminación *backward* y el método *stepwise*.

En la selección *forward*, partiendo de un modelo sin regresores en cada iteración se agrega al modelo la variable que genera la mayor mejora en eficiencia del modelo. Esto se realiza hasta que no queden variables para agregar o hasta que ya no existan mejoras en el modelo. En la eliminación *backward*, partiendo de

un modelo con todos los regresores en cada iteración se quita del modelo la variable que genera la mayor mejora en eficiencia del modelo. Esto se realiza hasta que no queden variables para quitar o hasta que ya no existan mejoras en el modelo. En un método *stepwise* se mezclan ambas alternativas. Partiendo de un modelo sin regresores, en cada iteración primero se realiza un paso de selección *forward* y luego otro paso de eliminación *backward*. Esto se realiza hasta que ya no existan variables para agregar/eliminar o hasta que ya no existan mejoras en el modelo en las nuevas iteraciones.

Para aplicar el método *stepwise* en el modelo *benchmark* se inicia el proceso iterativo sin regresores y el conjunto potencial de regresores o de variables candidatas corresponde a todas las variables de la base de datos de trabajo descrita en la sección 6.2.2.

Un elemento mencionado pero no especificado en la descripción de los métodos de selección secuencial de atributos es el criterio para evaluar la eficiencia o calidad del modelo. De acuerdo a lo recomendado por Venables y Ripley (2002), se evalúan dos criterios de calidad de los modelos. El primero es el *Akaike information criterion* (AIC) y el segundo es el *Bayesian information criterion* (BIC). Estos se definen de la siguiente manera.

$$AIC = -2\ln(\mathcal{L}(b/x)) + 2n \tag{6.1}$$

$$BIC = -2\ln(\mathcal{L}(b/x)) + \ln(I)n \tag{6.2}$$

Donde $\mathcal{L}(b/x)$ es el valor de la función de verosimilitud del modelo que se evalúa, n es el número de atributos considerados e I es el número de observaciones de la base de datos. Tanto el AIC como el BIC son medidas relativas de la calidad de un modelo, que buscan describir la relación entre la exactitud y complejidad de estos. El mejor modelo es el que tiene el menor AIC o BIC según sea el criterio utilizado.

El componente $-2\ln(\mathcal{L}(b/x))$ habla de la exactitud del modelo y se busca que sea lo más negativo posible (para encontrar los parámetros de un modelo se maximiza la función de verosimilitud). En el caso del AIC, el componente $2n$ crece según el número de atributos que contenga el modelo, por lo que al buscar el menor AIC se prefieren modelos con menos atributos (se prefiere el modelo con mayor nivel de parsimonia). En el caso de BIC, el componente $\ln(I)n$ también crece con el número de atributos, por lo que al buscar el menor BIC se prefieren modelos con menos atributos. La diferencia entre el AIC y el BIC se encuentra en el nivel de penalización sobre el número de atributos del modelo. El AIC penaliza el número de atributos a través de amplificar por dos el número de atributos utilizados, en cambio el BIC penaliza amplificando por $\ln(I)$ el número de atributos. El término $\ln(I)$ es mayor que 2 cuando el número de observaciones en la base de datos I es mayor que $e^2 \approx 7.39$, por lo prácticamente para cualquier base de datos el criterio BIC será mucho más estricto en la penalización del número de atributos utilizados por el modelo. Dado esto se espera que al utilizar un criterio BIC se obtenga un modelo con un número menor de atributos con respecto al obtenido utilizando un criterio AIC.

Debido a lo anterior, se realizarán dos procedimientos de selección de atributos *stepwise*, que entregarán dos subconjuntos de variables a utilizar. El primero aplicando el proceso *stepwise* utilizando el criterio BIC, que entrega un subconjunto de variables A. El segundo aplicando el proceso *stepwise* utilizando el criterio AIC, que entrega un subconjunto de variables B. Ambos conjuntos se traducirán en dos versiones para el

modelo *benchmark*; y para los modelos (1), (2) y (3). En caso que alguno de estos modelos requiera de forma obligatoria alguna variable no contenida en los conjuntos A ó B, estas serán agregadas manualmente.

6.3. Resultados

6.3.1. Resultados Generales

A continuación se describen los resultados de la aplicación de los modelos de acuerdo a la metodología descrita en la sección anterior. Los resultados detallados del modelo *benchmark* y los modelos (2), (3) y (4) $-$aplicados para los conjuntos de datos A y $B-$ se detallan en el Anexo J.

En el Anexo D se presentan los diferentes indicadores estadísticos que son utilizados para evaluar y comparar los resultados de cada modelo.

En el Anexo J.1 se detallan los resultados del proceso de selección de atributos. En la columna *Benchmark* de la tabla J.1 se identifican las variables que fueron seleccionadas por el algoritmo de selección de atributos con la marca "S". Se separan los resultados de utilizar el criterio BIC, que define el conjunto de atributos A, y el criterio AIC, que define el conjunto de atributos B.

Estos conjuntos A y B luego se utilizan para resolver los modelos (1), (2) y (3). Todas las variables requeridas por los modelos (1) y (2) se encuentran en los conjuntos A y B, por lo que no hay problemas para su estimación. En el caso del modelo (3), la variable *HUNTS_Dummy_CE* no se encuentra en los conjuntos A y B pero es requerida por el modelo. En este caso la variable es agregada manualmente y en la tabla J.1 estos casos quedan identificados con la marca "M". En esta tabla las variables que son obligatorias para alguno de los modelos propuestos se destacan con (*).

La solución de los modelos (1), (2) y (3) además requieren la especificación del parámetro μ. Éste es el parámetro de penalización en la función objetivo de la desviación de la solución con respecto a las condiciones de equilibrio del juego plantado en cada modelo. En particular para el modelo (3) además se requiere especificar un parámetro e que corresponde al costo marginal de los esfuerzos en marketing. En el caso de esta aplicación, como los esfuerzos en marketing del modelo (3) se representan a través de las variables *HUNTS_Dummy_CE* y *HEINZ_Dummy_CE*, el parámetro e representa el costo de exponer a un hogar a los comerciales de TV.

La determinación de los parámetros μ y e se encuentra fuera de los objetivos y alcances de este trabajo, por lo que para estimar los modelos planteados se asumirán como un dato conocido. En la tabla 6.11 se definen los parámetros a utilizar en la estimación de los modelos propuestos. Estos parámetros fueron escogidos arbitrariamente a partir de un análisis de sensibilidad de los resultados, donde se buscaba que los resultados cambiarán respecto del modelo *benchmark* pero sin deteriorar de forma radical los indicadores. Una línea de trabajo futuro corresponde a buscar una metodología para encontrar la mejor elección de estos parámetros.

Parámetro	Valor
μ	-0.2
e	0.1

Tabla 6.11: Parámetros utilizados en la estimación de los modelos propuestos. Fuente: Elaboración Propia.

En el anexo J.2 se detallan los resultados de la estimación de los modelos *benchmark*, (1), (2) y (3) utilizando el conjunto de atributos *A*. En el anexo J.3 se detallan los resultados de la estimación de estos modelos utilizando el conjunto de atributos *B*. Para cada una de las versiones de los modelos los resultados se dividen en tres grupos. En el primer grupo se muestran los resultados de la estimación del modelo sobre la base de entrenamiento detallando: los coeficientes obtenidos; su desviación estándar; los resultados de la prueba de Wald, incluyendo su *p*-valor; y el intervalo de confianza de los coeficientes al 95 % (Larose, 2006). En el segundo grupo, para cada modelo se detallan algunos estadísticos asociados a la estimación obtenida utilizando la base de entrenamiento: el logaritmo de la función de verosimilitud o *Log-Likelihood*; el valor de los estadísticos BIC y AIC; el estadístico *G* asociado a la prueba del *Log-Likelihood ratio* y su *p*-valor; y las versiones del R^2 de McFadden's, Cox & Snell y Nagelkerke. En el Anexo D.1 se describen las características de estos indicadores. En estos dos grupos de resultados se detallan estadísticos que permiten conocer la calidad de la estimación obtenida para cada modelo. Sin embargo, estos resultados no pueden ser concluyentes pues estos indicadores no explican la capacidad predictiva de los modelos y además sus resultados sólo se evalúan sobre la base de entrenamiento.

El tercer grupo de resultados es el más importante, porque para cada modelo se evalúan diferentes criterios de eficiencia que permiten comparar y concluir sobre el desempeño y la capacidad predictiva de los modelos tanto en la base entrenamiento como las bases prueba y validación. Son de particular interés los resultados que se obtengan en la base de validación, ya que evalúa el comportamiento del modelo en un periodo temporal fuera del horizonte de estimación. En este grupo se evalúan el estadístico de Kolmogórov-Smirnov (KS), el área bajo la curva ROC (AUC) y el error cuadrático medio (MSE) de la predicción del modelo con respecto a las decisiones de compra de los hogares. Estos indicadores son descritos en los Anexos D.2, D.3 y D.4 respectivamente. En relación a la curva ROC adicionalmente se detallan el estadístico J de Youden, el *cut-off* escogido a través de este indicador y las tasas de acierto y error asociadas a este *cut-off*.

6.3.2. Análisis Resultados Estadísticos

El estadístico KS mide la separación entre las distribuciones acumuladas de σ_i para los hogares que eligen la marca Hunt's y los que escogen la marca Heinz. En la sección D.2 se describe en detalle las características del estadístico KS. En la tabla 6.12 se detallan los resultados para estadístico KS para los diferentes modelos en las bases de entrenamiento, prueba y validación. Para la base de entrenamiento, se observa que el mejor nivel de KS se obtiene con el modelo (1) utilizando el conjunto de atributos *B*, que presenta un KS un 0.7 % mejor que en el caso del modelo *benchmark* en el mismo conjunto de atributos. Al observar sólo los resultados del conjunto de atributos *A*, el mejor nivel de KS también se obtiene con el modelo (1), que presenta un KS un

Set atributos	Modelo	Estadístico			Mejora c/r *Benchmark*		
		Training	***Test***	***Validation***	***Training***	***Test***	***Validation***
A	*Benchmark*	31.1 %	29.1 %	37.4 %			
	Modelo (1)	32.2 %	29.6 %	37.6 %	3.5 %	1.5 %	0.5 %
	Modelo (2)	27.1 %	25.5 %	39.2 %	-13.1 %	-12.4 %	4.8 %
	Modelo (3)	26.8 %	24.4 %	38.5 %	-13.8 %	-16.1 %	3.1 %
B	*Benchmark*	34.3 %	30.4 %	40.2 %			
	Modelo (1)	34.6 %	29.9 %	40.6 %	0.7 %	-1.5 %	0.9 %
	Modelo (2)	31.6 %	27.5 %	44.3 %	-8.0 %	-9.4 %	10.2 %
	Modelo (3)	31.6 %	26.5 %	40.9 %	-8.0 %	-12.9 %	1.7 %

Tabla 6.12: Resultados estadístico Kolmogórov-Smirnov (KS) por modelo y base de datos. Fuente: Elaboración Propia.

3.5 % mejor que en el caso del modelo *benchmark* en el mismo conjunto de atributos. Para ambos conjuntos de atributos y en términos del estadístico KS luego del modelo (1) le siguen el modelo *benchmark*, modelo (2) y modelo (3) respectivamente. Los KS de los modelos (2) y (3) son significativamente peores que los del modelo *benchmark*: en el caso del conjunto de atributos *A* son más bajos en aproximadamente un 13.5 %, mientras que en el conjunto de atributos *B* son más bajos en aproximadamente un 8 %.

En el caso de la base de prueba el mejor KS se obtiene con el modelo *benchmark* en el conjunto de atributos *B*. Con este conjunto de atributos, en términos de KS luego siguen el modelo (1), modelo (2) y modelo (3). Los KS del modelo (2) y modelo (3) también son significativamente más bajos que el modelo *benchmark*: un 12.4 % y 16.1 % respectivamente. Para el conjunto de atributos *A* el mejor KS se obtiene con el Modelo (1), seguido por el modelo *benchmark*, modelo (2) y modelo (3). Los KS del modelo (2) y modelo (3) también son significativamente más bajos que el modelo *benchmark*: un 9.4 % y 12.9 % respectivamente.

Para la base de validación el mejor KS se obtiene al utilizar el modelo (2) con el conjunto de datos *B*, configuración que entrega un KS un 10.2 % mejor que el obtenido con el modelo *benchmark* con el mismo conjunto de atributos. En este caso se pasa de un KS igual a 40.2 % para el modelo *benchmark* a un KS de 44.3 % para el modelo (2), un aumento que en magnitud es bastante considerable. Al observar sólo los resultados del conjunto de atributos *B*, luego del modelo (2) siguen el modelo (3), el modelo (1) y finalmente el modelo *benchmark* en términos de KS. El modelo (3) es un 3.1 % mejor que el modelo *benchmark*, mientras que el modelo (1) lo supera en a penas un 0.5 %. En el caso del conjunto de atributos *A*, el modelo (2) también es el modelo con mejor KS, presentando un mejor indicador de discriminación en un 4.8 % con respecto al modelo *benchmark*. Al observar sólo los resultados del conjunto de atributos *A*, luego del modelo (2) siguen también el modelo (3), el modelo (1) y finalmente el modelo *benchmark* en términos de KS. El modelo (3) es un 1.7 % mejor que el modelo *benchmark*, mientras que el modelo (1) lo supera en a penas un 0.9 %.

El estadístico AUC mide el area bajo la curva ROC. Entre mayor AUC tenga un modelo entonces se interpreta que tiene una mejor eficiencia en sus predicciones. En la sección D.3 se describe en detalle las características del estadístico AUC. En la tabla 6.13 se detallan los resultados para estadístico AUC para los

Set atributos	Modelo	Estadístico			Mejora c/r *Benchmark*		
		Training	***Test***	***Validation***	***Training***	***Test***	***Validation***
A	*Benchmark*	71.1 %	69.6 %	74.4 %			
	Modelo (1)	71.0 %	69.4 %	74.1 %	-0.1 %	-0.3 %	-0.4 %
	Modelo (2)	68.7 %	67.6 %	76.1 %	-3.4 %	-2.9 %	2.4 %
	Modelo (3)	68.3 %	67.1 %	75.4 %	-4.0 %	-3.7 %	1.4 %
B	*Benchmark*	73.9 %	70.8 %	76.2 %			
	Modelo (1)	73.6 %	70.1 %	76.6 %	-0.5 %	-1.0 %	0.5 %
	Modelo (2)	71.7 %	68.9 %	78.7 %	-3.0 %	-2.7 %	3.3 %
	Modelo (3)	71.4 %	68.4 %	78.1 %	-3.4 %	-3.4 %	2.5 %

Tabla 6.13: Resultados área bajo la curva (AUC) por modelo y base de datos. Fuente: Elaboración Propia.

diferentes modelos en las bases de entrenamiento, prueba y validación. Para la base de entrenamiento y para la base de prueba, se observa que el mejor nivel de AUC se obtiene en el modelo *benchmark* utilizando el conjunto de atributos *B*. Al observar sólo los resultados del conjunto de atributos *A*, el mejor nivel de AUC también se obtiene con el modelo *benchmark* para ambas bases. Para ambos conjuntos de atributos y en términos del estadístico AUC luego del modelo *benchmark* le siguen el modelo (1), modelo (2) y modelo (3) respectivamente.

En la base de validación el mejor AUC se obtiene con el modelo (2) utilizando el conjunto de atributos *B*, que resulta ser un 3.3 % mejor que el modelo *benchmark* en el mismo conjunto de atributos. Al observar sólo los resultados del conjunto de atributos *B*, luego del modelo (2) siguen el modelo (3), el modelo (1) y finalmente el modelo *benchmark* en términos de AUC. Al observar sólo los resultados del conjunto de atributos *A*, los mejores resultados de AUC se obtienen también con el modelo (2), que resulta ser un 2.4 % mejor que el modelo *benchmark* en el mismo conjunto de atributos. Luego del modelo (2) siguen el modelo (3), el modelo *benchmark* y finalmente el modelo (1), aunque los dos últimos con valores de AUC muy cercanos.

El error cuadrático medio (MSE) es una medida de la desviación que existe entre las predicciones de un modelo y las decisiones de los hogares. Se desea que un modelo cometa el menor error posible y que por lo tanto tenga el menor MSE. En la sección D.4 se describe en detalle las características del MSE. En la tabla 6.14 se detallan los resultados del MSE para los diferentes modelos en las bases de entrenamiento, prueba y validación. Para la base de entrenamiento y para la base de prueba, se observa que el menor MSE se obtiene en el modelo *benchmark* utilizando el conjunto de atributos *B*. Al observar sólo los resultados del conjunto de atributos *A*, el menor nivel de MSE también se obtiene con el modelo *benchmark* para ambas bases. Para ambos conjuntos de atributos y en términos del menor MSE luego del modelo *benchmark* le siguen el modelo (1), modelo (2) y modelo (3) respectivamente.

En la base de validación el menor MSE se obtiene con el modelo (2) utilizando el conjunto de atributos *B*, que resulta ser un 4.4 % mejor que el modelo *benchmark* en el mismo conjunto de atributos. Al observar sólo los resultados del conjunto de atributos *B*, luego del modelo (2) siguen el modelo (3), el modelo (1) y finalmente el modelo *benchmark*. Al observar sólo los resultados del conjunto de atributos *A*, los menores

Set atributos	Modelo	Estadístico			Mejora c/r *Benchmark*		
		Training	***Test***	***Validation***	***Training***	***Test***	***Validation***
A	*Benchmark*	18.1 %	17.9 %	19.9 %			
	Modelo (1)	18.3 %	18.1 %	20.1 %	-1.4 %	-1.3 %	-0.9 %
	Modelo (2)	18.8 %	18.5 %	19.4 %	-4.0 %	-3.6 %	3.0 %
	Modelo (3)	18.9 %	18.6 %	19.7 %	-4.5 %	-4.2 %	1.0 %
B	*Benchmark*	17.2 %	17.6 %	19.3 %			
	Modelo (1)	17.5 %	17.8 %	19.2 %	-1.6 %	-0.8 %	0.3 %
	Modelo (2)	18.0 %	18.2 %	18.4 %	-4.6 %	-3.2 %	4.4 %
	Modelo (3)	18.1 %	18.3 %	18.8 %	-5.1 %	-3.7 %	2.3 %

Tabla 6.14: Resultados error cuadrático medio (MSE) por modelo y base de datos. Fuente: Elaboración Propia.

MSE se obtienen también con el modelo (2), que resulta ser un 3.0 % mejor que el modelo *benchmark* en el mismo conjunto de atributos. Luego del modelo (2) siguen el modelo (3), el modelo *benchmark* y finalmente el modelo (1), aunque los dos últimos con valores de MSE muy cercanos.

Para los diferentes modelos en general se observa que para la base de prueba se tienen resultados similares al de la base de entrenamiento. Esto se explica ya que las dos bases se construyen al separar el total de observaciones aleatoriamente en dos. Por otro lado, en la base de validación se observan resultados significativamente diferentes a los que se obtienen en la base de entrenamiento o en la base de prueba. Esta diferencia existe porque la base de validación representa una muestra que se aleja de la distribución de probabilidad asociada a los datos utilizados en la estimación de los modelos: utiliza un periodo temporal diferente y posterior al periodo de entrenamiento. De esta forma, un modelo que se sobreajusta a la base de entrenamiento no tendrá la capacidad de predecir eficientemente el comportamiento de los hogares que participan de la base de validación. En este sentido, los resultados sobre la base de validación permiten concluir sobre la capacidad del modelo de comprender el comportamiento de compra de los hogares más allá de la base de entrenamiento.

El modelo *benchmark* en general es el que obtiene mejores resultados en la base de entrenamiento y en la base de prueba, para los diferentes estadísticos. Para la base de validación, en términos relativos al resto de los modelos, éste es el modelo que tiene peores resultados para esta base.

Cuando el modelo *benchmark* no es el que tiene mejor rendimiento en la base de entrenamiento o en la base de prueba, entonces el modelo (1) es el que obtiene mejores resultados para los diferentes estadísticos en estas bases. En efecto, el modelo (1) es el que presenta el mejor desempeño en la base de entrenamiento para el estadístico KS. Este es un elemento a destacar en el sentido que se esperaría que el modelo *benchmark* tenga los mejores resultados en la base de entrenamiento. Esto porque el modelo (1) utiliza la metodología propuesta en este trabajo, y por lo tanto incorpora restricciones en el espacio de soluciones factibles con respecto al modelo *benchmark*, disminuyendo en teoría su capacidad de ajustarse de forma eficiente a las observaciones de la base de estimación.

El modelo (1) incorpora restricciones de competencia en precios entre dos firmas que buscan aumentar su participación de mercado. Sin embargo, la incorporación de la competencia en precios no mejora de forma suficiente la capacidad de generalización del modelo con respecto al comportamiento de compra de los hogares. Consistentemente con lo que sucede para el modelo *benchmark*, el modelo (1) presenta malos resultados en la base de validación.

El modelo (2) siempre presenta malos resultados en la base de entrenamiento y en la base de prueba, para los diferentes estadísticos, sólo superando al modelo (3) en los diferentes casos. Esto se explica por las restricciones de competencia en precios y descuentos que se incorporan en la solución del modelo (2). Los resultados cambian cuando se evalúa la base de validación, donde el modelo (2) siempre es el modelo con mejores resultados en términos de calidad de los estadísticos. El modelo (2) entonces tiene la particularidad de tener estadísticos considerablemente peores al *benchmark* en la base de entrenamiento y en la base de prueba. Sin embargo, el mismo modelo tiene estadísticos significativamente mejores al *benchmark* en la base de validación.

El modelo (3) siempre es el que obtiene los peores resultados en la base de entrenamiento y en la base de prueba, para los diferentes estadísticos. Esta situación es consistente con que este modelo corresponde a la versión más compleja del esquema propuesto en este trabajo, donde se incluyen restricciones de competencia en precios, descuentos y esfuerzos en marketing entre las firmas del mercado. Los resultados cambian cuando se evalúa la base de validación, donde el modelo (3) siempre es el segundo mejor modelo en términos de calidad de los estadísticos. Al igual que el modelo (2), el modelo (3) siempre supera al modelo *benchmark* en la calidad de sus estadísticos en la base de validación.

En la base de validación el modelo (2) obtiene mejores resultados que el modelo (3) a pesar de que este último incorpora mayores elementos de competencia en su solución. Sin embargo, se debe considerar que para poder resolver el modelo (3), se tuvo que suponer el costo marginal de los esfuerzos de marketing que se representan a través del parámetro e. Este elemento fue escogido de forma arbitraria dado que no representaba el foco de este trabajo. Queda abierta entonces la interrogante sobre si es posible mejorar los resultados del modelo (3), superando los del modelo (2), dedicando mayores esfuerzos en la estimación del parámetro e.

Las conclusiones anteriores corresponden a la evidencia empírica recogida de las pruebas realizadas. Sin embargo, para asegurar con certeza estadística que las diferencias de los resultados de los modelos en cada base de datos son significativos, seria necesario realizar validación cruzada (generando varias instancias de las bases de entrenamiento, prueba y validación) y luego evaluando los resultados por ejemplo a través de una prueba χ^2.

6.3.3. Interpretación de Variables

Los modelos *Benchmark*, (1), (2) y (3) son estimados con los conjuntos de atributos A y B. Para las variables que pertenecen a ambos conjuntos de atributos, se puede verificar que para las dos versiones de cada modelo (sobre los conjuntos A y B) los coeficientes estimados son consistentes en su signo, a pesar de que se pueden observar cambios relevantes en la magnitud de los coeficientes. Esta situación es deseada porque para un mismo modelo se espera que cada variable tenga la misma interpretación en términos del sentido de su

influencia en la variable de respuesta. El cambio de magnitud se puede explicar por la presencia de otras variables, las que reparten su influencia sobre la variable de respuesta.

Los modelos planteados estiman la probabilidad de que el hogar evaluado compra la marca Hunt's. Considerando que el cambio marginal de la probabilidad ante cambios en algún atributo es directamente proporcional al coeficiente del atributo, entonces el signo del atributo indica el sentido del cambio (aumenta/disminuye) y su magnitud tiene implicancias sobre la magnitud del cambio de la probabilidad. El modelo *Benchmark* representa la estimación tradicional de regresión logística. En este modelo, para ambos conjuntos de atributos, la variable de precio asociada a la variable Hunt's tiene un signo negativo, lo que indica que aumentos de precio de la marca Hunt's se traducen en una menor probabilidad de elección de esta marca. La variable de descuento en precio de Hunt's tiene un signo positivo, lo que indica que los descuentos hacen más probable la elección de esta marca. Los precios y descuentos de la marca Heinz tiene signos positivos y negativos respectivamente. Esto indica que los aumentos en el precio de Heinz se traducen en una mayor probabilidad de elección Hunt's (o equivalentemente una menor probabilidad de elección de Heinz) y aumentos en los descuentos de Heinz se traducen en menor probabilidad de elección de Hunt's. En el modelo *Benchmark* sólo está presente la variable asociada a la exposición de los hogares a comerciales de TV de la marca Heinz, pues la respectiva variable asociada a la marca Hunt's no es escogida en el proceso de selección de atributos. Esto puede ser explicado porque la marca Heinz es la que más activamente expone a los hogares a los comerciales de TV. Sin embargo, el signo de su coeficiente es positivo lo cual no resulta intuitivo: cuando los hogares han sido expuestos a los comerciales de TV de Heinz aumenta la probabilidad de elección de la marca Hunt's. A pesar de esto el atributo es significativo, con un *p-value* igual a 0 %, aportando en la predicción de las probabilidades de elección.

La tabla 6.15 resume los coeficientes estimados para los modelos *Benchmark*, (1), (2) y (3) sobre el conjunto de atributos *A*. Además se muestra el cambio porcentual del valor absoluto de los coeficientes de los modelos (1), (2) y (3) con respecto al valor absoluto de los coeficientes del modelo *Benchmark*. Se consideran los cambios en valor absoluto porque se busca comparar el cambio en magnitud de los coeficientes aislando el efecto que puedan tener sus signos.

En la tabla 6.15 se puede observar que el coeficiente asociado a la constante mantiene su signo en todos los modelos, pero aumenta su magnitud en las estimaciones realizadas para los modelos (1), (2) y (3). Las variables de precios y descuentos mantienen su signo a lo largo de las estimaciones de todos los modelos. La variable de precio de la marca Hunt's disminuye su importancia en los modelos (1), (2), (3). En cambio, la variable de precio de la marca Heinz mantiene su importancia en los modelos (1), (2) y (3), ya que presenta aumentos bien marginales. Estos ajustes en los coeficientes de precio tienen relación con que estos modelos incorporan en su estimación restricciones de un juego de competencia que incluye la variable precio. La marca Hunt's tiene en promedio precios más altos que la marca Heinz, por lo que a partir del modelo *Benchmark* se podría concluir que cambios en el precio de Hunt's podrían ser más efectivos en términos del cambio en la probabilidad de su elección. Sin embargo, los modelos (1), (2) y (3) indican que la sensibilidad de las probabilidades de elección al precio de Hunt's en realidad es más baja y que un nivel de precio promedio más alto responde a las condiciones de equilibrio del mercado en que participan.

En el caso de las variables de descuentos, en el modelo (1) las variables de descuento de ambas marcas disminuyen su importancia. El modelo (1) no incorpora el descuento como una variable que participa del juego de competencia subyacente. Sin embargo en los modelos (2) y (3), los que si consideran los descuentos

Atributo	Coeficientes				% de Cambio de abs(Coef.)		
	Benchmark	**Mod.(1)**	**Mod.(2)**	**Mod.(3)**	**Mod.(1)**	**Mod.(2)**	**Mod.(3)**
constante	-0.4174	-1.8762	-1.5447	-1.5118	350 %	270 %	262 %
P_HUNTS_Unit_Extended_Price	-2.9094	-1.5526	-1.6602	-1.6686	-47 %	-43 %	-43 %
P_HEINZ_Unit_Extended_Price	2.4380	2.5669	2.4792	2.4695	5 %	2 %	1 %
P_HUNTS_Unit_TV_Coupons	0.7725	0.4809	1.1537	1.1502	-38 %	49 %	49 %
P_HEINZ_Unit_TV_Coupons	-0.7126	-0.6652	-2.0365	-2.0442	-7 %	186 %	187 %
HUNTS_Dummy_CE				0.4616			
HEINZ_Dummy_CE	0.6418	0.5482	0.3589	-0.1729	-15 %	-44 %	-73 %
Tipo_Baja	0.4477	0.3212	0.2973	0.2906	-28 %	-34 %	-35 %
Tipo_Normal	0.3769	0.2671	0.2353	0.2312	-29 %	-38 %	-39 %
Display	0.4445	0.4040	0.3388	0.3342	-9 %	-24 %	-25 %
Advertising	-1.3579	-0.9098	-0.6996	-0.7141	-33 %	-48 %	-47 %
Type	0.8487	0.6390	0.4422	0.4536	-25 %	-48 %	-47 %
Meter_Status	-0.2239	-0.1696	-0.1401	-0.1161	-24 %	-37 %	-48 %
Number_of_Cats	1.1011	0.7920	0.6538	0.6403	-28 %	-41 %	-42 %
Food_Processor__0	-0.2856	-0.2055	-0.1737	-0.1713	-28 %	-39 %	-40 %
Food_Processor__1	-0.4448	-0.3497	-0.2879	-0.2902	-21 %	-35 %	-35 %
Curling_Iron__0	0.2411	0.1039	0.1208	0.1187	-57 %	-50 %	-51 %
Curling_Iron__1	0.7645	0.3854	0.4024	0.4062	-50 %	-47 %	-47 %

Tabla 6.15: Coeficientes para modelos con atributos *A* y cambio porcentual de su valor absoluto respecto del *Benchmark*. Fuente: Elaboración Propia.

como herramientas competitivas en los juegos subyacentes, las variables de descuentos de ambas marcas aumentan de forma importante la magnitud de sus coeficientes. En el caso de los modelos (2) y (3) la variable de precio de la marca Hunt's baja la magnitud de su coeficiente, el precio de Heinz mantiene su magnitud y los coeficientes asociados a los descuentos de ambas firmas aumentan su magnitud de forma muy importante, en particular el coeficiente del descuento de Heinz. Bajo el esquema tradicional (modelo *Benchmark*) la importancia de los descuentos se diluye junto el resto de las variables. Al incorporar los descuentos como elemento competitivo, estas variables toman un rol más protagónico en la predicción.

La variable de marketing asociada a la marca Heinz disminuye su importancia en los modelos (1) y (2). La variable de la otra marca no está presente en estos modelos. El modelo (3) es el único de los modelos propuestos que introduce las variables de marketing en el juego de competencia. Para la estimación de este modelo se introducen las variables de marketing asociadas a ambas firmas. El coeficiente de la variable asociada a la marca Heinz cambia su signo y disminuye su magnitud respecto del modelo *Benchmark*. El cambio de signo de este coeficiente corrige la interpretación de los resultados: ahora cuando el hogar ha sido expuesto a los comerciales de TV de la firma Heinz disminuye la probabilidad de elección de la firma Hunt's. De la misma forma, el signo positivo de la variable asociada a la marca Hunt's, se interpreta como que es más probable la elección de esta marca cuando los hogares han sido expuestos a comerciales de TV de Hunt's.

Para el resto de las variables se puede observar que los signos de los coeficientes se mantienen pero la magnitud de los coeficientes disminuye de forma consistente. Esto se explica porque las variables de precio, descuentos y marketing aumentan su importancia relativa a costa del resto de las variables. Es importante

Atributo	Valor absoluto estadístico Z				% de Cambio de abs(Z)		
	Benchmark	**Mod.(1)**	**Mod.(2)**	**Mod.(3)**	**Mod.(1)**	**Mod.(2)**	**Mod.(3)**
constante	1.38	13.68	11.34	11.13	888 %	720 %	704 %
P_HUNTS_Unit_Extended_Price	14.04	20.09	20.42	20.58	43 %	45 %	47 %
P_HEINZ_Unit_Extended_Price	11.58	33.05	31.39	31.30	185 %	171 %	170 %
P_HUNTS_Unit_TV_Coupons	4.96	3.66	15.82	15.81	-26 %	219 %	218 %
P_HEINZ_Unit_TV_Coupons	4.63	5.12	27.31	27.46	11 %	490 %	494 %
HUNTS_Dummy_CE				5.97			
HEINZ_Dummy_CE	5.30	5.11	3.67	2.62	-4 %	-31 %	-51 %
Tipo_Baja	6.30	5.27	5.31	5.20	-16 %	-16 %	-17 %
Tipo_Normal	5.48	4.53	4.34	4.28	-17 %	-21 %	-22 %
Display	5.58	5.93	5.37	5.31	6 %	-4 %	-5 %
Advertising	12.93	11.29	9.71	9.96	-13 %	-25 %	-23 %
Type	7.14	7.00	5.43	5.59	-2 %	-24 %	-22 %
Meter_Status	3.93	3.49	3.16	2.64	-11 %	-20 %	-33 %
Number_of_Cats	3.72	3.03	2.72	2.67	-18 %	-27 %	-28 %
Food_Processor__0	3.96	3.40	3.15	3.12	-14 %	-20 %	-21 %
Food_Processor__1	4.15	3.79	3.39	3.42	-9 %	-18 %	-18 %
Curling_Iron__0	2.45	1.26	1.59	1.57	-49 %	-35 %	-36 %
Curling_Iron__1	8.86	5.57	6.28	6.36	-37 %	-29 %	-28 %

Tabla 6.16: Valor absoluto de estadístico Z para modelos con atributos A y cambio porcentual de su valor absoluto respecto del *Benchmark*. Fuente: Elaboración Propia.

notar que todas las conclusiones anteriores, si bien son sacadas a partir de los modelos estimados sobre el conjunto de atributos A, también se aplican para el caso de las estimaciones sobre el conjunto de atributos B.

La tabla 6.16 resume el valor absoluto de los estadísticos Z estimados para los modelos *Benchmark*, (1), (2) y (3) sobre el conjunto de atributos A. Además se muestra el cambio porcentual del valor absoluto del estadístico para los modelos (1), (2) y (3) con respecto a los del *Benchmark*. El estadístico Z corresponde al valor del coeficiente ajustado por su desviación estándar y sigue una distribución normal estandarizada. Este estadístico se utiliza para probar la hipótesis nula de que el coeficiente de cada variable es igual a cero, permitiendo medir la significancia de cada variable a través del *p-value*. Entre más grande es el valor absoluto de Z entonces más bajo es el *p-value* y por lo tanto más significativa resulta la variable.

En la tabla 6.16 se pueden observar como cambian los valores absolutos del estadístico Z y por lo tanto concluir sobre como cambian los niveles de significancia de las variables. La constante aumenta su nivel de significancia para los modelos (1), (2) y (3) con respecto al modelo *Benchmark*. Las variables precios también aumentan sus niveles de significancia en todos los modelos. En el caso de las variables de descuento, en el modelo (1) se observa una disminución en su nivel de significancia para la variable de descuento de la firma Hunt's y un leve aumento en la significancia de la variable de descuento de la firma Heinz. Para los modelos (2) y (3) se observan fuertes aumentos en el nivel de significancia de las variables de descuento asociadas a ambas firmas. En el caso de las variables de marketing, la asociada a la firma Heinz disminuye su nivel de significancia en todos los modelos aunque siempre se mantiene con *p-value* bajo el 1 %. Sin embargo, cuando en el modelo (3) ingresa la variable de marketing asociada a la firma Hunt's esta tiene un *p-value* de 0 % confirmando que es significativa para predecir las probabilidades de elección. Para el resto de las

variables se observa una disminución generalizada en su significancia, pero sólo en pocos casos las variables superan un *p-value* del 5%. Estas conclusiones se han sacado a partir de los resultados de los modelos sobre el conjunto de atributos A, sin embargo se puede verificar que estas mismas conclusiones se obtienen realizando los análisis sobre las estimaciones de los modelos en el conjunto de atributos B.

Parte IV

Conclusiones

Capítulo 7

Conclusiones

7.1. Conclusiones Generales

Las firmas requieren contar con herramientas que les permitan estimar probabilidades de fuga para su cartera de clientes y así decidir sobre qué clientes concentrar sus esfuerzos de retención de clientes. Si bien este problema ha sido estudiado a través de técnicas estadísticas tradicionales, como la Regresión Logística, estos modelos no recogen el razonamiento estratégico de los clientes al momento de decidir si fugarse o no, ni el de las firmas que compiten por retener o atraer al cliente.

En este trabajo se ha abordado el desafío de estudiar el problema de predicción de fuga de clientes incorporando aspectos competitivos del mercado en que las firmas participan. El objetivo fue plantear un esquema cuantitativo que potencie una herramienta estadística a través de la Teoría de Juegos.

Se ha propuesto estimar las probabilidades de fuga de clientes a partir de la forma funcional utilizada en modelos de Regresión Logística pero imponiendo que los parámetros del modelo cumplan con las condiciones de equilibrio de un modelo de Teoría de Juegos, donde las firmas utilizan sus herramientas competitivas para atraer a los clientes y estos escogen fugarse o no según estas mismas probabilidades estimadas endógenamente. La metodología propuesta requiere de plantear primero un modelo de Teoría de Juegos que caracterice la interacción entre las firmas y los clientes, y luego encontrar las condiciones de equilibrio de este juego. Luego se debe plantear un problema de optimización no lineal cuya función objetivo corresponde a maximizar la función de verosimilitud del modelo tradicional de Regresión Logística y cuyas restricciones corresponden al sistema de ecuaciones que determinan las condiciones de equilibrio del juego. Se ha propuesto un método de solución numérica para resolver el problema de optimización asociado a la metodología planteada. Para su solución las restricciones del problema se incorporan en la función objetivo penalizando las desviaciones que se produzcan en cada una de ellas. Se han planteado tres modelos de Teoría de Juegos para abordar la estimación de probabilidades de fuga de clientes a través de este esquema: el primero, donde las firmas compiten solamente en precios; el segundo, donde las firmas compiten en precios y descuentos; y finalmente, uno donde las firmas compiten en precios, descuentos y esfuerzos en marketing.

Las metodologías han sido implementadas computacionalmente y luego se han aplicado a un problema real. Los resultados demuestran que el modelo tradicional siempre supera a los modelos propuestos en las bases de entrenamiento y prueba. Esto sucede porque los modelos propuestos imponen condiciones sobre los parámetros estimados buscando que estos además satisfagan las condiciones de equilibrio de los juegos planteados. Sin embargo, la brecha existente en la base de entrenamiento se reduce notablemente en la base validación. En la base de validación, que se encuentran fuera del periodo temporal utilizado para estimar los modelos, se observa que los modelos propuestos tienen desempeños superiores al modelo tradicional explicado principalmente por su capacidad de comprender de forma más estructural el fenómeno estudiado.

Los modelos propuestos sacrifican precisión en la estimación de sus parámetros −reflejada en los peores resultados de los estadísticos en las bases de entrenamiento y la base de prueba− para mejorar su capacidad de entender el comportamiento de compra racional de los hogares −situación reflejada en los mejores estadísticos en la base de validación. Las restricciones de equilibrio recogen efectos de largo plazo de la base de datos, que quedan implícitos en el $\hat{b}$ estimado y que se pierden en la estimación tradicional de regresión logística. Al darle más importancia a comportamientos estables en el tiempo y consistentes con la racionalidad de los jugadores, se tiene que los modelos propuestos mejoran la capacidad de predicción en las bases de validación. Por otro lado, los modelos propuestos modifican la importancia relativa de las variables en la predicción de las probabilidades de elección de los clientes de acuerdo al juego de competencia que enfrentan. Esto permite mejorar la interpretación de las variables resultantes y su importancia para entender las elecciones de los clientes. Adicionalmente se observa que las variables que son parte del juego planteado, mejoran de forma importante sus niveles de significancia.

7.2. Conclusiones Específicas

7.2.1. Modelamiento del Proceso de Fuga de Clientes

En esta tesis ha sido posible estudiar y luego caracterizar el proceso de fuga de clientes a través de modelos de Teoría de Juegos. El problema de fuga de clientes puede entenderse como un problema de competencia donde las firmas atraen a los clientes de las otras firmas y realizan esfuerzos por retener a sus propios clientes. Los clientes, por otra parte deben decidir la firma que escogerán en cada periodo, evaluando los precios y descuentos que ofrecen las firmas además de las preferencias que han construido por cada marca. En el capítulo 3 se realiza una revisión completa de la literatura de Teoría de Juegos donde se caracteriza este tipo de problemas. Esta revisión de literatura permitió al autor de esta tesis proponer un modelo de Teoría de Juegos que estudia un mercado donde las firmas compiten en precios y en esfuerzos de marketing que definen las lealtades de los consumidores. Este trabajo fue presentado en una conferencia internacional y es citado en esta tesis como parte de la revisión bibliográfica. Si bien todos los modelos estudiados en el capítulo 3 permiten entender muy bien el problema microeconómico de las firmas y de los clientes, estos resultan difíciles de aplicar debido al alto componente teórico que presentan.

En la metodología propuesta en el capítulo 5 se utilizan modelos de Teoría de Juegos que caracterizan el problema de fuga de clientes de una forma más simplificada que los planteados en el capítulo 3. Se asume la simplificación de fenómeno para ganar en términos de su aplicabilidad al problema real, lo cuál se consigue

con éxito al aplicarlo a un problema real que considera la interacción de las firmas con miles de clientes. De acuerdo a los resultados obtenidos, estos modelos simplificados de Teoría de Juego le entregan al modelo estadístico la capacidad de comprender de forma estructural el fenómeno económico subyacente.

7.2.2. Probabilidad de Fuga de Clientes desde el Enfoque Tradicional

En el capítulo 4 se plantea el problema de Clasificación Binaria y en particular la metodología de Regresión Logística. Esta metodología es analizada profundamente detallando sus fundamentos, interpretación y su proceso de estimación. Así mismo, en el Anexo D se definen los principales indicadores y estadísticos utilizados para evaluar los resultados de este tipo de modelos.

Para evaluar la capacidad de los modelos propuestos, además de resolverlos se ha utilizado como *Benchmark* el enfoque tradicional de Regresión Logística. Respecto del modelo tradicional se puede concluir sobre la importancia del proceso de selección de atributos. Para resolver los modelos se han utilizado dos enfoques de selección de atributos que entregan dos conjuntos de atributos alternativos. Los mejores resultados en términos de los estadísticos se consiguen con el conjunto de atributo que utiliza el mayor número de variables. Sin embargo, la bondad de los modelos con un menor número de atributos radica en su parsimonia y su mayor facilidad en la interpretación.

7.2.3. Probabilidades de Fuga de Clientes según Metodología Propuesta

En el capítulo 5 se plantea la metodología propuesta en esta tesis, que incorpora conceptos de competencia y Teoría de Juegos en un modelo estadístico basado en Regresión Logística. Primero se propone el esquema general, el cual puede ser aplicado a cualquier problema de Clasificación Binaria donde exista de forma subyacente una situación donde los agentes interactúan de forma estratégica. Se propone también el enfoque de solución numérica general a través del método de Newton-Raphson, donde se concluye que para su estimación se debe conocer el gradiente y hessiano del problema. Posteriormente se plantean tres modelos de Teoría de Juegos para abordar la estimación de probabilidades de fuga de clientes a través de este esquema. Para cada uno de estos modelos se determinan las condiciones de equilibrio, se plantea el problema de optimización a resolver y luego se despejan las expresiones algebraicas que determinan su gradiente y hessiano.

La elección de la forma funcional de la probabilidad a través del modelo de Regresión Logística favorece la solución algebraica del modelo de teoría de juegos y luego del problema de optimización no lineal. En particular, ésta permite encontrar formulas cerradas para el gradiente y para el hessiano de cada problema, aunque estos tenga expresiones que no dejan de ser complejas.

Una de las desventajas del enfoque utilizado es que se requiere un desarrollo complejo para caracterizar el problema estudiado. La formulación y los esquemas de estimación de un modelo de regresión logística se pueden aplicar a cualquier problema de Clasificación Binaria. Sin embargo, la metodología propuesta requiere de complejos desarrollos matemáticos para caracterizar en particular el problema de predicción de fuga de clientes. Si se busca aplicar este esquema a otro problema de Clasificación Binaria se deberá pro-

poner primero un modelo de Teoría de Juegos, luego resolver sus condiciones de equilibrio, considerar estas condiciones como restricciones del problema de optimización que se requiere, relajar estas restricciones considerándolas como penalizaciones de la función objetivo y recién ahí despejar algebraicamente el gradiente y hessiano del problema. Sin embargo, utilizando este esquema existen ganancias en la correcta interpretación de las variables que explican las elecciones de los agentes y en la capacidad del modelo obtenido de adaptarse a cambios en las condiciones de mercado al recoger de forma implícita la racionalidad de los jugadores.

7.2.4. Aplicación de las Metodologías Propuestas a un Caso Real

En el capítulo 6 se aplican los modelos propuestos sobre un caso real. Para esto se utilizan datos de escáner de un panel de consumidores sobre un mercado de Ketchups. Una de las debilidades principales de las metodologías propuestas es suponer que sólo existen dos firmas en el mercado. En este sentido, para su aplicación se ha analizado la competencia existente entre las dos mayores firmas del mercado, quienes representan el 75 % de las compras de Ketchup en el formato de 32 oz. A partir de los datos disponibles se han podido identificar datos de precios, descuentos en precios, esfuerzos en marketing y datos demográficos de los hogares que realizan las compras. Los esfuerzos en marketing se construyen identificando los hogares que al momento de su compra han sido expuestos previamente a comerciales de TV de cada marca.

Los resultados permiten observar que en general todos los modelos tienen un mejor desempeño al utilizar el conjunto de atributos *B*, que corresponde al conjunto con mayor número de variables. Al utilizar el conjunto *B* en lugar del *A* los modelos tienen un mejor KS en aproximadamente un 9 % promedio, un mejor AUC en aproximadamente un 3 % promedio y un MSE más bajo en aproximadamente un 3.5 % promedio. El mayor número de variables permite recoger de forma más completa el comportamiento y las decisiones de compra de los hogares. Esto quiere decir que el proceso de selección de atributos es fundamental en la construcción de los modelos, sobretodo en la búsqueda del equilibrio entre su capacidad de predicción y el nivel de parsimonia. De acuerdo a los resultados obtenidos se puede concluir que al utilizar el criterio de selección de atributos bajo el criterio BIC se consigue disminuir la complejidad del modelo en términos del número de atributos, sin embargo esto se traduce en pérdidas relevantes del poder predictivo del modelo con respecto a la utilización del criterio AIC.

Los modelos que utilizan el esquema propuesto presentan resultados peores que el modelo *benchmark* en la base de entrenamiento y en la base de prueba. En efecto, estos pueden tener KS menores hasta en un 16.1 %, AUC menores hasta en un 4.0 % y MSE mayores hasta en un 5.1 % en estas bases. Las restricciones incorporadas en los modelos propuestos, provocan que inevitablemente las soluciones factibles disminuyan con respecto al caso sin restricciones y que por lo tanto empeore la calidad de sus estadísticos. Como el modelo *benchmark* no impone restricciones a sus soluciones, entonces el conjunto de soluciones factibles de cualquier modelo con restricciones es un subconjunto del conjunto de soluciones factibles del modelo *benchmark*.

Sin embargo, los modelos propuestos tienen mucho mejor rendimiento fuera de la base de entrenamiento. Los mejores resultados se obtienen particularmente en la base de validación, que corresponde a la base que se encuentra fuera del periodo temporal utilizado en la estimación de los modelos. Los resultados de KS mejoran hasta en un 10.2 %, el AUC puede mejorar hasta en un 3.3 % y el MSE puede disminuir hasta en un

4.4 %. Las restricciones de equilibrio recogen efectos de largo plazo de la base de datos, dándole más importancia a comportamientos que se encuentran alineados con la racionalidad de los jugadores, permitiendo que los modelos propuestos mejoren su capacidad de predicción en las bases de validación.

Por otro lado, los modelos propuestos permiten realizar un mejor análisis sobre la capacidad de predecir las probabilidades de elección de los hogares. Cuando se incorporan las variables de precio, descuentos o esfuerzos en marketing como herramientas competitivas para las firmas (se incorporan en los juegos de competencia subyacentes), entonces se generan cambios relevantes sobre los parámetros de las variables y su significancia. De acuerdo a los resultados obtenidos, los modelos propuestos regulan la importancia de los precios de ambas firmas, además de entregar un rol más protagónico a las variables de descuentos en relación al modelo *Benchmark*. En el caso de los esfuerzos en marketing, los modelos que no la utilizan como herramienta competitiva entregan resultados que son contradictorios con la intuición económica. Sin embargo, las estimaciones del modelo (3) corrigen esta situación entregando la correcta interpretación de las variables. Para los modelos propuestos en este trabajo, todas las variables que se utilizan como herramientas competitivas en los respectivos juegos de los modelos (1), (2) y (3) mejoran de forma importante su nivel de significancia.

La aplicación de las metodologías ha sido exitosa. Primero porque ha sido posible construir un modelo tradicional para caracterizar las probabilidades de elección de los consumidores. Segundo porque fue posible estimar los modelos propuestos sobre una base de datos que contiene las compras de miles de hogares, comprobando la aplicabilidad de estas metodologías. Finalmente, porque los resultados obtenidos con los modelos propuestos son competitivos respecto del enfoque tradicional. Esto sobretodo considerando que los modelos propuestos funcionan de mejor manera fuera de la muestra de calibración y en especial fuera del periodo temporal utilizado en la estimación.

7.2.5. Implementación de las Metodologías Propuestas

En el Anexo F se detalla la estrategia de implementación computacional utilizada para las metodologías propuestas. Los códigos desarrollados consideran una rutina del método de Newton-Raphson y las versiones del gradiente y hessiano de cada problema de optimización a resolver.

La implementación propia de la estimación del modelo de regresión logística tradicional entrega los mismos resultados, en términos de los parámetros y estadísticos, que cualquier otro paquete estadístico.

Una de las dificultades que se debió sortear fue la complejidad del cálculo del hessiano y de su matriz inversa en los casos de muestras con un alto número de registros. Para esto se propuso un esquema de cálculo por partes que permitió disminuir los tiempos de proceso y las necesidad de memoria virtual para el computo del hessiano.

Un elemento no abordado en este trabajo fue la implementación de técnicas numéricas para encontrar cuál es el mejor parámetro μ y e del problema. El parámetro μ penaliza en la función objetivo las desviaciones de la solución con respecto a las condiciones de equilibrio del juego planteado en cada modelo. El parámetro e corresponde al costo marginal de los esfuerzos en marketing que realizan las firmas y que se utiliza en el tercero de los modelos propuestos.

7.3. Desafíos Futuros

Este trabajo propone una primera aproximación de determinación endógena de las probabilidades de fuga de clientes de un modelo estadístico con consideraciones de competencia. Este enfoque puede ser mejorado en diferentes ámbitos, así como también puede ser extendido a otros problemas de Clasificación donde los agentes tienen comportamientos estratégicos. A continuación se proponen diferentes líneas de trabajo que nacen a partir de esta tesis.

- **Utilizar la metodología propuesta en otros problemas de Clasificación.** Como ya fue mencionado, el esquema propuesto en el capítulo 5 es general y podría ser aplicado a otros problemas de clasificación donde existan interacciones estratégicas entre los agentes participantes. En el problema de detección de fraudes existe un fiscalizador intentando identificar las operaciones fraudulentas de las que no lo son. Por otro lado existe un delincuente intentando realizar el fraude y para esto trata utilizar diversas técnicas para persuadir al fiscalizador. En el problema de *credit scoring* una entidad financiera busca máximizar sus utilidades a través de la entrega de créditos a los clientes que son buenos pagadores y rechazando las solicitudes de los malos pagadores. Por otro lado, los clientes buscan que la entidad financiera los vea como buenos pagadores para maximizar la probabilidad de que el crédito les sea otorgado. En estos ejemplos existe una interacción estratégica entre los agentes, quienes actúan racionalmente en la búsqueda de maximizar sus utilidades. Por lo tanto para estimar las probabilidades de fraude o de que un cliente sea buen pagador, se puede utilizar la metodología propuesta en este trabajo incorporando en el modelo estadístico las condiciones de equilibrio de los respectivos modelos de Teoría de Juegos.

- **Extender la metodología a un modelo de Clasificación Multiclase.** Una de las desventajas de la metodología propuesta en el capítulo 5 es que al utilizar una forma funcional de Regresión Logística es aplicable sólo a problemas de Clasificación Binaria. Esto genera limitaciones importantes a la hora de buscar potenciales aplicaciones. Por ejemplo, en el caso aplicado estudiado en este trabajo existían más de dos firmas en el mercado por lo que los datos tuvieron que adecuarse para ser abordados a través de la metodología propuesta. En este sentido se propone extender la metodología a una forma funcional que permita abordar un problema de Clasificación Multiclase. Se propone abordar este desafío utilizado como base el modelo de elección discreta Logit Multinomial, el cuál es una generalización del modelo de Regresión Logística y mantiene las bondades en el tratamiento algebraico. Este último punto es relevante porque permite encontrar formulas cerradas para el gradiente y el hessiano del problema.

- **Proponer una técnica de estimación de los parámetros μ y e.** Para resolver los modelos propuestos en el capítulo 5 se utilizan estos dos parámetros. El parámetro μ penaliza en la función objetivo las desviaciones de la solución con respecto a las condiciones de equilibrio del juego plantado en cada modelo. El parámetro e corresponde al costo marginal de los esfuerzos en marketing que realizan las firmas y que se utiliza en el tercero de los modelos propuestos. No fue el objetivo de esta tesis cuestionar o proponer una forma de determinar cuál es el mejor parámetro de penalización μ, ni cómo estimar el costo marginal de los esfuerzos en marketing e. En el caso del parámetro μ se pueden utilizar grillas de valores alternativos y luego encontrar para cuál de los valores propuestos se obtienen los mejores estadísticos en las bases de entrenamiento, prueba y validación. Un elemento que también

debe ser estudiado en este sentido es la sensibilidad y la estabilidad de los resultados obtenidos a estos parámetros.

- **Métodos alternativos de solución del problema de optimización.** Para resolver los modelos propuestos en el capítulo 5, el problema de optimización no lineal con restricciones es relajando incorporando en la función objetivo penalizaciones por el no cumplimiento de cada una de las restricciones. Siempre es posible encontrar metodologías de solución de este tipo de problemas de optimización que sean más adecuadas para el tipo de función objetivo o de restricciones. En este sentido, una línea de trabajo interesante es estudiar la solución numérica de los tipos de problemas de optimización resultantes a través de la metodología propuesta. En esta misma línea, para el cálculo de la matriz hessiana de cada problema de optimización, así como para calcular su inversa, se propuso un método de cálculo por partes. Como trabajo futuro se recomienda explorar la utilización del paquete *spam* en el lenguaje R propuesto por Furrer y Sain (2010), donde se plantean una serie de rutinas eficientes para el manejo de matrices de gran tamaño.

- **Mejorar los modelos de Teoría de Juegos utilizados.** Sin duda una línea de trabajo constante es mejorar los modelos de Teoría de Juegos utilizados para determinar las condiciones de equilibrio que deben satisfacer los parámetros del modelo estadístico. Existen dos aspectos que sin dudas resultan muy interesantes de estudiar tanto del punto de vista teórico como del aplicado. El primero, es estudiar cómo influye en el nivel de competencia del mercado las restricciones presupuestarias que las firmas tienen para los esfuerzos en marketing que destinan para atraer o retener a los consumidores. El segundo, es la incorporación de asimetrías en los parámetros utilizados en los modelos propuestos. La mayoría de los juegos planteados en los capítulos 3 y 5 utilizan parámetros simétricos para las firmas de forma de simplificar su resolución. Esto supone que las firmas tienen habilidades equivalentes, cuando en realidad las firmas son diferentes y cada una posee diferentes ventajas competitivas. Por ejemplo, el parámetro θ del modelo propuesto en el capítulo 3.4 representa el nivel base de lealtad de los clientes hacia las firmas y es natural plantear que cada firma posee diferentes niveles de lealtad base: θ_A y θ_B. De la misma forma, el parámetro e del modelo propuesto en el capítulo 5 representa el costo marginal de los esfuerzos en marketing que realizan las firmas y es natural plantear que cada firma tenga estructuras de costos diferentes: e_A y e_B.

Bibliografía

BITRAN, GABRIEL R. y MONDSCHEIN, SUSANA V.: «Mailing Decisions in the Catalog Sales Industry». *Management Science*, 1996, **42(9)**, pp. 1364–1381.

BOLTON, RUTH N.: «A Dynamic Model of the Duration of the Customer's Relationship with a Continuous Service Provider: The Role of Satisfaction». *Marketing Science*, 1998, **17(1)**, pp. 45–65.

BOULDING, WILLIAM; STAELIN, RICHARD; EHRET, MICHAEL y JOHNSTON, WESLEY J.: «A Customer Relationship Management Roadmap: What Is Known, Potential Pitfalls, and Where to Go». *Journal of Marketing*, 2005, **69**, p. 155–166.

BUTTLE, FRANCIS: *Customer Relationship Management*. Butterworth Heinemann. Butterworth-Heinemann, 2008.

CAMINAL, RAMON y MATUTES, CARMEN: «Endogenous switching costs in a duopoly model». *International Journal of Industrial Organization*, 1990, **8(3)**, pp. 353–373.

CHEN, YONGMIN: «Paying Customers to Switch». *Journal of Economics & Management Strategy*, 1997, **6(4)**, pp. 877–897.

CHEN, YUXIN; NARASIMHAN, CHAKRAVARTHI y ZHANG, Z. JOHN: «Individual Marketing with Imperfect Targetability». *Marketing Science*, 2001, **20(1)**, pp. 23–41.

CHIANG, DING-AN; WANG, YI-FAN; LEE, SHAO-LUN y LIN, CHENG-JUNG: «Goal-oriented sequential pattern for network banking churn analysis». *Expert Systems with Applications*, 2003, **25(3)**, pp. 293–302.

COUSSEMENT, KRISTOF y DEN POEL, DIRK VAN: «Churn prediction in subscription services: An application of support vector machines while comparing two parameter-selection techniques». *Expert Systems with Applications*, 2008, **34(1)**, pp. 313 – 327.

DEBONIS, J.N.; BALINSKI, E.W. y ALLEN, P.: *Value-based marketing for bottom-line success: 5 steps to creating customer value*. An American Marketing Association Title Series. McGraw-Hill, 2003.

DWYER, F. ROBERT: «Customer lifetime valuation to support marketing decision making». *Journal of Direct Marketing*, 1997, **11(4)**, pp. 6–13.

FISCHER, RONALD: *Curso de Organización Industrial*. CEA-DII, Universidad de Chile, 2006.

FRUCHTER, GILA E. y ZHANG, Z. JOHN: «Dynamic Targeted Promotions: A Customer Retention and Acquisition Perspective». *Journal of Service Research*, 2004, **7(1)**, pp. 3–19.

FUDENBERG, DREW y TIROLE, JEAN: *Game Theory*. MIT Press, 1991. ISBN 9780262061414.

——: «Customer poaching and brand switching». *Journal of Economics*, 2000, **31(4)**, pp. 634–657.

FURRER, REINHARD y SAIN, STEPHAN R.: «spam: A Sparse Matrix R Package with Emphasis on MCMC Methods for Gaussian Markov Random Fields». *Journal of Statistical Software*, 2010, **36(10)**, pp. 1–25. ISSN 1548-7660.

GIBBONS, R.: *Game Theory for Applied Economists*. Princeton paperbacks. Princeton University Press, 1992. ISBN 9780691003955.

GLADY, N; BAESENS, BART y CROUX, CHRISTOPHE: «Modeling customer loyalty using customer lifetime value». *Open access publications from katholieke universiteit leuven*, Katholieke Universiteit Leuven, 2006.

GLADY, NICOLAS; BAESENS, BART y CROUX, CHRISTOPHE: «Modeling churn using customer lifetime value». *European Journal of Operational Research*, 2009, **197(1)**, pp. 402 – 411.

GRAND, PHILIP: «Achieving excellent customer relationships». *Informe técnico*, BWCS Ltd, 2002.

GUPTA, SUNIL; HANSSENS, DOMINIQUE; HARDIE, BRUCE; KAHN, WILIAM; KUMAR, V.; LIN, NATHANIEL; RAVISHANKER, NALINI y SRIRAM, S.: «Modeling Customer Lifetime Value». *Journal of Service Research*, 2006, **9(2)**, pp. 139–155.

GUPTA, SUNIL y STEENBURGH, THOMAS J.: «Allocating Marketing Resources». *Marketing Mix Decisions: New Perspectives and Practices, Roger A. Kerin and Rob O'Regan, eds., American Marketing Association, Chicago, IL, 2008*, 2005.

HADDENA, JOHN; TIWARIA, ASHUTOSH; ROYA, RAJKUMAR y RUTAB, DYMITR: «Computer assisted customer churn management: State-of-the-art and future trends.» *Computers & Operations Research*, 2007, **34(10)**, pp. 2902–2917.

HAENLEIN, MICHAEL; KAPLAN, ANDREAS M y SCHODER, DETLEF: «Valuing the Real Option of Abandoning Unprofitable Customers When Calculating Customer Lifetime Value». *Journal of Marketing*, 2006, **70(3)**, pp. 5–20.

HOSMER, D.W. y LEMESHOW, S.: *Applied Logistic Regression*. Wiley Series in Probability and Statistics: Texts and References Section. Wiley, 2004. ISBN 9780471356325.

HOTELLING, HAROLD: «Stability in Competition». *Economic Journal*, 1929, **39**, pp. 41–57.

HSIEH, NAN-CHEN: «An integrated data mining and behavioral scoring model for analyzing bank customers». *Expert Systems with Applications*, 2004, **27(4)**, pp. 623–633.

HUNG, SHIN-YUAN; YEN, DAVID C. y WANG, HSIU-YU: «Applying data mining to telecom churn management». *Expert Systems with Applications*, 2006, **31(3)**, pp. 515 – 524.

JONKER, JEDID-JAH; PIERSMA, NANDA y DEN POEL, DIRK VAN: «Joint optimization of customer segmentation and marketing policy to maximize long-term profitability». *Expert Systems with Applications*, 2004, **27(2)**, pp. 159 – 168.

KIM, BYUNG-DO; SHI, MENGZE y SRINIVASAN, KANNAN: «Reward Programs and Tacit Collusion». *Marketing Science*, 2001, **20(2)**, pp. 99–120.

KLEMPERER, PAUL: «The Competitiveness of Markets with Switching Costs». *The RAND Journal of Economics*, 1987a, **18(1)**, pp. 138–150.

——: «Markets with Consumer Switching Costs». *The Quarterly Journal of Economics*, 1987b, **102(2)**, pp. 375–394.

——: «Competition when Consumers have Switching Costs: An Overview with Applications to Industrial Organization, Macroeconomics, and International Trade». *Review of Economic Studies*, 1995, **62**, pp. 515–539.

LAROSE, DANIEL T.: *Data Mining Methods and Models*. Willey-Interscience, 2006.

LEE, JONGSU; KIM, YEONBAE; LEE, JEONG-DONG y PARK, YURI: «Estimating the extent of potential competition in the Korean mobile telecommunications market: Switching costs and number portability». *International Journal of Industrial Organization*, 2006, **24(1)**, pp. 107 – 124.

LEMON, KATHERINE N.; WHITE, TIFFANY BARNETT y WINER, RUSSELL S.: «Dynamic Customer Relationship Management: Incorporating Future Considerations into the Service Retention Decision». *The Journal of Marketing*, 2002, **66(1)**, pp. 1–14.

LEWIS, MICHAEL: «Incorporating Strategic Consumer Behavior into Customer Valuation». *The Journal of Marketing*, 2005, **69(4)**, pp. 230–238.

LÓPEZ, LEONARDO; FIGUEROA, NICOLÁS y WEBER, RICHARD: «Competencia en Precios y Esfuerzos de Marketing en Modelos Dinámicos». XIV Latin-Ibero-American Congress on Operation Research (CLAIO 2008). Cartagena de Indias, Colombia., 2008.

MANSKI, CHARLES F.: «The structure of random utility models». *Theory and Decision*, 1977, **8**, pp. 229–254. ISSN 0040-5833.

MCFADDEN, DANIEL: «Conditional logit analysis of qualitative choice behavior». *Frontiers in Econometrics*, 1974, pp. 105–142.

——: «The Revealed Preferences of a Government Bureaucracy: Empirical Evidence». *The Bell Journal of Economics*, 1976, **7(1)**, pp. 55–72.

MIRANDA, JAIME; REY, PABLO y WEBER, RICHARD: «Predicción de Fugas de Clientes para una Institución Financiera mediante Support Vector Machines». *Revista Ingeniería de Sistemas*, 2005, **XIX**.

MIZUNO, MAKOTO; SAJI, AKIRA; SUMITA, USHIO y SUZUKI, HIDEO: «Optimal threshold analysis of segmentation methods for identifying target customers». *European Journal of Operational Research*, 2008, **186(1)**, pp. 358 – 379.

MUSALEM, ANDRÉS y JOSHI, YOGESH: «How Much Should You Invest In Each Customer Relationship? A Competitive Strategic Approach». *Forthcomming in Marketing Science*, 2007.

NESLIN, SCOTT A; GUPTA, SUNIL; KAMAKURA, WAGNER; LU, JUNXIANG y MASON, CHARLOTTE H: «Defection Detection: Measuring and Understanding the Predictive Accuracy of Customer Churn Models». *Journal of Marketing Research*, 2006, **43(2)**, pp. 204–211.

NGAI, E.W.T.; XIU, LI y CHAU, D.C.K.: «Application of data mining techniques in customer relationship management: A literature review and classification». *Expert Systems with Applications*, 2009, **36(2, Part 2)**, pp. 2592 – 2602.

OSBORNE, MARTIN J. y RUBINSTEIN, ARIEL: *A Course in Game Theory*. The MIT Press, first[a] edición, 1994. ISBN 0262650401.

PANDA, TAPAN K.: «Creating Customer Life Time Value Through Effective CRM in Financial Services Industry». *Journal of Services Research*, 2003, **2(2)**.

PEPPERS, D. y ROGERS, M.: *Managing Customer Relationships: A Strategic Framework*. John Wiley & Sons, 2011.

PRINZIE, ANITA y DEN POEL, DIRK VAN: «Investigating purchasing-sequence patterns for financial services using Markov, MTD and MTDg models». *European Journal of Operational Research*, 2006, **170(3)**, pp. 710–734.

R DEVELOPMENT CORE TEAM: *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008. ISBN 3-900051-07-0.
`http://www.R-project.org`

REICHHELD, FREDERICK F. y SASSER, W. EARL: «Zero Defections: Quality Comes to Services». *Harvard Business School Review*, 2000, **5**, pp. 105–111.

REIKO, AOKI y JOHN, SMALL: «The Economics of Number Portability: Switching Costs and Two-Part Tariffs». *PIE/CIS Discussion Paper 483*, Center for Intergenerational Studies, Institute of Economic Research, Hitotsubashi University, 2010.

ROSSET, SAHARON; NEUMANN, EINAT; EICK, URI y VATNIK, NURIT: «Customer Lifetime Value Models for Decision Support». *Data Mining and Knowledge Discovery*, 2003, **7**, pp. 321–339.

RYALS, LYNETTE J y KNOX, SIMON: «Measuring risk-adjusted customer lifetime value and its impact on relationship marketing strategies and shareholder value». *European Journal of Marketing*, 2005, **39(5-6)**, pp. 456–472.

SHAFFER, GREG y ZHANG, Z. JOHN: «Pay to Switch or Pay to Stay: Preference-Based Price Discrimination in Markets with Switching Costs». *Journal of Economics & Management Strategy*, 2000, **9(3)**, pp. 397–424.

——: «Competitive One-to-One Promotions». *Management Science*, 2002, **48(9)**, pp. 1143–1160.

SHI, MENGZE; CHIANG, JEONGWEN y RHEE, BYONG-DUK: «Price Competition with Reduced Consumer Switching Costs: The Case of Wireless Number Portability in the Cellular Phone Industry». *Management Science*, 2006, **52(1)**, pp. 27–38.

THOMPSON, H.: *Who stole my customer?: winning strategies for creating and sustaining customer loyalty*. Pearson Prentice Hall, 2004.

TRAIN, KENNETH: *Discrete Choice Methods with Simulation*. Discrete Choice Methods with Simulation. Cambridge University Press, 2003. ISBN 9780521017152.

VAN DEN POEL, DIRK y LARIVIÈRE, BART: «Customer attrition analysis for financial services using proportional hazard models». *European Journal of Operational Research*, 2004, **157(1)**, pp. 196–217.

VAPNIK, VLADIMIR: *The Nature of Statistical Learning Theory*. Statistics for Engineering and Information Science. Springer, 2ª edición, 2000.

VENABLES, BILL y RIPLEY, BRIAN: *Modern applied statistics with S*. Statistics and computing. Springer, 2002. ISBN 9780387954578.

VON NEUMANN, JOHN y MORGENSTERN, OSKAR: *Theory of games and economic behavior*. Princeton university press, 1944.

ZHAO, YU; LI, BING; LI, XIU; LIU, WENHUANG y REN, SHOUJU: «Customer Churn Prediction Using Improved One-Class Support Vector Machine». En: Xue Li; Shuliang Wang y Zhao Dong (Eds.), *Advanced Data Mining and Applications*, volumen 3584 de *Lecture Notes in Computer Science*, pp. 731–731. Springer Berlin / Heidelberg, 2005.

Anexo A

Conceptos de Teoría de Juegos

A.1. El Dilema del Prisionero

A.1.1. Definición del Juego

A continuación se presenta a modo de ejemplo uno de los juegos más estudiados en al literatura de Teoría de Juegos: *El Dilema del Prisionero*.

Este juego se origina cuando dos personas son arrestados por un crimen. La policía carece de evidencia suficiente para condenar a cualquiera de los sospechosos, y por lo tanto necesitan que alguno de ellos confiese el crimen. La policía pone a ambos sospechosos en celdas diferentes para evitar que se puedan comunicar. La policía habla por separado con ellos explicándoles que si confiesan serán dejados en libertad, siempre y cuando el otro sospechoso no haya confesado. Si ninguno confiesa, entonces ambos sólo estarán 1 año en la cárcel. Si sólo uno de ellos confiesa, entonces el otro irá a la cárcel por 10 años; si ambos confiesan el crimen, entonces ambos irán a la cárcel por 9 años (recibiendo una disminución de 1 año en su condena por declarar). Ambos sospechosos, en adelante los jugadores, deben escoger simultáneamente entre sus dos acciones posibles: confesar(C) o no confesar (NC). En el juego los pagos que reciben los jugadores corresponden a los años de cárcel que recibirían en cada situación. Si ambos jugadores confiesan (C) entonces recibirán un pago de -9 cada uno. Si ambos jugadores no confiesan (NC) entonces recibirán un pago de -1 cada uno. Si uno de ellos confiesa (C) y el otro no confiesa (NC), entonces el primero recibe un pago de 0 y el segundo recibe un pago de -10.

La situación anterior supone que los jugadores sólo toman en cuenta los años que puedan pasar en la cárcel. En caso que los jugadores valoren otros elementos, como por ejemplo la amistad y la lealtad, entonces será necesario plantear otro esquema de pagos para cada combinación de acciones.

A.1.2. Representación en Forma Estratégica o Normal

En la tabla A.1 se representa el juego *El Dilema del Prisionero* en su forma estratégica o normal. En las filas de la tabla se encuentran las alternativas de acción del jugador 1. En las columnas se encuentran las alternativas de acción del jugador 2. En cada coordenada de la tabla se encuentra un par ordenado que representa los pagos que reciben el jugador 1 y el jugador 2 respectivamente, según sean las acciones que escojan los jugadores. Por ejemplo, si el jugador 1 escoge confesar (C) y el jugador 2 escoge no confesar (NC) entonces el jugador 1 recibe un pago de 0 y el jugador 2 recibe un pago de -10.

Jugador 1 \ Jugador 2	**C**	**NC**
C	-9,-9	0,-10
NC	-10,0	-1,-1

Tabla A.1: Juego *Dilema del Prisionero* en su forma estratégica o normal

A.1.3. Representación en Forma Extensiva

En la figura A.1 se representa como ejemplo el juego del *Dilema del Prisionero* en su forma extensiva a través de un árbol. El primer nodo corresponde al jugador 1 quien tiene dos acciones posibles: confesar (C) o no confesar (NC). Los siguientes dos nodos corresponden al jugador 2, quien nuevamente tiene dos acciones posibles: confesar (C) o no confesar (NC). La línea punteada entre los dos nodos del jugador 2 significa que ambos nodos pertenecen al mismo conjunto de información, y que por lo tanto, cuando le toca decidir al jugador 2 no puede distinguir en cuál de los dos nodos se encuentra: no sabe cuál fue la acción que tomo el jugador 1. Esto quiere decir que ambos jugadores toman sus decisiones de forma simultánea. En las hojas del árbol se encuentran los pagos que reciben los jugadores: arriba el pago del jugador 1 y luego el pago del jugador 2. Por ejemplo, si el jugador 1 escoge confesar (C) y el jugador 2 escoge no confesar (NC) entonces el jugador 1 recibe un pago de 0 y el jugador 2 recibe un pago de -10.

A.1.4. Equilibrios en el Juego

Analizando el caso del *Dilema del Prisionero*, consideremos la situación en que se encuentra el jugador 1. Supongamos que este jugador cree que el jugador 2 no confesará (NC): si el jugador 1 decide confesar (C) entonces quedará en libertad, lo que es preferible a la opción de no confesar (NC) que se traduce en 1 año de cárcel (ya que se está suponiendo que el otro jugador escoge la acción NC). Si el jugador 1 cree que, por el contrario, el jugador 2 confesará (C): si el jugador 1 decide confesar (C) entonces deberá pasar 9 años en la cárcel, lo que es mejor que asumir completamente las culpas y pasar 10 años de cárcel al escoger no confesar (NC). Por lo tanto, sin importar lo que decida el jugador 2, el jugador 1 estará mejor si decide confesar (C) transformando esta alternativa en su estrategia dominante. El mismo análisis se puede realizar para el jugador 2 concluyendo que, dada la simetría del juego, la alternativa de confesar (C) es su estrategia dominante sin importar lo que hace el primer jugador. En conclusión, la estrategia donde ambos jugadores confiesan (C) es

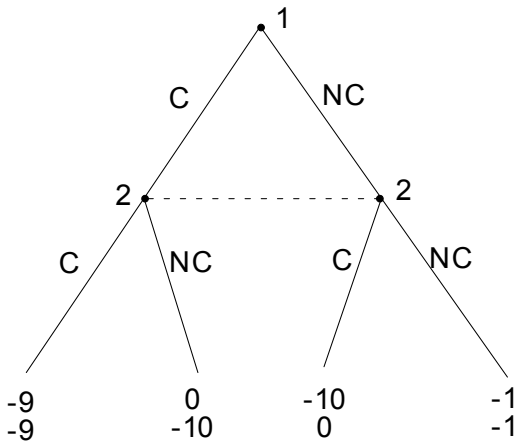

Figura A.1: Juego *Dilema del Prisionero* en su forma extensiva

un equilibrio en estrategias dominantes. Lo notable de este juego es que ambos jugadores deciden confesar (C), teniendo que pagar con 9 años de cárcel cada uno, a pesar de que si hubiesen cooperado tomando ambos la decisión de no confesar (NC), habrían terminado en una mejor posición pagando sólo con 1 año de cárcel cada uno. Este dilema que enfrentan ambos jugadores es el que le da el nombre al tradicional juego. En la literatura se hace referencia al *Dilema del Prisionero* cuando los jugadores de cualquier juego enfrentan la disyuntiva de que cooperando con el resto de los jugadores podrían estar mejor, pero al no representar ésta una situación de equilibrio deciden no cooperar terminando en una situación peor.

Anexo B

Retención de Clientes y Competencia

B.1. Competencia con Discriminación en Precios y Costos de Cambio

Chen (1997) plantea un juego entre dos firmas (A y B), que producen un producto homogéneo con costo marginal $c \geq 0$. Existe un continuo de consumidores de tamaño uno, cada uno de los cuales posee una demanda unitaria por periodo con un valor de reserva R, donde $R > c + \theta$ y con $\theta > 0$. En el primer periodo, las firmas compiten en precio y como resultado una proporción α ($0 \leq \alpha \leq 1$) de los consumidores compra en la firma A y el resto, $1 - \alpha$, compra en la firma B. En el segundo periodo, las firmas conocen en que firma compró cada consumidor en el primer periodo. Si un consumidor se cambia de la firma donde compro en el primer periodo, incurrirá en un costo de cambio s, conocido sólo por él al inicio del segundo periodo. La variable s de cada consumidor se considera un realización independiente de una variable aleatoria S que se distribuye uniformemente en $[0, \theta]$. Posteriormente, las firmas compiten nuevamente en precios en el segundo periodo. Bajo un regimen de discriminación en precios, cada firma puede proveer un pago monetario o descuento para que un consumidor se cambie desde la firma rival. Bajo un regimen de precios uniformes, este pago no está permitido, y por lo tanto las firmas cobran un único precio.

Chen (1997) comienza su análisis en el segundo periodo bajo un regimen de discriminación en precios. Para cada firma $i \in \{A, B\}$ se define: p_{i2} el precio de la firma i en el segundo periodo, m_i el descuento en precio que realiza la firma i a un nuevo cliente que se cambia de la firma rival y q_{ij} el número de consumidores que compraron en la firma j en el primer periodo pero que comprarán en el segundo periodo en la firma i.

En este caso las utilidades de ambas firmas en el segundo periodo se pueden escribir de la siguiente forma.

$$\pi_{A2} = (p_{A2} - c)q_{AA} + (p_{A2} - c - m_A)q_{AB} \quad \text{(B.1)}$$

$$\pi_{B2} = (p_{B2} - c)q_{BB} + (p_{B2} - c - m_B)q_{BA} \quad \text{(B.2)}$$

Un consumidor que en el primer periodo compró en A volverá a escoger A si su excedente personal es superior al que obtiene cuando escoge la firma B, o sea cuando se cumple que $R - p_{A2} \geq R - (p_{B2} + s - m_B)$ ó equivalentemente si $s \geq p_{A2} - p_{B2} + m_B$. Y se cambiará a la firma B en el caso contrario: $s < p_{A2} - p_{B2} + m_B$. De la misma forma, un consumidor que en el primer periodo compró en B, volverá a escoger B si su excedente es superior al que obtiene escogiendo la firma A, o sea cuando se cumple que $R - (p_{A2} + s - m_A) \leq R - p_{B2}$, o equivalentemente si $s \geq p_{B2} - p_{A2} + m_A$. Y se cambiará a la firma A en el caso contrario: $s < p_{B2} - p_{A2} + m_A$. De esta forma, es posible caracterizar la demanda del segundo periodo a través de los q_{ij}, que representan la transición de las decisiones de los consumidores desde el primer al segundo periodo.

$$q_{AA} = \alpha \int_{p_{A2} - p_{B2} + m_B}^{\theta} \frac{1}{\theta} ds = \alpha \left(1 - \frac{p_{A2} - p_{B2} + m_B}{\theta} \right) \tag{B.3}$$

$$q_{BA} = \frac{\alpha}{\theta} (p_{A2} - p_{B2} + m_B) \tag{B.4}$$

$$q_{BB} = (1 - \alpha) \left(1 - \frac{p_{B2} - p_{A2} + m_A}{\theta} \right) \tag{B.5}$$

$$q_{AB} = \frac{(1 - \alpha)}{\theta} (p_{B2} - p_{A2} + m_A) \tag{B.6}$$

Se puede demostrar la existencia de un único equilibrio de Nash en el segundo periodo bajo un régimen de discriminación en precios. En este equilibrio el precio que escoge cada firma es independiente de la participación de mercado conseguida en el primer periodo, mayor al costo marginal y menor que el precio monopólico R. Los precios y descuentos de equilibrio son los siguientes.

$$p_{A2}^* = p_{B2}^* = c + \frac{2}{3}\theta \tag{B.7}$$

$$m_A^* = m_B^* = \frac{1}{3}\theta \tag{B.8}$$

Estos resultados implican que $q_{AB} = (1 - \alpha)/3$ y $q_{BA} = \alpha/3$. Es decir, un tercio de los consumidores deciden cambiar de proveedor en el segundo periodo. Notemos que si los costos de cambio se aproximan a cero (cuando θ se aproxima a cero) el precio de equilibrio del mercado se aproxima al costo marginal, que corresponde precio competitivo, y el descuento para los nuevos clientes se aproxima a cero. Este resultado es consistente con resultados encontrados previamente en la literatura (Klemperer, 1995).

Es posible realizar el mismo análisis anterior para el segundo periodo pero bajo un régimen de precios uniformes. Chen (1997) concluye que bajo un régimen sin discriminación en precios, la firma que obtiene la mayor participación de mercado en el primer periodo, en el segundo periodo decide un precio mayor que bajo un régimen con discriminación. En efecto, bajo un regimen de precios uniformes los mayores precios posibles se dan si en el primer periodo las firmas consiguen la misma participación de mercado. El excedente de los consumidores es menor en un regimen de precios uniformes que cuando las firmas pueden discriminar en precios. De la misma forma, las utilidades de equilibrio del segundo periodo para ambas firmas son menores en un régimen con discriminación en precios que en un régimen de precios uniformes. Los precios de equilibrio bajo régimen de precios uniformes son los siguientes.

$$p_{A2}^{u^*} = \begin{cases} c + \frac{1+\alpha}{3\alpha}\theta, & \text{si } \alpha \geq 1/2 \\ c + \frac{1+\alpha}{3(1-\alpha)}\theta, & \text{si } \alpha < 1/2 \end{cases} \tag{B.9}$$

$$p_{B2}^{u^*} = \begin{cases} c + \frac{2-\alpha}{3\alpha}\theta, & \text{si } \alpha \geq 1/2 \\ c + \frac{2-\alpha}{3(1-\alpha)}\theta, & \text{si } \alpha < 1/2 \end{cases} \tag{B.10}$$

Para el primer periodo, se define p_{i1} como el precio de la firma $i \in \{A, B\}$ en el primer periodo. Chen (1997) determina que existe un único equilibrio perfecto en el subjuego bajo un régimen de discriminación en precios, en donde los precios del primer periodo son los siguientes.

$$p_{A1}^* = p_{B1}^* = c - \frac{\delta}{3}\theta \tag{B.11}$$

Se define el parámetro $\delta \in (0, 1]$ como el factor de descuento para ambas firmas. Se puede notar que las firmas escogen para el primer periodo un mismo precio que se encuentra debajo del costo marginal del producto. Esto quiere decir que ambas firmas deciden intensificar la competencia en el primer periodo para asegurar una buena participación de mercado, ya que la única forma de aumentar participación de mercado en el segundo periodo es a través del pago de descuento a los clientes que en el primer periodo escogen la firma rival.

En el caso bajo régimen de precios uniformes, también existe un único equilibrio perfecto en el subjuego, en donde el precio del primer periodo es el siguiente.

$$p_{A1}^{u*} = p_{B1}^{u*} = c + \frac{2}{3}\delta\theta \tag{B.12}$$

Las firmas deciden precios sobre el costo marginal en el primer y segundo periodo. En el segundo periodo las firmas saben que tienen una fracción del mercado asegurada y otra fracción a la que no podrán seducir, tal como sucede en el modelo de competencia de Hotelling. En efecto, dado que en el primer periodo las firmas deciden el mismo precio ambas consiguen la misma participación de mercado. Dado esto, en el segundo periodo se replican los resultados del modelo de competencia de Hotelling: $p_{A2}^{u*} = p_{B2}^{u*} = c + \theta$, cuando $\alpha = 1/2$.

B.2. Competencia con Discriminación en Precios y Preferencias de los Consumidores

Fudenberg y Tirole (2000) proponen modelo de dos firmas (A y B) y dos periodos. Las firmas producen productos con un costo marginal $c \geq 0$. Existe un continuo de consumidores que tienen demandas unitarias en

cada periodo pero que, a diferencia de Chen (1997), no enfrentan costos por cambiarse de firma. En cambio, los consumidores poseen preferencias relativas por cada firma. La preferencia relativa de un consumidor por la firma B por sobre la firma A está dada por $\theta \in [-\overline{\theta}, \overline{\theta}]$ con $\overline{\theta} > 0$ y es constante en ambos periodos. θ tiene una función de distribución acumulada F conocida. Si un consumidor con preferencia θ paga p_t en el periodo t, obtendrá en ese periodo una utilidad según la siguiente expresión.

$$u_t(\theta) = \begin{cases} v - \theta/2 - p_t, & \text{Si escoge a la firma } A \text{ en } t; \\ v + \theta/2 - p_t, & \text{Si escoge a la firma } B \text{ en } t. \end{cases} \tag{B.13}$$

La dinámica del juego propuesto por Fudenberg y Tirole (2000) es la misma descrita anteriormente para Chen (1997) bajo el régimen de discriminación en precios: las firmas A y B escogen precios a y b respectivamente para todos los consumidores en el primer periodo. En el siguiente periodo, las firmas pueden discriminar entre sus propios clientes y los de la competencia. De esta forma, la firma A escoge precios α para sus clientes y $\widehat{\alpha}$ para los de su competencia y la firma B escoge precios β para sus clientes y $\widehat{\beta}$ para los de su competencia. Los autores encuentran que en equilibrio las firmas otorgan descuentos en el segundo periodo a los consumidores que en el primer periodo escogieron la firma rival. Además, una fracción positiva de consumidores se cambia de firma en el segundo periodo a pesar de que se estén cambiando a una firma que ellos no prefieren.

Fudenberg y Tirole (2000) determinan que si θ se distribuye uniformemente entre $-\overline{\theta}$ y $\overline{\theta}$ en el segundo periodo, cuando está permitida la discriminación en precios, se obtiene un precio menor que el precio obtenido cuando la discriminación está prohibida. Esta situación se encuentra en línea con los resultados que se obtienen en Chen (1997).

Los resultados de Fudenberg y Tirole (2000) difieren a los de Chen (1997) en el caso del primer periodo. Los precios de equilibrio para el primer periodo son simétricos para ambas firmas y están dados por la siguiente expresión.

$$a = b = c + \left(1 + \frac{\delta}{3}\right)\overline{\theta} \tag{B.14}$$

Este precio es mayor al obtenido cuando la discriminación en precios no está permitida, precio que está dado por $c + \overline{\theta}$. Estos resultados son contrarios a los obtenidos por Chen (1997) pues, bajo un régimen discriminación en precios, se obtiene un precio que es menor al caso de precios uniformes y que es incluso es menor que el costo marginal del producto (B.11).

B.3. Pagar para Cambiar o Pagar para Retener

En el modelo planteado por Shaffer y Zhang (2000) existen dos firmas (A y B). El mercado está divido en dos: el grupo α de consumidores (una fracción $\theta \in [\frac{1}{2}, 1]$ del total de consumidores) que comprarían en la firma A si los precios de ambas firmas fuesen los mismos y el grupo β con el resto de los consumidores

(una fracción $1-\theta$ del total de consumidores) que comprarían en la firma B si los precios de ambas firmas fuesen los mismos. Las firmas deben decidir los precios P_A y P_B que cobran al grupo α y los precios $\widetilde{P}_A$ y $\widetilde{P}_B$ que cobran al grupo β. Los costos de cambios son representados a través de una medida de lealtad, definida como la mínima diferencia de precio necesaria para inducir a un consumidor para cambiar de firma. Un consumidor del grupo α con lealtad l elige la firma A si $P_A \leq P_B + l$, o equivalentemente si $l \geq P_A - P_B$; y elige la firma B si $P_A > P_B + l$, o equivalentemente si $l < P_A - P_B$. De la misma forma, un consumidor del grupo β con lealtad l elige la firma B si $\widetilde{P}_B \leq \widetilde{P}_A + l$, o equivalentemente si $l \geq \widetilde{P}_B - \widetilde{P}_A$; y elige la firma A si $\widetilde{P}_B > \widetilde{P}_A + l$, o equivalentemente si $l < \widetilde{P}_B - \widetilde{P}_A$.

Cada consumidor del grupo $k \in \{\alpha, \beta\}$ está caracterizado por su lealtad l distribuida uniformemente entre $[0, l_k]$. Se define la función $F_k(x)$, como la fracción de consumidores del grupo k con lealtad menor o igual a x. De esta forma, $F_\alpha(P_A - P_B)$ representa la fracción de consumidores del grupo α con lealtad menor o igual a $P_A - P_B$ y por lo tanto la fracción de consumidores del grupo α que escoge la firma B. Normalizando en uno el total de consumidores en el mercado, es posible escribir las demandas de cada firma como un promedio ponderado de las demandas de cada grupo de consumidores.

$$D_A(P_A, \widetilde{P}_A, P_B, \widetilde{P}_B) = \theta[1 - F_\alpha(P_A - P_B)] + (1-\theta)F_\beta(\widetilde{P}_B - \widetilde{P}_A) \quad \text{(B.15)}$$

$$D_B(P_A, \widetilde{P}_A, P_B, \widetilde{P}_B) = \theta F_\alpha(P_A - P_B) + (1-\theta)[1 - F_\beta(\widetilde{P}_B - \widetilde{P}_A)] \quad \text{(B.16)}$$

B.4. Competencia con Promociones Uno a Uno

Shaffer y Zhang (2002) propone un juego de dos firmas (A y B) que ofrecen un producto con costo marginal $c \geq 0$. Cada consumidor tiene una demanda unitaria por el producto y está dispuesto a pagar hasta $V \geq c$ por el producto de su firma preferida. Los consumidores se diferencian en cuanto están dispuestos a pagar por el producto de su firma menos preferida. La lealtad de los consumidores se definen de forma asimétrica como en Shaffer y Zhang (2000), pero ahora la lealtad l de un consumidor es tal que $l \in [-l_B, l_A]$ siguiendo una distribución uniforme, donde $0 \leq l_i < V$ para $i \in \{A, B\}$. Los consumidores con lealtad positiva prefieren la firma A y consumidores con lealtad negativa prefieren la firma B. Si $l_A = l_B$ se tiene que las lealtades son simétricas y quedan definidas como en Fudenberg y Tirole (2000). Si $l_B = 0$ y $l_A > 0$ quiere decir que todos consumidores observan que la firma A tiene una mayor calidad, por lo tanto la prefieren, si $l_A = 0$ y $l_B > 0$ quiere decir que todos los consumidores observan que la firma B tiene una mayor calidad. Se asume que las firmas conocen las lealtades de todos los consumidores. Las firmas participan de un juego de dos periodos, en donde los precios regulares P_A y P_B son escogidos en la primera etapa y luego en la segunda se escogen las estrategias de promociones. En el segundo periodo, las firmas deben decidir qué segmentos de consumidores serán objeto de promociones y el valor de los descuentos que se les ofrecerán. El costo de segmentar cada consumidor es de $z \geq 0$ independiente si este escoge la firma o no. Sea $d_i(l)$ el descuento que realiza la firma i a un consumidor con lealtad l. Este consumidor escogerá a la firma A si y solo si el precio neto cobrado por A no excede al precio neto cobrado por B por más que l, es decir, si $P_A - d_A(l) - [P_B - d_B(l)] \leq l$.

Una firma utilizará una estrategia de *promoción uno a uno* siempre que sea rentable hacerlo al asumir el costos de segmentación, es decir, dependerá de la relación entre z y $P_i - c$. En la práctica existen cuatro

situaciones que se pueden presentar: para ninguna firma es rentable una estrategia de *promoción uno a uno* (1), sólo es rentable para la firma A (2), sólo es rentable para la firma B (3) y que la estrategia sea rentable para ambas firmas (4).

Si para ninguna de las firmas es rentable una estrategia de *promoción uno a uno*, entonces los consumidores sólo tendrán la opción de pagar efectivamente P_A ó P_B, sin descuentos. Un consumidor con lealtad l escoge la firma A si $P_A - P_B \leq l$, por lo tanto, el nivel de lealtad $\widehat{l}$ del consumidor que queda indiferente entre una y otra firma es $\widehat{l} = P_A - P_B$. Entonces los consumidores ubicados a la derecha de $\widehat{l}$ comprarán en A y los que quedan a la izquierda lo harán en B. De esta forma la demanda de A será $(l_A - \widehat{l})/(l_A + l_B)$ y la demanda de B será $(\widehat{l} + l_B)/(l_A + l_B)$, y las utilidades son las siguientes.

$$\pi_{A2}^{(1)} = \frac{(P_A - c)(l_A - \widehat{l})}{l_A + l_B} \qquad \pi_{B2}^{(1)} = \frac{(P_B - c)(\widehat{l} + l_B)}{l_A + l_B} \tag{B.17}$$

Si en cambio, sólo para la firma A es rentable la estrategia de *promoción uno a uno*, ésta no aplicará esta estrategia sobre todos los consumidores. Todos los consumidores ubicados a la derecha de $\widehat{l}$ prefieren la firma A dados los precios P_A y P_B, sin requerir incentivos adicionales para escogerla, por lo que A no ofrecerá descuentos adicionales para ellos. Para cada consumidor a la izquierda de $\widehat{l}$, la firma A aplicará como máximo un descuento de $\overline{d}_A = P_A - c - z$, pues de otra forma no sería rentable realizar descuentos. Como estos consumidores sólo reciben descuentos de la firma A, la escogerán sólo si $[P_A - d_A(l)] - P_B \leq l$. Si un consumidor con lealtad $l \leq \widehat{l}$ (ubicado a la izquierda de $\widehat{l}$) recibe el máximo descuento $\overline{d}_A$, escogerá A si $[P_A - \overline{d}_A] - P_B = c + z - P_B \leq l$. Por lo tanto, el ultimo consumidor, hacia la izquierda de $\widehat{l}$, al que conviene dar descuento es el que tiene una lealtad $\widehat{l}_B = c + z - P_B$. En definitiva, si sólo para A es rentable una estrategia de *promoción uno a uno*, aplicará esta estrategia sólo en el segmento de consumidores con lealtades entre $\widehat{l}_B$ y $\widehat{l}$. Los consumidores entre $\widehat{l}$ y l_A escogerán de todas formas A y los consumidores entre $-l_B$ y $\widehat{l}_B$ escogerán B. La firma A ofrecerá el menor descuento posible al segmento objeto de promociones, por lo tanto, la promoción ofrecida al consumidor de lealtad l, con $\widehat{l}_B \leq l \leq \widehat{l}$, es $d_A(l) = P_A - P_B - l$ y las utilidades de equilibrio en el segundo periodo son las siguientes.

$$\pi_{A2}^{(2)} = \frac{1}{l_A + l_B}\left((P_A - c)(l_A - \widehat{l}) + \int_{\widehat{l}_B}^{\widehat{l}} (P_B + l - c - z)\,dl\right) \tag{B.18}$$

$$\pi_{B2}^{(2)} = \frac{1}{l_A + l_B}(P_B - c)(\widehat{l}_B + l_B) \tag{B.19}$$

En el caso cuando sólo para B es rentable una estrategia de *promoción uno a uno* se resuelven de una manera análoga al caso anterior. Cuando para ambas firmas es rentable una estrategia de *promoción uno a uno* se generan cuatro zonas de demanda: una zona que siempre escoge a la firma A, una zona que siempre escoge a la firma B, una zona que escoge a la firma A ya que esta firma le ofrece un descuento de acuerdo a su nivel de lealtad y una zona que escoge la firma B ya que esta firma le ofrece un descuento de acuerdo a su nivel de lealtad. Se puede demostrar que en este caso las utilidades de equilibrio en el segundo periodo están dadas por la siguiente expresión.

$$\pi_{A2}^{(4)} = \frac{(P_A - c)(l_A - \widehat{l_A})}{l_A + l_B} + \frac{\widehat{l_A^2}}{2(l_A + l_B)} \quad \text{(B.20)}$$

$$\pi_{B2}^{(4)} = \frac{(P_B - c)(l_B + \widehat{l_B})}{l_A + l_B} + \frac{\widehat{l_B^2}}{2(l_A + l_B)} \quad \text{(B.21)}$$

Shaffer y Zhang (2002) concluyen que $\forall z \geq 0$, y suponiendo que $l_A \geq l_B$, para cada caso existe un único equilibrio perfecto en el subjuego:

1. Si $z \geq (2l_A + l_B)/3$, los precios regulares de equilibrio son $P_A = (2l_A + l_B)/3 + c$ y $P_B = (l_A + 2l_B)/3 + c$, y ninguna de las firmas ofrecerá una estrategia de *promoción uno a uno*.

2. Si $l_B \leq z < (2l_A + l_B)/3$, los precios regulares de equilibrio son $P_A = (2l_A + l_B - z)/2 + c$ y $P_B = (l_B + z)/2 + c$, y sólo la firma A ofrecerá una estrategia de *promoción uno a uno*.

3. Si $z < l_B$, los precios regulares de equilibrio son $P_A = l_A + c$ y $P_B = l_B + c$, y ambas firmas ofrecerán una estrategia de *promoción uno a uno*.

B.5. Competencia en Esfuerzos de Marketing

Musalem y Joshi (2007) proponen un juego de dos periodos donde dos firmas (A y B) compiten por adquirir y luego retener a un único consumidor. Cada firma $i \in \{A, B\}$ decide el esfuerzo a_{i1} que realiza por adquirir al consumidor en el primer periodo. En el segundo periodo, la firma que captó al consumidor en el primer periodo compite por retenerlo y la otra firma por robarlo a su competidor. La firma $i \in \{A, B\}$ que no haya podido captar al cliente en el primer periodo, decide el esfuerzo a_{i2} que realizará en el segundo periodo por adquirir al consumidor. La firma $i \in \{A, B\}$ que captó al cliente en el primer periodo, decide el esfuerzo q_{i2} que realizará por retener al consumidor en el segundo periodo. El consumidor está caracterizado por su lealtad l_i a la firma $i \in \{A, B\}$, y escoge a la firma que le entrega la mayor utilidad. La utilidad que obtiene el consumidor en el primer periodo por escoger la firma $i \in \{A, B\}$ se modela a través de la siguiente función de utilidad aleatoria.

$$U_{i1} = V_{i1} + \varepsilon_{i1} = \ln(l_i + \theta_i a_{i1}) + \varepsilon_{i1} \quad \text{(B.22)}$$

La expresión $V_{i1} = \ln(l_i + \theta_i a_{i1})$ recoge la utilidad que le entrega al consumidor su grado de lealtad hacia la firma y el esfuerzo de marketing que realiza la firma por atraerlo. El parámetro θ_i representa la eficiencia de los esfuerzos de cada firma y el logaritmo natural en la función recoge que el esfuerzo de las firmas exhiben retornos decrecientes. ε_{i1} captura el impacto de los factores que explican la utilidad de los consumidores y que las firmas no pueden observar.

Los autores suponen que los términos ε_{i1} son variables aleatorias independientes e idénticamente distribuidas a través de una distribución doble exponencial o Gumbel, tal como se plantea en el Capítulo 4.3. Esto permite determinar explícitamente la probabilidad p_{i1} de que el consumidor escoja a la firma i en el primer periodo.

$$p_{i1} = \frac{e^{V_{i1}}}{e^{V_{A1}} + e^{V_{B1}}} = \frac{l_i + \theta_i a_{i1}}{l_A + \theta_A a_{A1} + l_B + \theta_B a_{B1}} \tag{B.23}$$

Como resultado del primer periodo del juego hay dos posibles escenarios: la firma A adquiere al consumidor en el primer periodo, o lo hace la firma B. Estos dos escenarios entonces se traducen en dos subjuegos en el segundo periodo: el primero asociado al escenario donde la firma A adquiere al cliente y el segundo al escenario donde la firma B adquiere al cliente. Si la firma $i \in \{A,B\}$ fue la que captó al consumidor en el primer periodo, entonces obtiene un margen m_i. Sea la firma $k \in \{A,B\}$ la que captura al consumidor en el primer periodo y sea $j \in \{A,B\} \setminus \{k\}$ la otra firma. Para el segundo periodo se tendrá para la firma k que:

- $\widetilde{m}_k = f_m \cdot m_k$ será el margen que obtendrá si el consumidor la escoge nuevamente en el segundo periodo.
- $\widetilde{\theta}_k = f_\theta \cdot \theta_k$ será la eficiencia del esfuerzo de retención que realice en el segundo periodo.
- $\widetilde{l}_k = f_l \cdot l_k$ será la lealtad que tenga el consumidor por esta firma en el segundo periodo.

Con los parámetros f_m, f_θ y f_l se pueden manipular diferentes efectos que provoca la permanencia de un consumidor como cliente de una misma firma en el tiempo. Así por ejemplo, el margen puede aumentar como consecuencia de una disminución en los costos operaciones, la lealtad del consumidor en el segundo periodo puede aumentar por aumentos en los costos de cambio y la eficiencia de los esfuerzos de la firma por retener a su cliente puede aumentar como consecuencia del mayor conocimiento que tiene la firma sobre las preferencias y necesidades de su cliente. Para la firma que no captura al consumidor en el primer periodo: el margen m_j, la eficiencia θ_j y la lealtad l_j; para el segundo periodo se mantienen iguales a los valores del primer periodo.

En el primer subjuego del segundo periodo, la firma A debe decidir el esfuerzo de retención q_{2A} que realizará por el consumidor, mientras que la firma B debe decidir el esfuerzo de adquisición a_{2A} que realizará para conseguir que el consumidor se cambie a la firma B. Análogamente a lo realizado para el primer periodo y bajo la misma forma funcional para la utilidad aleatoria del consumidor, Musalem y Joshi (2007) determinan que las probabilidades de que el consumidor escoja a la firma A o a la firma B en este subjuego.

$$p_{A2|A} = \frac{\widetilde{l}_A + \widetilde{\theta}_A q_{2A}}{\widetilde{l}_A + \widetilde{\theta}_A q_{2A} + l_B + \theta_B a_{2A}} \qquad p_{B2|A} = \frac{l_B + \theta_B a_{2A}}{\widetilde{l}_A + \widetilde{\theta}_A q_{2A} + l_B + \theta_B a_{2A}} \tag{B.24}$$

Por lo tanto, las utilidades esperadas para ambas firmas en el primer subjuego están dadas por la siguiente expresión.

$$\pi_{A2|A} = \left(\frac{\widetilde{l}_A + \widetilde{\theta}_A q_{2A}}{\widetilde{l}_A + \widetilde{\theta}_A q_{2A} + l_B + \theta_B a_{2A}}\right)\widetilde{m}_A - q_{2A} \tag{B.25}$$

$$\pi_{B2|A} = \left(\frac{l_B + \theta_B a_{2A}}{\widetilde{l}_A + \widetilde{\theta}_A q_{2A} + l_B + \theta_B a_{2A}}\right) m_B - a_{B2} \tag{B.26}$$

En el segundo subjuego, la firma B debe decidir el esfuerzo de retención q_{2B} que realizará por el consumidor, mientras que la firma A debe decidir el esfuerzo de adquisición a_{2B} que realizará para conseguir que el consumidor se cambie a la firma A. Las probabilidades de que el consumidor escoja a la firma A o a la firma B en este subjuego son las siguientes.

$$p_{A2|B} = \frac{l_A + \theta_A a_{2B}}{l_A + \theta_A a_{2B} + \widetilde{l}_B + \widetilde{\theta}_B q_{2B}} \qquad p_{B2|B} = \frac{\widetilde{l}_B + \widetilde{\theta}_B q_{2B}}{l_A + \theta_A a_{2B} + \widetilde{l}_B + \widetilde{\theta}_B q_{2B}} \tag{B.27}$$

Por lo tanto, las utilidades esperadas para ambas firmas en el segundo subjuego están dadas por la siguiente expresión.

$$\pi_{A2|B} = \left(\frac{l_A + \theta_A a_{2B}}{l_A + \theta_A a_{2B} + \widetilde{l}_B + \widetilde{\theta}_B q_{2B}}\right) m_A - a_{2B} \tag{B.28}$$

$$\pi_{B2|B} = \left(\frac{\widetilde{l}_B + \widetilde{\theta}_B q_{2B}}{l_A + \theta_A a_{2B} + \widetilde{l}_B + \widetilde{\theta}_B q_{2B}}\right)\widetilde{m}_B - q_{2B} \tag{B.29}$$

Para obtener el valor presente de las utilidades de las firmas durante los dos periodos se deben considerar los posibles escenarios. Para la firma A se puede deducir el valor presente de las utilidades de la siguiente forma. Con probabilidad p_{A1} el consumidor escoge a la firma A en el primer periodo: en este caso la firma A obtiene una utilidad de $m_A - a_{A1}$ en el primer periodo y una utilidad esperada de $\pi_{A2|A}$ en el segundo periodo como resultado de aplicar una estrategia de retención de su cliente. Con probabilidad p_{B1} el consumidor escoge a la firma B en el primer periodo: en este caso la firma A obtiene una pérdida de $-a_{A1}$ en el primer periodo, ya que realiza sin éxito un esfuerzo por atraer al cliente, y una utilidad esperada de $\pi_{A2|B}$ en el segundo periodo como resultado de su estrategia para hacer que el consumidor cambie de firma. Para la firma B se puede realizar el mismo análisis de forma análoga. En definitiva, el valor presente de las utilidades totales de ambas firmas están dadas por la siguiente expresión.

$$\pi_A = p_{A1}(m_A + \beta\pi_{A2|A}) + p_{B1}\beta\pi_{A2|B} - a_{A1} \tag{B.30}$$

$$\pi_B = p_{B1}(m_B + \beta\pi_{B2|B}) + p_{A1}\beta\pi_{B2|A} - a_{B1} \tag{B.31}$$

Donde $\beta \in (0,1]$ es el factor de descuento de un periodo a otro de ambas firmas.

Aprovechando la simetría de los resultados es posible concentrar los análisis, sin perdida de generalidad, sólo en el primer subjuego del segundo periodo. Musalem y Joshi (2007) encuentran que cuando las lealtades del consumidor por ambas firmas son suficientemente pequeñas[1] entonces el único equilibrio de Nash para el primer subjuego del segundo periodo es el siguiente.

$$q_2^* = \widetilde{m}_A \frac{\widetilde{m}_A \widetilde{\theta}_A m_B \theta_B}{(\widetilde{m}_A \widetilde{\theta}_A + m_B \theta_B)^2} - \frac{\widetilde{l}_A}{\widetilde{\theta}_A} \tag{B.32}$$

$$a_2^* = m_B \frac{\widetilde{m}_A \widetilde{\theta}_A m_B \theta_B}{(\widetilde{m}_A \widetilde{\theta}_A + m_B \theta_B)^2} - \frac{l_B}{\theta_B} \tag{B.33}$$

De (B.32) es posible concluir que, *ceteris paribus*, el equilibrio de esfuerzos de retención de la firma A crece con su margen de contribución, decrece con la lealtad del cliente, y exhibe una relación no monotónica (primero crece y luego decrece) con la eficiencia de su esfuerzo de retención.

Por otro lado, de (B.33) es posible concluir que cuando el consumidor es relativamente menos (más) atractivo para la firma A, es decir, $\widetilde{m}_A \widetilde{\theta}_A < m_B \theta_B$ ($\widetilde{m}_A \widetilde{\theta}_A > m_B \theta_B$) entonces el equilibrio del esfuerzo de atracción de la firma B crece (decrece) tanto con el margen de contribución de la firma A, como con la eficiencia de los esfuerzos de retención de la firma A.

Los autores también concluyen que cuando un consumidor es relativamente más atractivo para la firma B ($\widetilde{m}_A \widetilde{\theta}_A < m_B \theta_B$), un aumento en el margen de contribución de la firma A provoca un aumento en los esfuerzos de ambas firmas. Por otro lado, cuando un cliente es relativamente más atractivo para la firma A ($\widetilde{m}_A \widetilde{\theta}_A > m_B \theta_B$), un aumento en el margen de contribución de la firma A provoca un aumento en los esfuerzos de la firma A, pero una disminución en los esfuerzos de la firma B.

Bajo supuestos de simetría, es decir $\theta_A = \theta_B = \theta$, $m_A = m_B = m$ y $l_A = l_B = l$, los autores llegan a un expresión para el equilibrio de Nash del primer periodo del juego. Cuando las lealtades por cada firma son suficientemente pequeñas[2] entonces el único equilibrio de Nash para el juego de dos periodos es el siguiente.

$$a_{A1}^* = a_{B1}^* = \frac{(1+f_m f_\theta)^2(-(4+\beta) f_\theta l + \beta f_l l + \beta f_\theta m \theta) + \beta f_\theta m \theta (f_m^3 f_\theta^2 - 1)}{4 f_\theta \theta (1+f_m f_\theta)^2} \tag{B.35}$$

El esfuerzo que realiza cada firma por captar el consumidor en el primer periodo siempre es creciente con

[1] Si las lealtades del consumidor por ambas firmas cumplen la siguiente condición:

$$\widetilde{l}_A < \widetilde{m}_A \widetilde{\theta}_A \frac{\widetilde{m}_A \widetilde{\theta}_A m_B \theta_B}{(\widetilde{m}_A \widetilde{\theta}_A + m_B \theta_B)^2} \qquad l_B < m_B \theta_B \frac{\widetilde{m}_A \widetilde{\theta}_A m_B \theta_B}{(\widetilde{m}_A \widetilde{\theta}_A + m_B \theta_B)^2}$$

[2] Si las lealtades del consumidor por cada firma cumplen la siguiente condición:

$$l < min\{\frac{f_m f_\theta m \theta}{(1+f_m f_\theta)^2}, \frac{f_m^2 f_\theta^2 m \theta}{f_l (1+f_m f_\theta)^2}, \frac{\beta f_\theta (f_m^3 f_\theta^2 - 1) + f_\theta m \theta (1+f_m f_\theta)^2}{((4+\beta) f_\theta - \beta f_l)(1+f_m f_\theta)^2}\} \tag{B.34}$$

f_m y con f_l. También crece con f_θ cuando $f_\theta > \frac{1}{8}$. Por otro lado, la utilidad total descontada de cada firma π^* crece con f_l y exhibe una relación no monotónica (primero decreciendo y luego creciendo) con f_m y f_θ.

Anexo C

Modelo de Elección Discreta Logit Multinomial: Estimación de las Probabilidades Elección

En la expresión (4.6) se define una formula cerrada para la probabilidad P_{ik} de que el agente i escoja la alternativa k. Esta probabilidad es función de las utilidades observadas $V_{ij} = V(b, x_{ij})$ que el agente obtiene de cada alternativa, las que se determinan a partir de los atributos asociados x_{ij} y del vector de parámetros b. El objetivo del investigador entonces es estimar el vector de parámetros b.

Para cada agente se conoce el vector de atributos $x_{ij} \in \mathbb{R}^n$ que observa sobre cada alternativa y su decisión recogida a través de la variable $y_{ij} \in \{0,1\}$, la que se define de la siguiente forma.

$$y_{ij} = \begin{cases} 1 & \text{Si el agente } i \text{ escoge la alternativa } j \\ 0 & \sim \end{cases} \tag{C.1}$$

Esta variable debe cumplir que $\sum_j y_{ij} = 1$, ya que las alternativas son mutuamente excluyentes y el agente siempre se decide por al menos una de ellas.

De esta forma, la probabilidad de que el agente i escoja la alternativa observada por el investigador está dada por $\prod_{k=1}^{J} (P_{ik})^{y_{ik}}$. Si se asume que las decisiones de los agentes son independientes entonces la probabilidad conjunta de las decisiones observadas están dadas por la siguiente expresión, que corresponde a la función de verosimilitud del problema planteado.

$$\mathcal{L}(b/x) = \prod_{i=1}^{I} \prod_{k=1}^{J} (P_{ik})^{y_{ik}} \tag{C.2}$$

Los parámetros b que mejor explican la realidad que observa el investigador, dados x_{ij} e y_{ij}, son los que

maximizan la probabilidad conjunta de las decisiones observadas representadas a través de la función de verosimilitud $\mathcal{L}(b/x)$. Dado que el logaritmo natural es una función monótona creciente, la maximización de la función de verosimilitud con respecto a b es equivalente a la maximización del logaritmo de la función de verosimilitud con respecto a b. Esto es un elemento esencial ya que en este caso el logaritmo de la función de verosimilitud es mucho más fácil de tratar desde el punto de vista de las expresiones algebráicas y de la complejidad computacional. Se define el logaritmo de la función de verosimilitud como $\ell(b/x) = \ln(\mathcal{L}(b/x))$. A partir de las propiedades de la función $\ln(\cdot)$ es posible representar $\ell(b/x)$ según la siguiente expresión.

$$\ell(b) = \sum_{i=1}^{I} \sum_{k=1}^{J} y_{ik} \ln(P_{ik}) \tag{C.3}$$

De esta forma, el estimador de máxima verosimilitud del vector de parámetros está dado por la solución del siguiente problema de optimización.

$$\hat{b} = \underset{b \in \mathbb{R}^n}{\arg\max} \{\ell(b/x)\} \tag{C.4}$$

Por lo tanto, las probabilidades estimadas por el modelo Logit corresponden a las determinadas a partir de los parámetros de máxima verosimilitud.

$$\hat{P}_{ik} = \frac{e^{V(\hat{b},x_{ik})}}{\sum_j e^{V(\hat{b},x_{ij})}} \tag{C.5}$$

Para encontrar $\hat{b}$ entonces se debe resolver el problema de optimización planteado en (C.4). La solución debe satisfacer las condiciones de primer orden del problema, las que se obtienen derivando $\ell(b/x)$ con respecto a cada componente del vector b e igualando a cero.

$$\nabla_b \ell(b/x) = 0 \tag{C.6}$$

Tal como plantea Train (2003), la expresión $\ell(b/x)$ puede ser modificada para facilitar luego el cálculo de su derivada de la siguiente forma.

$$\begin{aligned} \ell(b/x) &= \sum_{i=1}^{I} \sum_{k=1}^{J} y_{ik} \ln(P_{ik}) = \sum_{i=1}^{I} \sum_{k=1}^{J} y_{ik} \ln\left(\frac{e^{V_{ik}}}{\sum_j e^{V_{ij}}}\right) \\ &= \sum_{i=1}^{I} \sum_{k=1}^{J} y_{ik} V_{ik} - \sum_{i=1}^{I} \sum_{k=1}^{J} y_{ik} \ln\left(\sum_{j=1}^{J} e^{V_{ij}}\right) \end{aligned} \tag{C.7}$$

Así, el gradiente del logaritmo de la función de verosimilitud se puede desarrollar de la siguiente forma según plantea Train (2003).

$$\begin{aligned}
\nabla_b \ell(b/x) &= \sum_{i=1}^{I}\sum_{k=1}^{J} y_{ik}\nabla_b V_{ik} - \sum_{i=1}^{I}\sum_{k=1}^{J} y_{ik}\left(\frac{1}{\sum_j e^{V_{ij}}}\sum_{j=1}^{J} e^{V_{ij}}\nabla_b V_{ij}\right) \\
&= \sum_{i=1}^{I}\sum_{k=1}^{J} y_{ik}\nabla_b V_{ij} - \sum_{i=1}^{I}\sum_{k=1}^{J} y_{ik}\left(\sum_{j=1}^{J} P_{ij}\nabla_b V_{ij}\right) \\
&= \sum_{i=1}^{I}\sum_{k=1}^{J} y_{ik}\nabla_b V_{ij} - \sum_{i=1}^{I}\left(\sum_{j=1}^{J} P_{ij}\nabla_b V_{ij}\right)\sum_{k=1}^{J} y_{ik} \\
&= \sum_{i=1}^{I}\sum_{k=1}^{J} y_{ik}\nabla_b V_{ij} - \sum_{i=1}^{I}\sum_{j=1}^{J} P_{ij}\nabla_b V_{ij} \\
&= \sum_{i=1}^{I}\sum_{k=1}^{J} (y_{ik} - P_{ik})\nabla_b V_{ij} \qquad \text{(C.8)}
\end{aligned}$$

Cuando la utilidad observada sigue una forma lineal, es decir $V_{ik} = b^T x_{ik}$, entonces $\nabla_b V_{ij} = x_{ik}$. De esta forma, las condiciones de primer orden del problema se obtienen al igualar a cero el gradiente de $\ell(b)$ con respecto a b, $\nabla_b \ell(b) = 0$, resultando el siguiente sistema de ecuaciones.

$$\sum_{i=1}^{I}\sum_{k=1}^{J} (y_{ik} - P_{ik})x_{ik} = 0 \qquad \text{(C.9)}$$

Si se reordenan los términos y se divide por el número de agentes I en cada lado, entonces la expresión anterior es equivalente al siguiente sistema de ecuaciones.

$$\frac{1}{I}\sum_{i=1}^{I}\sum_{k=1}^{J} y_{ik}\, x_{ik} = \frac{1}{I}\sum_{i=1}^{I}\sum_{k=1}^{J} P_{ik}\, x_{ik} \qquad \text{(C.10)}$$

El término de la izquierda se puede interpretar como el promedio observado de cada atributo sobre las alternativas escogidas por los agentes. El término de la derecha, en cambio, corresponde al valor esperado de cada atributo sobre las alternativas escogidas por los agentes de acuerdo a las probabilidades estimadas. Por lo tanto, de acuerdo a las condiciones de primer orden, el estimador de máxima verosimilitud $\hat{b}$ hace que para cada variable explicativa el promedio observado en la muestra utilizada para estimar el modelo sea igual al promedio que luego predice el modelo. Estas condiciones inducen al modelo a reproducir los promedios observados en la muestra.

No es posible despejar una formula cerrada para encontrar los parámetros $\hat{b}$. Sin embargo, a través de métodos numéricos se puede resolver el sistema de ecuaciones que plantean las condiciones de primer orden. Algunos de los algoritmos comúnmente utilizados para estos fines son (Train, 2003): Newton-Raphson; Berndt-B. Hall-R. Hall-Hausman (BHHH) y su variante BHHH-2; Davidon–Fletcher–Powell; y Broyden–Fletcher–Goldfarb–Shanno, entre otros.

Anexo D

Criterios de Evaluación de los Modelos

D.1. Bondad de Ajuste del Modelo

D.1.1. Razón de Verosimilitud

Para los modelos resueltos a través del método de máxima verosimilitud comúnmente se utiliza el estadístico llamado razón de verosimilitud (*likelihood ratio*) para medir la bondad de ajuste del modelo estimado. Este estadístico compara qué tan bueno es el modelo que utiliza los parámetros de máxima verosimilitud con respecto a un modelo cuyos parámetros asociados a los atributos son todos iguales a cero y donde sólo se utiliza una constante. Este último caso se denotará como modelo nulo. Esta comparación se realiza evaluando cuánto mejora la función de verosimilitud con el modelo que utiliza los parámetros estimados en relación al modelo nulo. Este estadístico también es conocido como el pseudo R^2 de McFadden. En adelante se denotará al estadístico pseudo R^2 como $\widetilde{R}^2$.

$$\widetilde{R}^2_{\text{McFadden's}} = 1 - \frac{\ell(\hat{b})}{\ell(b_{\text{nulo}})} \tag{D.1}$$

Donde $\ell(\hat{b})$ corresponde al logaritmo de la verosimilitud con los parámetros estimados $\hat{b}$ y $\ell(b_{\text{nulo}})$ al asociado al modelo nulo. Si el logaritmo de la verosimilitud no mejora con los parámetros estimados, con respecto al caso nulo, entonces $\ell(\hat{b}) = \ell(b_{\text{nulo}})$ y por lo tanto $\widetilde{R}^2 = 0$. Éste es el mínimo valor que el $\widetilde{R}^2$ puede tomar porque de lo contrario $\hat{b}$ no sería el estimador de máxima verosimilitud. Si por el contrario los parámetros estimados son tan buenos que son capaces de predecir a la perfección las decisiones de los agentes, entonces la verosimilitud sería igual a uno y por lo tanto el logaritmo de la verosimilitud sería igual a cero. De esta forma cuando $\ell(\hat{b}) = 0$ entonces $\widetilde{R}^2 = 1$, que corresponde al máximo valor de $\widetilde{R}^2$.

En resumen, la razón de verosimilitud $\widetilde{R}^2$ se mueve desde cero, cuando el modelo es malo, hasta uno, cuando el modelo es perfecto. Cuando para un mismo problema a resolver (los mismos datos y alternativas de de-

cisión) se tienen dos modelos o estimaciones para los parámetros, se puede decir que el modelo con mayor $\widetilde{R}^2$ es el modelo superior ya que se ajusta mejor a los datos de la muestra. Como el valor de la razón de verosimilitud depende del logaritmo de la verosimilitud es correcto comparar dos modelos a través de este estadístico cuando estos corresponden al mismo problema. Cuando dos modelos son estimados sobre muestras no idénticas o sobre alternativas de decisión diferentes, entonces los $\widetilde{R}^2$ no son comparables totalmente ya que se determinan a través de funciones de verosimilitud construidas de diferente forma. En este último caso, las medidas son sólo referenciales.

Adicionalmente existen otras medidas equivalentes al $\widetilde{R}^2$ de McFadden, entre ellas:

- $\widetilde{R}^2$ ajustado de McFadden, que a partir del $\widetilde{R}^2$ de McFadden penaliza al modelo que utiliza muchos atributos para realizar su predicción. Si para un mismo problema dos modelos son iguales en términos del valor del logaritmo de la verosimilitud entonces el modelo que utilice el menor número de atributos será el que tenga mejor nivel de $\widetilde{R}^2$. Si el número de atributos es n, incluyendo la constante, entonces el estadístico se define de la siguiente forma.

$$\widetilde{R}^2_{\text{McFadden's Adjusted}} = 1 - \frac{\ell(\hat{b}) - (n-1)}{\ell(b_{\text{nulo}})} \tag{D.2}$$

- $\widetilde{R}^2$ de Cox & Snell, que es otra medida de mejora en términos de verosimilitud ente los modelos con los parámetros estimados versus el modelo nulo. Este estadístico adicionalmente toma en consideración el tamaño poblacional de la muestra utilizada en la estimación. A mayor valor del estadístico mejor el modelo en términos de ajuste. Sin embargo, el problema del estadístico es que su máximo valor se acerca a uno pero nunca lo alcanza.

$$\widetilde{R}^2_{\text{Cox \& Snell}} = 1 - \exp\left(\frac{-2\left[\ell(\hat{b}) - \ell(b_{\text{nulo}})\right]}{I}\right) \tag{D.3}$$

- $\widetilde{R}^2$ de Nagelkerke o también conocido como de Cragg & Uhler's, toma el estadístico propuesto por Cox & Snell y lo escala para que su rango se encuentre efectivamente entre cero y uno. De esta forma existe un máximo teórico igual a uno para el modelo que logre el ajuste perfecto.

$$\widetilde{R}^2_{\text{Nagelkerke}} = \frac{R^2_{\text{Cox \& Snell}}}{1 - \exp\left(\frac{2\ell(b_{\text{nulo}})}{I}\right)} \tag{D.4}$$

D.1.2. Significancia del Modelo y de los Atributos del Modelo

Supongamos que se estima un modelo utilizando n regresores incluyendo una constante. Luego interesa conocer la significancia completa del modelo, considerando los $n-1$ atributos o predictores que utiliza

el modelo. Para esto se puede construir el estadístico G y realizar la prueba de la razón de verosimilitud (*likelihood ratio test*) cuya hipótesis nula H_0 es que todos los parámetros asociados a algún atributo son iguales a cero.

$$H_0 : b_2 = b_3 = \ldots = b_n = 0 \tag{D.5}$$

La hipótesis nula no considera al parámetro b_1 ya que es coeficiente asociado a la constante del modelo. La prueba se basa en el estadístico G el que se define de la siguiente forma.

$$G = -2\left[\ell(b_{\text{nulo}}) - \ell(\hat{b})\right] \tag{D.6}$$

Donde $\ell(\hat{b})$ corresponde al logaritmo de la verosimilitud con los parámetros estimados $\hat{b}$ y $\ell(b_{\text{nulo}})$ el asociado al modelo nulo. Se puede demostrar que G sigue una distribución χ^2 con $n-1$ grados de libertad. De esta forma, a menor valor del *p-value* más probable el rechazo de la hipótesis nula H_0 y por lo tanto más probable concluir que, a determinado nivel de confianza, al menos uno o más de uno de los $n-1$ parámetros asociados a los atributos del modelo son diferentes a cero.

En caso de querer evaluar la significancia del u-ésimo atributo en particular se puede replicar la misma prueba anterior. La hipótesis nula H_0 en este caso es que el parámetro en cuestión es igual a cero $H_0 : b_u = 0$ y el estadístico G tendría que redefinirse como sigue.

$$G = -2\left[\ell(b_{\tilde{u}}) - \ell(\hat{b})\right] \tag{D.7}$$

Donde $b_{\tilde{u}}$ son los parámetros estimados sin utilizar el atributo u, $\ell(b_{\tilde{u}})$ es el logaritmo de la verosimilitud del modelo que no utiliza el atributo u y $\ell(\hat{b})$ es el logaritmo de la verosimilitud del modelo estimado. En este caso G sigue una distribución χ^2 con 1 grado de libertad y se puede concluir sobre la significancia del atributo u de acuerdo al *p-value* obtenido.

Alternativamente también se puede realizar la prueba de Wald. La hipótesis nula también es $H_0 : b_u = 0$ y en este caso el estadístico a evaluar es el siguiente.

$$Z_{\text{Wald}} = \frac{b_u}{\text{StDev}(b_u)} \tag{D.8}$$

Z_{Wald} sigue una distribución normal estándar y $\text{StDev}(b_u)$ corresponde a la desviación estándar asociada al parámetro estimado u.

Es importante notar que para los modelos no lineales resueltos a través del Método de Máxima Verosimilitud, la desviación estándar de los coeficientes (incluyendo la constante) se puede obtener a través de la raíz cuadrada de los elementos de la diagonal del valor negativo del inverso de matriz Hessiana asociada al respectivo modelo.

$$\text{StDev}(\hat{b}_j) = \sqrt{-\mathbf{H}_{\ell}(b/x)^{-1}_{j,j}} \quad \forall j \in \{0,...,n-1\} \tag{D.9}$$

D.2. Estadístico de Kolmogórov-Smirnov (KS)

El estadístico de Kolmogórov-Smirnov (KS) permite contrastar la hipótesis de que, para una variable unidimensional como por ejemplo la probabilidad σ_i estimada por un modelo, dos muestras proceden de una misma distribución de probabilidad. Para esto se comparan las funciones de distribución acumuladas empíricas de ambas muestras, calculando la distancia vertical máxima que existe entre las dos distribuciones para la variable de interés.

La utilidad de estadístico KS radica en su capacidad de determinar si existen diferencias estadísticamente significativas en las distribuciones de ambas muestras para la variable de interés. Si F_A y F_B corresponden a las funciones de distribución empíricas de las muestras A y B respectivamente, entonces el estadístico KS está dado por la siguiente expresión.

$$KS = \sup_x |F_A(x) - F_B(x)| \tag{D.10}$$

Este estadístico se puede utilizar para evaluar la capacidad de discriminación de un modelo que estima la probabilidad $\sigma_i = \Pr(y_i = 1)$ de que cada observación pertenezca a la clase positiva. De esta forma es posible definir F_A y F_B como las funciones de distribución empíricas de los conjuntos A y B, donde $A = \{i/y_i = 1\}$ (que corresponde a las observaciones que efectivamente pertenecen a la clase positiva) y $B = \{i/y_i = 0\}$ (que corresponde a las observaciones que efectivamente pertenecen a la clase negativa). De esta forma, el estadístico KS medirá qué tan diferentes son las estimaciones de σ_i para las observaciones que pertenecen a la clase positiva versus los que pertenecen a la clase negativa. Se espera que σ_i tenga valores más cercanos a 1 para las observaciones que pertenecen a la clase positiva y que tenga valores más cercanos a 0 para las que pertenecen a la clase negativa. Los modelos con mayores valores de KS tienen una mayor capacidad de discriminación pues son capaces de separar con mayor eficiencia a las observaciones de las diferentes clases: las observaciones con σ_i más cercano a 1, efectivamente son observaciones de la clase positiva; y las observaciones con σ_i más cercano a 0, efectivamente pertenecen a la clase negativa.

En la figura D.1 se puede ver un ejemplo del estadístico KS. La distribución en línea punteada comienza a acumular un mayor número de casos para valores bajos de σ_i, mientras que la otra distribución acumula más casos en valores altos de σ_i. Cuando la distribución en línea punteada ha acumulado el 75 %, la otra distribución recién ha acumulado el 27 %. Como en este punto es donde se alcanza la mayor diferencia entre ambas distribuciones acumuladas entonces el estadístico KS para este caso es de un 48 %.

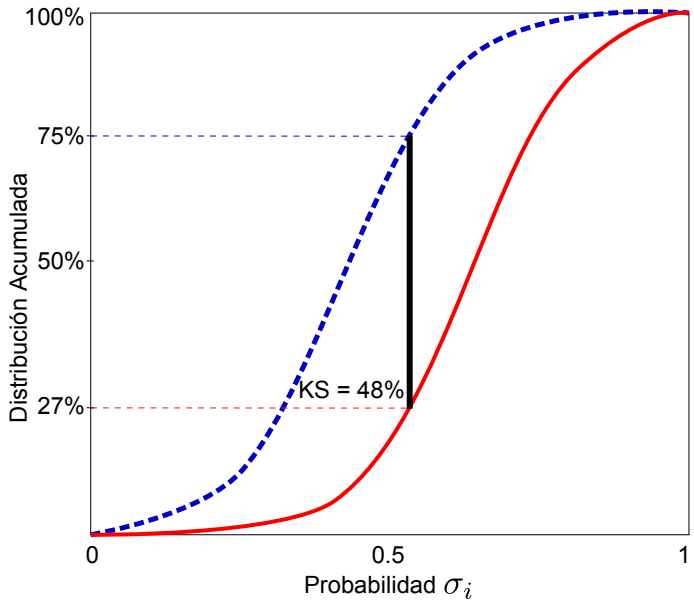

Figura D.1: Ejemplo estadístico Kolmogórov-Smirnov (KS)

D.3. Curva ROC

D.3.1. Área Bajo la Curva ROC

La curva ROC (*Receiver Operating Characteristic*) ilustra la eficiencia de un clasificador binario a medida que su umbral de decisión o *cut-off* se modifica. Esta curva se crea graficando la tasa de verdaderos positivos (TPR, *True Positive Rate*) en el eje Y versus la tasa de falsos positivos (FPR, *False Positive Rate*) en el eje X, a medida que se modifica el umbral de decisión. La definición de Verdaderos Positivos (TP), Falsos Positivos (FP), Verdaderos Negativos (TN) y Falsos Negativos (FN) ya fue realizada en la sección 4.2. En la figura D.2 se grafican tres diferentes curvas ROC.

Supongamos que un modelo estima una probabilidad σ_i de que el caso sea positivo o que $y_i = 1$. Sea $\widehat{y}_i$ la predicción que hace un modelo sobre la variable dependiente y_i. Si el *cut-off* que se define es c entonces: el modelo predice $\widehat{y}_i = 1$ ó que el caso es positivo cuando $\sigma_i > c$; y el modelo predice $\widehat{y}_i = 0$ ó que el caso es negativo en el caso contrario.

La TPR mide la eficiencia de un modelo para detectar o clasificar correctamente los casos positivos ($y_i = 1$): número de casos positivos correctamente clasificados sobre el total de casos positivos disponibles en la muestra. La FPR mide el nivel de error que comete el modelo al clasificar los casos negativos ($y_i = 0$): número de casos negativos mal clasificados sobre el total de casos negativos disponibles en la muestra.

$$TPR = \frac{TP}{TP+FN} \qquad FPR = \frac{FP}{TN+FP} \tag{D.11}$$

La curva ROC entonces es construida a partir de los valores de TPR y FPR para diferentes valores del *cut-off* c. Para conseguir esto generalmente se construye un grilla de valores para c, desde 0 hasta 1, y para cada valor de c se estiman la TPR y la FRP. Si se define un *cut-off* $c = 0$, entonces todas las predicciones del modelo serán positivas. Por lo tanto los casos sólo pueden ser verdaderos positivos o falsos positivos; así que el TPR y el FPR serán iguales a 100 %. Este caso se ubicaría en el vértice superior derecho del gráfico ROC, como se observa en la figura D.2. En el otro caso extremo, cuando se define un *cut-off* $c = 1$ todas las predicciones del modelo serán negativas. Por lo tanto ningún caso será verdaderos positivos ni falsos positivos; así que el TPR y el FPR serán iguales a 0 %. Este caso se ubicaría en el vértice inferior izquierdo del gráfico ROC, como se observa en la figura D.2. Para cualquier otro valor de c, el modelo se encontrará en cualquier punto del espacio ROC que se encuentra por sobre la diagonal que va desde el vértice inferior izquierdo al superior derecho. En la figura D.2 se pueden ver las curvas ROC *A*, *B* y *C*, que corresponden a diferentes modelos. Todos los modelos se deben encontrar sobre esta diagonal, ya que si un predictor se encuentra bajo la diagonal basta con invertir sus predicciones para que se encuentre sobre la diagonal. La diagonal se dice que corresponde al resultado de un predictor aleatorio.

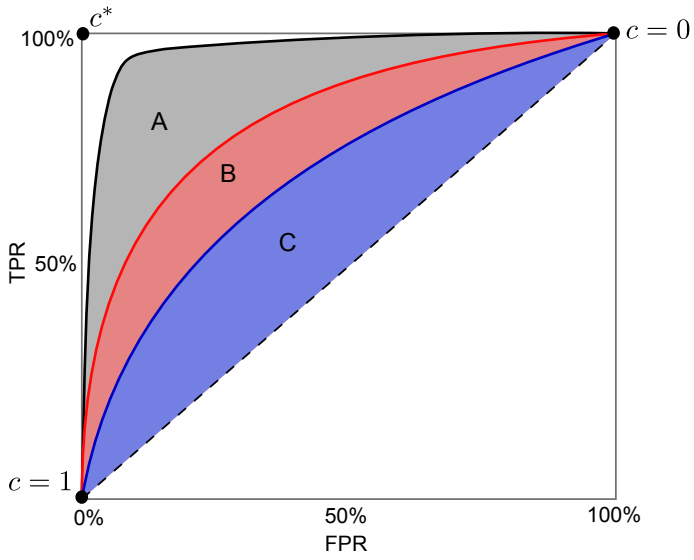

Figura D.2: Ejemplo estadístico área bajo la curva ROC (AUC)

Si existe un modelo perfecto entonces este tendrá una TPR del 100 % y una FPR del 0 %. Esto quiere decir que el modelo predice correctamente todos los casos positivos y sin errores todos los casos negativos. En el gráfico ROC entonces este modelo se encontraría sobre el vértice superior izquierdo, destacado como c^* en la figura D.2.

Una forma de determinar la capacidad predictiva de varios modelos es comparando el área bajo la curvas ROC (AUC, *area under curve*). La curva ROC con mayor área corresponde a la que se acerca más al modelo perfecto. El AUC será de un 50 % para el peor modelo o modelo aleatorio, correspondiente al área del triángulo bajo la diagonal; y será de 100 % para el modelo perfecto, correspondiente al área del cuadrado que forma todo el espacio ROC. En la figura D.2 el modelo con mayor AUC es el modelo *A*, luego el *B* y finalmente el *C*.

Se dice que el modelo con mayor AUC es el mejor modelo en términos de la eficiencia en predicción. Sean σ_i y σ_j las probabilidades de ser positivo que asigna un modelo a un caso que es positivo y otro que es negativo respectivamente, ambos casos escogidos aleatoriamente. El área bajo la curva ROC de este modelo se puede interpretar entonces como la probabilidad que $\sigma_i > \sigma_j$.

D.3.2. Estadístico J de Youden

El estadístico J de Youden se define como la máxima distancia vertical que existe entre algún punto de la curva ROC y la diagonal asociada al predictor aleatorio. De la figura D.3 se puede ver que para cualquier punto sobre la curva ROC la distancia vertical a la diagonal está dada por la diferencia entre la TPR y la FPR. El estadístico J de Youden entonces se determina a partir de la máxima diferencia que se alcanza entre la TPR y la FPR. Para este cálculo se puede aprovechar la grilla de valores c que se construye para el *cut-off* y que se utiliza para construir la curva ROC. A partir de esto el estadístico J se puede calcular como:

$$J = \text{máx}_{0 \leq c \leq 1}\{TPR(c) - FPR(c)\} \tag{D.12}$$

Este estadístico es utilizado comúnmente para determinar el *cut-off* óptimo asociado a un modelo. El clasificador que utiliza el *cut-off* asociado al estadístico J, es el que tiene la mejor relación entre altas tasas de verdaderos positivos y bajos niveles de falsos positivos.

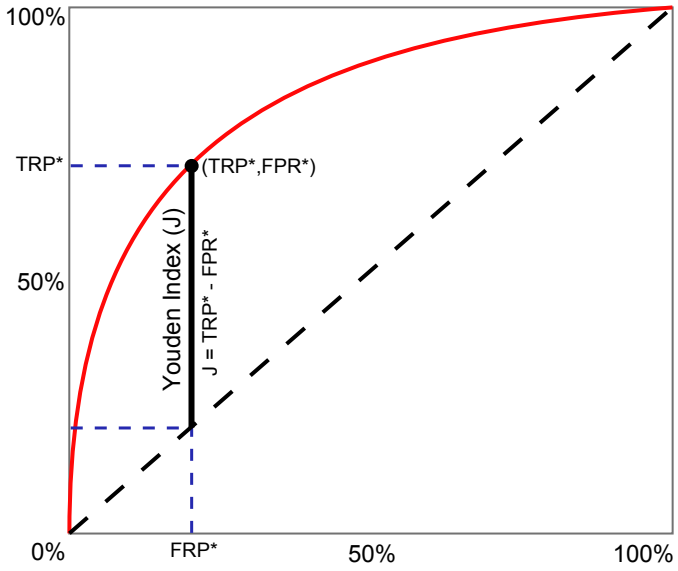

Figura D.3: Ejemplo estadístico J de Younden

D.4. Error Cuadrático Medio (MSE)

La calidad de cada modelo se puede medir también comparando la predicción que hace sobre cada observación versus la clase a la que pertenecen. Supongamos que un modelo calcula una probabilidad σ_i de que la observación *i* pertenezca a la clase positiva. La variable y_i toma el valor 1 cuando la observación *i* pertenece a la clase positiva y 0 cuando no. Entonces es posible medir el error que comete el modelo en su predicción sobre todos los clientes través del Error Cuadrático Medio o el *Mean Squared Error* (MSE).

$$MSE = \frac{1}{I}\sum_{i=1}^{I}(\sigma_i - y_i)^2 \tag{D.13}$$

Se desea que un modelo cometa el menor error posible y que por lo tanto tenga el menor MSE.

Anexo E

Modelos de Fuga de Clientes con Restricciones de Competencia: Detalle de Cálculos

E.1. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios

E.1.1. Cálculo del Gradiente $\nabla_b \tilde{\ell}(b)$

La expresión (5.20) para la función de verosimilitud se puede trabajar algebraicamente. Al expandir los términos al cuadrado y luego agrupar algunos términos semejantes, se llega a la siguiente fórmula.

$$\begin{aligned}\tilde{\ell}(b/x) &= \sum_{i=1}^{I}(y_i \ln(\sigma_i) + (1-y_i)\ln(1-\sigma_i)) + \mu \sum_{i=1}^{I}\{(\beta_A^2 p_{Ai}^2 + \beta_B^2 p_{Bi}^2)\sigma_i^2(1-\sigma_i)^2 \\ &\quad +[1-2\sigma_i(1-\sigma_i)] + 2\beta_A\sigma_i^2(1-\sigma_i)p_{Ai} - 2\beta_B\sigma_i(1-\sigma_i)^2 p_{Bi}\}\end{aligned} \tag{E.1}$$

En este caso, se supone que los atributos contenidos en los vectores x_i son conocidos. Para poder llegar a una expresión del gradiente es útil conocer primero los siguientes resultados.

$$\frac{\partial \sigma_i}{\partial b_k} = \sigma_i(1-\sigma_i)x_{ik} \quad \forall i, \forall k \tag{E.2}$$

$$\frac{\partial(\sigma_i(1-\sigma_i))}{\partial b_k} = \sigma_i(1-\sigma_i)(1-2\sigma_i)x_{ik} \quad \forall i, \forall k \tag{E.3}$$

$$\frac{\partial(\sigma_i^2(1-\sigma_i)^2)}{\partial b_k} = 2\sigma_i^2(1-\sigma_i)^2(1-2\sigma_i)x_{ik} \quad \forall i, \forall k \tag{E.4}$$

$$\frac{\partial(\sigma_i^2(1-\sigma_i))}{\partial b_k} = \sigma_i^2(1-\sigma_i)(2-3\sigma_i)x_{ik} \quad \forall i, \forall k \tag{E.5}$$

$$\frac{\partial(\sigma_i(1-\sigma_i)^2)}{\partial b_k} = \sigma_i(1-\sigma_i)^2(1-3\sigma_i)x_{ik} \quad \forall i, \forall k \tag{E.6}$$

El primer resultado se obtiene fácilmente a partir de la definición de σ_i. Los siguientes resultados utilizan el primero como base y son interesantes de conocer porque son derivadas que se requieren calcular al derivar la función $\tilde{\ell}(b/x)$ con respecto a los componentes de b. El gradiente de $\tilde{\ell}(b/x)$ entonces está compuesto por cada una de estas derivadas parciales.

$$\begin{aligned}\frac{\partial\tilde{\ell}(b/x)}{\partial b_k} &= \sum_{i=1}^{I}(y_i-\sigma_i)x_{ik} + 2\mu\sum_{i=1}^{I}\left[(\beta_A^2 p_{Ai}^2+\beta_B^2 p_{Bi}^2)(1-2\sigma_i)\sigma_i(1-\sigma)-(1-2\sigma_i)\right. \\ &\quad \left.+\beta_A\sigma_i(2-3\sigma_i)p_{Ai}-\beta_B(1-\sigma_i)(1-3\sigma_i)p_{Bi}\right]\sigma_i(1-\sigma_i)x_{ik} \quad \forall k\end{aligned} \tag{E.7}$$

Notemos que la expresión anterior es válida para todos los componentes del vector b con excepción del segundo y tercer componentes β_A y β_B. Estos dos parámetros son los que están asociados a los precios p_{Ai} y p_{Bi} que las firmas cobran y ambos se encuentran explicitamente en la fórmula de la función $\tilde{\ell}(b/x)$, y no sólo a través de σ_i como el resto de los componentes de b. Para adecuar la fórmula (E.7) para obtener las derivadas correctas de $\tilde{\ell}(b/x)$ con respecto a β_A y β_B ($k=2$ y $k=3$ en la fórmula respectivamente), se deben incorporar algnunos términos adicionales por la presencia de estos parámetros en el resto de la fórmula. Por ejemplo cuando se deriva la función $\tilde{\ell}(b/x)$ con respecto a $b_{k=2}=\beta_A$, para el término $2\beta_A\sigma_i^2(1-\sigma_i)p_{Ai}$ en la fórmula (E.7) se considera que su derivada sólo es $2\beta_A\sigma_i^2(1-\sigma_i)(2-3\sigma_i)p_{Ai}^2$, cuando esta derivada en realidad debería ser $2\beta_A\sigma_i^2(1-\sigma_i)(2-3\sigma_i)p_{Ai}^2+2\sigma_i^2(1-\sigma_i)p_{Ai}$. Las derivadas corregidas para los dos casos mencionados entonces serían las siguientes.

$$\frac{\partial\tilde{\ell}(b/x)}{\partial\beta_A} = \frac{\partial\tilde{\ell}(b/x)}{\partial b_2} + 2\mu\sum_{i=1}^{I}\left[\beta_A p_{Ai}(1-\sigma_i)+1\right]\sigma_i^2(1-\sigma_i)p_{Ai} \tag{E.8}$$

$$\frac{\partial\tilde{\ell}(b/x)}{\partial\beta_B} = \frac{\partial\tilde{\ell}(b/x)}{\partial b_3} + 2\mu\sum_{i=1}^{I}\left[\beta_B p_{Bi}\sigma_i-1\right]\sigma_i(1-\sigma_i)^2 p_{Bi} \tag{E.9}$$

E.1.2. Cálculo del Hessiano $\mathbf{H}_{\tilde{\ell}}(b)$

Para determinar el hessiano de $\tilde{\ell}(b/x)$ es necesario caracterizar sus segundas derivadas parciales con respecto a los componentes del vector b. Éstas se pueden representar de forma genérica excepto para las derivadas que incluyen cruces con β_A y β_B. Esto porque estas dos variables se encuentran de forma explicita en la expresión de $\tilde{\ell}(b/x)$, a diferencia del resto de los componentes de b que solo están presentes indirectamente a través de σ_i.

Para contrurir el hessiano se presenta primero la expresión genérica, obviando que β_A y β_B se encuentran de forma explicita en $\tilde{\ell}(b/x)$, y luego se realizarán los ajustes necesarios a las derivadas cruzadas que contengan las variables β_A y β_B. El componente (j,k) de la matriz hessiana está dado por la siguiente expresión.

$$\begin{aligned}\frac{\partial^2\tilde{\ell}(b/x)}{\partial b_j\,\partial b_k} &= -\sum_{i=1}^{I}\sigma_i(1-\sigma_i)x_{ik}x_{ij}+2\mu\sum_{i=1}^{I}\left[(\beta_A^2p_{Ai}^2+\beta_B^2p_{Bi}^2)(1-6\sigma_i+6\sigma_i^2)\right.\\ &\quad +2\beta_A(1-3\sigma_i)p_{Ai}+2\beta_B(2-3\sigma_i)p_{Bi}+2\left]\sigma_i^2(1-\sigma_i)^2x_{ik}x_{ij}\right.\\ &\quad +2\mu\sum_{i=1}^{I}\left[(\beta_A^2p_{Ai}^2+\beta_B^2p_{Bi}^2)(1-2\sigma_i)\sigma_i(1-\sigma_i)-(1-2\sigma_i)+\beta_A\sigma_i(2-3\sigma_i)p_{Ai}\right.\\ &\quad \left.-\beta_B(1-\sigma_i)(1-3\sigma_i)p_{Bi}\right](1-2\sigma_i)(1-\sigma_i)\sigma_ix_{ik}x_{ij}\quad\forall k,j\end{aligned}\tag{E.10}$$

Para llegar a la expresión anterior se utilizaron los resultados de las fórmulas (E.2) a la (E.6) y adicionalmente se utilizan los siguientes resultados.

$$\frac{\partial(\sigma_i(2-3\sigma_i))}{\partial b_k}=2\sigma_i(1-\sigma_i)(1-3\sigma_i)x_{ik}\quad\forall i,\forall k\tag{E.11}$$

$$\frac{\partial((1-\sigma_i)(1-3\sigma_i))}{\partial b_k}=-2\sigma_i(1-\sigma_i)(2-3\sigma_i)x_{ik}\quad\forall i,\forall k\tag{E.12}$$

$$\frac{\partial(\sigma_i(1-\sigma_i)(1-2\sigma_i))}{\partial b_k}=\sigma_i(1-\sigma_i)(1-6\sigma_i+6\sigma_i^2)x_{ik}\quad\forall i,\forall k\tag{E.13}$$

Las derivadas cruzadas que contienen las variables β_A y β_B son: $\frac{\partial^2\tilde{\ell}}{\partial\beta_A\partial b_k}$, $\frac{\partial^2\tilde{\ell}}{\partial\beta_B\partial b_k}$, $\frac{\partial^2\tilde{\ell}}{\partial\beta_A^2}$, $\frac{\partial^2\tilde{\ell}}{\partial\beta_B^2}$ y $\frac{\partial^2\tilde{\ell}}{\partial\beta_A\partial\beta_B}$. Los siguientes son los ajustes para corregir la matriz hessiana.

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial b_k}=\frac{\partial^2\tilde{\ell}(b/x)}{\partial b_2\partial b_k}+2\mu\sum_{i=1}^{I}\Big[2\beta_Ap_{Ai}(1-2\sigma_i)(1-\sigma_i)+(2-3\sigma_i)\Big]\sigma_i^2(1-\sigma_i)x_{ik}p_{Ai}\quad\forall k\neq 2,3\tag{E.14}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial b_k}=\frac{\partial^2\tilde{\ell}(b/x)}{\partial b_3\partial b_k}+2\mu\sum_{i=1}^{I}\Big[2\beta_Bp_{Bi}(1-2\sigma_i)\sigma_i-(1-3\sigma_i)\Big]\sigma_i(1-\sigma_i)^2x_{ik}p_{Bi}\quad\forall k\neq 2,3\tag{E.15}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A^2}=\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial b_2}+2\mu\sum_{i=1}^{I}\Big[2\beta_Ap_{Ai}(1-2\sigma_i)(1-\sigma_i)+(1-\sigma)+(2-3\sigma_i)\Big]\sigma_i^2(1-\sigma_i)p_{Ai}^2\tag{E.16}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B^2}=\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial b_3}+2\mu\sum_{i=1}^{I}\Big[2\beta_Bp_{Bi}(1-2\sigma_i)\sigma_i+\sigma-(1-3\sigma_i)\Big]\sigma_i(1-\sigma_i)^2p_{Bi}^2\tag{E.17}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial\beta_B}=\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial b_3}+2\mu\sum_{i=1}^{I}\Big[2\beta_Bp_{Bi}(1-2\sigma_i)\sigma_i-(1-3\sigma_i)\Big]\sigma_i(1-\sigma_i)^2p_{Ai}p_{Bi}\tag{E.18}$$

De esta forma la matriz hessiana $\mathbf{H}_{\tilde{\ell}}(b)$ se construye inicialmente a partir de los componentes de la expresión (E.10) y luego realizando los ajustes a la derivadas cruzadas que incluyen las variables β_A y β_B a partir de las expresiones de la (E.14) a la (E.18).

E.2. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios y Descuentos en Precios

E.2.1. Cálculo del Gradiente $\nabla_b \tilde{\ell}(b)$

La expresión (5.31) para la función de verosimilitud se puede trabajar algebraicamente. Al expandir los términos al cuadrado y luego agrupar algunos términos semejantes, se puede llegar a la siguiente fórmula.

$$\begin{aligned}\tilde{\ell}(b/x) \;=\; & \sum_{i=1}^{I}(y_i\ln(\sigma_i)+(1-y_i)\ln(1-\sigma_i))+\mu\sum_{i=1}^{I}\Big\{\big[(\beta_A^2+\theta_A^2)(p_{Ai}-d_{Ai})^2 \\ & +(\beta_B^2+\theta_B^2)(p_{Bi}-d_{Bi})^2\big]\,\sigma_i^2(1-\sigma_i)^2+2[1-2\sigma_i(1-\sigma_i)] \\ & +2(\beta_A-\theta_A)\sigma_i^2(1-\sigma_i)(p_{Ai}-d_{Ai})-2(\beta_B-\theta_B)\sigma_i(1-\sigma_i)^2(p_{Bi}-d_{Bi})\Big\}\end{aligned} \tag{E.19}$$

En este caso, se supone que los atributos contenidos en los vectores x_i son conocidos.

El gradiente de $\tilde{\ell}(b/x)$ entonces está compuesto por cada una de estas derivadas parciales de esta función con respecto a los componentes del vector b.

$$\begin{aligned}\frac{\partial\tilde{\ell}(b/x)}{\partial b_k} = & \sum_{i=1}^{I}(y_i-\sigma_i)x_{ik}+2\mu\sum_{i=1}^{I}\Big\{\big[(\beta_A^2+\theta_A^2)(p_{Ai}-d_{Ai})^2 \\ & +(\beta_B^2+\theta_B^2)(p_{Bi}-d_{Bi})^2\big]\,(1-2\sigma_i)\sigma_i(1-\sigma)-2(1-2\sigma_i)+(\beta_A-\theta_A)\sigma_i(2-3\sigma_i)(p_{Ai}-d_{Ai}) \\ & -(\beta_B-\theta_B)(1-\sigma_i)(1-3\sigma_i)(p_{Bi}-p_{Bi})\Big\}\,\sigma_i(1-\sigma_i)x_{ik} \quad \forall k\end{aligned} \tag{E.20}$$

Notemos que la expresión anterior es válida para todos los componentes del vector b con excepción de los componentes β_A, β_B, θ_A y θ_B. Estos parámetros son los que están asociados a los precios y descuentos que las firmas deciden y además se encuentran explicitamente en la fórmula de la función $\tilde{\ell}(b/x)$ recien definida, y no sólo a través de σ_i como el resto de los componentes de b. Esta situación es equivalente a la que se plantea en la sección (5.3.2), al presentar la primera expresión para la derivada $\frac{\partial\tilde{\ell}}{\partial b_k}$ en la fórmula (E.7).

Para adecuar la fórmula (E.20) para obtener las derivadas correctas de $\tilde{\ell}(b/x)$ con respecto a β_A, β_B, θ_A y θ_B ($k=2$, $k=3$, $k=4$ y $k=5$ en la fórmula respectivamente), se deben incorporar algunos términos adicionales por la presencia de estos parámetros en el resto de la fórmula tal como se hace la sección 5.3.2. Las derivadas corregidas para estos casos entonces serían las siguientes.

$$\frac{\partial\tilde{\ell}(b/x)}{\partial\beta_A} = \frac{\partial\tilde{\ell}(b/x)}{\partial b_2} + 2\mu\sum_{i=1}^{I}\left[\beta_A(p_{Ai}-d_{Ai})(1-\sigma_i)+1\right]\sigma_i^2(1-\sigma_i)(p_{Ai}-d_{Ai}) \tag{E.21}$$

$$\frac{\partial\tilde{\ell}(b/x)}{\partial\beta_B} = \frac{\partial\tilde{\ell}(b/x)}{\partial b_3} + 2\mu\sum_{i=1}^{I}\left[\beta_B(p_{Bi}-d_{Bi})\sigma_i-1\right]\sigma_i(1-\sigma_i)^2(p_{Bi}-d_{Bi}) \tag{E.22}$$

$$\frac{\partial\tilde{\ell}(b/x)}{\partial\theta_A} = \frac{\partial\tilde{\ell}(b/x)}{\partial b_4} + 2\mu\sum_{i=1}^{I}\left[\theta_A(p_{Ai}-d_{Ai})(1-\sigma_i)-1\right]\sigma_i^2(1-\sigma_i)(p_{Ai}-d_{Ai}) \tag{E.23}$$

$$\frac{\partial\tilde{\ell}(b/x)}{\partial\theta_B} = \frac{\partial\tilde{\ell}(b/x)}{\partial b_5} + 2\mu\sum_{i=1}^{I}\left[\theta_B(p_{Bi}-d_{Bi})\sigma_i+1\right]\sigma_i(1-\sigma_i)^2(p_{Bi}-d_{Bi}) \tag{E.24}$$

E.2.2. Cálculo del Hessiano $\mathbf{H}_{\tilde{\ell}}(b)$

Para determinar el hessiano de $\tilde{\ell}(b/x)$ es necesario caracterizar sus segundas derivadas parciales con respecto a los componentes del vector b. Éstas se pueden representar de forma genérica excepto para las derivadas que incluyen cruces con β_A, β_B, θ_A y θ_B. Esto porque estos cuatro parámetros se encuentran de forma explicita en la expresión de $\tilde{\ell}(b/x)$, a diferencia del resto de los componentes de b que solo están presentes indirectamente a través de σ_i. Esta situación es equivalente a la planteada en la sección 5.3.2, tanto para la primera expresión del gradiente en la fórmula (E.7) como para la primera expresión del hessiano en la fórmula (E.10).

Para construrir el hessiano se presenta primero la expresión genérica, obviando que β_A, β_B, θ_A y θ_B se encuentran de forma explicita en $\tilde{\ell}(b/x)$, y luego se realizarán los ajustes necesarios a las derivadas cruzadas que contengan estos componentes de b. El componente (j,k) de la matriz hessiana está dado por la siguiente expresión.

$$
\begin{aligned}
\frac{\partial^2\tilde{\ell}(b/x)}{\partial b_j\,\partial b_k} = &-\sum_{i=1}^{I}\sigma_i(1-\sigma_i)x_{ik}\,x_{ij}+2\mu\sum_{i=1}^{I}\Big\{\left[(\beta_A^2+\theta_A^2)(p_{Ai}-d_{Ai})^2\right.\\
&\left.+(\beta_B^2+\theta_B^2)(p_{Bi}-d_{Bi})^2\right](1-6\sigma_i+6\sigma_i^2)+2(\beta_A-\theta_A)(1-3\sigma_i)(p_{Ai}-d_{Ai})\\
&+2(\beta_B-\theta_B)(2-3\sigma_i)(p_{Bi}-d_{Bi})+4\Big\}\,\sigma_i^2(1-\sigma_i)^2x_{ik}\,x_{ij}\\
&+2\mu\sum_{i=1}^{I}\Big\{\left[(\beta_A^2+\theta_A^2)(p_{Ai}-d_{Ai})^2+(\beta_B^2+\theta_B^2)(p_{Bi}-d_{Bi})^2\right](1-2\sigma_i)\sigma_i(1-\sigma_i)\\
&-2(1-2\sigma_i)+(\beta_A-\theta_A)\sigma_i(2-3\sigma_i)(p_{Ai}-d_{Ai})\\
&-(\beta_B-\theta_B)(1-\sigma_i)(1-3\sigma_i)(p_{Bi}-d_{Bi})\Big\}(1-2\sigma_i)(1-\sigma_i)\sigma_i\,x_{ik}\,x_{ij}\quad\forall k,j
\end{aligned}
\tag{E.25}
$$

Los siguientes son los ajustes para corregir la matriz hessiana.

$$
\begin{aligned}
\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial b_k} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial b_2\partial b_k}+2\mu\sum_{i=1}^{I}[2\beta_A(p_{Ai}-d_{Ai})(1-2\sigma_i)(1-\sigma_i)\\
&\quad+(2-3\sigma_i)]\,\sigma_i^2(1-\sigma_i)x_{ik}(p_{Ai}-d_{Ai})\quad\forall k\neq 2,3,4,5
\end{aligned}
\tag{E.26}
$$

$$
\begin{aligned}
\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial b_k} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial b_3\partial b_k}+2\mu\sum_{i=1}^{I}[2\beta_B(p_{Bi}-d_{Bi})(1-2\sigma_i)\sigma_i\\
&\quad-(1-3\sigma_i)]\,\sigma_i(1-\sigma_i)^2x_{ik}(p_{Bi}-d_{Bi})\quad\forall k\neq 2,3,4,5
\end{aligned}
\tag{E.27}
$$

$$
\begin{aligned}
\frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_A\partial b_k} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial b_4\partial b_k}+2\mu\sum_{i=1}^{I}[2\theta_A(p_{Ai}-d_{Ai})(1-2\sigma_i)(1-\sigma_i)\\
&\quad-(2-3\sigma_i)]\,\sigma_i^2(1-\sigma_i)x_{ik}(p_{Ai}-d_{Ai})\quad\forall k\neq 2,3,4,5
\end{aligned}
\tag{E.28}
$$

$$
\begin{aligned}
\frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_B\partial b_k} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial b_5\partial b_k}+2\mu\sum_{i=1}^{I}[2\theta_B(p_{Bi}-d_{Bi})(1-2\sigma_i)\sigma_i\\
&\quad+(1-3\sigma_i)]\,\sigma_i(1-\sigma_i)^2x_{ik}(p_{Bi}-d_{Bi})\quad\forall k\neq 2,3,4,5
\end{aligned}
\tag{E.29}
$$

$$
\begin{aligned}
\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A^2} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial b_2}+2\mu\sum_{i=1}^{I}[2\beta_A(p_{Ai}-d_{Ai})p_{Ai}(1-2\sigma_i)(1-\sigma_i)\\
&\quad+(p_{Ai}-d_{Ai})(1-\sigma_i)+p_{Ai}(2-3\sigma_i)]\,\sigma_i^2(1-\sigma_i)(p_{Ai}-d_{Ai})
\end{aligned}
\tag{E.30}
$$

$$
\begin{aligned}
\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B^2} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial b_3}+2\mu\sum_{i=1}^{I}[2\beta_B(p_{Bi}-d_{Bi})p_{Bi}(1-2\sigma_i)\sigma_i\\
&\quad+(p_{Bi}-d_{Bi})\sigma_i-p_{Bi}(1-3\sigma_i)]\,\sigma_i(1-\sigma_i)^2(p_{Bi}-d_{Bi})
\end{aligned}
\tag{E.31}
$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_A^2} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_A\partial b_4}+2\mu\sum_{i=1}^{I}[2\theta_A(p_{Ai}-d_{Ai})d_{Ai}(1-2\sigma_i)(1-\sigma_i) +(p_{Ai}-d_{Ai})(1-\sigma_i)-d_{Ai}(2-3\sigma_i)]\,\sigma_i^2(1-\sigma_i)(p_{Ai}-d_{Ai}) \tag{E.32}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_B^2} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_B\partial b_5}+2\mu\sum_{i=1}^{I}[2\theta_B(p_{Bi}-d_{Bi})d_{Bi}(1-2\sigma_i)\sigma_i +(p_{Bi}-d_{Bi})\sigma_i-d_{Bi}(1-3\sigma_i)]\,\sigma_i(1-\sigma_i)^2(p_{Bi}-d_{Bi}) \tag{E.33}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial\beta_B} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial b_3}+2\mu\sum_{i=1}^{I}[2\beta_B(p_{Bi}-d_{Bi})(1-2\sigma_i)\sigma_i -(1-3\sigma_i)]\,\sigma_i(1-\sigma_i)^2p_{Ai}(p_{Bi}-d_{Bi}) \tag{E.34}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial\theta_A} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial b_4}+2\mu\sum_{i=1}^{I}[2\theta_A(p_{Ai}-d_{Ai})(1-2\sigma_i)(1-\sigma_i) -(2-3\sigma_i)]\,\sigma_i^2(1-\sigma_i)p_{Ai}(p_{Ai}-d_{Ai}) \tag{E.35}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial\theta_B} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial b_5}+2\mu\sum_{i=1}^{I}[2\theta_B(p_{Bi}-d_{Bi})(1-2\sigma_i)\sigma_i +(1-3\sigma_i)]\,\sigma_i(1-\sigma_i)^2p_{Ai}(p_{Bi}-d_{Bi}) \tag{E.36}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial\theta_A} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial b_4}+2\mu\sum_{i=1}^{I}[2\theta_A(p_{Ai}-d_{Ai})(1-2\sigma_i)(1-\sigma_i) -(2-3\sigma_i)]\,\sigma_i^2(1-\sigma_i)p_{Bi}(p_{Ai}-d_{Ai}) \tag{E.37}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial\theta_B} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial b_5}+2\mu\sum_{i=1}^{I}[2\theta_B(p_{Bi}-d_{Bi})(1-2\sigma_i)\sigma_i +(1-3\sigma_i)]\,\sigma_i(1-\sigma_i)^2p_{Ai}(p_{Bi}-d_{Bi}) \tag{E.38}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_A\partial\theta_B} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_A\partial b_5}+2\mu\sum_{i=1}^{I}[2\theta_B(p_{Bi}-d_{Bi})(1-2\sigma_i)\sigma_i +(1-3\sigma_i)]\,\sigma_i(1-\sigma_i)^2d_{Ai}(p_{Bi}-d_{Bi}) \tag{E.39}$$

De esta forma la matriz hessiana $\mathbf{H}_{\tilde{\ell}}(b)$ se construye inicialmente a partir de los componentes de la expresión (E.25) y luego realizando los ajustes a esta expresión a partir de las derivadas cruzadas que incluyen a las variables β_A, β_B, θ_A y θ_B, las que se detallan en las expresiones de la (E.26) a la (E.39).

E.3. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios, Descuentos en Precios y Esfuerzos en Marketing

E.3.1. Cálculo del Gradiente $\nabla_b \tilde{\ell}(b)$

La expresión (5.44) para la función de verosimilitud se puede trabajar algebraicamente. Al expandir los términos al cuadrado y luego agrupar algunos términos semejantes, se puede llegar a la siguiente fórmula.

$$\begin{aligned}\tilde{\ell}(b/x) \;=\; & \sum_{i=1}^{I}(y_i\ln(\sigma_i)+(1-y_i)\ln(1-\sigma_i))+\mu\sum_{i=1}^{I}\Big\{\big[(\beta_A^2+\theta_A^2+\phi_A^2)(p_{Ai}-d_{Ai})^2 \\ & +(\beta_B^2+\theta_B^2+\phi_B^2)(p_{Bi}-d_{Bi})^2\big]\,\sigma_i^2(1-\sigma_i)^2+2[1-2\sigma_i(1-\sigma_i)] \\ & +[2(\beta_A-\theta_A)\sigma_i-2\phi_A e]\sigma_i(1-\sigma_i)(p_{Ai}-d_{Ai}) \\ & -[2(\beta_B-\theta_B)(1-\sigma_i)-2\phi_B e]\sigma_i(1-\sigma_i)(p_{Bi}-d_{Bi})+2e^2\Big\}\end{aligned} \tag{E.40}$$

En este caso, se supone que los atributos contenidos en los vectores x_i son conocidos.

El gradiente de $\tilde{\ell}(b/x)$ entonces está compuesto por cada una de estas derivadas parciales de esta función con respecto a los componentes del vector b.

$$\begin{aligned}\frac{\partial\tilde{\ell}(b/x)}{\partial b_k} = & \sum_{i=1}^{I}(y_i-\sigma_i)x_{ik}+2\mu\sum_{i=1}^{I}\Big\{\big[(\beta_A^2+\theta_A^2+\phi_A^2)(p_{Ai}-d_{Ai})^2 \\ & +(\beta_B^2+\theta_B^2+\phi_B^2)(p_{Bi}-d_{Bi})^2\big]\,(1-2\sigma_i)\sigma_i(1-\sigma)-2(1-2\sigma_i) \\ & +(\beta_A-\theta_A)\sigma_i(2-3\sigma_i)(p_{Ai}-d_{Ai})-(\beta_B-\theta_B)(1-\sigma_i)(1-3\sigma_i)(p_{Bi}-p_{Bi}) \\ & -\phi_A\,e\,(1-2\sigma_i)(p_{Ai}-d_{Ai})+\phi_B\,e\,(1-2\sigma_i)(p_{Bi}-d_{Bi})\Big\}\,\sigma_i(1-\sigma_i)x_{ik} \quad \forall k\end{aligned} \tag{E.41}$$

Notemos que la expresión anterior es válida para todos los componentes del vector b con excepción de los componentes β_A, β_B, θ_A, θ_B, ϕ_A y ϕ_B. Estos parámetros son los que están asociados a los precios, descuentos y esfuerzos en marketing que las firmas deciden y además se encuentran explicitamente en la fórmula de la función $\tilde{\ell}(b/x)$ recien definida, y no sólo a través de σ_i como el resto de los componentes de b. Esta situación es equivalente a la que se plantea en las secciones 5.3.2 y 5.4.2, al presentar la primera expresión para la derivada $\frac{\partial\tilde{\ell}}{\partial b_k}$ en las fórmulas (E.7) y (E.20) respectivamente.

Para adecuar la fórmula (E.41) para obtener las derivadas correctas de $\tilde{\ell}(b/x)$ con respecto a β_A, β_B, θ_A, θ_B, ϕ_A y ϕ_B ($k=2$, $k=3$, $k=4$, $k=5$, $k=6$ y $k=7$ en la fórmula respectivamente), se

deben incorporar algunos términos adicionales por la presencia de estos parámetros en el resto de la fórmula tal como se hace la secciones 5.3.2 y 5.4.2.

Los ajustes que se deben realizar a las derivadas de $\tilde{\ell}(b/x)$ con respecto a β_A, β_B, θ_A y θ_B son exactamente los mismos que los que se realizan en la versión del modelo planteado en la sección 5.4.2 a través de las fórmulas (E.21), (E.22), (E.23) y (E.24) respectivamente. Para completar la corrección del gradiente de $\tilde{\ell}(b/x)$ es necesario plantear también los ajustes a las derivadas de esta función con respcto a ϕ_A y ϕ_B.

$$\frac{\partial\tilde{\ell}(b/x)}{\partial\phi_A} = \frac{\partial\tilde{\ell}(b/x)}{\partial b_6} + 2\mu\sum_{i=1}^{I}\left[\phi_A(p_{Ai}-d_{Ai})\sigma(1-\sigma_i)-e\right]\sigma_i(1-\sigma_i)(p_{Ai}-d_{Ai}) \tag{E.42}$$

$$\frac{\partial\tilde{\ell}(b/x)}{\partial\phi_B} = \frac{\partial\tilde{\ell}(b/x)}{\partial b_7} + 2\mu\sum_{i=1}^{I}\left[\phi_B(p_{Bi}-d_{Bi})\sigma(1-\sigma_i)+e\right]\sigma_i(1-\sigma_i)(p_{Bi}-d_{Bi}) \tag{E.43}$$

E.3.2. Cálculo del Hessiano $\mathbf{H}_{\tilde{\ell}}(b)$

Para determinar el hessiano de $\tilde{\ell}(b/x)$ es necesario caracterizar sus segundas derivadas parciales con respecto a los componentes del vector b. Éstas se pueden representar de forma genérica excepto para las derivadas que incluyen cruces con β_A, β_B, θ_A, θ_B, ϕ_A y ϕ_B. Esto porque estos seis parámetros se encuentran de forma explicita en la expresión de $\tilde{\ell}(b/x)$, a diferencia del resto de los componentes de b que solo están presentes indirectamente a través de σ_i. Esta situación es equivalente a la planteada en las secciones 5.3.2 y 5.4.2, tanto para las primera expresiones del gradiente como para las primeras expresiones del hessiano.

Para construrir el hessiano se presenta primero la expresión genérica, obviando que β_A, β_B, θ_A, θ_B, ϕ_A y ϕ_B se encuentran de forma explicita en $\tilde{\ell}(b/x)$, y luego se realizarán los ajustes necesarios a las derivadas cruzadas que contengan estos componentes de b. El componente (j,k) de la matriz hessiana está dado por la siguiente expresión.

$$\begin{aligned}
\frac{\partial^2\tilde{\ell}(b/x)}{\partial b_j\,\partial b_k} = &-\sum_{i=1}^{I}\sigma_i(1-\sigma_i)x_{ik}x_{ij} + 2\mu\sum_{i=1}^{I}\Big\{\left[(\beta_A^2+\theta_A^2+\phi_A^2)(p_{Ai}-d_{Ai})^2\right.\\
&\left.+(\beta_B^2+\theta_B^2+\phi_B^2)(p_{Bi}-d_{Bi})^2\right](1-6\sigma_i+6\sigma_i^2)+2(\beta_A-\theta_A)(1-3\sigma_i)(p_{Ai}-d_{Ai})\\
&+2(\beta_B-\theta_B)(2-3\sigma_i)(p_{Bi}-d_{Bi})+2\phi_A e(p_{Ai}-d_{Ai})-2\phi_B e(p_{Bi}-d_{Bi})+4\Big\}\sigma_i^2(1-\sigma_i)^2x_{ik}x_{ij}\\
&+2\mu\sum_{i=1}^{I}\Big\{\left[(\beta_A^2+\theta_A^2+\phi_A^2)(p_{Ai}-d_{Ai})^2+(\beta_B^2+\theta_B^2+\phi_B^2)(p_{Bi}-d_{Bi})^2\right](1-2\sigma_i)\sigma_i(1-\sigma_i)\\
&-2(1-2\sigma_i)+(\beta_A-\theta_A)\sigma_i(2-3\sigma_i)(p_{Ai}-d_{Ai})-(\beta_B-\theta_B)(1-\sigma_i)(1-3\sigma_i)(p_{Bi}-d_{Bi})\\
&-\phi_A e(1-2\sigma)(p_{Ai}-d_{Ai})+\phi_B e(1-2\sigma)(p_{Bi}-d_{Bi})\Big\}(1-2\sigma_i)(1-\sigma_i)\sigma_i x_{ik}x_{ij}\quad\forall k,j
\end{aligned} \tag{E.44}$$

Los ajustes que se deben realizar al hessiano de $\tilde{\ell}(b/x)$ en las segundas derivadas parciales con respecto a los cruces que incluyen las variables β_A, β_B, θ_A y θ_B son exactamente los mismos que los que se realizan en la versión del modelo planteado en la sección 5.4.2. Estos ajustes recién mencionados son los detallados en las fórmulas desde la (E.26) a la (E.39) en la sección 5.4.2. Para terminar de corregir el hessiano falta incorporar adicionalmente los ajustes a las segundas derivadas parciales con respecto a los cruces que incluyen a las variables ϕ_A y ϕ_B, que son los nuevos parámetros que se introducen en esta versión del modelo al incorporar los esfuerzos en marketing en las decisiones de las firmas. Estos ajustes se detallan desde la formula (E.45) a la (E.57).

$$\begin{aligned}\frac{\partial^2\tilde{\ell}(b/x)}{\partial\phi_A^2} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial\phi_A\partial b_6}+2\mu\sum_{i=1}^{I}[2\phi_A(p_{Ai}-d_{Ai})m_{Ai}(1-2\sigma_i)\sigma_i(1-\sigma_i)\\ &\quad +(p_{Ai}-d_{Ai})\sigma_i(1-\sigma_i)-e(1-2\sigma_i)m_{Ai}]\,\sigma_i(1-\sigma_i)(p_{Ai}-d_{Ai})\end{aligned} \tag{E.45}$$

$$\begin{aligned}\frac{\partial^2\tilde{\ell}(b/x)}{\partial\phi_B^2} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial\phi_B\partial b_7}+2\mu\sum_{i=1}^{I}[2\phi_B(p_{Bi}-d_{Bi})m_{Bi}(1-2\sigma_i)\sigma_i(1-\sigma_i)\\ &\quad +(p_{Bi}-d_{Bi})\sigma_i(1-\sigma_i)+e(1-2\sigma_i)m_{Bi}]\,\sigma_i(1-\sigma_i)(p_{Bi}-d_{Bi})\end{aligned} \tag{E.46}$$

$$\begin{aligned}\frac{\partial^2\tilde{\ell}(b/x)}{\partial\phi_A\partial b_k} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial b_6\partial b_k}+2\mu\sum_{i=1}^{I}[2\phi_A(p_{Ai}-d_{Ai})\sigma_i(1-\sigma_i)\\ &\quad -e]\,\sigma_i(1-\sigma_i)(1-2\sigma_i)x_{ik}(p_{Ai}-d_{Ai}) \quad \forall k\neq 2,3,4,5,6,7\end{aligned} \tag{E.47}$$

$$\begin{aligned}\frac{\partial^2\tilde{\ell}(b/x)}{\partial\phi_B\partial b_k} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial b_7\partial b_k}+2\mu\sum_{i=1}^{I}[2\phi_B(p_{Bi}-d_{Bi})\sigma_i(1-\sigma_i)\\ &\quad +e]\,\sigma_i(1-\sigma_i)(1-2\sigma_i)x_{ik}(p_{Bi}-d_{Bi}) \quad \forall k\neq 2,3,4,5,6,7\end{aligned} \tag{E.48}$$

$$\begin{aligned}\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial\phi_A} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial b_6}+2\mu\sum_{i=1}^{I}[2\phi_A(p_{Ai}-d_{Ai})\sigma_i(1-\sigma_i)\\ &\quad -e]\,\sigma_i(1-\sigma_i)(1-2\sigma_i)p_{Ai}(p_{Ai}-d_{Ai})\end{aligned} \tag{E.49}$$

$$\begin{aligned}\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial\phi_B} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_A\partial b_7}+2\mu\sum_{i=1}^{I}[2\phi_B(p_{Bi}-d_{Bi})\sigma_i(1-\sigma_i)\\ &\quad +e]\,\sigma_i(1-\sigma_i)(1-2\sigma_i)p_{Ai}(p_{Bi}-d_{Bi})\end{aligned} \tag{E.50}$$

$$\begin{aligned}\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial\phi_A} &= \frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial b_6}+2\mu\sum_{i=1}^{I}[2\phi_A(p_{Ai}-d_{Ai})\sigma_i(1-\sigma_i)\\ &\quad -e]\,\sigma_i(1-\sigma_i)(1-2\sigma_i)p_{Bi}(p_{Ai}-d_{Ai})\end{aligned} \tag{E.51}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial\phi_B} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\beta_B\partial b_7}+2\mu\sum_{i=1}^{I}[2\phi_B(p_{Bi}-d_{Bi})\sigma_i(1-\sigma_i) +e]\,\sigma_i(1-\sigma_i)(1-2\sigma_i)p_{Bi}(p_{Bi}-d_{Bi}) \tag{E.52}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_A\partial\phi_A} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_A\partial b_6}+2\mu\sum_{i=1}^{I}[2\phi_A(p_{Ai}-d_{Ai})\sigma_i(1-\sigma_i) -e]\,\sigma_i(1-\sigma_i)(1-2\sigma_i)d_{Ai}(p_{Ai}-d_{Ai}) \tag{E.53}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_A\partial\phi_B} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_A\partial b_7}+2\mu\sum_{i=1}^{I}[2\phi_B(p_{Bi}-d_{Bi})\sigma_i(1-\sigma_i) +e]\,\sigma_i(1-\sigma_i)(1-2\sigma_i)d_{Ai}(p_{Bi}-d_{Bi}) \tag{E.54}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_B\partial\phi_A} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_B\partial b_6}+2\mu\sum_{i=1}^{I}[2\phi_A(p_{Ai}-d_{Ai})\sigma_i(1-\sigma_i) -e]\,\sigma_i(1-\sigma_i)(1-2\sigma_i)d_{Bi}(p_{Ai}-d_{Ai}) \tag{E.55}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_B\partial\phi_B} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\theta_B\partial b_7}+2\mu\sum_{i=1}^{I}[2\phi_B(p_{Bi}-d_{Bi})\sigma_i(1-\sigma_i) +e]\,\sigma_i(1-\sigma_i)(1-2\sigma_i)d_{Bi}(p_{Bi}-d_{Bi}) \tag{E.56}$$

$$\frac{\partial^2\tilde{\ell}(b/x)}{\partial\phi_A\partial\phi_B} = \frac{\partial^2\tilde{\ell}(b/x)}{\partial\phi_A\partial b_7}+2\mu\sum_{i=1}^{I}[2\phi_A(p_{Bi}-d_{Bi})\sigma_i(1-\sigma_i) +e]\,\sigma_i(1-\sigma_i)(1-2\sigma_i)m_{Ai}(p_{Bi}-d_{Bi}) \tag{E.57}$$

De esta forma la matriz hessiana $\mathbf{H}_{\tilde{\ell}}(b)$ se construye inicialmente a partir de los componentes de la expresión (E.44) y luego realizando los ajustes a esta expresión a partir de las derivadas cruzadas que incluyen a las variables β_A, β_B, θ_A, θ_B, ϕ_A y ϕ_B.

Anexo F

Implementación Computacional de las Metodologías

F.1. Modelo de Regresión Logística

Como se plantea en capítulo 4, para resolver el Modelo de Regresión Logística se utiliza el método Newton-Raphson. Este método plantea una recurrencia que permite avanzar hacia la solución óptima a partir de los valores del gradiente y del hessiano de la función objetivo. A continuación se presentan las estrategias de implementación del gradiente, el hessiano y luego del método de Newton-Raphson, así como sus códigos fuente en el lenguaje R.

F.1.1. Cálculo del Gradiente $\nabla_b \ell(b/x)$

El cálculo del gradiente $\nabla_b \ell(b/x)$ se muestra en la expresión (4.29). Para fines de la implementación del cálculo del gradiente, a continuación se re-escribe esta expresión pero ahora de forma matricial. Para esto se requiere definir primero la matriz x, el vector σ y el vector y.

Se define la matriz $x \in \mathbb{R}^{I \times n}$, la cual tiene I filas (cada una asociada a un objeto) y n columnas (cada una asociada a uno de los atributos que explican las características de los objetos). Para construir esta matriz en cada fila se ubican los vectores $x_i \in \mathbb{R}^n$ de forma transpuesta. Al transponer el vector x_i, en la fila i queda una sub-matriz de 1 fila y n columnas cada una asociada a los n atributos que explican las características del objeto. Al considerar los I vectores x_i se consigue formar la matriz x. Como se ha definido que el primer componente de los vectores x_i es un valor uno (así el parámetro asociado corresponde a la constante) entonces la matriz x tiene una columna de unos.

$$x = \begin{pmatrix} (x_1)^T \\ \vdots \\ (x_i)^T \\ \vdots \\ (x_I)^T \end{pmatrix} = \begin{pmatrix} x_{11} & \cdots & x_{1k} & \cdots & x_{1n} \\ \vdots & & \vdots & & \vdots \\ x_{i1} & \cdots & x_{ik} & \cdots & x_{in} \\ \vdots & & \vdots & & \vdots \\ x_{I1} & \cdots & x_{Ik} & \cdots & x_{In} \end{pmatrix} = \begin{pmatrix} 1 & x_{12} & \cdots & x_{1k} & \cdots & x_{1n} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{i2} & \cdots & x_{ik} & \cdots & x_{in} \\ \vdots & \vdots & & \vdots & & \vdots \\ 1 & x_{I2} & \cdots & x_{Ik} & \cdots & x_{In} \end{pmatrix} \tag{F.1}$$

Se definen el vector σ y el vector y, ambos de dimensión I, cuyos componentes son las variables σ_i e y_i respectivamente.

$$\sigma = \begin{pmatrix} \sigma_1 \\ \vdots \\ \sigma_i \\ \vdots \\ \sigma_I \end{pmatrix} \qquad y = \begin{pmatrix} y_1 \\ \vdots \\ y_i \\ \vdots \\ y_I \end{pmatrix} \tag{F.2}$$

Las variables σ_i se pueden calcular a partir de la expresión (4.26). Para fines de la implementación computacional se utiliza la función **`get_prob`**, que recibe como insumos b y x, cuyo código R se encuentra en el Código Fuente del Anexo G.5.

El gradiente $\nabla_b \ell(b/x)$ de la función $\ell(b/x)$ se puede construir de forma matricial partiendo de la expresión (4.29) para las derivadas parciales de $\ell(b/x)$ con respecto a cada componente del vector b, quedando expresado de la siguiente forma.

$$\nabla_b \ell(b/x) = x^T \cdot (y - \sigma) \tag{F.3}$$

Para fines de la implementación computacional es posible crear la función **`gradiente_f`**, que realiza el cálculo matricial anterior y que utiliza tres insumos: el vector de parámetros de los atributos b, la matriz x y el vector y. El código R de esta función se detalla el Código Fuente del Anexo G.3.

F.1.2. Cálculo del Hessiano $\mathbf{H}_\ell(b/x)$

El cálculo del hessiano $\mathbf{H}_\ell(b/x)$ se muestra en la expresión (4.35). Para fines de la implementación del cálculo del hessiano, a continuación se re-escribe esta expresión pero ahora de forma matricial. Para esto se requiere definir además la matriz $W \in \mathbb{R}^{I \times I}$, cuyos componentes están dado por la siguiente expresión.

$$W_{ij} = \begin{cases} \sigma_i(1-\sigma_i) & \text{Si } i = j \quad \forall i, j \in \{1, ..., I\} \\ 0 & \sim \end{cases} \tag{F.4}$$

De esta forma, y partiendo de la expresión (4.35) que detalla cada uno de sus componentes, el hessiano de $\ell(b/x)$ se puede expresar de la siguiente forma matricial.

$$\mathbf{H}_\ell(b/x) = -x^T \cdot W \cdot x \tag{F.5}$$

Notemos que en particular la matriz W, de dimensiones $I \times I$, puede llegar a tener un alto número de componentes cuando el problema tiene un número alto de observaciones I. Si I es muy grande entonces existen problemas de memoria computacional para almacenar todos los componentes de esta matriz. Otro problema es el aumento en los tiempos de ejecución al calcular el hessiano de $\ell(b/x)$, explicado por el número de cómputos que se debe realizar debido al número de componentes de la matriz W. Para disminuir la complejidad computacional del cálculo de $\mathbf{H}_\ell(b/x)$ se propone realizarlo por partes. Para esto es necesario generar particiones de las matrices x y W.

La matriz x será particionada a lo largo de las filas de la matriz. Cada una de las particiones corresponde a una sub-matriz compuesta por hasta I_{max} filas y por las n columnas de x. El parámetro I_{max} se define como el número máximo de filas que se desea almacenar en las particiones, de un total de I filas ($I_{max} < I$). Así, el número de particiones requeridas corresponde al parámetro $s = \lceil I/I_{max} \rceil$. Como I_{max} no tiene que ser necesariamente múltiplo del número total de observaciones I, la última partición tendrá las últimas $u = I - I_{max}(s-1)$ filas de la matriz x original. De esta forma se definen s particiones x^j, con $j = 1, \ldots, s$, tales que las $s-1$ primeras particiones están dadas por $x^j \in \mathbb{R}^{I_{max} \times n}$ $\forall j < s$ y la última partición corresponde a $x^s \in \mathbb{R}^{u \times n}$.

$$x^j = \begin{pmatrix} x_{(j-1)I_{max}+1,1} & \cdots & x_{(j-1)I_{max}+1,n} \\ \vdots & \ddots & \vdots \\ x_{jI_{max},1} & \cdots & x_{jI_{max},n} \end{pmatrix} \forall j < s; \quad x^s = \begin{pmatrix} x_{(s-1)I_{max}+1,1} & \cdots & x_{(s-1)I_{max}+1,n} \\ \vdots & \ddots & \vdots \\ x_{I,1} & \cdots & x_{I,n} \end{pmatrix} \tag{F.6}$$

La matriz W es particionada en varias sub-matrices cuadradas de dimensiones de hasta $I_{max} \times I_{max}$. El parámetro I_{max} corresponde al número máximo de filas y columnas que se desea almacenar en las particiones, de un total de I filas y columnas de la matriz W. El número de particiones requeridas corresponde a $s \times s$ tal que $s = \lceil I/I_{max} \rceil$. Como W es una matriz sólo con elementos en la diagonal y con ceros en el resto de la matriz, entonces las particiones de la diagonal tendrán esta misma característica y el resto de las particiones sólo tendrán ceros en sus componentes. Dado esto, para expresar la matriz W en particiones basta con definir s particiones W^j, con $j = 1, \ldots, s$, que componen la diagonal de la matriz original. El resto de la matriz se compone de sub-matrices con ceros en todos sus componentes. Las primeras $s-1$ particiones están dadas por $W^j \in \mathbb{R}^{I_{max} \times I_{max}}$

$\forall j < s$. En la última partición se almacenarán las últimas $u = I - I_{max}(s-1)$ filas y columnas de la matriz W original, por lo que esta se define como $W^s \in \mathbb{R}^{u \times u}$.

$$W = \begin{pmatrix} W^1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & W^s \end{pmatrix} \tag{F.7}$$

Cada partición W^j es una sub-matriz cuyos componentes (i,h) están dados de acuerdo a la siguiente expresión.

$$W^j_{ih} = \begin{cases} \sigma_i(1-\sigma_i) & \text{Si } i = h \quad \forall i,h \\ 0 & \sim \end{cases} \tag{F.8}$$

Cuando $j < s$ entonces $i,h \in \{(j-1)I_{max}+1, ..., jI_{max}\}$. Para la última partición, cuando $j = s$ entonces $i,h \in \{(s-1)I_{max}+1, ..., I\}$.

A partir de las particiones x^j y W^j definidas anteriormente, es posible re-escribir la expresión matricial $-x^T \cdot W \cdot x$ para el cálculo del hessiano $\mathbf{H}_\ell(b/x)$. Utilizando las propiedades de la multiplicación por partes de matrices, se puede llegar a la siguiente expresión.

$$\begin{aligned} \mathbf{H}_\ell(b/x) &= -x^T \cdot W \cdot x = -\begin{pmatrix} (x^1)^T & \dots & (x^s)^T \end{pmatrix} \cdot \begin{pmatrix} W^1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & W^s \end{pmatrix} \cdot \begin{pmatrix} x^1 \\ \vdots \\ x^s \end{pmatrix} \\ &= -\begin{pmatrix} (x^1)^T & \dots & (x^s)^T \end{pmatrix} \cdot \begin{pmatrix} W^1 \cdot x^1 \\ \vdots \\ W^s \cdot x^q \end{pmatrix} = -\sum_{j=1}^{s} (x^j)^T \cdot W^j \cdot x^j \end{aligned} \tag{F.9}$$

De esta forma, el hessiano se puede calcular alternativamente a partir de la suma de las s matrices $-(x^j)^T \cdot W^j \cdot x^j$. Notemos que es posible sumar las matrices anteriores porque, independiente del tamaño de las particiones, la matriz $-(x^j)^T \cdot W^j \cdot x^j$ siempre tiene dimensiones de $n \times n$. Por ejemplo, para $j < s$ las dimensiones de $(x^j)^T$, W^j y x^j siempre son $(n \times I_{max})$, $(I_{max} \times I_{max})$ y $(I_{max} \times n)$ respectivamente por lo que la multiplicación $-(x^j)^T \cdot W^j \cdot x^j$ tiene dimensiones de $n \times n$.

La propiedad anterior permite implementar una función de cálculo del hessiano más eficiente en términos de complejidad computacional con ahorros importantes respecto de la memoria requerida y del número de cómputos necesarios. Primero se define una variable auxiliar hf_sum y luego se itera recorriendo cada partición $j = 1, \dots, s$. En la j-ésima iteración se crean las particiones x^j y

W^j. Para definir x^j se necesita la matriz x, según (F.6), y para definir W^j se necesita el vector σ, según (F.8). Luego, en la misma iteración se actualiza la variable auxiliar de acuerdo a la expresión (F.10); y luego se destruyen las particiones x^j y W^j liberando la memoria utilizada.

$$hf_sum = hf_sum - (x^j)^T \cdot W^j \cdot x^j \tag{F.10}$$

Al finalizar las iteraciones el hessiano estará dado por la variable auxiliar: $\mathbf{H}_\ell(b/x) = hf_sum$.

Para fines de la implementación computacional es posible crear la función **`hessiano_f`** que utiliza dos insumos: el vector de parámetros de los atributos b y la matriz x. El código R asociado a esta función se encuentra en el Código Fuente del Anexo G.4.

F.1.3. Método de Newton-Raphson

De acuerdo a lo planteado en el capítulo 4, para resolver el modelo de Regresión Logística utilizando el método de Newton-Raphson se requiere construir la siguiente recurrencia que permite en cada iteración ir avanzando hacia la solución óptima.

$$b^{(q+1)} = b^{(q)} - \mathbf{H}_\ell^{-1}(b^{(q)})\,\nabla_b \ell(b^{(q)}) \tag{F.11}$$

Para comenzar la recurrencia anterior se necesita de una solución inicial $b^{(0)}$ y para finalizarla de un criterio de parada. En cada iteración es necesario determinar el gradiente $\nabla_b \ell(b^{(q)})$ y el hessiano $\mathbf{H}_\ell(b^{(q)})$.

Para implementar lo anterior, se ha desarrollado la función **`newton_raphson`** la que recibe tres insumos: una solución inicial $b^{(0)}$, la matriz x que contiene los atributos de todas las observaciones que son parte de la muestra y el vector y que indica la clase a la que pertenece cada una de las observaciones. A partir de estos insumos, más la definición de un criterio de parada, es posible aplicar la recurrencia anteriormente definida. La función **`newton_raphson`** utiliza las funciones **`gradiente_f`** y **`hessiano_f`** para determinar el gradiente $\nabla_b \ell(b^{(q)})$ y el hessiano $\mathbf{H}_\ell(b^{(q)})$ respectivamente, las cuales ya fueron definidas en la secciones anteriores. En cada iteración, la función **`gradiente_f`** recibe como insumo la solución candidata $b^{(q)}$, la matriz x y el vector y. En el caso de la función **`hessiano_f`**, en cada iteración recibe como insumo la solución candidata $b^{(q)}$ y la matriz x.

El criterio de parada implementado en la función **`newton_raphson`** corresponde a la tercera alternativa planteada en la sección 4.4.4: $\|b^{(q+1)} - b^{(q)}\|_1 < \varepsilon$. Utilizando este criterio el método de Newton-Raphson finalizará sus iteraciones cuando la solución encontrada no cambia de una iteración a otra. Lo anterior implica que el valor de la función objetivo no mejorará en las nuevas

iteraciones y por lo tanto el gradiente es aproximadamente cero. Este criterio de parada es el más simple de evaluar en términos computacionales según lo planteado en la sección 4.4.4. Para fines de la implementación se ha utilizado un valor para la variable $\varepsilon = 0{,}000000001$, el cual asegura un nivel de precisión bastante exigente.

El código R asociado a la función **`newton_raphson`** se encuentra en el Código Fuente del Anexo G.2.

Adicionalmente se ha construído la función **`run_logistica`** que tiene como objetivo administrar la estimación del modelo de Regresión Logística. Esta función recibe los insumos que permiten construir la matriz x y el vector y. Luego, determina una solución inicial $b^{(0)}$ utilizando la función **`newton_raphson`** para estimar el modelo de Regresión Logística sólo utilizando una constante (sin variables explicativas) y partiendo de una solución inicial con todos los parámetros iguales a cero. De esta forma, el $b^{(0)}$ resultante del proceso anterior es igual a cero para todos los parámetros asociados a alguno de los atributos y diferente de cero para la constante. Finalmente, partiendo de $b^{(0)}$ como solución inicial, se utiliza nuevamente la función **`newton_raphson`** para obtener la solución definitiva del modelo de Regresión Logística. El código R asociado a la función **`run_logistica`** se encuentra en el Código Fuente del Anexo G.1.

F.2. Modelo de Regresión Logística con Restricciones de Competencia

F.2.1. Método de Newton-Raphson

Para utilizar el método de Newton-Raphson para resolver los modelos de Fuga de Clientes con Restricciones de Competencia propuestos en el capítulo 5, se requiere construir la siguiente recurrencia que permite en cada iteración ir avanzando hacia la solución óptima.

$$b^{(q+1)} = b^{(q)} - \mathbf{H}_{\tilde{\ell}}^{-1}(b^{(q)})\,\nabla_b \tilde{\ell}(b^{(q)}) \tag{F.12}$$

Para comenzar la recurrencia anterior se necesita de una solución inicial $b^{(0)}$ y para finalizarla de un criterio de parada. En cada iteración es necesario determinar el gradiente $\nabla_b \tilde{\ell}(b^{(q)})$ y el hessiano $\mathbf{H}_{\tilde{\ell}}(b^{(q)})$, cuyo cálculo depende de cuál de los modelos propuestos se busca resolver.

Para implementar lo anterior, se ha desarrollado la función **`newton_raphson_modificado`** la que recibe seis insumos: una solución inicial $b^{(0)}$, la matriz x, el vector y, el parámetro μ, la variable *modelo* y el parámetro e. El parámetro μ es el multiplicador que utilizan los modelos propuestos en el capítulo 5 que pondera el peso relativo del cumplimiento de las condiciones de equilibrio con

respecto a la maximización de la función de verosimilitud. La variable *modelo* indica cuál de los tres modelos propuestos en el capítulo 5 se desea resolver: $modelo = 1$ corresponde al modelo de Fuga de Clientes con Restricciones de Competencia en Precios (sección 5.3); $modelo = 2$ corresponde al modelo de Fuga de Clientes con Restricciones de Competencia en Precios y Descuentos en Precios (sección 5.4); y $modelo = 3$ corresponde al modelo de Fuga de Clientes con Restricciones de Competencia en Precios, Descuentos en Precios y Esfuerzos en Marketing (sección 5.5). El parámetro *e* corresponde al parámetro utilizado por el tercer modelo propuesto y que representa el costo marginal de los esfuerzos en marketing que realizan las firmas (si $modelo \neq 3$ entonces este parámetro es ignorado).

El criterio de parada implementado en la función **newton_raphson_modificado** corresponde al mismo que se propone para la función **newton_raphson** que resuelve el modelo de Regresión Logística tradicional: $\|b^{(q+1)} - b^{(q)}\|_1 < \varepsilon$. Para fines de la implementación también se ha utilizado un valor para la variable $\varepsilon = 0{,}000000001$, el cual asegura un nivel de precisión bastante exigente.

La función **newton_raphson_modificado** utiliza el valor de la variable *modelo* para identificar el modelo que se desea resolver y escoger las funciones que se deben utilizar para determinar el gradiente y el hessiano utilizados por el método de Newton-Rapshon. Las funciones **gradiente_L_01** y **hessiano_L_01** son utilizadas cuando $modelo = 1$. La implementación de estas funciones se presenta en la sección F.2.2. Las funciones **gradiente_L_02** y **hessiano_L_02** son utilizadas cuando $modelo = 2$. La implementación de estas funciones se presenta en la sección F.2.3. Las funciones **gradiente_L_03** y **hessiano_L_03** son utilizadas cuando $modelo = 3$. La implementación de estas funciones se presenta en la sección F.2.4. En cada iteración las funciones asociadas al gradiente utilizan como insumos a la solución candidata $b^{(q)}$, la matriz x, el vector y, el parámetro μ y, sólo cuando $modelo = 3$, el parámetro e. Las funciones asociadas al hessiano utilizan como insumos a la solución candidata $b^{(q)}$, la matriz x, el parámetro μ y, sólo cuando $modelo = 3$, el parámetro e.

El código R asociado a la función **newton_raphson_modificado** se encuentra en el Código Fuente del Anexo G.7.

Adicionalmente se ha construído la función **run_logistica_modificada** que tiene como objetivo administrar la estimación de los modelos de Fuga de Clientes con Restricciones de Competencia propuestos en este trabajo. Esta función recibe los insumos que permiten construir la matriz x, el vector y, la variable *modelo*, el parámetro μ y el parámetro e. Luego, determina una solución inicial $b^{(0)}$ utilizando la función **newton_raphson** para estimar un modelo de Regresión Logística sólo utilizando una constante (sin variables explicativas) y partiendo de una solución inicial con todos los parámetros iguales a cero. De esta forma, el $b^{(0)}$ resultante del proceso anterior es igual a cero para todos los parámetros asociados a alguno de los atributos y diferente de cero para la constante. Finalmente, partiendo de $b^{(0)}$ como solución inicial, se utiliza ahora la función **newton_raphson_modificado** para obtener la solución definitiva del modelo de Fuga de Clientes con Restricciones de Competencia según sea el valor de la variable *modelo*. El código R asociado a la función **run_logistica_modificada** se encuentra en el Código Fuente del

Anexo G.6.

F.2.2. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios

Función `gradiente_L_01` para cálculo de $\nabla_b \tilde{\ell}(b/x)$

La función **`gradiente_L_01`** tiene por objetivo el cálculo de $\nabla_b \ell(b/x)$ para el modelo propuesto de Fuga de Clientes con Restricciones de Competencia en Precios. En cada iteración esta función utiliza como insumos a la solución candidata $b^{(q)}$, la matriz x, el vector y y el parámetro μ.

La función implementa el cálculo del gradiente $\nabla_b \tilde{\ell}(b/x)$ de la misma forma en que lo hace la función **`gradiente_f`**. Se realiza el cálculo matricial de la formula (E.7), que tiene la expresión básica del cálculo de las derivadas parciales de este gradiente. Pero luego se realizan los ajustes a las derivadas parciales β_A y β_B que son detallados en las expresiones (E.8) y (E.9).

Cálculo `hessiano_L_01` para cálculo de $\mathbf{H}_{\tilde{\ell}}(b/x)$

La función **`hessiano_L_01`** tiene por objetivo el cálculo de $\mathbf{H}_{\tilde{\ell}}(b/x)$ para el modelo propuesto de Fuga de Clientes con Restricciones de Competencia en Precios. En cada iteración esta función utiliza como insumos a la solución candidata $b^{(q)}$, la matriz x y el parámetro μ.

La función implementa el cálculo del hessiano $\mathbf{H}_{\tilde{\ell}}(b/x)$ de la misma forma en que lo hace la función **`hessiano_f`**. Se realiza el cálculo matricial y por partes de la formula (E.10), que tiene la expresión básica del cálculo de las derivadas parciales de este gradiente. Pero luego se realizan los ajustes a las segundas derivadas parciales que contienen cruces con las variables β_A y β_B y que son detallados en las expresiones de la (E.14) a la (E.18).

F.2.3. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios y Descuentos en Precios

Función `gradiente_L_02` para cálculo de $\nabla_b \tilde{\ell}(b/x)$

La función **`gradiente_L_02`** tiene por objetivo el cálculo de $\nabla_b \ell(b/x)$ para el modelo propuesto de Fuga de Clientes con Restricciones de Competencia en Precios y Descuentos en Precios. En cada iteración esta función utiliza como insumos a la solución candidata $b^{(q)}$, la matriz x, el vector y y el parámetro μ.

La función implementa el cálculo del gradiente $\nabla_b \tilde{\ell}(b/x)$ de la misma forma en que lo hace la función **`gradiente_f`**. Se realiza el cálculo matricial de la formula (E.20), que tiene la expresión básica del cálculo de las derivadas parciales de este hessiano. Pero luego se realizan los ajustes a las derivadas parciales β_A, β_B, θ_A y θ_B que son detallados en las expresiones (E.21) a la (E.24).

Cálculo `hessiano_L_02` para cálculo de $\mathbf{H}_{\tilde{\ell}}(b/x)$

La función **`hessiano_L_02`** tiene por objetivo el cálculo de $\mathbf{H}_{\tilde{\ell}}(b/x)$ para el modelo propuesto de Fuga de Clientes con Restricciones de Competencia en Precios y Descuentos en Precios. En cada iteración esta función utiliza como insumos a la solución candidata $b^{(q)}$, la matriz x y el parámetro μ.

La función implementa el cálculo del hessiano $\mathbf{H}_{\tilde{\ell}}(b/x)$ de la misma forma en que lo hace la función **`hessiano_f`**. Se realiza el cálculo matricial y por partes de la formula (E.25), que tiene la expresión básica del cálculo de las derivadas parciales de este hessiano. Pero luego se realizan los ajustes a las segundas derivadas parciales que contienen cruces con las variables β_A, β_B, θ_A y θ_B y que son detallados en las expresiones de la (E.26) a la (E.39).

F.2.4. Modelo de Fuga de Clientes con Restricciones de Competencia en Precios, Descuentos en Precios y Esfuerzos en Marketing

Función `gradiente_L_03` para cálculo de $\nabla_b \tilde{\ell}(b/x)$

La función **`gradiente_L_03`** tiene por objetivo el cálculo de $\nabla_b \ell(b/x)$ para el modelo propuesto de Fuga de Clientes con Restricciones de Competencia en Precios, Descuentos en Precios y Esfuerzos en Marketing. En cada iteración esta función utiliza como insumos a la solución candidata $b^{(q)}$, la matriz x, el vector y, el parámetro μ y el parámetro e.

La función implementa el cálculo del gradiente $\nabla_b \tilde{\ell}(b/x)$ de la misma forma en que lo hace la función **`gradiente_f`**. Se realiza el cálculo matricial de la formula (E.41), que tiene la expresión básica del cálculo de las derivadas parciales de este gradiente. Pero luego se realizan los ajustes a las derivadas parciales β_A, β_B, θ_A, θ_B, ϕ_A y ϕ_B. Los ajustes sobre las derivadas parciales β_A, β_B, θ_A y θ_B son los mismos que realizan sobre la función **`gradiente_L_02`**, es decir, las expresiones desde la (E.21) a la (E.24). Y los ajustes sobre las derivadas parciales ϕ_A y ϕ_B son los que se detallan en las expresiones (E.42) y (E.43).

Cálculo `hessiano_L_03` para cálculo de $\mathbf{H}_{\tilde{\ell}}(b/x)$

La función **`hessiano_L_03`** tiene por objetivo el cálculo de $\mathbf{H}_{\tilde{\ell}}(b/x)$ para el modelo propuesto de Fuga de Clientes con Restricciones de Competencia en Precios, Descuentos en Precios y Esfuerzos en Marketing. En cada iteración esta función utiliza como insumos a la solución candidata $b^{(q)}$, la matriz x, el parámetro μ y el parámetro e.

La función implementa el cálculo del hessiano $\mathbf{H}_{\tilde{\ell}}(b/x)$ de la misma forma en que lo hace la función **`hessiano_f`**. Se realiza el cálculo matricial y por partes de la formula (E.44), que tiene la expresión básica del cálculo de las derivadas parciales de este hessiano. Pero luego se realizan los ajustes a las segundas derivadas parciales que contienen cruces con las variables β_A, β_B, θ_A, θ_B, ϕ_A y ϕ_B. Los ajustes sobre las segundas derivadas parciales que contienen cruces con las variables β_A, β_B, θ_A y θ_B son los mismos que realizan sobre la función **`hessiano_L_02`**, es decir, las expresiones desde la (E.26) a la (E.39). Y los ajustes sobre las segundas derivadas parciales que contienen cruces con las variables ϕ_A y ϕ_B son los que se detallan en las expresiones de la (E.45) a la (E.57).

Anexo G

Implementación Computacional de las Metodologías: Códigos de Fuente

G.1. Regresión Logística

```
run_logistica <- function(datos,y) {

#datos tiene todas las variables a utilizar
I <- dim(datos)[1] #numero de objetos
m <- dim(datos)[2] #numero de atributos

X <- cbind(matrix(1, nrow=I, ncol=1),as.matrix.data.frame(datos[,1:m])) #set
    de atributos incluyendo vector de 1 al principio
y <- as.matrix(y) #variable dependiente

#eliminamos los nombres de columnas y filas
colnames(X)<-NULL
rownames(X)<-NULL
colnames(y)<-NULL
rownames(y)<-NULL

b_cte <- newton_raphson(as.matrix(0),as.matrix(X[,1]),y)
b_inicial <- matrix(0, nrow=(m+1), ncol=1) #m atributos + la cte
b_inicial[1,1] <- b_cte

run_logistica <- newton_raphson(b_inicial,X,y)

}
```

Código Fuente G.1: Código R para estimación de modelo de Regresión Logística

```
newton_raphson <- function(b_inicial, X, y) {

converged <- 0
iter <- 0
max_iter <- 100
epsilon <- 0.000000001

b_next <- b_inicial

while(iter < max_iter && converged==0){
    iter <- iter+1
    b <- b_next

    grad_f <- gradiente_f(b, X, y)
    Hf <- hessiano_f(b, X)
    b_next <- b - solve(Hf) %*% grad_f

    if( sum( abs(b_next - b)) < epsilon ) converged <- 1
}

newton_raphson <- b_next

}
```

Código Fuente G.2: Código R con método Newton-Raphson

```
gradiente_f <- function(b, X, y) {

sigma<-get_prob(b, X)

gradiente_f <- t(X) %*% (y-sigma)

}
```

Código Fuente G.3: Código R para cálculo de $\nabla_b \ell(b/x)$

```
hessiano_f <- function(b, X) {

I <- dim(X)[1] #número de objetos
n <- length(b) #número de atributos (incluye cte)
I_max <- 1000
s <- ceiling(I/I_max) #número de particiones de w

I_j<-I_max #tamaño de la particion j, para j<s
I_acum<-0 #para identificar en que I comienza y termina cada partición

sigma<-get_prob(b, X)
hf_sum <- matrix(0, nrow=n, ncol=n) #suma acumulada de las matrices hessianas
    que aporta cada partición
```

```
13
14 for(j in 1:s){ #iteramos sobre las particiones de w
15     if(j==s){
16         I_j<- I - I_max*(s-1) #la ultima particion tiene los elementos que
               sobran
17     }
18
19     w_j<- matrix(0, nrow=I_j, ncol=I_j)
20     diag(w_j)<- sigma[(I_acum+1):(I_acum+I_j)]*(1-sigma[(I_acum+1):(I_acum+I_j
           )]) # se genera particio de W
21
22     hf_sum <- hf_sum - t(X[(I_acum+1):(I_acum+I_j),]) %*% w_j %*% X[(I_acum+1)
           :(I_acum+I_j),]
23
24     rm(w_j) #se elimina partición
25     I_acum<-I_acum+I_j #coordenada de inicio de sgte particion
26 }
27
28 hessiano_f <- hf_sum
29
30 }
```

Código Fuente G.4: Código R para cálculo de $\mathbf{H}_\ell(b/x)$

```
1 get_prob <- function(b, X) {
2
3 get_prob <- 1/(1+exp(-X%*%b))
4
5 }
```

Código Fuente G.5: Código R para cálculo de σ

G.2. Modelos Propuestos

```
1  run_logistica_modificada <- function(datos,y,modelo,mu,e) {
2
3  #datos tiene todas las variables a utilizar
4  I <- dim(datos)[1]
5  m <- dim(datos)[2] #numero de atributos
6
7  X <- cbind(matrix(1, nrow=I, ncol=1),as.matrix.data.frame(datos[,1:m])) #set
       de atributos incluyendo vector de 1 al principio
8  y <- as.matrix(y) #clasificacion
9
10 #eliminamos los nombres de columnas y filas
11 colnames(X)<-NULL
```

```
rownames(X)<-NULL
colnames(y)<-NULL
rownames(y)<-NULL

b_cte <- newton_raphson(as.matrix(0),as.matrix(X[,1]),y)
b_inicial <- matrix(0, nrow=(m+1), ncol=1) #m atributos + la cte
b_inicial[1,1] <- b_cte

run_logistica_modificada <- newton_raphson_modificado(b_inicial, X, y, mu,
    modelo, e)

}
```

Código Fuente G.6: Código R para estimación de modelos propuestos

```
newton_raphson_modificado <- function(b_inicial, X, y, mu, modelo, e) {

converged <- 0
iter <- 0
max_iter <- 100
epsilon <- 0.000000001

b_next <- b_inicial

while(iter < max_iter && converged==0){
    iter <- iter+1
    b <- b_next

    if(modelo==1){
        grad_L <- gradiente_L_01(b, X, y, mu)
        H_L <- hessiano_L_01(b, X, mu)
    }else if(modelo==2){
        grad_L <- gradiente_L_02(b, X, y, mu)
        H_L <- hessiano_L_02(b, X, mu)
    }else if(modelo==3){
        grad_L <- gradiente_L_03(b, X, y, mu, e)
        H_L <- hessiano_L_03(b, X, mu, e)
    }

    b_next <- b - solve(H_L) %*% grad_L

    if( sum( abs(b_next - b) ) < epsilon ) converged <- 1

}

newton_raphson_modificado <- b_next

}
```

Código Fuente G.7: Código R con método Newton-Raphson modificado para modelos propuestos

```
gradiente_L_01 <- function(b, X, y, mu) {

beta_A <- b[2] #coef asociado a p_A
beta_B <- b[3] #coef asociado a p_B
p_A <- X[,2] #precio firma A
p_B <- X[,3] #precio firma B

sigma <- get_prob(b, X)

grad <- t(X) %*% (y - sigma) + 2*mu* ( t(X) %*% ( (  (beta_A^2 * p_A^2 +
    beta_B^2 * p_B^2)*sigma*(1-sigma)*(1-2*sigma) - (1-2*sigma) + beta_A*sigma
    *(2-3*sigma)*p_A - beta_B*(1-sigma)*(1-3*sigma)*p_B  )*sigma*(1-sigma) ) )

#ajustes en componentes asociados a p_A y p_B
grad[2] <- grad[2] + 2*mu*sum( ( beta_A*p_A*(1-sigma) + 1 )*p_A*sigma^2*(1-
    sigma) )
grad[3] <- grad[3] + 2*mu*sum( ( beta_B*p_B*sigma - 1 )*p_B*sigma*(1-sigma)^2
    )

gradiente_L_01 <- grad

}
```

Código Fuente G.8: Código R para cálculo de $\nabla_b \tilde{\ell}(b/x)$ para primer modelo propuesto

```
hessiano_L_01 <- function(b, X, mu) {

beta_A <- b[2]
beta_B <- b[3]
p_A <- X[,2]
p_B <- X[,3]

I <- dim(X)[1]
n <- length(b) #número de atributos (incluye cte)
I_max <- 1000
s <- ceiling(I/I_max) #número de particiones de w

I_j<-I_max #tamaño de la particion j, para j<s
I_acum<-0 #para identificar en que I comienza y termina cada partición

sigma <- get_prob(b, X)
hL_sum <- matrix(0, nrow=n, ncol=n) #suma acumulada de las matrices hessianas
    que aporta cada partición

V1 <- (beta_A^2 * p_A^2 + beta_B^2 * p_B^2)*(1 - 6*sigma + 6*sigma^2) + beta_A
    *(2-6*sigma)*p_A + beta_B*(4-6*sigma)*p_B + 2
V2 <- (beta_A^2 * p_A^2 + beta_B^2 * p_B^2)*(1-2*sigma)*sigma*(1-sigma) -
    (1-2*sigma) + beta_A*sigma*(2-3*sigma)*p_A - beta_B*(1-sigma)*(1-3*sigma)*
    p_B
```

```
21
22 for(j in 1:s){ #iteramos sobre las particiones de w
23     if(j==s){
24         I_j<- I - I_max*(s-1) #la ultima particion tiene los elementos que
               sobran
25     }
26
27     W_j<- matrix(0, nrow=I_j, ncol=I_j)
28     Z_j<- matrix(0, nrow=I_j, ncol=I_j)
29     Q_j<- matrix(0, nrow=I_j, ncol=I_j)
30
31     diag(W_j)<- sigma[(I_acum+1):(I_acum+I_j)]*(1-sigma[(I_acum+1):(I_acum+I_j
           )])
32     diag(Z_j)<- 2*mu * V1[(I_acum+1):(I_acum+I_j)] * sigma[(I_acum+1):(I_acum+
           I_j)]^2 * (1-sigma[(I_acum+1):(I_acum+I_j)])^2
33     diag(Q_j)<- 2*mu * V2[(I_acum+1):(I_acum+I_j)] * sigma[(I_acum+1):(I_acum+
           I_j)] * (1-sigma[(I_acum+1):(I_acum+I_j)]) * (1-2*sigma[(I_acum+1):(
           I_acum+I_j)])
34
35     hL_sum <- hL_sum - t(X[(I_acum+1):(I_acum+I_j),]) %*% W_j %*% X[(I_acum+1)
           :(I_acum+I_j),] + t(X[(I_acum+1):(I_acum+I_j),]) %*% Z_j %*% X[(I_acum
           +1):(I_acum+I_j),] + t(X[(I_acum+1):(I_acum+I_j),]) %*% Q_j %*% X[(
           I_acum+1):(I_acum+I_j),]
36
37     rm(W_j) #se elimina partición
38     rm(Z_j) #se elimina partición
39     rm(Q_j) #se elimina partición
40
41     I_acum<-I_acum+I_j #coordenada de inicio de sgte particion
42 }
43
44 #ajustar dervidas parciales con beta_A y beta_B
45 #hL_sum ya corresponde al hessiano sin considerar derivadas de beta
46 for(u in 1:n){
47     #ajuste por beta_A
48     hL_sum[2,u] <- hL_sum[2,u] + t(X[,u]) %*% ( 2*mu* ( 2*beta_A*p_A^2*sigma
           *(1-sigma)*(1-2*sigma) + sigma*(2-3*sigma)*p_A )*sigma *(1-sigma) )
49     if(u!=3) hL_sum[u,2] <- hL_sum[2,u]
50
51     #ajuste por beta_B
52     hL_sum[3,u] <- hL_sum[3,u] + t(X[,u]) %*% ( 2*mu* ( 2*beta_B*p_B^2*sigma
           *(1-sigma)*(1-2*sigma) - (1-sigma)*(1-3*sigma)*p_B )*sigma *(1-sigma) )
53     if(u!=2) hL_sum[u,3] <- hL_sum[3,u]
54 }
55
56 #ajuste por derivada beta_A vs beta_B
57 hL_sum[2,3] <- hL_sum[2,3] + t(X[,2]) %*% ( 2*mu*( 2*beta_B*p_B*sigma*(1-sigma
       )*(1-2*sigma) - (1-sigma)*(1 -3*sigma) )*sigma*(1-sigma)*p_B )
58 hL_sum[3,2] <- hL_sum[3,2] + t(X[,3]) %*% ( 2*mu*( 2*beta_A*p_A*sigma*(1-sigma
       )*(1-2*sigma) + sigma*(2 -3*sigma) )*sigma*(1-sigma)*p_A )
59
```

```
#ajuste final para derivadas beta_A-beta_A y beta_B-beta_B
hL_sum[2,2] <- hL_sum[2,2] + t(X[,2]) %*% ( 2*mu*( 2*beta_A*p_A^2*sigma*(1-
    sigma)*(1-2*sigma) + p_A*sigma*(1-sigma) +sigma*(2 -3*sigma)*p_A )*sigma
    *(1-sigma) )
hL_sum[3,3] <- hL_sum[3,3] + t(X[,3]) %*% ( 2*mu*( 2*beta_B*p_B^2*sigma*(1-
    sigma)*(1-2*sigma) + p_B*sigma*(1-sigma) -(1-sigma)*(1 -3*sigma)*p_B )*
    sigma*(1-sigma) )

#hL_sum ya contiene el hessiano de los coeficientes logit (b's)
hessiano_L_01 <- hL_sum

}
```

Código Fuente G.9: Código R para cálculo de $\mathbf{H}_{\tilde{\ell}}(b/x)$ para primer modelo propuesto

```
gradiente_L_02 <- function(b, X, y, mu) {

beta_A <- b[2] #coef asociado a p_A
beta_B <- b[3] #coef asociado a p_B
theta_A <- b[4] #coef asociado a d_A
theta_B <- b[5] #coef asociado a d_B
p_A <- X[,2] #precio firma A
p_B <- X[,3] #precio firma B
d_A <- X[,4] #descuento firma A
d_B <- X[,5] #descuento firma B

sigma <- get_prob(b, X)

grad <- t(X) %*% (y - sigma) + 2*mu* ( t(X) %*% ( (  ((beta_A^2 + theta_A^2) *
     (p_A - d_A)^2 + (beta_B^2 + theta_B^2) * (p_B - d_B)^2)*sigma*(1-sigma)
    *(1-2*sigma) - 2*(1-2*sigma) + (beta_A - theta_A)*sigma*(2-3*sigma)*(p_A -
    d_A) - (beta_B - theta_B)*(1-sigma)*(1-3*sigma)*(p_B - d_B) )*sigma*(1-
    sigma) ) )

#ajustes en componentes asociados a p_A, p_B, theta_A y theta_B
grad[2] <- grad[2] + 2*mu*sum( ( beta_A*(p_A - d_A)*(1-sigma) + 1 )*(p_A - d_A
    )*sigma^2*(1-sigma) )
grad[3] <- grad[3] + 2*mu*sum( ( beta_B*(p_B - d_B)*sigma - 1 )*(p_B - d_B)*
    sigma*(1-sigma)^2 )
grad[4] <- grad[4] + 2*mu*sum( ( theta_A*(p_A - d_A)*(1-sigma) - 1 )*(p_A -
    d_A)*sigma^2*(1-sigma) )
grad[5] <- grad[5] + 2*mu*sum( ( theta_B*(p_B - d_B)*sigma + 1 )*(p_B - d_B)*
    sigma*(1-sigma)^2 )

gradiente_L_02 <- grad

}
```

Código Fuente G.10: Código R para cálculo de $\nabla_b \tilde{\ell}(b/x)$ para segundo modelo propuesto

```
hessiano_L_02 <- function(b, X, mu) {

beta_A <- b[2]
beta_B <- b[3]
theta_A <- b[4]
theta_B <- b[5]
p_A <- X[,2]
p_B <- X[,3]
d_A <- X[,4]
d_B <- X[,5]

I <- dim(X)[1]
n <- length(b)
I_max <- 1000
s <- ceiling(I/I_max) #número de particiones de w

I_j <- I_max #tamaño de la particion j, para j<s
I_acum <- 0 #para identificar en que I comienza y termina cada partición

sigma <- get_prob(b, X)
hL_sum <- matrix(0, nrow=n, ncol=n) #suma acumulada de las matrices hessianas
    que aporta cada partición

V1 <- ((beta_A^2 + theta_A^2) * (p_A - d_A)^2 + (beta_B^2 + theta_B^2) * (p_B
    - d_B)^2)*(1 - 6*sigma + 6*sigma^2) + 2*(beta_A - theta_A)*(1-3*sigma)*(p_A
     - d_A) + 2*(beta_B - theta_B)*(2-3*sigma)*(p_B - d_B) + 4
V2 <- ((beta_A^2 + theta_A^2) * (p_A - d_A)^2 + (beta_B^2 + theta_B^2) * (p_B
    - d_B)^2)*(1-2*sigma)*sigma*(1-sigma) - 2*(1-2*sigma) + (beta_A - theta_A)*
    sigma*(2-3*sigma)*(p_A - d_A) - (beta_B - theta_B)*(1-sigma)*(1-3*sigma)*(
    p_B - d_B)

for(j in 1:s){ #iteramos sobre las particiones de w
    if(j==s){
        I_j<- I - I_max*(s-1) #la ultima particion tiene los elementos que
            sobran
    }

    W_j<- matrix(0, nrow=I_j, ncol=I_j)
    Z_j<- matrix(0, nrow=I_j, ncol=I_j)
    Q_j<- matrix(0, nrow=I_j, ncol=I_j)

    diag(W_j)<- sigma[(I_acum+1):(I_acum+I_j)]*(1-sigma[(I_acum+1):(I_acum+I_j
        )])
    diag(Z_j)<- 2*mu * V1[(I_acum+1):(I_acum+I_j)] * sigma[(I_acum+1):(I_acum+
        I_j)]^2 * (1-sigma[(I_acum+1):(I_acum+I_j)])^2
    diag(Q_j)<- 2*mu * V2[(I_acum+1):(I_acum+I_j)] * sigma[(I_acum+1):(I_acum+
        I_j)] * (1-sigma[(I_acum+1):(I_acum+I_j)]) * (1-2*sigma[(I_acum+1):(
        I_acum+I_j)])

    hL_sum <- hL_sum - t(X[(I_acum+1):(I_acum+I_j),]) %*% W_j %*% X[(I_acum+1)
```

```
        :(I_acum+I_j),] + t(X[(I_acum+1):(I_acum+I_j),]) %*% Z_j %*% X[(I_acum
        +1):(I_acum+I_j),] + t(X[(I_acum+1):(I_acum+I_j),]) %*% Q_j %*% X[(
        I_acum+1):(I_acum+I_j),]

    rm(W_j) #se elimina partición
    rm(Z_j) #se elimina partición
    rm(Q_j) #se elimina partición

    I_acum<-I_acum+I_j #coordenada de inicio de sgte particion
}

#ajustar dervidas parciales con beta_A, beta_B, theta_A y theta_B
#hL_sum ya corresponde al hessiano sin considerar derivadas de beta y theta
for(u in 1:n){
    #ajuste por beta_A
    hL_sum[2,u] <- hL_sum[2,u] + t(X[,u]) %*% ( 2*mu* ( 2*beta_A*(1-sigma)
        *(1-2*sigma)*(p_A - d_A) + (2-3*sigma) )*sigma^2*(1-sigma)*(p_A - d_A)
        )
    if(u!=3 && u!=4 && u!=5) hL_sum[u,2] <- hL_sum[2,u]

    #ajuste por beta_B
    hL_sum[3,u] <- hL_sum[3,u] + t(X[,u]) %*% ( 2*mu* ( 2*beta_B*sigma*(1-2*
        sigma)*(p_B - d_B) - (1-3*sigma) )*sigma*(1-sigma)^2*(p_B - d_B) )
    if(u!=2 && u!=4 && u!=5) hL_sum[u,3] <- hL_sum[3,u]

    #ajuste por theta_A
    hL_sum[4,u] <- hL_sum[4,u] + t(X[,u]) %*% ( 2*mu* ( 2*theta_A*(1-sigma)
        *(1-2*sigma)*(p_A - d_A) - (2-3*sigma) )*sigma^2*(1-sigma)*(p_A - d_A)
        )
    if(u!=2 && u!=3 && u!=5) hL_sum[u,4] <- hL_sum[4,u]

    #ajuste por theta_B
    hL_sum[5,u] <- hL_sum[5,u] + t(X[,u]) %*% ( 2*mu* ( 2*theta_B*sigma*(1-2*
        sigma)*(p_B - d_B) + (1-3*sigma) )*sigma*(1-sigma)^2*(p_B - d_B) )
    if(u!=2 && u!=3 && u!=4) hL_sum[u,5] <- hL_sum[5,u]

}

#ajuste por derivadas cruzadas beta_A vs beta_B
hL_sum[2,3] <- hL_sum[2,3] + t(X[,2]) %*% ( 2*mu*( 2*beta_B*sigma^2*(1-sigma)
    ^2*(1-2*sigma)*(p_B - d_B)^2 - sigma*(1-sigma)^2*(1 -3*sigma)*(p_B - d_B) )
    )
hL_sum[3,2] <- hL_sum[2,3]

#ajuste por derivadas cruzadas beta_A vs theta_A
hL_sum[2,4] <- hL_sum[2,4] + t(X[,2]) %*% ( 2*mu*( 2*theta_A*sigma^2*(1-sigma)
    ^2*(1-2*sigma)*(p_A - d_A)^2 - sigma^2*(1-sigma)*(2 -3*sigma)*(p_A - d_A) )
    )
hL_sum[4,2] <- hL_sum[2,4]

#ajuste por derivadas cruzadas beta_A vs theta_B
```

```
hL_sum[2,5] <- hL_sum[2,5] + t(X[,2]) %*% ( 2*mu*( 2*theta_B*sigma^2*(1-sigma)^2*(1-2*sigma)*(p_B - d_B)^2 + sigma*(1-sigma)^2*(1 -3*sigma)*(p_B - d_B) ) )
hL_sum[5,2] <- hL_sum[2,5]

#ajuste por derivadas cruzadas beta_B vs theta_A
hL_sum[3,4] <- hL_sum[3,4] + t(X[,3]) %*% ( 2*mu*( 2*theta_A*sigma^2*(1-sigma)^2*(1-2*sigma)*(p_A - d_A)^2 - sigma^2*(1-sigma)*(2 -3*sigma)*(p_A - d_A) ) )
hL_sum[4,3] <- hL_sum[3,4]

#ajuste por derivadas cruzadas beta_B vs theta_B
hL_sum[3,5] <- hL_sum[3,5] + t(X[,3]) %*% ( 2*mu*( 2*theta_B*sigma^2*(1-sigma)^2*(1-2*sigma)*(p_B - d_B)^2 + sigma*(1-sigma)^2*(1 -3*sigma)*(p_B - d_B) ) )
hL_sum[5,3] <- hL_sum[3,5]

#ajuste por derivadas cruzadas theta_A vs theta_B
hL_sum[4,5] <- hL_sum[4,5] + t(X[,4]) %*% ( 2*mu*( 2*theta_B*sigma^2*(1-sigma)^2*(1-2*sigma)*(p_B - d_B)^2 + sigma*(1-sigma)^2*(1 -3*sigma)*(p_B - d_B) ) )
hL_sum[5,4] <- hL_sum[4,5]

#ajuste final para derivadas beta-beta y theta-theta
hL_sum[2,2] <- hL_sum[2,2] + t(X[,2] - X[,4]) %*% ( 2*mu*( (p_A - d_A)*(1-sigma) + 2*beta_A*(p_A - d_A)*(1-sigma)*(1-2*sigma)*p_A + (2 -3*sigma)*p_A )*sigma^2*(1-sigma) )
hL_sum[3,3] <- hL_sum[3,3] + t(X[,3] - X[,5]) %*% ( 2*mu*( (p_B - d_B)*sigma + 2*beta_B*(p_B - d_B)*sigma*(1-2*sigma)*p_B - (1 -3*sigma)*p_B )*sigma*(1-sigma)^2 )
hL_sum[4,4] <- hL_sum[4,4] + t(X[,2] - X[,4]) %*% ( 2*mu*( (p_A - d_A)*(1-sigma) + 2*theta_A*(p_A - d_A)*(1-sigma)*(1-2*sigma)*d_A - (2 -3*sigma)*d_A )*sigma^2*(1-sigma) )
hL_sum[5,5] <- hL_sum[5,5] + t(X[,3] - X[,5]) %*% ( 2*mu*( (p_B - d_B)*sigma + 2*theta_B*(p_B - d_B)*sigma*(1-2*sigma)*d_B + (1 -3*sigma)*d_B )*sigma*(1-sigma)^2 )

#hL_sum ya contiene el hessiano de los coeficientes logit (b's)
hessiano_L_02 <- hL_sum

}
```

Código Fuente G.11: Código R para cálculo de $\mathbf{H}_{\tilde{\ell}}(b/x)$ para segundo modelo propuesto

```
gradiente_L_03 <- function(b, X, y, mu, e) {

beta_A <- b[2]
beta_B <- b[3]
theta_A <- b[4]
theta_B <- b[5]
```

```
phi_A <- b[6]
phi_B <- b[7]
p_A <- X[,2]
p_B <- X[,3]
d_A <- X[,4]
d_B <- X[,5]
w_A <- X[,6]
w_B <- X[,7]

sigma <- get_prob(b, X)

grad <- t(X) %*% (y - sigma) + 2*mu* ( t(X) %*% ( (  ((beta_A^2 + theta_A^2 +
    phi_A^2) * (p_A - d_A)^2 + (beta_B^2 + theta_B^2 + phi_B^2) * (p_B - d_B)
    ^2)*sigma*(1-sigma)*(1-2*sigma) - 2*(1-2*sigma) + (beta_A - theta_A)*sigma
    *(2-3*sigma)*(p_A - d_A) - (beta_B - theta_B)*(1-sigma)*(1-3*sigma)*(p_B -
    d_B) - phi_A*e*(1-2*sigma)*(p_A - d_A) + phi_B*e*(1-2*sigma)*(p_B - d_B) )*
    sigma*(1-sigma) ) )

#ajustes en componentes asociados a p_A, p_B, theta_A, theta_B, phi_A y phi_B
grad[2] <- grad[2] + 2*mu*sum( ( beta_A*(p_A - d_A)*(1-sigma) + 1 )*(p_A - d_A
    )*sigma^2*(1-sigma) )
grad[3] <- grad[3] + 2*mu*sum( ( beta_B*(p_B - d_B)*sigma - 1 )*(p_B - d_B)*
    sigma*(1-sigma)^2 )
grad[4] <- grad[4] + 2*mu*sum( ( theta_A*(p_A - d_A)*(1-sigma) - 1 )*(p_A -
    d_A)*sigma^2*(1-sigma) )
grad[5] <- grad[5] + 2*mu*sum( ( theta_B*(p_B - d_B)*sigma + 1 )*(p_B - d_B)*
    sigma*(1-sigma)^2 )
grad[6] <- grad[6] + 2*mu*sum( ( phi_A*(p_A - d_A)*sigma*(1-sigma) - e )*(p_A
    - d_A)*sigma*(1-sigma) )
grad[7] <- grad[7] + 2*mu*sum( ( phi_B*(p_B - d_B)*sigma*(1-sigma) + e )*(p_B
    - d_B)*sigma*(1-sigma) )

gradiente_L_03 <- grad

}
```

Código Fuente G.12: Código R para cálculo de $\nabla_b \tilde{\ell}(b/x)$ para tercer modelo propuesto

```
hessiano_L_03 <- function(b, X, mu, e) {

beta_A <- b[2]
beta_B <- b[3]
theta_A <- b[4]
theta_B <- b[5]
phi_A <- b[6]
phi_B <- b[7]
p_A <- X[,2]
p_B <- X[,3]
d_A <- X[,4]
d_B <- X[,5]
```

```
w_A <- X[,6]
w_B <- X[,7]

I<-dim(X)[1]
n<-length(b)
I_max<-1000
s<-ceiling(I/I_max) #número de particiones de w

I_j<-I_max #tamaño de la particion j, para j<s
I_acum<-0 #para identificar en que I comienza y termina cada partición

sigma <- get_prob(b, X)
hL_sum <- matrix(0, nrow=n, ncol=n) #suma acumulada de las matrices hessianas
    que aporta cada partición

V1 <- ((beta_A^2 + theta_A^2 + phi_A^2) * (p_A - d_A)^2 + (beta_B^2 + theta_B
    ^2 + phi_B^2) * (p_B - d_B)^2)*(1 - 6*sigma + 6*sigma^2) + 2*(beta_A -
    theta_A)*(1-3*sigma)*(p_A - d_A) + 2*(beta_B - theta_B)*(2-3*sigma)*(p_B -
    d_B) + 2*phi_A*e*(p_A - d_A) - 2*phi_B*e*(p_B - d_B) + 4
V2 <- ((beta_A^2 + theta_A^2 + phi_A^2) * (p_A - d_A)^2 + (beta_B^2 + theta_B
    ^2 + phi_B^2) * (p_B - d_B)^2)*(1-2*sigma)*sigma*(1-sigma) - 2*(1-2*sigma)
    + (beta_A - theta_A)*sigma*(2-3*sigma)*(p_A - d_A) - (beta_B - theta_B)*(1-
    sigma)*(1-3*sigma)*(p_B - d_B) - phi_A*e*(1-2*sigma)*(p_A - d_A) + phi_B*e
    *(1-2*sigma)*(p_B - d_B)

for(j in 1:s){ #iteramos sobre las particiones de w
    if(j==s){
        I_j<- I - I_max*(s-1) #la ultima particion tiene los elementos que
            sobran
    }

    W_j<- matrix(0, nrow=I_j, ncol=I_j)
    Z_j<- matrix(0, nrow=I_j, ncol=I_j)
    Q_j<- matrix(0, nrow=I_j, ncol=I_j)

    diag(W_j)<- sigma[(I_acum+1):(I_acum+I_j)]*(1-sigma[(I_acum+1):(I_acum+I_j
        )])
    diag(Z_j)<- 2*mu * V1[(I_acum+1):(I_acum+I_j)] * sigma[(I_acum+1):(I_acum+
        I_j)]^2 * (1-sigma[(I_acum+1):(I_acum+I_j)])^2
    diag(Q_j)<- 2*mu * V2[(I_acum+1):(I_acum+I_j)] * sigma[(I_acum+1):(I_acum+
        I_j)] * (1-sigma[(I_acum+1):(I_acum+I_j)]) * (1-2*sigma[(I_acum+1):(
        I_acum+I_j)])

    hL_sum <- hL_sum - t(X[(I_acum+1):(I_acum+I_j),]) %*% W_j %*% X[(I_acum+1)
        :(I_acum+I_j),] + t(X[(I_acum+1):(I_acum+I_j),]) %*% Z_j %*% X[(I_acum
        +1):(I_acum+I_j),] + t(X[(I_acum+1):(I_acum+I_j),]) %*% Q_j %*% X[(
        I_acum+1):(I_acum+I_j),]

    rm(W_j) #se elimina partición
    rm(Z_j) #se elimina partición
    rm(Q_j) #se elimina partición
```

```
48
49      I_acum<-I_acum+I_j
50  }
51
52  #ajustar derivadas parciales con beta_A, beta_B, theta_A, theta_B, phi_A y
        phi_B
53  #hL_sum ya corresponde al hessiano sin considerar derivadas de beta, theta y
        phi
54  for(u in 1:n){
55      #ajuste por beta_A
56      hL_sum[2,u] <- hL_sum[2,u] + t(X[,u]) %*% ( 2*mu* ( 2*beta_A*(1-sigma)
            *(1-2*sigma)*(p_A - d_A) + (2-3*sigma) )*sigma^2*(1-sigma)*(p_A - d_A)
            )
57      if(u!=3 && u!=4 && u!=5 && u!=6 && u!=7) hL_sum[u,2] <- hL_sum[2,u]
58
59      #ajuste por beta_B
60      hL_sum[3,u] <- hL_sum[3,u] + t(X[,u]) %*% ( 2*mu* ( 2*beta_B*sigma*(1-2*
            sigma)*(p_B - d_B) - (1-3*sigma) )*sigma*(1-sigma)^2*(p_B - d_B) )
61      if(u!=2 && u!=4 && u!=5 && u!=6 && u!=7) hL_sum[u,3] <- hL_sum[3,u]
62
63      #ajuste por theta_A
64      hL_sum[4,u] <- hL_sum[4,u] + t(X[,u]) %*% ( 2*mu* ( 2*theta_A*(1-sigma)
            *(1-2*sigma)*(p_A - d_A) - (2-3*sigma) )*sigma^2*(1-sigma)*(p_A - d_A)
            )
65      if(u!=2 && u!=3 && u!=5 && u!=6 && u!=7) hL_sum[u,4] <- hL_sum[4,u]
66
67      #ajuste por theta_B
68      hL_sum[5,u] <- hL_sum[5,u] + t(X[,u]) %*% ( 2*mu* ( 2*theta_B*sigma*(1-2*
            sigma)*(p_B - d_B) + (1-3*sigma) )*sigma*(1-sigma)^2*(p_B - d_B) )
69      if(u!=2 && u!=3 && u!=4 && u!=6 && u!=7) hL_sum[u,5] <- hL_sum[5,u]
70
71      #ajuste por phi_A
72      hL_sum[6,u] <- hL_sum[6,u] + t(X[,u]) %*% ( 2*mu* ( 2*phi_A*sigma*(1-sigma
            )*(p_A - d_A) - e )*sigma*(1-sigma)*(1-2*sigma)*(p_A - d_A) )
73      if(u!=2 && u!=3 && u!=4 && u!=5 && u!=7) hL_sum[u,6] <- hL_sum[6,u]
74
75      #ajuste por phi_B
76      hL_sum[7,u] <- hL_sum[7,u] + t(X[,u]) %*% ( 2*mu* ( 2*phi_B*sigma*(1-sigma
            )*(p_B - d_B) + e )*sigma*(1-sigma)*(1-2*sigma)*(p_B - d_B) )
77      if(u!=2 && u!=3 && u!=4 && u!=5 && u!=6) hL_sum[u,7] <- hL_sum[7,u]
78  }
79
80  #ajuste por derivadas cruzadas beta_A vs beta_B
81  hL_sum[2,3] <- hL_sum[2,3] + t(X[,2]) %*% ( 2*mu*( 2*beta_B*sigma^2*(1-sigma)
        ^2*(1-2*sigma)*(p_B - d_B)^2 - sigma*(1-sigma)^2*(1 -3*sigma)*(p_B - d_B) )
         )
82  hL_sum[3,2] <- hL_sum[2,3]
83
84  #ajuste por derivadas cruzadas beta_A vs theta_A
85  hL_sum[2,4] <- hL_sum[2,4] + t(X[,2]) %*% ( 2*mu*( 2*theta_A*sigma^2*(1-sigma)
        ^2*(1-2*sigma)*(p_A - d_A)^2 - sigma^2*(1-sigma)*(2 -3*sigma)*(p_A - d_A) )
```

```
    )
hL_sum[4,2] <- hL_sum[2,4]

#ajuste por derivadas cruzadas beta_A vs theta_B
hL_sum[2,5] <- hL_sum[2,5] + t(X[,2]) %*% ( 2*mu*( 2*theta_B*sigma^2*(1-sigma)
    ^2*(1-2*sigma)*(p_B - d_B)^2 + sigma*(1-sigma)^2*(1 -3*sigma)*(p_B - d_B) )
    )
hL_sum[5,2] <- hL_sum[2,5]

#ajuste por derivadas cruzadas beta_A vs phi_A
hL_sum[2,6] <- hL_sum[2,6] + t(X[,2]) %*% ( 2*mu*( 2*phi_A*sigma^2*(1-sigma)
    ^2*(1-2*sigma)*(p_A - d_A)^2 - e*sigma*(1-sigma)*(1 -2*sigma)*(p_A - d_A) )
    )
hL_sum[6,2] <- hL_sum[2,6]

#ajuste por derivadas cruzadas beta_A vs phi_B
hL_sum[2,7] <- hL_sum[2,7] + t(X[,2]) %*% ( 2*mu*( 2*phi_B*sigma^2*(1-sigma)
    ^2*(1-2*sigma)*(p_B - d_B)^2 + e*sigma*(1-sigma)*(1 -2*sigma)*(p_B - d_B) )
    )
hL_sum[7,2] <- hL_sum[2,7]

#ajuste por derivadas cruzadas beta_B vs theta_A
hL_sum[3,4] <- hL_sum[3,4] + t(X[,3]) %*% ( 2*mu*( 2*theta_A*sigma^2*(1-sigma)
    ^2*(1-2*sigma)*(p_A - d_A)^2 - sigma^2*(1-sigma)*(2 -3*sigma)*(p_A - d_A) )
    )
hL_sum[4,3] <- hL_sum[3,4]

#ajuste por derivadas cruzadas beta_B vs theta_B
hL_sum[3,5] <- hL_sum[3,5] + t(X[,3]) %*% ( 2*mu*( 2*theta_B*sigma^2*(1-sigma)
    ^2*(1-2*sigma)*(p_B - d_B)^2 + sigma*(1-sigma)^2*(1 -3*sigma)*(p_B - d_B) )
    )
hL_sum[5,3] <- hL_sum[3,5]

#ajuste por derivadas cruzadas beta_B vs phi_A
hL_sum[3,6] <- hL_sum[3,6] + t(X[,3]) %*% ( 2*mu*( 2*phi_A*sigma^2*(1-sigma)
    ^2*(1-2*sigma)*(p_A - d_A)^2 - e*sigma*(1-sigma)*(1 -2*sigma)*(p_A - d_A) )
    )
hL_sum[6,3] <- hL_sum[3,6]

#ajuste por derivadas cruzadas beta_B vs phi_B
hL_sum[3,7] <- hL_sum[3,7] + t(X[,3]) %*% ( 2*mu*( 2*phi_B*sigma^2*(1-sigma)
    ^2*(1-2*sigma)*(p_B - d_B)^2 + e*sigma*(1-sigma)*(1 -2*sigma)*(p_B - d_B) )
    )
hL_sum[7,3] <- hL_sum[3,7]

#ajuste por derivadas cruzadas theta_A vs theta_B
hL_sum[4,5] <- hL_sum[4,5] + t(X[,4]) %*% ( 2*mu*( 2*theta_B*sigma^2*(1-sigma)
    ^2*(1-2*sigma)*(p_B - d_B)^2 + sigma*(1-sigma)^2*(1 -3*sigma)*(p_B - d_B) )
    )
hL_sum[5,4] <- hL_sum[4,5]

```

```
#ajuste por derivadas cruzadas theta_A vs phi_A
hL_sum[4,6] <- hL_sum[4,6] + t(X[,4]) %*% ( 2*mu*( 2*phi_A*sigma^2*(1-sigma)^2*(1-2*sigma)*(p_A - d_A)^2 - e*sigma*(1-sigma)*(1 -2*sigma)*(p_A - d_A) ) )
hL_sum[6,4] <- hL_sum[4,6]

#ajuste por derivadas cruzadas theta_A vs phi_B
hL_sum[4,7] <- hL_sum[4,7] + t(X[,4]) %*% ( 2*mu*( 2*phi_B*sigma^2*(1-sigma)^2*(1-2*sigma)*(p_B - d_B)^2 + e*sigma*(1-sigma)*(1 -2*sigma)*(p_B - d_B) ) )
hL_sum[7,4] <- hL_sum[4,7]

#ajuste por derivadas cruzadas theta_B vs phi_A
hL_sum[5,6] <- hL_sum[5,6] + t(X[,5]) %*% ( 2*mu*( 2*phi_A*sigma^2*(1-sigma)^2*(1-2*sigma)*(p_A - d_A)^2 - e*sigma*(1-sigma)*(1 -2*sigma)*(p_A - d_A) ) )
hL_sum[6,5] <- hL_sum[5,6]

#ajuste por derivadas cruzadas theta_B vs phi_B
hL_sum[5,7] <- hL_sum[5,7] + t(X[,5]) %*% ( 2*mu*( 2*phi_B*sigma^2*(1-sigma)^2*(1-2*sigma)*(p_B - d_B)^2 + e*sigma*(1-sigma)*(1 -2*sigma)*(p_B - d_B) ) )
hL_sum[7,5] <- hL_sum[5,7]

#ajuste por derivadas cruzadas phi_A vs phi_B
hL_sum[6,7] <- hL_sum[6,7] + t(X[,6]) %*% ( 2*mu*( 2*phi_B*sigma^2*(1-sigma)^2*(1-2*sigma)*(p_B - d_B)^2 + e*sigma*(1-sigma)*(1 -2*sigma)*(p_B - d_B) ) )
hL_sum[7,6] <- hL_sum[6,7]

#ajuste final para derivadas beta-beta, theta-theta y phi-phi
hL_sum[2,2] <- hL_sum[2,2] + t(X[,2] - X[,4]) %*% ( 2*mu*( (p_A - d_A)*(1-sigma) + 2*beta_A*(p_A - d_A)*(1-sigma)*(1-2*sigma)*p_A + (2 -3*sigma)*p_A )*sigma^2*(1-sigma) )
hL_sum[3,3] <- hL_sum[3,3] + t(X[,3] - X[,5]) %*% ( 2*mu*( (p_B - d_B)*sigma + 2*beta_B*(p_B - d_B)*sigma*(1-2*sigma)*p_B - (1 -3*sigma)*p_B )*sigma*(1-sigma)^2 )
hL_sum[4,4] <- hL_sum[4,4] + t(X[,2] - X[,4]) %*% ( 2*mu*( (p_A - d_A)*(1-sigma) + 2*theta_A*(p_A - d_A)*(1-sigma)*(1-2*sigma)*d_A - (2 -3*sigma)*d_A )*sigma^2*(1-sigma) )
hL_sum[5,5] <- hL_sum[5,5] + t(X[,3] - X[,5]) %*% ( 2*mu*( (p_B - d_B)*sigma + 2*theta_B*(p_B - d_B)*sigma*(1-2*sigma)*d_B + (1 -3*sigma)*d_B )*sigma*(1-sigma)^2 )
hL_sum[6,6] <- hL_sum[6,6] + t(X[,2] - X[,4]) %*% ( 2*mu*( (p_A - d_A)*sigma*(1-sigma) + 2*phi_A*(p_A - d_A)*sigma*(1-sigma)*(1-2*sigma)*w_A - e*(1 -2*sigma)*w_A )*sigma*(1-sigma) )
hL_sum[7,7] <- hL_sum[7,7] + t(X[,3] - X[,5]) %*% ( 2*mu*( (p_B - d_B)*sigma*(1-sigma) + 2*phi_B*(p_B - d_B)*sigma*(1-sigma)*(1-2*sigma)*w_B + e*(1 -2*sigma)*w_B )*sigma*(1-sigma) )

#hL_sum ya contiene el hessiano de los coeficientes logit (b's)
```

```
149 hessiano_L_03 <- hL_sum
150
151 }
```

Código Fuente G.13: Código R para cálculo de $\mathbf{H}_{\tilde{\ell}}(b/x)$ para tercer modelo propuesto

Anexo H

Diccionario de variables

H.1. Variables asociadas a elementos competitivos

Atributo	**Definición**	**Ámbito**
Total_Units_Purchased	Unidades compradas	Precio
Extended_Price	Precio total pagado antes del descuento de cupones	Precio
Total_Value_of_Store_Coupons	Valor de los cupones de la tienda	Precio
Total_Value_of_Manufacturer_Coupons	Valor de los cupones del fabricante	Precio
End_Aisle_Display_Code	N = No; Y = Yes	Display
Front_Aisle_Display_Code	N = No; Y = Yes	Display
In-Aisle_Display_Code	N = No; Y = Yes	Display
Other_Display_Code	N = No; Y = Yes	Display
Advertising_Code	N = No; Y = Yes	Advertising
Ad_Type	L = Line Ad; M = Major Ad; Blank = No Ad	Advertising
In-Ad_Coupon_Code	N = No; Y = Yes	Coupon
Point-of-Purchase_Code	N = No; Y = Yes	Coupon
Special_Price_Code	N = No; Y = Yes	Precio

Tabla H.1: Definición de variables asociadas a elementos competitivos

H.2. Variables demográficas

Atributo	Definición
Cable_Status	0 = No Cable; 1 = Has Cable
Meter_Status	0 = Non-Metered; 1 = Metered
Panelist_Status	0 = Unassigned Panelist 1 = Non-Cancelled 2 = Non-Cancelled 3 = Non-Cancelled 4 = Reinstated 5 = Non-Called 6 = Mystery Household 7 = Cancelled Panelist 8 = Automatic Preliminary Cancellation 9 = Automatic Final Cancellation 99 = Cancelled Mystery Household
Number_of_Cats	Number of Cats
Number_of_Dogs	Number of Dogs
Number_of_TVs	Number of TVs
Residence_Type	1 = Apartment 2 = Condo 3 = Single Family 4 = Multiple Family 5 = Mobile Home 6 = Other
Residence_Status	1 = Owned 2 = Rented 3 = Other
Residence_Duration	1 = 1/2 year 2 = 1 year 3 = 1-2 years 4 = 3-4 years 5 = 5-10 years 6 = 11-15 years 7 = 16-20 years 8 = Over 20 years
Household_Income	1 = Less than $5,000 2 = $5,000 - $9,000 3 = $10,000 - $14,999

Atributo	Definición
	4 = $15,000 - $19,999 5 = $20,000 - $24,999 6 = $25,000 - $29,999 7 = $30,000 - $34,999 8 = $35,000 - $39,999 9 = $40,000 - $44,999 10 = $45,000 - $49,999 11 = $50,000 - $59,999 12 = $60,000 - $74,999 13 = $75,000 - $99,999 14 = $100,000 or More
Number_of_Members_in_Household	Number of Members in Household
Male_Head_Working_Status	1 = Employed Full Time 2 = Employed Part Time 3 = Retired (Employed) 4 = Temporary Unemployed 5 = Disabled, Student, etc. 6 = Full-Time Homemaker 7 = Retired (Not Employed)
Male_Head_Average_Working_Hours	Male Head Average Working Hours
Male_Head_Occupation	1 = Farming, Forestry, Fishing 2 = Operator, Fabricator, Laborer 3 = Precision, Production, Craft, Repairs 4 = Protective Service, Private HH Worker 5 = Service Occupation 6 = Administrative Support, Clerical 7 = Sales 8 = Professional Specialty 9 = Technical 10 = Executive, Managerial
Male_Head_Education	1 = Did Not Attend School 2 = Some Grade School 3 = Graduate Grade School 4 = Some High School 5 = Graduated High School 6 = Some Trade or Business School 7 = Graduated Trade or Business School 8 = Some College 9 = Graduated College 10 = Attended Graduate School 11 = Post-Graduate Degree

Atributo	Definición
Male_Head_SpanishHispanic	0 = Not Spanish 1 = Mexican American 2 = Puerto Rican 3 = Cuban 4 = Other Hispanic
Male_Head_Race	1 = White 2 = Black 3 = Asian 4 = Native American 5 = Other
Female_Head_Working_Status	1 = Employed Full Time 2 = Employed Part Time 3 = Retired (Employed) 4 = Temporary Unemployed 5 = Disabled, Student, etc. 6 = Full-Time Homemaker 7 = Retired (Not Employed)
Female_Head_Average_Working_Hours	Female Head Average Working Hours
Female_Head_Occupation	1 = Farming, Forestry, Fishing 2 = Operator, Fabricator, Laborer 3 = Precision, Production, Craft, Repairs 4 = Protective Service, Private HH Worker 5 = Service Occupation 6 = Administrative Support, Clerical 7 = Sales 8 = Professional Specialty 9 = Technical 10 = Executive, Managerial
Female_Head_Education	1 = Did Not Attend School 2 = Some Grade School 3 = Graduate Grade School 4 = Some High School 5 = Graduated High School 6 = Some Trade or Business School 7 = Graduated Trade or Business School 8 = Some College 9 = Graduated College 10 = Attended Graduate School 11 = Post-Graduate Degree
Female_Head_SpanishHispanic	0 = Not Spanish

Atributo	Definición
	1 = Mexican American 2 = Puerto Rican 3 = Cuban 4 = Other Hispanic
Female_Head_Race	1 = White 2 = Black 3 = Asian 4 = Native American 5 = Other
Washing_Machine	0 = No Response; 1 = No; 2 = Yes
Clothes_Dryer	0 = No Response; 1 = No; 2 = Yes
Dishwasher	0 = No Response; 1 = No; 2 = Yes
Freezer	0 = No Response; 1 = No; 2 = Yes
Toaster	0 = No Response; 1 = No; 2 = Yes
Toaster_Broiler_Oven	0 = No Response; 1 = No; 2 = Yes
Blender	0 = No Response; 1 = No; 2 = Yes
Food_Processor	0 = No Response; 1 = No; 2 = Yes
Microwave	0 = No Response; 1 = No; 2 = Yes
Convection_Oven	0 = No Response; 1 = No; 2 = Yes
Coffee_Maker	0 = No Response; 1 = No; 2 = Yes
Trash_Compactor	0 = No Response; 1 = No; 2 = Yes
Garbage_Disposal	0 = No Response; 1 = No; 2 = Yes
Hair_Dryer	0 = No Response; 1 = No; 2 = Yes
Curling_Iron	0 = No Response; 1 = No; 2 = Yes
Hair_Rollers	0 = No Response; 1 = No; 2 = Yes
Vacuum_Cleaner	0 = No Response; 1 = No; 2 = Yes
Water_Softner	0 = No Response; 1 = No; 2 = Yes
VCR	0 = No; 1 = Yes

Tabla H.2: Definición de variables demográficas

H.3. Variables construidas

Atributo	Definición
Display	1 si se cumple (End_Aisle_Display_Code=Y OR Front_Aisle_Display_Code=Y OR In-Aisle_Display_Code=Y OR Other_Display_Code=Y); 0 en caso contrario.
Advertising	1 si se cumple Advertising_Code=Y; 0 en caso contrario.
Type	1 si se cumple (Ad_Type=M OR Ad_Type=L); 0 en caso contrario.
Coupon	1 si se cumple (In-Ad_Coupon_Code=Y OR Point-of-Purchase_Code=Y); 0 en caso contrario.
Special_Price	1 si se cumple Special_Price_Code=Y; 0 en caso contrario.

Tabla H.3: Definición de variables construidas I

Atributo	Definición
Frecuencia_Baja	1 = Hogar es de frecuencia baja de compra; 0 = No
Frecuencia_Normal	1 = Hogar es de frecuencia normal de compra; 0 = No
Frecuencia_Alta	1 = Hogar es de frecuencia alta de compra; 0 = No
P_HUNTS_Unit_Extended_Price	Precio unitario de Hunt's antes del descuento de cupones
P_HEINZ_Unit_Extended_Price	Precio unitario de Heinz antes del descuento de cupones
Dif_Unit_Extended_Price	Diferencia entre precio de la marca comprada versus otra marca
Dif2_Unit_Extended_Price	Diferencia entre precio Hunt's versus precio Heinz
P_HUNTS_Unit_TV_Coupons	Valor unitario de los cupones de descuento de Hunt's (considera cupones de tienda y fabricante)
P_HEINZ_Unit_TV_Coupons	Valor unitario de los cupones de descuento de Heinz (considera cupones de tienda y fabricante)
HUNTS_Dummy_CE	1 = Hogar ha visto comercial de TV de Hunt's antes de la fecha de compra; 0 = No
HEINZ_Dummy_CE	1 = Hogar ha visto comercial de TV de Heinz antes de la fecha de compra; 0 = No

Tabla H.4: Definición de variables construidas II

Anexo I

Análisis descriptivo de variables

I.1. Variables asociadas a elementos competitivos

I.1.1. Variables continuas

Atributo	Valores	N	Mean	S.D.[1]	95 % C.I.[2]		Min	Max
					Lower	Upper		
Units_Purchased	0	9834	1.05	0.26	1.04	1.05	1	6
	1	4433	1.06	0.33	1.05	1.07	1	8
	Total	14267	1.05	0.29	1.05	1.06	1	8
Extended_Price	0	9834	1.15	0.29	1.14	1.16	0.74	5.94
	1	4433	1.14	0.35	1.13	1.15	0.79	8.72
	Total	14267	1.15	0.31	1.14	1.15	0.74	8.72
TV_Store_Coupons	0	9834	0.07	0.17	0.07	0.08	0.00	0.70
	1	4433	0.02	0.10	0.02	0.02	0.00	3.36
	Total	14267	0.06	0.16	0.05	0.06	0.00	3.36
TV_Manufacturer_Coupons	0	9834	0.07	0.20	0.07	0.07	0.00	2.09
	1	4433	0.09	0.23	0.09	0.10	0.00	5.00
	Total	14267	0.08	0.21	0.07	0.08	0.00	5.00

1. Standard Deviation
2. Confidence Interval

Tabla I.1: Análisis descriptivo

Atributo	Tipo	Sum of Squares	df	Mean Square	F	Sig.
Units_Purchased	Between Groups	0.60	1	0.60	7.35	0.67 %
	Within Groups	1161.92	14265	0.08		
	Total	1162.51	14266			
Extended_Price	Between Groups	0.18	1	0.18	1.79	18.04 %
	Within Groups	1399.21	14265	0.10		
	Total	1399.38	14266			
TV_Store_Coupons	Between Groups	9.07	1	9.07	386.88	0.00 %
	Within Groups	334.50	14265	0.02		
	Total	343.57	14266			
TV_Manufacturer_Coupons	Between Groups	1.73	1	1.73	39.23	0.00 %
	Within Groups	629.29	14265	0.04		
	Total	631.02	14266			

Tabla I.2: Análisis de Varianza (ANOVA)

I.1.2. Variables categóricas

Atributo	Valores	y					
		0		1		Total	
		N	N (%)	N	N (%)	N	N (%)
End_Aisle_DC	N	8931	90.8	3925	88.5	12856	90.1
	Y	903	9.2	508	11.5	1411	9.9
	Total	9834	100.0	4433	100.0	14267	100.0
Front_Aisle_DC	N	9797	99.6	4425	99.8	14222	99.7
	Y	37	0.4	8	0.2	45	0.3
	Total	9834	100.0	4433	100.0	14267	100.0
InAisle_DC	N	9691	98.5	4316	97.4	14007	98.2
	Y	143	1.5	117	2.6	260	1.8
	Total	9834	100.0	4433	100.0	14267	100.0
Other_DC	N	9642	98.0	4278	96.5	13920	97.6
	Y	192	2.0	155	3.5	347	2.4
	Total	9834	100.0	4433	100.0	14267	100.0
Advertising_Code	N	5763	58.6	3215	72.5	8978	62.9
	Y	4071	41.4	1218	27.5	5289	37.1
	Total	9834	100.0	4433	100.0	14267	100.0
Ad_Type		7781	79.1	3514	79.3	11295	79.2
	L	122	1.2	102	2.3	224	1.6
	M	1931	19.6	817	18.4	2748	19.3
	Total	9834	100.0	4433	100.0	14267	100.0
InAd_Coupon	N	7779	79.1	4120	92.9	11899	83.4
	Y	2055	20.9	313	7.1	2368	16.6
	Total	9834	100.0	4433	100.0	14267	100.0
PointOfPurchase	N	9834	100.0	4378	98.8	14212	99.6
	Y	0	0.0	55	1.2	55	0.4
	Total	9834	100.0	4433	100.0	14267	100.0
Special_Price	N	7352	74.8	3138	70.8	10490	73.5
	Y	2482	25.2	1295	29.2	3777	26.5
	Total	9834	100.0	4433	100.0	14267	100.0

Tabla I.3: Análisis descriptivo

Atributo	**Estadístico**	**y**
End_Aisle_DC	Chi-square	17.7781
	df	1
	Sig.	0.00 %
Front_Aisle_DC	Chi-square	3.7250
	df	1
	Sig.	5.36 %
InAisle_DC	Chi-square	23.9880
	df	1
	Sig.	0.00 %
Other_DC	Chi-square	30.7000
	df	1
	Sig.	0.00 %
Advertising_Code	Chi-square	253.8493
	df	1
	Sig.	0.00 %
Ad_Type	Chi-square	24.1963
	df	2
	Sig.	0.00 %
InAd_Coupon	Chi-square	422.5738
	df	1
	Sig.	0.00 %
PointOfPurchase	Chi-square	122.4821
	df	1
	Sig.	0.00 %
Special_Price	Chi-square	24.7879
	df	1
	Sig.	0.00 %

Tabla I.4: Pearson Chi-Square tests

I.2. Variables construidas

I.2.1. Variables continuas

Atributo	Valores	N	Mean	S.D.[1]	95 % C.I.[2]		Min	Max
					Lower	Upper		
P_HUNTS_Unit_Extended_Price	0	9267	1.13	0.16	1.12	1.13	0.79	1.79
	1	4115	1.08	0.18	1.07	1.08	0.79	1.79
	Total	13382	1.11	0.17	1.11	1.11	0.79	1.79
P_HEINZ_Unit_Extended_Price	0	9267	1.10	0.15	1.10	1.10	0.74	1.69
	1	4115	1.16	0.15	1.15	1.16	0.79	1.59
	Total	13382	1.12	0.15	1.11	1.12	0.74	1.69
Dif_Unit_Extended_Price	0	9267	-0.03	0.19	-0.03	-0.02	-0.80	0.73
	1	4115	-0.08	0.21	-0.08	-0.07	-0.70	0.80
	Total	13382	-0.04	0.20	-0.05	-0.04	-0.80	0.80
Dif2_Unit_Extended_Price	0	9267	0.03	0.19	0.02	0.03	-0.73	0.80
	1	4115	-0.08	0.21	-0.08	-0.07	-0.70	0.80
	Total	13382	0.00	0.20	-0.01	0.00	-0.73	0.80
P_HUNTS_Unit_TV_Coupons	0	9267	0.10	0.16	0.10	0.11	0.00	1.09
	1	4115	0.10	0.19	0.10	0.11	0.00	1.29
	Total	13382	0.10	0.17	0.10	0.11	0.00	1.29
P_HEINZ_Unit_TV_Coupons	0	9267	0.14	0.23	0.13	0.14	0.00	1.59
	1	4115	0.08	0.16	0.07	0.08	0.00	1.49
	Total	13382	0.12	0.21	0.12	0.12	0.00	1.59

1. Standard Deviation
2. Confidence Interval

Tabla I.5: Análisis descriptivo

Atributo	Tipo	Sum of Squares	df	Mean Square	F	Sig.
P_HUNTS_Unit_Extended_Price	Between Groups	6.28	1	6.28	225.71	0.00 %
	Within Groups	372.16	13380	0.03		
	Total	378.44	13381			
P_HEINZ_Unit_Extended_Price	Between Groups	8.95	1	8.95	400.57	0.00 %
	Within Groups	299.09	13380	0.02		
	Total	308.04	13381			
Dif_Unit_Extended_Price	Between Groups	6.96	1	6.96	183.25	0.00 %
	Within Groups	508.25	13380	0.04		
	Total	515.21	13381			
Dif2_Unit_Extended_Price	Between Groups	30.04	1	30.04	790.82	0.00 %
	Within Groups	508.25	13380	0.04		
	Total	538.29	13381			
P_HUNTS_Unit_TV_Coupons	Between Groups	0.00	1	0.00	0.02	88.79 %
	Within Groups	394.24	13380	0.03		
	Total	394.24	13381			
P_HEINZ_Unit_TV_Coupons	Between Groups	9.90	1	9.90	224.15	0.00 %
	Within Groups	590.97	13380	0.04		
	Total	600.87	13381			

Tabla I.6: Análisis de Varianza (ANOVA)

I.2.2. Variables categóricas

Atributo	Valores	y					
		0		1		Total	
		N	N (%)	N	N (%)	N	N (%)
Display	0	8104	87.5	3403	82.7	11507	86.0
	1	1163	12.5	712	17.3	1875	14.0
	Total	9267	100.0	4115	100.0	13382	100.0
Advertising	0	5432	58.6	2978	72.4	8410	62.8
	1	3835	41.4	1137	27.6	4972	37.2
	Total	9267	100.0	4115	100.0	13382	100.0
Type	0	7343	79.2	3251	79.0	10594	79.2
	1	1924	20.8	864	21.0	2788	20.8
	Total	9267	100.0	4115	100.0	13382	100.0
Coupon	0	7323	79.0	3795	92.2	11118	83.1
	1	1944	21.0	320	7.8	2264	16.9
	Total	9267	100.0	4115	100.0	13382	100.0
Special_Price	0	6930	74.8	2896	70.4	9826	73.4
	1	2337	25.2	1219	29.6	3556	26.6
	Total	9267	100.0	4115	100.0	13382	100.0
HUNTS_Dummy_CE	0	8690	93.8	3738	90.8	12428	92.9
	1	577	6.2	377	9.2	954	7.1
	Total	9267	100.0	4115	100.0	13382	100.0
HEINZ_Dummy_CE	0	7923	85.5	3230	78.5	11153	83.3
	1	1344	14.5	885	21.5	2229	16.7
	Total	9267	100.0	4115	100.0	13382	100.0
Frecuencia	ALTA	4932	53.2	1828	44.4	6760	50.5
	BAJA	2121	22.9	1120	27.2	3241	24.2
	NORMAL	2214	23.9	1167	28.4	3381	25.3
	Total	9267	100.0	4115	100.0	13382	100.0

Tabla I.7: Análisis descriptivo

Atributo	**Estadístico**	**y**
Display	Chi-square	53.4241
	df	1
	Sig.	0.00 %
Advertising	Chi-square	230.8257
	df	1
	Sig.	0.00 %
Type	Chi-square	0.0950
	df	1
	Sig.	75.79 %
Coupon	Chi-square	353.3101
	df	1
	Sig.	0.00 %
Special_Price	Chi-square	28.3365
	df	1
	Sig.	0.00 %
HUNTS_Dummy_CE	Chi-square	37.0816
	df	1
	Sig.	0.00 %
HEINZ_Dummy_CE	Chi-square	100.6866
	df	1
	Sig.	0.00 %
Frecuencia	Chi-square	88.2459
	df	2
	Sig.	0.00 %

Tabla I.8: Pearson Chi-Square tests

I.3. Variables demográficas

I.3.1. Variables continuas

Atributo	Valores	N	Mean	S.D.[1]	95 % C.I.[2]		Min	Max
					Lower	Upper		
Number_of_Cats	0	9267	0.32	0.71	0.30	0.33	0	8
	1	4115	0.39	0.80	0.37	0.42	0	8
	Total	13382	0.34	0.74	0.33	0.35	0	8
Number_of_Dogs	0	9267	0.56	0.75	0.55	0.58	0	9
	1	4115	0.58	0.72	0.56	0.60	0	6
	Total	13382	0.57	0.74	0.55	0.58	0	9
Number_of_TVs	0	9267	2.47	1.13	2.45	2.49	1	8
	1	4115	2.45	1.13	2.41	2.48	1	8
	Total	13382	2.46	1.13	2.44	2.48	1	8
Number_of_Members_in_ Household	0	9267	3.51	1.37	3.48	3.53	1	8
	1	4115	3.51	1.35	3.47	3.55	1	8
	Total	13382	3.51	1.36	3.48	3.53	1	8
Male_Head_Average_ Working_Hours	0	9267	34.48	19.92	34.08	34.89	0	80
	1	4115	33.95	20.45	33.32	34.57	0	80
	Total	13382	34.32	20.09	33.98	34.66	0	80
Female_Head_Average_ Working_Hours	0	9267	24.65	18.50	24.27	25.02	0	90
	1	4115	23.32	19.17	22.73	23.91	0	90
	Total	13382	24.24	18.72	23.92	24.56	0	90

1. Standard Deviation
2. Confidence Interval

Tabla I.9: Análisis descriptivo

Atributo	Tipo	Sum of Squares	df	Mean Square	F	Sig.
Number_of_Cats	Between Groups	15.8	1	15.76	29.07	0.00%
	Within Groups	7254.0	13380	0.54		
	Total	7269.8	13381			
Number_of_Dogs	Between Groups	1.4	1	1.41	2.56	10.98%
	Within Groups	7387.4	13380	0.55		
	Total	7388.8	13381			
Number_of_TVs	Between Groups	1.0	1	0.98	0.76	38.26%
	Within Groups	17183.6	13380	1.28		
	Total	17184.6	13381			
Number_of_Members_in_ Household	Between Groups	0.2	1	0.16	0.09	76.89%
	Within Groups	24824.6	13380	1.86		
	Total	24824.8	13381			
Male_Head_Average_ Working_Hours	Between Groups	818.6	1	818.63	2.03	15.44%
	Within Groups	5399083.2	13380	403.52		
	Total	5399901.9	13381			
Female_Head_Average_ Working_Hours	Between Groups	5017.4	1	5017.42	14.33	0.02%
	Within Groups	4683233.5	13380	350.02		
	Total	4688250.9	13381			

Tabla I.10: Análisis de Varianza (ANOVA)

I.3.2. Variables categóricas

Atributo	Valores	y					
		0		1		Total	
		N	N (%)	N	N (%)	N	N (%)
Cable_Status	0	2811	30.3	1309	31.8	4120	30.8
	1	6456	69.7	2806	68.2	9262	69.2
	Total	9267	100.0	4115	100.0	13382	100.0
Meter_Status	0	4863	52.5	2301	55.9	7164	53.5
	1	4404	47.5	1814	44.1	6218	46.5
	Total	9267	100.0	4115	100.0	13382	100.0
Panelist_Status	1	8737	94.3	3822	92.9	12559	93.8
	2	25	0.3	28	0.7	53	0.4
	3	4	0.0	0	0.0	4	0.0
	4	55	0.6	33	0.8	88	0.7
	5	274	3.0	115	2.8	389	2.9
	8	27	0.3	4	0.1	31	0.2
	9	145	1.6	113	2.7	258	1.9
	Total	9267	100.0	4115	100.0	13382	100.0
Residence_Type	1	254	2.7	101	2.5	355	2.7
	2	64	0.7	46	1.1	110	0.8
	3	8597	92.8	3781	91.9	12378	92.5
	4	169	1.8	82	2.0	251	1.9
	5	141	1.5	84	2.0	225	1.7
	6	42	0.5	21	0.5	63	0.5
	Total	9267	100.0	4115	100.0	13382	100.0
Residence_Status	1	8326	89.8	3670	89.2	11996	89.6
	2	866	9.3	418	10.2	1284	9.6
	3	75	0.8	27	0.7	102	0.8
	Total	9267	100.0	4115	100.0	13382	100.0
Residence_Duration	1	226	2.4	102	2.5	328	2.5
	2	156	1.7	90	2.2	246	1.8
	3	696	7.5	362	8.8	1058	7.9
	4	1340	14.5	674	16.4	2014	15.1
	5	2970	32.0	1353	32.9	4323	32.3
	6	1632	17.6	595	14.5	2227	16.6
	7	902	9.7	329	8.0	1231	9.2
	8	1345	14.5	610	14.8	1955	14.6
	Total	9267	100.0	4115	100.0	13382	100.0
Household_Income	1	159	1.7	86	2.1	245	1.8
	2	446	4.8	172	4.2	618	4.6

Atributo	Valores	y					
		0		1		Total	
		N	N (%)	N	N (%)	N	N (%)
	3	674	7.3	343	8.3	1017	7.6
	4	946	10.2	524	12.7	1470	11.0
	5	1656	17.9	650	15.8	2306	17.2
	6	1443	15.6	720	17.5	2163	16.2
	7	1192	12.9	546	13.3	1738	13.0
	8	922	9.9	299	7.3	1221	9.1
	9	622	6.7	270	6.6	892	6.7
	10	465	5.0	201	4.9	666	5.0
	11	413	4.5	162	3.9	575	4.3
	12	211	2.3	81	2.0	292	2.2
	13	47	0.5	20	0.5	67	0.5
	14	71	0.8	41	1.0	112	0.8
	Total	9267	100.0	4115	100.0	13382	100.0
Male_Head_Working_Status	0	1084	11.7	488	11.9	1572	11.7
	1	6524	70.4	2901	70.5	9425	70.4
	2	279	3.0	87	2.1	366	2.7
	3	132	1.4	48	1.2	180	1.3
	4	322	3.5	102	2.5	424	3.2
	5	89	1.0	58	1.4	147	1.1
	7	837	9.0	431	10.5	1268	9.5
	Total	9267	100.0	4115	100.0	13382	100.0
Male_Head_Occupation	0	2010	21.7	977	23.7	2987	22.3
	1	45	0.5	7	0.2	52	0.4
	2	1413	15.2	556	13.5	1969	14.7
	3	1033	11.1	476	11.6	1509	11.3
	4	117	1.3	76	1.8	193	1.4
	5	149	1.6	91	2.2	240	1.8
	6	667	7.2	299	7.3	966	7.2
	7	991	10.7	416	10.1	1407	10.5
	8	952	10.3	341	8.3	1293	9.7
	9	275	3.0	160	3.9	435	3.3
	10	1615	17.4	716	17.4	2331	17.4
	Total	9267	100.0	4115	100.0	13382	100.0
Male_Head_Education	0	1084	11.7	488	11.9	1572	11.7
	1	1	0.0	5	0.1	6	0.0
	2	32	0.3	29	0.7	61	0.5
	3	313	3.4	182	4.4	495	3.7
	4	465	5.0	207	5.0	672	5.0
	5	2040	22.0	836	20.3	2876	21.5
	6	480	5.2	204	5.0	684	5.1
	7	1007	10.9	381	9.3	1388	10.4
	8	1522	16.4	740	18.0	2262	16.9

Atributo	Valores	y					
		0		1		Total	
		N	N (%)	N	N (%)	N	N (%)
	9	1391	15.0	677	16.5	2068	15.5
	10	296	3.2	123	3.0	419	3.1
	11	636	6.9	243	5.9	879	6.6
	Total	9267	100.0	4115	100.0	13382	100.0
Male_Head_SpanishHispanic	0	9257	99.9	4094	99.5	13351	99.8
	1	6	0.1	11	0.3	17	0.1
	4	4	0.0	10	0.2	14	0.1
	Total	9267	100.0	4115	100.0	13382	100.0
Male_Head_Race	0	1084	11.7	488	11.9	1572	11.7
	1	8092	87.3	3556	86.4	11648	87.0
	2	31	0.3	43	1.0	74	0.6
	3	22	0.2	9	0.2	31	0.2
	4	27	0.3	7	0.2	34	0.3
	5	11	0.1	12	0.3	23	0.2
	Total	9267	100.0	4115	100.0	13382	100.0
Female_Head_Working_Status	0	83	0.9	43	1.0	126	0.9
	1	4237	45.7	1763	42.8	6000	44.8
	2	2181	23.5	898	21.8	3079	23.0
	3	59	0.6	26	0.6	85	0.6
	4	66	0.7	65	1.6	131	1.0
	5	77	0.8	35	0.9	112	0.8
	6	1931	20.8	998	24.3	2929	21.9
	7	633	6.8	287	7.0	920	6.9
	Total	9267	100.0	4115	100.0	13382	100.0
Female_Head_Occupation	0	2724	29.4	1363	33.1	4087	30.5
	2	159	1.7	88	2.1	247	1.8
	3	54	0.6	37	0.9	91	0.7
	4	62	0.7	55	1.3	117	0.9
	5	1044	11.3	415	10.1	1459	10.9
	6	2819	30.4	1052	25.6	3871	28.9
	7	543	5.9	243	5.9	786	5.9
	8	1210	13.1	550	13.4	1760	13.2
	9	205	2.2	101	2.5	306	2.3
	10	447	4.8	211	5.1	658	4.9
	Total	9267	100.0	4115	100.0	13382	100.0
Female_Head_Education	0	83	0.9	43	1.0	126	0.9
	1	1	0.0	4	0.1	5	0.0
	2	13	0.1	23	0.6	36	0.3
	3	355	3.8	154	3.7	509	3.8
	4	411	4.4	275	6.7	686	5.1
	5	2485	26.8	965	23.5	3450	25.8
	6	776	8.4	315	7.7	1091	8.2

Atributo	Valores	y					
		0		1		Total	
		N	N (%)	N	N (%)	N	N (%)
	7	1361	14.7	537	13.0	1898	14.2
	8	1814	19.6	784	19.1	2598	19.4
	9	1410	15.2	703	17.1	2113	15.8
	10	285	3.1	166	4.0	451	3.4
	11	273	2.9	146	3.5	419	3.1
	Total	9267	100.0	4115	100.0	13382	100.0
Female_Head_SpanishHispanic	0	9247	99.8	4112	99.9	13359	99.8
	1	4	0.0	0	0.0	4	0.0
	4	16	0.2	3	0.1	19	0.1
	Total	9267	100.0	4115	100.0	13382	100.0
Female_Head_Race	0	83	0.9	43	1.0	126	0.9
	1	9089	98.1	4029	97.9	13118	98.0
	2	24	0.3	25	0.6	49	0.4
	3	21	0.2	1	0.0	22	0.2
	4	15	0.2	15	0.4	30	0.2
	5	35	0.4	2	0.0	37	0.3
	Total	9267	100.0	4115	100.0	13382	100.0
Washing_Machine	0	352	3.8	135	3.3	487	3.6
	1	2541	27.4	1629	39.6	4170	31.2
	2	6374	68.8	2351	57.1	8725	65.2
	Total	9267	100.0	4115	100.0	13382	100.0
Clothes_Dryer	0	410	4.4	199	4.8	609	4.6
	1	2462	26.6	1573	38.2	4035	30.2
	2	6395	69.0	2343	56.9	8738	65.3
	Total	9267	100.0	4115	100.0	13382	100.0
Dishwasher	0	3501	37.8	1559	37.9	5060	37.8
	1	1893	20.4	1211	29.4	3104	23.2
	2	3873	41.8	1345	32.7	5218	39.0
	Total	9267	100.0	4115	100.0	13382	100.0
Freezer	0	2357	25.4	1105	26.9	3462	25.9
	1	1916	20.7	1139	27.7	3055	22.8
	2	4994	53.9	1871	45.5	6865	51.3
	Total	9267	100.0	4115	100.0	13382	100.0
Toaster	0	250	2.7	140	3.4	390	2.9
	1	2343	25.3	1566	38.1	3909	29.2
	2	6674	72.0	2409	58.5	9083	67.9
	Total	9267	100.0	4115	100.0	13382	100.0
Toaster_Broiler_Oven	0	6377	68.8	2736	66.5	9113	68.1
	1	1208	13.0	770	18.7	1978	14.8
	2	1682	18.2	609	14.8	2291	17.1
	Total	9267	100.0	4115	100.0	13382	100.0
Blender	0	1112	12.0	477	11.6	1589	11.9

Atributo	Valores	y					
		0		1		Total	
		N	N (%)	N	N (%)	N	N (%)
	1	2398	25.9	1539	37.4	3937	29.4
	2	5757	62.1	2099	51.0	7856	58.7
	Total	9267	100.0	4115	100.0	13382	100.0
Food_Processor	0	5725	61.8	2412	58.6	8137	60.8
	1	1414	15.3	823	20.0	2237	16.7
	2	2128	23.0	880	21.4	3008	22.5
	Total	9267	100.0	4115	100.0	13382	100.0
Microwave	0	1139	12.3	515	12.5	1654	12.4
	1	2360	25.5	1516	36.8	3876	29.0
	2	5768	62.2	2084	50.6	7852	58.7
	Total	9267	100.0	4115	100.0	13382	100.0
Convection_Oven	0	5675	61.2	2499	60.7	8174	61.1
	1	1560	16.8	874	21.2	2434	18.2
	2	2032	21.9	742	18.0	2774	20.7
	Total	9267	100.0	4115	100.0	13382	100.0
Coffee_Maker	0	2139	23.1	812	19.7	2951	22.1
	1	2118	22.9	1389	33.8	3507	26.2
	2	5010	54.1	1914	46.5	6924	51.7
	Total	9267	100.0	4115	100.0	13382	100.0
Trash_Compactor	0	8899	96.0	3971	96.5	12870	96.2
	1	200	2.2	83	2.0	283	2.1
	2	168	1.8	61	1.5	229	1.7
	Total	9267	100.0	4115	100.0	13382	100.0
Garbage_Disposal	0	5106	55.1	2201	53.5	7307	54.6
	1	2011	21.7	1205	29.3	3216	24.0
	2	2150	23.2	709	17.2	2859	21.4
	Total	9267	100.0	4115	100.0	13382	100.0
Hair_Dryer	0	579	6.2	226	5.5	805	6.0
	1	2472	26.7	1609	39.1	4081	30.5
	2	6216	67.1	2280	55.4	8496	63.5
	Total	9267	100.0	4115	100.0	13382	100.0
Curling_Iron	0	960	10.4	408	9.9	1368	10.2
	1	2219	23.9	1538	37.4	3757	28.1
	2	6088	65.7	2169	52.7	8257	61.7
	Total	9267	100.0	4115	100.0	13382	100.0
Hair_Rollers	0	4858	52.4	2062	50.1	6920	51.7
	1	1707	18.4	1087	26.4	2794	20.9
	2	2702	29.2	966	23.5	3668	27.4
	Total	9267	100.0	4115	100.0	13382	100.0
Vacuum_Cleaner	0	58	0.6	33	0.8	91	0.7
	1	2440	26.3	1617	39.3	4057	30.3
	2	6769	73.0	2465	59.9	9234	69.0

Atributo	**Valores**	**y**					
		0		**1**		**Total**	
		N	**N (%)**	**N**	**N (%)**	**N**	**N (%)**
	Total	9267	100.0	4115	100.0	13382	100.0
Water_Softner	0	4679	50.5	2561	62.2	7240	54.1
	1	428	4.6	163	4.0	591	4.4
	2	4160	44.9	1391	33.8	5551	41.5
	Total	9267	100.0	4115	100.0	13382	100.0
VCR	0	4307	46.5	1851	45.0	6158	46.0
	1	4960	53.5	2264	55.0	7224	54.0
	Total	9267	100.0	4115	100.0	13382	100.0

Tabla I.11: Análisis descriptivo

Atributo	Estadístico	y
Cable_Status	Chi-square	2.9174
	df	1
	Sig.	8.76%
Meter_Status	Chi-square	13.5630
	df	1
	Sig.	0.02%
Panelist_Status	Chi-square	41.9112
	df	6
	Sig.	0.00%
Residence_Type	Chi-square	12.6622
	df	5
	Sig.	2.68%
Residence_Status	Chi-square	2.9786
	df	2
	Sig.	22.55%
Residence_Duration	Chi-square	44.0563
	df	7
	Sig.	0.00%
Household_Income	Chi-square	66.9890
	df	13
	Sig.	0.00%
Male_Head_Working_Status	Chi-square	30.2550
	df	6
	Sig.	0.00%
Male_Head_Occupation	Chi-square	51.5087
	df	10
	Sig.	0.00%
Male_Head_Education	Chi-square	47.8400
	df	11
	Sig.	0.00%
Male_Head_SpanishHispanic	Chi-square	20.1298
	df	2
	Sig.	0.00%
Male_Head_Race	Chi-square	32.9899
	df	5
	Sig.	0.00%
Female_Head_Working_Status	Chi-square	46.7589
	df	7

Atributo	Estadístico	y
	Sig.	0.00 %
Female_Head_Occupation	Chi-square	62.7922
	df	9
	Sig.	0.00 %
Female_Head_Education	Chi-square	90.5533
	df	11
	Sig.	0.00 %
Female_Head_SpanishHispanic	Chi-square	3.7791
	df	2
	Sig.	15.11 %
Female_Head_Race	Chi-square	33.6137
	df	5
	Sig.	0.00 %
Washing_Machine	Chi-square	196.7861
	df	2
	Sig.	0.00 %
Clothes_Dryer	Chi-square	193.1002
	df	2
	Sig.	0.00 %
Dishwasher	Chi-square	160.1822
	df	2
	Sig.	0.00 %
Freezer	Chi-square	102.8488
	df	2
	Sig.	0.00 %
Toaster	Chi-square	240.2566
	df	2
	Sig.	0.00 %
Toaster_Broiler_Oven	Chi-square	83.0756
	df	2
	Sig.	0.00 %
Blender	Chi-square	188.9796
	df	2
	Sig.	0.00 %
Food_Processor	Chi-square	46.1723
	df	2
	Sig.	0.00 %

Atributo	**Estadístico**	**y**
Microwave	Chi-square	192.7277
	df	2
	Sig.	0.00 %
Convection_Oven	Chi-square	51.3904
	df	2
	Sig.	0.00 %
Coffee_Maker	Chi-square	175.0614
	df	2
	Sig.	0.00 %
Trash_Compactor	Chi-square	2.1535
	df	2
	Sig.	34.07 %
Garbage_Disposal	Chi-square	117.0818
	df	2
	Sig.	0.00 %
Hair_Dryer	Chi-square	208.0998
	df	2
	Sig.	0.00 %
Curling_Iron	Chi-square	261.5097
	df	2
	Sig.	0.00 %
Hair_Rollers	Chi-square	123.7634
	df	2
	Sig.	0.00 %
Vacuum_Cleaner	Chi-square	230.6207
	df	2
	Sig.	0.00 %
Water_Softner	Chi-square	159.8905
	df	2
	Sig.	0.00 %
VCR	Chi-square	2.5638
	df	1
	Sig.	10.93 %

Tabla I.12: Pearson Chi-Square tests

Anexo J

Resultados

J.1. Selección de atributos

Atributos	***Benchmark***		**Modelo (1)**		**Modelo (2)**		**Modelo (3)**	
	A	**B**	**A**	**B**	**A**	**B**	**A**	**B**
	BIC	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**	**BIC**	**AIC**
Tipo	S	S	S	S	S	S	S	S
Display	S	S	S	S	S	S	S	S
Advertising	S	S	S	S	S	S	S	S
Type	S	S	S	S	S	S	S	S
Coupon		S		S		S		S
Special_Price								
HUNTS_Dummy_CE(*)							M	M
HEINZ_Dummy_CE(*)	S	S	S	S	S	S	S	S
Cable_Status		S		S		S		S
Meter_Status	S	S	S	S	S	S	S	S
Panelist_Status								
Number_of_Cats	S	S	S	S	S	S	S	S
Number_of_Dogs								
Number_of_TVs								
Residence_Type		S		S		S		S
Residence_Status		S		S		S		S
Residence_Duration		S		S		S		S
Household_Income		S		S		S		S
Number_of_Members_in_Household		S		S		S		S
Male_Head_Working_Status								
Male_Head_Average_Working_Hours								
Male_Head_Occupation		S		S		S		S
Male_Head_Education								
Male_Head_SpanishHispanic								
Male_Head_Race		S		S		S		S
Female_Head_Working_Status		S		S		S		S
Female_Head_Average_Working_Hours								
Female_Head_Occupation		S		S		S		S

Atributos	*Benchmark*		Modelo (1)		Modelo (2)		Modelo (3)	
	A BIC	B AIC	A BIC	B AIC	A BIC	B AIC	A BIC	B AIC
Female_Head_Education		S		S		S		S
Female_Head_SpanishHispanic								
Female_Head_Race								
Washing_Machine		S		S		S		S
Clothes_Dryer		S		S		S		S
Dishwasher		S		S		S		S
Freezer		S		S		S		S
Toaster								
Toaster_Broiler_Oven								
Blender								
Food_Processor	S	S	S	S	S	S	S	S
Microwave								
Convection_Oven								
Coffee_Maker		S		S		S		S
Trash_Compactor								
Garbage_Disposal		S		S		S		S
Hair_Dryer								
Curling_Iron	S	S	S	S	S	S	S	S
Hair_Rollers		S		S		S		S
Vacuum_Cleaner		S		S		S		S
Water_Softner								
VCR								
P_HUNTS_Unit_Extended_Price(*)	S	S	S	S	S	S	S	S
P_HEINZ_Unit_Extended_Price(*)	S	S	S	S	S	S	S	S
Dif_Unit_Extended_Price		S		S		S		S
Dif2_Unit_Extended_Price								
P_HUNTS_Unit_TV_Coupons(*)	S	S	S	S	S	S	S	S
P_HEINZ_Unit_TV_Coupons(*)	S	S	S	S	S	S	S	S

Tabla J.1: Resultado selección de atributos. S: variables seleccionadas por algoritmo. M: variables agregadas manualmente.

J.2. Conjunto de atributos A

J.2.1. Modelo Regresión Logística (*Benchmark*)

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
constante	-0.4174	0.3016	-1.38	1.92	16.6 %	-1.0084	0.1737
P_HUNTS_Unit_Extended_Price	-2.9094	0.2072	-14.04	197.14	0.0 %	-3.3156	-2.5033
P_HEINZ_Unit_Extended_Price	2.4380	0.2105	11.58	134.11	0.0 %	2.0254	2.8506
P_HUNTS_Unit_TV_Coupons	0.7725	0.1557	4.96	24.63	0.0 %	0.4674	1.0775
P_HEINZ_Unit_TV_Coupons	-0.7126	0.1541	-4.63	21.39	0.0 %	-1.0146	-0.4106
HEINZ_Dummy_CE	0.6418	0.1211	5.30	28.08	0.0 %	0.4044	0.8792
Tipo_Baja	0.4477	0.0710	6.30	39.72	0.0 %	0.3085	0.5869
Tipo_Normal	0.3769	0.0688	5.48	30.01	0.0 %	0.2420	0.5117
Display	0.4445	0.0797	5.58	31.09	0.0 %	0.2883	0.6008
Advertising	-1.3579	0.1050	-12.93	167.10	0.0 %	-1.5638	-1.1520
Type	0.8487	0.1189	7.14	50.92	0.0 %	0.6156	1.0819
Meter_Status	-0.2239	0.0570	-3.93	15.42	0.0 %	-0.3356	-0.1121
Number_of_Cats	1.1011	0.2958	3.72	13.86	0.0 %	0.5213	1.6808
Food_Processor__0	-0.2856	0.0721	-3.96	15.71	0.0 %	-0.4269	-0.1444
Food_Processor__1	-0.4448	0.1071	-4.15	17.24	0.0 %	-0.6547	-0.2349
Curling_Iron__0	0.2411	0.0983	2.45	6.02	1.4 %	0.0485	0.4336
Curling_Iron__1	0.7645	0.0863	8.86	78.43	0.0 %	0.5953	0.9337

(*) 95 % confidence interval

Tabla J.2: Set atributos A. Resultado Modelo de Regresión Logística (*Benchmark*).

Estadístico	Valor
N	7117
Log-Likelihood	-3835.1
AIC	7702.3
BIC	7812.2
G statistic	822.6
p-value(G statistic)	0
$\widetilde{R}^2$ (McFadden's)	9.69 %
$\widetilde{R}^2$ (McFadden's adjusted)	9.33 %
$\widetilde{R}^2$ (Cox & Snell)	10.92 %
$\widetilde{R}^2$ (Nagelkerke)	15.66 %

Tabla J.3: Set atributos A. Modelo *Benchmark*, Estadísticos de Training.

Estadístico	Training	Test	Validation
N	7117	3008	3257
KS	31.14 %	29.13 %	37.38 %
AUC (ROC)	71.07 %	69.64 %	74.35 %
Cut-off (ROC)	0.2850	0.2850	0.2850
TP rate	66.25 %	62.56 %	84.56 %
FN rate	33.75 %	37.44 %	15.44 %
FP rate	34.93 %	33.56 %	49.87 %
TN rate	65.07 %	66.44 %	50.13 %
J de Youden	31.32 %	29.00 %	34.69 %
MSE	18.07 %	17.86 %	19.94 %

Tabla J.4: Set atributos A. Modelo *Benchmark*, Estadísticos de Eficiencia por Base de datos.

J.2.2. Modelo propuesto (1)

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
constante	-1.8762	0.1372	-13.68	187.11	0.0 %	-2.1450	-1.6074
P_HUNTS_Unit_Extended_Price	-1.5526	0.0773	-20.09	403.64	0.0 %	-1.7040	-1.4011
P_HEINZ_Unit_Extended_Price	2.5669	0.0777	33.05	1092.06	0.0 %	2.4147	2.7192
P_HUNTS_Unit_TV_Coupons	0.4809	0.1312	3.66	13.43	0.0 %	0.2237	0.7382
P_HEINZ_Unit_TV_Coupons	-0.6652	0.1299	-5.12	26.22	0.0 %	-0.9199	-0.4106
HEINZ_Dummy_CE	0.5482	0.1073	5.11	26.12	0.0 %	0.3380	0.7584
Tipo_Baja	0.3212	0.0610	5.27	27.75	0.0 %	0.2017	0.4407
Tipo_Normal	0.2671	0.0590	4.53	20.50	0.0 %	0.1515	0.3827
Display	0.4040	0.0681	5.93	35.22	0.0 %	0.2706	0.5375
Advertising	-0.9098	0.0806	-11.29	127.51	0.0 %	-1.0677	-0.7518
Type	0.6390	0.0912	7.00	49.05	0.0 %	0.4602	0.8179
Meter_Status	-0.1696	0.0486	-3.49	12.20	0.1 %	-0.2648	-0.0744
Number_of_Cats	0.7920	0.2610	3.03	9.21	0.2 %	0.2805	1.3035
Food_Processor__0	-0.2055	0.0604	-3.40	11.58	0.1 %	-0.3238	-0.0871
Food_Processor__1	-0.3497	0.0922	-3.79	14.37	0.0 %	-0.5304	-0.1689
Curling_Iron__0	0.1039	0.0827	1.26	1.58	20.9 %	-0.0581	0.2660
Curling_Iron__1	0.3854	0.0692	5.57	31.00	0.0 %	0.2497	0.5210

(*) 95 % confidence interval

Tabla J.5: Set atributos A. Resultado Modelo (1): Competencia en precios.

Estadístico	Valor
μ	-0.2
N	7117
Log-Likelihood	-4035.2
AIC	8102
BIC	8212
G statistic	2111.5
p-value(G statistic)	0
$\widetilde{R}^2$ (McFadden's)	8.47 %
$\widetilde{R}^2$ (McFadden's adjusted)	8.12 %
$\widetilde{R}^2$ (Cox & Snell)	9.62 %
$\widetilde{R}^2$ (Nagelkerke)	13.80 %

Tabla J.6: Set atributos A. Modelo (1), Estadísticos de Training.

Estadístico	Training	Test	Validation
N	7117	3008	3257
KS	32.23 %	29.58 %	37.55 %
AUC (ROC)	71.02 %	69.42 %	74.07 %
Cut-off (ROC)	0.3020	0.3020	0.3020
TP rate	68.68 %	65.15 %	87.99 %
FN rate	31.32 %	34.85 %	12.01 %
FP rate	36.15 %	35.97 %	54.63 %
TN rate	63.85 %	64.03 %	45.37 %
J de Youden	32.53 %	29.18 %	33.36 %
MSE	18.33 %	18.10 %	20.13 %

Tabla J.7: Set atributos A. Modelo (1), Estadísticos de Eficiencia por Base de datos.

J.2.3. Modelo propuesto (2)

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
constante	-1.5447	0.1362	-11.34	128.68	0.0%	-1.8116	-1.2778
P_HUNTS_Unit_Extended_Price	-1.6602	0.0813	-20.42	417.11	0.0%	-1.8195	-1.5008
P_HEINZ_Unit_Extended_Price	2.4792	0.0790	31.39	985.62	0.0%	2.3244	2.6339
P_HUNTS_Unit_TV_Coupons	1.1537	0.0729	15.82	250.31	0.0%	1.0108	1.2966
P_HEINZ_Unit_TV_Coupons	-2.0365	0.0746	-27.31	745.77	0.0%	-2.1827	-1.8903
HEINZ_Dummy_CE	0.3589	0.0978	3.67	13.47	0.0%	0.1672	0.5506
Tipo_Baja	0.2973	0.0560	5.31	28.18	0.0%	0.1875	0.4070
Tipo_Normal	0.2353	0.0542	4.34	18.87	0.0%	0.1291	0.3415
Display	0.3388	0.0631	5.37	28.82	0.0%	0.2151	0.4626
Advertising	-0.6996	0.0721	-9.71	94.22	0.0%	-0.8409	-0.5583
Type	0.4422	0.0815	5.43	29.47	0.0%	0.2826	0.6019
Meter_Status	-0.1401	0.0444	-3.16	9.96	0.2%	-0.2272	-0.0531
Number_of_Cats	0.6538	0.2405	2.72	7.39	0.7%	0.1824	1.1252
Food_Processor__0	-0.1737	0.0551	-3.15	9.94	0.2%	-0.2817	-0.0657
Food_Processor__1	-0.2879	0.0850	-3.39	11.48	0.1%	-0.4544	-0.1214
Curling_Iron__0	0.1208	0.0758	1.59	2.54	11.1%	-0.0279	0.2695
Curling_Iron__1	0.4024	0.0641	6.28	39.45	0.0%	0.2768	0.5280

(*) 95 % confidence interval

Tabla J.8: Set atributos A. Resultado Modelo (2): Competencia en precios y descuentos.

Estadístico	Valor
μ	-0.2
N	7117
Log-Likelihood	-4385.4
AIC	8803
BIC	8913
G statistic	3100.4
p-value(G statistic)	0
$\widetilde{R}^2$ (McFadden's)	6.61%
$\widetilde{R}^2$ (McFadden's adjusted)	6.26%
$\widetilde{R}^2$ (Cox & Snell)	7.59%
$\widetilde{R}^2$ (Nagelkerke)	10.89%

Tabla J.9: Set atributos A. Modelo (2), Estadísticos de Training.

Estadístico	Training	Test	Validation
N	7117	3008	3257
KS	27.06 %	25.51 %	39.18 %
AUC (ROC)	68.66 %	67.61 %	76.12 %
Cut-off (ROC)	0.3160	0.3160	0.3160
TP rate	68.73 %	65.64 %	86.35 %
FN rate	31.27 %	34.36 %	13.65 %
FP rate	41.54 %	40.98 %	48.30 %
TN rate	58.46 %	59.02 %	51.70 %
J de Youden	27.19 %	24.66 %	38.05 %
MSE	18.80 %	18.51 %	19.35 %

Tabla J.10: Set atributos A. Modelo (2), Estadísticos de Eficiencia por Base de datos.

J.2.4. Modelo propuesto (3)

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
constante	-1.5118	0.1358	-11.13	123.98	0.0 %	-1.7779	-1.2457
P_HUNTS_Unit_Extended_Price	-1.6686	0.0811	-20.58	423.47	0.0 %	-1.8276	-1.5097
P_HEINZ_Unit_Extended_Price	2.4695	0.0789	31.30	979.93	0.0 %	2.3149	2.6241
P_HUNTS_Unit_TV_Coupons	1.1502	0.0728	15.81	249.81	0.0 %	1.0076	1.2929
P_HEINZ_Unit_TV_Coupons	-2.0442	0.0744	-27.46	754.26	0.0 %	-2.1901	-1.8983
HUNTS_Dummy_CE	0.4616	0.0773	5.97	35.68	0.0 %	0.3101	0.6131
HEINZ_Dummy_CE	-0.1729	0.0661	-2.62	6.84	0.9 %	-0.3024	-0.0433
Tipo_Baja	0.2906	0.0558	5.20	27.08	0.0 %	0.1811	0.4000
Tipo_Normal	0.2312	0.0540	4.28	18.33	0.0 %	0.1253	0.3370
Display	0.3342	0.0630	5.31	28.17	0.0 %	0.2108	0.4576
Advertising	-0.7141	0.0717	-9.96	99.13	0.0 %	-0.8547	-0.5735
Type	0.4536	0.0812	5.59	31.24	0.0 %	0.2945	0.6126
Meter_Status	-0.1161	0.0440	-2.64	6.95	0.8 %	-0.2023	-0.0298
Number_of_Cats	0.6403	0.2400	2.67	7.12	0.8 %	0.1698	1.1107
Food_Processor__0	-0.1713	0.0550	-3.12	9.71	0.2 %	-0.2791	-0.0636
Food_Processor__1	-0.2902	0.0847	-3.42	11.73	0.1 %	-0.4563	-0.1241
Curling_Iron__0	0.1187	0.0757	1.57	2.46	11.7 %	-0.0296	0.2670
Curling_Iron__1	0.4062	0.0639	6.36	40.40	0.0 %	0.2809	0.5314

(*) 95 % confidence interval

Tabla J.11: Set atributos A. Resultado Modelo (3): Competencia en precios, descuentos y marketing.

Estadístico	Valor
μ	-0.2
e	0.1
N	7117
Log-Likelihood	-4422.2
AIC	8878
BIC	8995
G statistic	3083.7
p-value(G statistic)	0
$\widetilde{R}^2$ (McFadden's)	6.31 %
$\widetilde{R}^2$ (McFadden's adjusted)	5.93 %
$\widetilde{R}^2$ (Cox & Snell)	7.25 %
$\widetilde{R}^2$ (Nagelkerke)	10.41 %

Tabla J.12: Set atributos B. Modelo (3), Estadísticos de Training.

Estadístico	Training	Test	Validation
N	7117	3008	3257
KS	26.83 %	24.44 %	38.52 %
AUC (ROC)	68.26 %	67.07 %	75.38 %
Cut-off (ROC)	0.3210	0.3210	0.3210
TP rate	67.24 %	62.93 %	86.51 %
FN rate	32.76 %	37.07 %	13.49 %
FP rate	40.27 %	39.66 %	47.80 %
TN rate	59.73 %	60.34 %	52.20 %
J de Youden	26.97 %	23.27 %	38.71 %
MSE	18.88 %	18.61 %	19.75 %

Tabla J.13: Set atributos B. Modelo (3), Estadísticos de Eficiencia por Base de datos.

J.3. Conjunto de atributos B

J.3.1. Modelo Regresión Logística (*Benchmark*)

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
constante	-0.5994	0.7664	-0.78	0.61	43.4 %	-2.1014	0.9026
P_HUNTS_Unit_Extended_Price	-3.0487	0.2193	-13.90	193.26	0.0 %	-3.4786	-2.6189
P_HEINZ_Unit_Extended_Price	2.6992	0.2271	11.89	141.31	0.0 %	2.2542	3.1443

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
P_HUNTS_Unit_TV_Coupons	0.7630	0.1610	4.74	22.46	0.0%	0.4475	1.0786
P_HEINZ_Unit_TV_Coupons	-0.8034	0.1597	-5.03	25.29	0.0%	-1.1164	-0.4903
HEINZ_Dummy_CE	0.6763	0.1257	5.38	28.96	0.0%	0.4299	0.9226
Dif_Unit_Extended_Price	-0.2255	0.1707	-1.32	1.75	18.6%	-0.5600	0.1090
Tipo_Baja	0.5110	0.0805	6.35	40.30	0.0%	0.3533	0.6688
Tipo_Normal	0.3798	0.0743	5.11	26.12	0.0%	0.2341	0.5254
Display	0.4435	0.0829	5.35	28.63	0.0%	0.2810	0.6060
Advertising	-2.3172	0.3585	-6.46	41.77	0.0%	-3.0199	-1.6144
Type	1.7265	0.3551	4.86	23.64	0.0%	1.0305	2.4226
Coupon	0.8971	0.3421	2.62	6.88	0.9%	0.2266	1.5675
Cable_Status	-0.1232	0.0653	-1.89	3.56	5.9%	-0.2512	0.0049
Meter_Status	-0.2356	0.0620	-3.80	14.45	0.0%	-0.3570	-0.1141
Number_of_Cats	1.1646	0.3212	3.63	13.14	0.0%	0.5350	1.7942
Residence_Type__1	0.2627	0.5463	0.48	0.23	63.1%	-0.8080	1.3335
Residence_Type__2	1.0915	0.5803	1.88	3.54	6.0%	-0.0459	2.2288
Residence_Type__3	0.0575	0.4984	0.12	0.01	90.8%	-0.9193	1.0343
Residence_Type__4	0.4201	0.5405	0.78	0.60	43.7%	-0.6392	1.4795
Residence_Type__5	0.4692	0.5438	0.86	0.74	38.8%	-0.5967	1.5351
Residence_Status	0.2763	0.1281	2.16	4.65	3.1%	0.0252	0.5274
Residence_Duration__1	0.1798	0.2123	0.85	0.72	39.7%	-0.2363	0.5960
Residence_Duration__2	0.5253	0.2311	2.27	5.17	2.3%	0.0724	0.9782
Residence_Duration__3	0.3395	0.1417	2.40	5.74	1.7%	0.0617	0.6172
Residence_Duration__4	0.2076	0.1201	1.73	2.99	8.4%	-0.0278	0.4430
Residence_Duration__5	0.2190	0.1045	2.10	4.39	3.6%	0.0142	0.4237
Residence_Duration__6	-0.1828	0.1146	-1.60	2.54	11.1%	-0.4074	0.0418
Residence_Duration__7	-0.2233	0.1332	-1.68	2.81	9.4%	-0.4844	0.0378
Household_Income__1	0.5190	0.3884	1.34	1.79	18.2%	-0.2422	1.2803
Household_Income__2	0.3081	0.3489	0.88	0.78	37.7%	-0.3757	0.9919
Household_Income__3	0.3429	0.3309	1.04	1.07	30.0%	-0.3057	0.9915
Household_Income__4	0.3925	0.3191	1.23	1.51	21.9%	-0.2330	1.0180
Household_Income__5	-0.0479	0.3154	-0.15	0.02	87.9%	-0.6661	0.5702
Household_Income__6	0.4150	0.3159	1.31	1.73	18.9%	-0.2041	1.0341
Household_Income__7	0.2754	0.3177	0.87	0.75	38.6%	-0.3473	0.8981
Household_Income__8	-0.1229	0.3234	-0.38	0.14	70.4%	-0.7567	0.5109
Household_Income__9	0.0995	0.3265	0.30	0.09	76.1%	-0.5404	0.7395
Household_Income__10	0.1266	0.3313	0.38	0.15	70.2%	-0.5227	0.7758
Household_Income__11	-0.1288	0.3351	-0.38	0.15	70.1%	-0.7856	0.5280
Household_Income__12	-0.0473	0.3556	-0.13	0.02	89.4%	-0.7442	0.6496
Household_Income__13	0.3841	0.5130	0.75	0.56	45.4%	-0.6213	1.3895
Number_of_Members_in_Household	0.2808	0.1914	1.47	2.15	14.2%	-0.0944	0.6560
Male_Head_Occupation__1	-0.5668	0.5250	-1.08	1.17	28.0%	-1.5958	0.4623
Male_Head_Occupation__2	-0.2643	0.1019	-2.59	6.73	1.0%	-0.4640	-0.0646
Male_Head_Occupation__3	-0.1807	0.1109	-1.63	2.66	10.3%	-0.3980	0.0366

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
Male_Head_Occupation__4	0.2392	0.2336	1.02	1.05	30.6%	-0.2187	0.6971
Male_Head_Occupation__5	-0.0081	0.2159	-0.04	0.00	97.0%	-0.4312	0.4151
Male_Head_Occupation__6	-0.0323	0.1250	-0.26	0.07	79.6%	-0.2772	0.2127
Male_Head_Occupation__7	-0.2104	0.1124	-1.87	3.51	6.1%	-0.4306	0.0098
Male_Head_Occupation__8	-0.2692	0.1180	-2.28	5.20	2.3%	-0.5004	-0.0379
Male_Head_Occupation__9	0.3086	0.1640	1.88	3.54	6.0%	-0.0128	0.6300
Male_Head_Race	0.2469	0.1104	2.24	5.00	2.5%	0.0305	0.4632
Female_Head_Working_Status__0	0.0611	0.3617	0.17	0.03	86.6%	-0.6477	0.7700
Female_Head_Working_Status__1	0.1158	0.1898	0.61	0.37	54.2%	-0.2562	0.4879
Female_Head_Working_Status__2	-0.0037	0.1975	-0.02	0.00	98.5%	-0.3908	0.3834
Female_Head_Working_Status__3	0.0185	0.4449	0.04	0.00	96.7%	-0.8534	0.8905
Female_Head_Working_Status__4	1.2927	0.3350	3.86	14.89	0.0%	0.6361	1.9492
Female_Head_Working_Status__5	-0.4999	0.3372	-1.48	2.20	13.8%	-1.1608	0.1610
Female_Head_Working_Status__6	0.2235	0.1424	1.57	2.46	11.7%	-0.0556	0.5025
Female_Head_Occupation__2	0.2567	0.2527	1.02	1.03	31.0%	-0.2385	0.7520
Female_Head_Occupation__3	-0.1279	0.3827	-0.33	0.11	73.8%	-0.8779	0.6222
Female_Head_Occupation__4	0.9746	0.3297	2.96	8.74	0.3%	0.3285	1.6208
Female_Head_Occupation__5	0.0806	0.1673	0.48	0.23	63.0%	-0.2473	0.4086
Female_Head_Occupation__6	-0.1541	0.1494	-1.03	1.06	30.2%	-0.4469	0.1387
Female_Head_Occupation__7	0.2049	0.1848	1.11	1.23	26.8%	-0.1573	0.5671
Female_Head_Occupation__8	0.0372	0.1688	0.22	0.05	82.5%	-0.2935	0.3680
Female_Head_Occupation__9	0.3868	0.2269	1.71	2.91	8.8%	-0.0578	0.8315
Female_Head_Education__1	0.8706	1.2729	0.68	0.47	49.4%	-1.6242	3.3654
Female_Head_Education__2	0.8880	0.6080	1.46	2.13	14.4%	-0.3036	2.0796
Female_Head_Education__3	-0.3135	0.2411	-1.30	1.69	19.4%	-0.7861	0.1591
Female_Head_Education__4	-0.0261	0.2251	-0.12	0.01	90.8%	-0.4672	0.4151
Female_Head_Education__5	-0.5755	0.1965	-2.93	8.58	0.3%	-0.9606	-0.1903
Female_Head_Education__6	-0.4195	0.2116	-1.98	3.93	4.7%	-0.8342	-0.0047
Female_Head_Education__7	-0.4170	0.1988	-2.10	4.40	3.6%	-0.8067	-0.0274
Female_Head_Education__8	-0.4082	0.1931	-2.11	4.47	3.5%	-0.7866	-0.0298
Female_Head_Education__9	-0.2824	0.1849	-1.53	2.33	12.7%	-0.6448	0.0800
Female_Head_Education__10	-0.0886	0.2243	-0.40	0.16	69.3%	-0.5282	0.3509
Washing_Machine	0.5083	0.2653	1.92	3.67	5.5%	-0.0118	1.0283
Clothes_Dryer	-0.4001	0.2665	-1.50	2.25	13.3%	-0.9225	0.1223
Dishwasher__0	-0.0300	0.0865	-0.35	0.12	72.9%	-0.1996	0.1397
Dishwasher__1	0.2205	0.1333	1.65	2.73	9.8%	-0.0409	0.4818
Freezer__0	-0.3170	0.0927	-3.42	11.70	0.1%	-0.4986	-0.1354
Freezer__1	-0.3872	0.1280	-3.03	9.16	0.3%	-0.6380	-0.1364
Food_Processor__0	-0.2464	0.0775	-3.18	10.12	0.2%	-0.3982	-0.0946
Food_Processor__1	-0.4488	0.1175	-3.82	14.59	0.0%	-0.6792	-0.2185
Coffee_Maker__0	-0.2275	0.0904	-2.52	6.33	1.2%	-0.4047	-0.0503
Coffee_Maker__1	0.0504	0.1203	0.42	0.18	67.6%	-0.1854	0.2861
Garbage_Disposal__0	0.0812	0.0884	0.92	0.85	35.8%	-0.0919	0.2544

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
Garbage_Disposal__1	-0.4108	0.1388	-2.96	8.76	0.3 %	-0.6829	-0.1387
Curling_Iron__0	0.2148	0.1231	1.74	3.04	8.1 %	-0.0265	0.4562
Curling_Iron__1	0.5560	0.1633	3.40	11.59	0.1 %	0.2359	0.8761
Hair_Rollers__0	0.1063	0.0797	1.33	1.78	18.2 %	-0.0499	0.2625
Hair_Rollers__1	0.2698	0.1148	2.35	5.53	1.9 %	0.0448	0.4949
Vacuum_Cleaner	-0.5471	0.2134	-2.56	6.57	1.0 %	-0.9654	-0.1288

(*) 95 % confidence interval

Tabla J.14: Set atributos A. Resultado Modelo de Regresión Logística (*Benchmark*).

Estadístico	Valor
N	7117
Log-Likelihood	-3691.3
AIC	7571
BIC	8216
G statistic	1110.3
p-value(G statistic)	0
$\widetilde{R}^2$ (McFadden's)	13.07 %
$\widetilde{R}^2$ (McFadden's adjusted)	10.88 %
$\widetilde{R}^2$ (Cox & Snell)	14.44 %
$\widetilde{R}^2$ (Nagelkerke)	20.73 %

Tabla J.15: Set atributos B. Modelo *Benchmark*, Estadísticos de Training.

Estadístico	Training	Test	Validation
N	7117	3008	3257
KS	34.34 %	30.40 %	40.22 %
AUC (ROC)	73.91 %	70.81 %	76.18 %
Cut-off (ROC)	0.3340	0.3340	0.3340
TP rate	58.34 %	53.45 %	77.77 %
FN rate	41.66 %	46.55 %	22.23 %
FP rate	23.74 %	23.95 %	38.58 %
TN rate	76.26 %	76.05 %	61.42 %
J de Youden	34.60 %	29.50 %	39.19 %
MSE	17.21 %	17.61 %	19.27 %

Tabla J.16: Set atributos B. Modelo *Benchmark*, Estadísticos de Eficiencia por Base de datos.

J.3.2. Modelo propuesto (1)

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
constante	-2.1198	0.5871	-3.61	13.04	0.0 %	-3.2704	-0.9692
P_HUNTS_Unit_Extended_Price	-1.5895	0.0791	-20.10	404.13	0.0 %	-1.7445	-1.4345
P_HEINZ_Unit_Extended_Price	2.6313	0.0800	32.89	1081.71	0.0 %	2.4745	2.7881
P_HUNTS_Unit_TV_Coupons	0.4568	0.1352	3.38	11.41	0.1 %	0.1917	0.7219
P_HEINZ_Unit_TV_Coupons	-0.6982	0.1339	-5.22	27.20	0.0 %	-0.9605	-0.4358
HEINZ_Dummy_CE	0.5692	0.1099	5.18	26.82	0.0 %	0.3538	0.7847
Dif_Unit_Extended_Price	-0.3267	0.1447	-2.26	5.10	2.4 %	-0.6103	-0.0431
Tipo_Baja	0.3716	0.0685	5.43	29.47	0.0 %	0.2375	0.5058
Tipo_Normal	0.2755	0.0634	4.35	18.90	0.0 %	0.1513	0.3997
Display	0.3845	0.0704	5.46	29.81	0.0 %	0.2465	0.5225
Advertising	-1.6670	0.3070	-5.43	29.49	0.0 %	-2.2687	-1.0653
Type	1.3163	0.3046	4.32	18.67	0.0 %	0.7192	1.9133
Coupon	0.7012	0.2955	2.37	5.63	1.8 %	0.1220	1.2804
Cable_Status	-0.1001	0.0557	-1.80	3.23	7.2 %	-0.2092	0.0090
Meter_Status	-0.1801	0.0526	-3.42	11.70	0.1 %	-0.2832	-0.0769
Number_of_Cats	0.8651	0.2812	3.08	9.46	0.2 %	0.3140	1.4162
Residence_Type__1	0.2223	0.4514	0.49	0.24	62.3 %	-0.6626	1.1071
Residence_Type__2	0.8138	0.4850	1.68	2.82	9.3 %	-0.1367	1.7644
Residence_Type__3	0.0442	0.4091	0.11	0.01	91.4 %	-0.7575	0.8459
Residence_Type__4	0.3086	0.4452	0.69	0.48	48.8 %	-0.5639	1.1812
Residence_Type__5	0.3702	0.4508	0.82	0.67	41.2 %	-0.5133	1.2537
Residence_Status	0.2069	0.1099	1.88	3.55	6.0 %	-0.0084	0.4222
Residence_Duration__1	0.1776	0.1833	0.97	0.94	33.3 %	-0.1818	0.5369
Residence_Duration__2	0.3855	0.1994	1.93	3.74	5.3 %	-0.0053	0.7764
Residence_Duration__3	0.2482	0.1217	2.04	4.16	4.1 %	0.0097	0.4867
Residence_Duration__4	0.1574	0.1024	1.54	2.36	12.4 %	-0.0433	0.3580
Residence_Duration__5	0.1703	0.0891	1.91	3.65	5.6 %	-0.0044	0.3449
Residence_Duration__6	-0.1335	0.0968	-1.38	1.90	16.8 %	-0.3233	0.0562
Residence_Duration__7	-0.1447	0.1119	-1.29	1.67	19.6 %	-0.3640	0.0746
Household_Income__1	0.3597	0.3362	1.07	1.14	28.5 %	-0.2992	1.0187
Household_Income__2	0.2281	0.3010	0.76	0.57	44.9 %	-0.3618	0.8180
Household_Income__3	0.2366	0.2870	0.82	0.68	41.0 %	-0.3258	0.7990
Household_Income__4	0.2903	0.2769	1.05	1.10	29.5 %	-0.2525	0.8330
Household_Income__5	-0.0411	0.2734	-0.15	0.02	88.1 %	-0.5769	0.4947
Household_Income__6	0.2945	0.2741	1.07	1.15	28.3 %	-0.2427	0.8318
Household_Income__7	0.1821	0.2756	0.66	0.44	50.9 %	-0.3580	0.7222
Household_Income__8	-0.0978	0.2797	-0.35	0.12	72.7 %	-0.6461	0.4504
Household_Income__9	0.0600	0.2827	0.21	0.05	83.2 %	-0.4940	0.6141
Household_Income__10	0.0763	0.2869	0.27	0.07	79.0 %	-0.4860	0.6385
Household_Income__11	-0.1021	0.2896	-0.35	0.12	72.4 %	-0.6697	0.4655
Household_Income__12	-0.0654	0.3077	-0.21	0.05	83.2 %	-0.6686	0.5377

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
Household_Income__13	0.2571	0.4500	0.57	0.33	56.8%	-0.6250	1.1391
Number_of_Members_in_Household	0.2379	0.1629	1.46	2.13	14.4%	-0.0814	0.5572
Male_Head_Occupation__1	-0.4145	0.4152	-1.00	1.00	31.8%	-1.2283	0.3994
Male_Head_Occupation__2	-0.1892	0.0868	-2.18	4.75	2.9%	-0.3593	-0.0191
Male_Head_Occupation__3	-0.1073	0.0943	-1.14	1.29	25.5%	-0.2922	0.0776
Male_Head_Occupation__4	0.1770	0.2023	0.88	0.77	38.2%	-0.2195	0.5735
Male_Head_Occupation__5	-0.0145	0.1832	-0.08	0.01	93.7%	-0.3736	0.3446
Male_Head_Occupation__6	-0.0103	0.1072	-0.10	0.01	92.4%	-0.2204	0.1998
Male_Head_Occupation__7	-0.1545	0.0955	-1.62	2.62	10.6%	-0.3418	0.0327
Male_Head_Occupation__8	-0.2029	0.0994	-2.04	4.17	4.1%	-0.3977	-0.0081
Male_Head_Occupation__9	0.2180	0.1409	1.55	2.39	12.2%	-0.0582	0.4941
Male_Head_Race	0.1625	0.0937	1.73	3.01	8.3%	-0.0211	0.3461
Female_Head_Working_Status__0	0.0420	0.3135	0.13	0.02	89.3%	-0.5724	0.6565
Female_Head_Working_Status__1	0.0732	0.1620	0.45	0.20	65.1%	-0.2443	0.3908
Female_Head_Working_Status__2	-0.0076	0.1683	-0.05	0.00	96.4%	-0.3375	0.3223
Female_Head_Working_Status__3	0.0262	0.3757	0.07	0.00	94.4%	-0.7101	0.7626
Female_Head_Working_Status__4	0.9743	0.2956	3.30	10.86	0.1%	0.3949	1.5537
Female_Head_Working_Status__5	-0.3818	0.2815	-1.36	1.84	17.5%	-0.9335	0.1700
Female_Head_Working_Status__6	0.1596	0.1207	1.32	1.75	18.6%	-0.0769	0.3961
Female_Head_Occupation__2	0.1784	0.2193	0.81	0.66	41.6%	-0.2515	0.6082
Female_Head_Occupation__3	-0.0836	0.3361	-0.25	0.06	80.4%	-0.7423	0.5752
Female_Head_Occupation__4	0.7745	0.2921	2.65	7.03	0.8%	0.2021	1.3469
Female_Head_Occupation__5	0.0635	0.1433	0.44	0.20	65.8%	-0.2173	0.3443
Female_Head_Occupation__6	-0.1081	0.1279	-0.84	0.71	39.8%	-0.3588	0.1426
Female_Head_Occupation__7	0.1375	0.1582	0.87	0.76	38.5%	-0.1725	0.4475
Female_Head_Occupation__8	0.0392	0.1450	0.27	0.07	78.7%	-0.2451	0.3235
Female_Head_Occupation__9	0.2983	0.1962	1.52	2.31	12.9%	-0.0864	0.6829
Female_Head_Education__1	0.7546	1.1949	0.63	0.40	52.8%	-1.5874	3.0966
Female_Head_Education__2	0.8243	0.5427	1.52	2.31	12.9%	-0.2394	1.8880
Female_Head_Education__3	-0.2056	0.2087	-0.99	0.97	32.4%	-0.6146	0.2034
Female_Head_Education__4	0.0077	0.1965	0.04	0.00	96.9%	-0.3775	0.3930
Female_Head_Education__5	-0.3923	0.1703	-2.30	5.31	2.1%	-0.7260	-0.0585
Female_Head_Education__6	-0.2990	0.1827	-1.64	2.68	10.2%	-0.6572	0.0592
Female_Head_Education__7	-0.2791	0.1722	-1.62	2.63	10.5%	-0.6166	0.0584
Female_Head_Education__8	-0.2793	0.1677	-1.67	2.78	9.6%	-0.6080	0.0493
Female_Head_Education__9	-0.1880	0.1608	-1.17	1.37	24.3%	-0.5032	0.1272
Female_Head_Education__10	-0.0643	0.1965	-0.33	0.11	74.4%	-0.4495	0.3209
Washing_Machine	0.3804	0.2206	1.72	2.97	8.5%	-0.0520	0.8128
Clothes_Dryer	-0.2637	0.2237	-1.18	1.39	23.8%	-0.7021	0.1747
Dishwasher__0	-0.0296	0.0721	-0.41	0.17	68.2%	-0.1709	0.1117
Dishwasher__1	0.1830	0.1147	1.60	2.54	11.1%	-0.0418	0.4079
Freezer__0	-0.2252	0.0774	-2.91	8.47	0.4%	-0.3769	-0.0736
Freezer__1	-0.2838	0.1095	-2.59	6.72	1.0%	-0.4983	-0.0693

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
Food_Processor__0	-0.1681	0.0648	-2.60	6.74	0.9 %	-0.2951	-0.0412
Food_Processor__1	-0.3161	0.1006	-3.14	9.87	0.2 %	-0.5133	-0.1189
Coffee_Maker__0	-0.1452	0.0750	-1.93	3.74	5.3 %	-0.2923	0.0019
Coffee_Maker__1	0.0468	0.1022	0.46	0.21	64.7 %	-0.1536	0.2472
Garbage_Disposal__0	0.0434	0.0733	0.59	0.35	55.4 %	-0.1003	0.1872
Garbage_Disposal__1	-0.3347	0.1198	-2.79	7.81	0.5 %	-0.5694	-0.1000
Curling_Iron__0	0.1334	0.1037	1.29	1.66	19.8 %	-0.0699	0.3367
Curling_Iron__1	0.3702	0.1402	2.64	6.97	0.8 %	0.0954	0.6449
Hair_Rollers__0	0.0755	0.0663	1.14	1.30	25.5 %	-0.0545	0.2055
Hair_Rollers__1	0.1920	0.0980	1.96	3.84	5.0 %	0.0000	0.3840
Vacuum_Cleaner	-0.2565	0.1817	-1.41	1.99	15.8 %	-0.6126	0.0995

(*) 95 % confidence interval

Tabla J.17: Set atributos B. Resultado Modelo (1): Competencia en precios.

Estadístico	Valor
μ	-0.2
N	7117
Log-Likelihood	-3924.2
AIC	8036
BIC	8682
G statistic	2333.6
p-value(G statistic)	0
$\widetilde{R}^2$ (McFadden's)	11.67 %
$\widetilde{R}^2$ (McFadden's adjusted)	9.48 %
$\widetilde{R}^2$ (Cox & Snell)	13.00 %
$\widetilde{R}^2$ (Nagelkerke)	18.66 %

Tabla J.18: Set atributos B. Modelo (1), Estadísticos de Training.

Estadístico	Training	Test	Validation
N	7117	3008	3257
KS	34.58 %	29.93 %	40.60 %
AUC (ROC)	73.56 %	70.13 %	76.57 %
Cut-off (ROC)	0.3130	0.3130	0.3130
TP rate	67.05 %	62.44 %	85.65 %
FN rate	32.95 %	37.56 %	14.35 %
FP rate	32.22 %	33.24 %	49.72 %
TN rate	67.78 %	66.76 %	50.28 %
J de Youden	34.83 %	29.20 %	35.93 %
MSE	17.48 %	17.76 %	19.21 %

Tabla J.19: Set atributos B. Modelo (1), Estadísticos de Eficiencia por Base de datos.

J.3.3. Modelo propuesto (2)

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
constante	-1.7925	0.5378	-3.33	11.11	0.1 %	-2.8465	-0.7385
P_HUNTS_Unit_Extended_Price	-1.6989	0.0830	-20.46	418.79	0.0 %	-1.8616	-1.5362
P_HEINZ_Unit_Extended_Price	2.5430	0.0809	31.42	987.32	0.0 %	2.3844	2.7016
P_HUNTS_Unit_TV_Coupons	1.1640	0.0741	15.70	246.49	0.0 %	1.0187	1.3093
P_HEINZ_Unit_TV_Coupons	-2.0717	0.0758	-27.35	747.90	0.0 %	-2.2202	-1.9233
HEINZ_Dummy_CE	0.3775	0.1000	3.77	14.24	0.0 %	0.1814	0.5735
Dif_Unit_Extended_Price	-0.3121	0.1338	-2.33	5.44	2.0 %	-0.5743	-0.0498
Tipo_Baja	0.3393	0.0627	5.41	29.30	0.0 %	0.2165	0.4622
Tipo_Normal	0.2369	0.0580	4.08	16.68	0.0 %	0.1232	0.3506
Display	0.3218	0.0651	4.95	24.46	0.0 %	0.1943	0.4493
Advertising	-1.1663	0.2873	-4.06	16.48	0.0 %	-1.7294	-0.6032
Type	0.8503	0.2854	2.98	8.88	0.3 %	0.2910	1.4096
Coupon	0.4290	0.2775	1.55	2.39	12.2 %	-0.1149	0.9730
Cable_Status	-0.1016	0.0511	-1.99	3.96	4.7 %	-0.2017	-0.0015
Meter_Status	-0.1592	0.0481	-3.31	10.95	0.1 %	-0.2534	-0.0649
Number_of_Cats	0.7144	0.2586	2.76	7.63	0.6 %	0.2076	1.2212
Residence_Type__1	0.2870	0.4095	0.70	0.49	48.3 %	-0.5155	1.0896
Residence_Type__2	0.7883	0.4406	1.79	3.20	7.4 %	-0.0753	1.6518
Residence_Type__3	0.0912	0.3715	0.25	0.06	80.6 %	-0.6370	0.8194
Residence_Type__4	0.3414	0.4041	0.84	0.71	39.8 %	-0.4506	1.1334
Residence_Type__5	0.3612	0.4112	0.88	0.77	38.0 %	-0.4447	1.1672
Residence_Status	0.1775	0.1014	1.75	3.06	8.0 %	-0.0212	0.3763
Residence_Duration__1	0.1632	0.1687	0.97	0.94	33.4 %	-0.1676	0.4939
Residence_Duration__2	0.2849	0.1808	1.58	2.48	11.5 %	-0.0694	0.6391
Residence_Duration__3	0.1807	0.1117	1.62	2.62	10.6 %	-0.0382	0.3997

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
Residence_Duration__4	0.1340	0.0939	1.43	2.04	15.4%	-0.0500	0.3179
Residence_Duration__5	0.1422	0.0820	1.73	3.01	8.3%	-0.0185	0.3029
Residence_Duration__6	-0.1193	0.0886	-1.35	1.81	17.8%	-0.2930	0.0545
Residence_Duration__7	-0.1311	0.1015	-1.29	1.67	19.6%	-0.3300	0.0677
Household_Income__1	0.3959	0.3052	1.30	1.68	19.5%	-0.2024	0.9942
Household_Income__2	0.2783	0.2733	1.02	1.04	30.9%	-0.2573	0.8138
Household_Income__3	0.2872	0.2600	1.10	1.22	26.9%	-0.2225	0.7969
Household_Income__4	0.3375	0.2510	1.34	1.81	17.9%	-0.1545	0.8294
Household_Income__5	0.0254	0.2476	0.10	0.01	91.8%	-0.4599	0.5107
Household_Income__6	0.3390	0.2485	1.36	1.86	17.2%	-0.1480	0.8260
Household_Income__7	0.2254	0.2497	0.90	0.81	36.7%	-0.2641	0.7148
Household_Income__8	-0.0300	0.2530	-0.12	0.01	90.6%	-0.5259	0.4659
Household_Income__9	0.1306	0.2563	0.51	0.26	61.0%	-0.3717	0.6330
Household_Income__10	0.0939	0.2607	0.36	0.13	71.9%	-0.4170	0.6047
Household_Income__11	0.0058	0.2623	0.02	0.00	98.2%	-0.5083	0.5199
Household_Income__12	0.0567	0.2783	0.20	0.04	83.9%	-0.4889	0.6022
Household_Income__13	0.3207	0.4158	0.77	0.59	44.1%	-0.4943	1.1357
Number_of_Members_in_Household	0.1788	0.1489	1.20	1.44	23.0%	-0.1131	0.4707
Male_Head_Occupation__1	-0.3721	0.3761	-0.99	0.98	32.3%	-1.1091	0.3650
Male_Head_Occupation__2	-0.1316	0.0791	-1.66	2.77	9.6%	-0.2866	0.0234
Male_Head_Occupation__3	-0.0645	0.0864	-0.75	0.56	45.6%	-0.2337	0.1048
Male_Head_Occupation__4	0.1685	0.1839	0.92	0.84	36.0%	-0.1920	0.5290
Male_Head_Occupation__5	0.0271	0.1689	0.16	0.03	87.3%	-0.3040	0.3582
Male_Head_Occupation__6	0.0188	0.0990	0.19	0.04	84.9%	-0.1751	0.2127
Male_Head_Occupation__7	-0.1257	0.0872	-1.44	2.08	14.9%	-0.2966	0.0452
Male_Head_Occupation__8	-0.1561	0.0901	-1.73	3.00	8.3%	-0.3328	0.0205
Male_Head_Occupation__9	0.1949	0.1285	1.52	2.30	12.9%	-0.0568	0.4467
Male_Head_Race	0.1600	0.0850	1.88	3.54	6.0%	-0.0066	0.3266
Female_Head_Working_Status__0	0.0459	0.2889	0.16	0.03	87.4%	-0.5203	0.6121
Female_Head_Working_Status__1	0.0355	0.1488	0.24	0.06	81.1%	-0.2562	0.3272
Female_Head_Working_Status__2	-0.0192	0.1545	-0.12	0.02	90.1%	-0.3221	0.2836
Female_Head_Working_Status__3	0.0258	0.3550	0.07	0.01	94.2%	-0.6699	0.7215
Female_Head_Working_Status__4	0.7505	0.2697	2.78	7.74	0.5%	0.2218	1.2792
Female_Head_Working_Status__5	-0.2833	0.2518	-1.13	1.27	26.1%	-0.7767	0.2102
Female_Head_Working_Status__6	0.1517	0.1099	1.38	1.91	16.8%	-0.0637	0.3672
Female_Head_Occupation__2	0.1471	0.2009	0.73	0.54	46.4%	-0.2466	0.5408
Female_Head_Occupation__3	-0.1098	0.3071	-0.36	0.13	72.1%	-0.7116	0.4920
Female_Head_Occupation__4	0.6499	0.2683	2.42	5.87	1.5%	0.1240	1.1757
Female_Head_Occupation__5	0.0712	0.1321	0.54	0.29	59.0%	-0.1877	0.3300
Female_Head_Occupation__6	-0.0594	0.1181	-0.50	0.25	61.5%	-0.2909	0.1722
Female_Head_Occupation__7	0.1220	0.1454	0.84	0.70	40.2%	-0.1630	0.4069
Female_Head_Occupation__8	0.0649	0.1338	0.49	0.24	62.8%	-0.1973	0.3270
Female_Head_Occupation__9	0.2562	0.1798	1.43	2.03	15.4%	-0.0961	0.6086

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
Female_Head_Education__1	0.5821	1.0915	0.53	0.28	59.4%	-1.5572	2.7214
Female_Head_Education__2	0.5226	0.4989	1.05	1.10	29.5%	-0.4551	1.5003
Female_Head_Education__3	-0.1871	0.1933	-0.97	0.94	33.3%	-0.5659	0.1917
Female_Head_Education__4	-0.0642	0.1812	-0.35	0.13	72.3%	-0.4193	0.2908
Female_Head_Education__5	-0.3751	0.1576	-2.38	5.67	1.7%	-0.6841	-0.0662
Female_Head_Education__6	-0.2620	0.1689	-1.55	2.41	12.1%	-0.5929	0.0689
Female_Head_Education__7	-0.2731	0.1592	-1.72	2.94	8.6%	-0.5852	0.0389
Female_Head_Education__8	-0.2726	0.1554	-1.75	3.07	8.0%	-0.5772	0.0321
Female_Head_Education__9	-0.1910	0.1491	-1.28	1.64	20.0%	-0.4832	0.1012
Female_Head_Education__10	-0.0904	0.1813	-0.50	0.25	61.8%	-0.4458	0.2649
Washing_Machine	0.3399	0.1977	1.72	2.96	8.6%	-0.0476	0.7274
Clothes_Dryer	-0.2128	0.2008	-1.06	1.12	28.9%	-0.6064	0.1809
Dishwasher__0	-0.0405	0.0653	-0.62	0.39	53.5%	-0.1684	0.0874
Dishwasher__1	0.1366	0.1054	1.30	1.68	19.5%	-0.0700	0.3432
Freezer__0	-0.1775	0.0697	-2.54	6.47	1.1%	-0.3141	-0.0408
Freezer__1	-0.2262	0.0999	-2.26	5.13	2.4%	-0.4220	-0.0304
Food_Processor__0	-0.1411	0.0589	-2.40	5.74	1.7%	-0.2565	-0.0257
Food_Processor__1	-0.2668	0.0924	-2.89	8.34	0.4%	-0.4479	-0.0858
Coffee_Maker__0	-0.1072	0.0679	-1.58	2.50	11.4%	-0.2402	0.0258
Coffee_Maker__1	0.0738	0.0933	0.79	0.63	42.9%	-0.1091	0.2568
Garbage_Disposal__0	0.0121	0.0662	0.18	0.03	85.5%	-0.1176	0.1419
Garbage_Disposal__1	-0.3073	0.1099	-2.80	7.82	0.5%	-0.5227	-0.0919
Curling_Iron__0	0.1142	0.0950	1.20	1.45	22.9%	-0.0719	0.3003
Curling_Iron__1	0.3201	0.1302	2.46	6.04	1.4%	0.0648	0.5754
Hair_Rollers__0	0.0548	0.0601	0.91	0.83	36.2%	-0.0629	0.1725
Hair_Rollers__1	0.1814	0.0895	2.03	4.10	4.3%	0.0059	0.3568
Vacuum_Cleaner	-0.2914	0.1660	-1.76	3.08	7.9%	-0.6168	0.0340

(*) 95% confidence interval

Tabla J.20: Set atributos B. Resultado Modelo (2): Competencia en precios y descuentos.

Estadístico	Valor
μ	-0.2
N	7117
Log-Likelihood	-4288.3
AIC	8765
BIC	9410
G statistic	3294.5
p-value(G statistic)	0
$\widetilde{R}^2$ (McFadden's)	9.58 %
$\widetilde{R}^2$ (McFadden's adjusted)	7.39 %
$\widetilde{R}^2$ (Cox & Snell)	10.81 %
$\widetilde{R}^2$ (Nagelkerke)	15.51 %

Tabla J.21: Set atributos B. Modelo (2), Estadísticos de Training.

Estadístico	Training	Test	Validation
N	7117	3008	3257
KS	31.59 %	27.53 %	44.34 %
AUC (ROC)	71.70 %	68.89 %	78.66 %
Cut-off (ROC)	0.3360	0.3360	0.3360
TP rate	65.96 %	61.95 %	82.61 %
FN rate	34.04 %	38.05 %	17.39 %
FP rate	34.28 %	35.15 %	42.03 %
TN rate	65.72 %	64.85 %	57.97 %
J de Youden	31.68 %	26.80 %	40.58 %
MSE	18.00 %	18.17 %	18.42 %

Tabla J.22: Set atributos B. Modelo (2), Estadísticos de Eficiencia por Base de datos.

J.3.4. Modelo propuesto (3)

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
constante	-1.7892	0.5367	-3.33	11.11	0.1 %	-2.8412	-0.7372
P_HUNTS_Unit_Extended_Price	-1.7067	0.0828	-20.61	424.69	0.0 %	-1.8690	-1.5443
P_HEINZ_Unit_Extended_Price	2.5317	0.0808	31.33	981.39	0.0 %	2.3733	2.6901
P_HUNTS_Unit_TV_Coupons	1.1604	0.0740	15.68	246.00	0.0 %	1.0154	1.3054
P_HEINZ_Unit_TV_Coupons	-2.0797	0.0756	-27.51	756.66	0.0 %	-2.2279	-1.9315
HUNTS_Dummy_CE	0.4729	0.0786	6.02	36.23	0.0 %	0.3189	0.6269
HEINZ_Dummy_CE	-0.1717	0.0673	-2.55	6.51	1.1 %	-0.3036	-0.0398
Dif_Unit_Extended_Price	-0.2945	0.1333	-2.21	4.88	2.7 %	-0.5557	-0.0332

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
Tipo_Baja	0.3338	0.0625	5.34	28.49	0.0%	0.2112	0.4563
Tipo_Normal	0.2342	0.0578	4.05	16.41	0.0%	0.1209	0.3475
Display	0.3178	0.0649	4.90	23.98	0.0%	0.1906	0.4450
Advertising	-1.1722	0.2864	-4.09	16.75	0.0%	-1.7335	-0.6108
Type	0.8555	0.2845	3.01	9.04	0.3%	0.2978	1.4132
Coupon	0.4194	0.2767	1.52	2.30	13.0%	-0.1230	0.9617
Cable_Status	-0.1018	0.0509	-2.00	4.00	4.6%	-0.2016	-0.0020
Meter_Status	-0.1345	0.0477	-2.82	7.95	0.5%	-0.2279	-0.0410
Number_of_Cats	0.6957	0.2580	2.70	7.27	0.7%	0.1900	1.2014
Residence_Type__1	0.2981	0.4091	0.73	0.53	46.6%	-0.5038	1.0999
Residence_Type__2	0.7916	0.4403	1.80	3.23	7.2%	-0.0714	1.6545
Residence_Type__3	0.1008	0.3712	0.27	0.07	78.6%	-0.6268	0.8284
Residence_Type__4	0.3456	0.4038	0.86	0.73	39.2%	-0.4458	1.1370
Residence_Type__5	0.3626	0.4109	0.88	0.78	37.8%	-0.4428	1.1679
Residence_Status	0.1830	0.1012	1.81	3.27	7.1%	-0.0154	0.3813
Residence_Duration__1	0.1580	0.1686	0.94	0.88	34.9%	-0.1723	0.4884
Residence_Duration__2	0.2894	0.1805	1.60	2.57	10.9%	-0.0645	0.6432
Residence_Duration__3	0.1846	0.1115	1.65	2.74	9.8%	-0.0340	0.4032
Residence_Duration__4	0.1400	0.0935	1.50	2.24	13.5%	-0.0433	0.3233
Residence_Duration__5	0.1507	0.0818	1.84	3.40	6.5%	-0.0096	0.3111
Residence_Duration__6	-0.1078	0.0884	-1.22	1.49	22.3%	-0.2811	0.0655
Residence_Duration__7	-0.1241	0.1012	-1.23	1.50	22.0%	-0.3224	0.0743
Household_Income__1	0.4002	0.3044	1.31	1.73	18.9%	-0.1964	0.9967
Household_Income__2	0.2869	0.2725	1.05	1.11	29.2%	-0.2472	0.8209
Household_Income__3	0.2951	0.2593	1.14	1.29	25.5%	-0.2132	0.8033
Household_Income__4	0.3399	0.2502	1.36	1.85	17.4%	-0.1505	0.8304
Household_Income__5	0.0258	0.2469	0.10	0.01	91.7%	-0.4580	0.5096
Household_Income__6	0.3439	0.2477	1.39	1.93	16.5%	-0.1416	0.8294
Household_Income__7	0.2256	0.2490	0.91	0.82	36.5%	-0.2624	0.7136
Household_Income__8	-0.0253	0.2523	-0.10	0.01	92.0%	-0.5197	0.4691
Household_Income__9	0.1282	0.2555	0.50	0.25	61.6%	-0.3726	0.6289
Household_Income__10	0.0892	0.2600	0.34	0.12	73.2%	-0.4203	0.5987
Household_Income__11	0.0105	0.2615	0.04	0.00	96.8%	-0.5019	0.5230
Household_Income__12	0.0603	0.2777	0.22	0.05	82.8%	-0.4840	0.6046
Household_Income__13	0.2890	0.4158	0.70	0.48	48.7%	-0.5259	1.1040
Number_of_Members_in_Household	0.1749	0.1485	1.18	1.39	23.9%	-0.1162	0.4660
Male_Head_Occupation__1	-0.3774	0.3748	-1.01	1.01	31.4%	-1.1121	0.3573
Male_Head_Occupation__2	-0.1381	0.0789	-1.75	3.06	8.0%	-0.2927	0.0165
Male_Head_Occupation__3	-0.0591	0.0861	-0.69	0.47	49.2%	-0.2279	0.1096
Male_Head_Occupation__4	0.1632	0.1831	0.89	0.79	37.3%	-0.1957	0.5221
Male_Head_Occupation__5	0.0300	0.1684	0.18	0.03	85.9%	-0.3002	0.3601
Male_Head_Occupation__6	0.0113	0.0987	0.11	0.01	90.9%	-0.1821	0.2047
Male_Head_Occupation__7	-0.1252	0.0869	-1.44	2.08	15.0%	-0.2956	0.0451

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
Male_Head_Occupation__8	-0.1583	0.0899	-1.76	3.10	7.8%	-0.3345	0.0179
Male_Head_Occupation__9	0.1734	0.1282	1.35	1.83	17.6%	-0.0779	0.4247
Male_Head_Race	0.1682	0.0847	1.99	3.94	4.7%	0.0022	0.3343
Female_Head_Working_Status__0	0.0332	0.2882	0.12	0.01	90.8%	-0.5317	0.5982
Female_Head_Working_Status__1	0.0376	0.1485	0.25	0.06	80.0%	-0.2534	0.3286
Female_Head_Working_Status__2	-0.0198	0.1542	-0.13	0.02	89.8%	-0.3220	0.2824
Female_Head_Working_Status__3	0.0270	0.3536	0.08	0.01	93.9%	-0.6661	0.7200
Female_Head_Working_Status__4	0.7487	0.2689	2.78	7.75	0.5%	0.2217	1.2757
Female_Head_Working_Status__5	-0.3122	0.2515	-1.24	1.54	21.5%	-0.8050	0.1807
Female_Head_Working_Status__6	0.1493	0.1097	1.36	1.85	17.3%	-0.0656	0.3643
Female_Head_Occupation__2	0.1478	0.2006	0.74	0.54	46.1%	-0.2454	0.5409
Female_Head_Occupation__3	-0.1030	0.3051	-0.34	0.11	73.6%	-0.7009	0.4950
Female_Head_Occupation__4	0.6221	0.2680	2.32	5.39	2.0%	0.0968	1.1474
Female_Head_Occupation__5	0.0603	0.1317	0.46	0.21	64.7%	-0.1979	0.3184
Female_Head_Occupation__6	-0.0598	0.1178	-0.51	0.26	61.2%	-0.2908	0.1712
Female_Head_Occupation__7	0.1150	0.1451	0.79	0.63	42.8%	-0.1694	0.3993
Female_Head_Occupation__8	0.0611	0.1334	0.46	0.21	64.7%	-0.2003	0.3226
Female_Head_Occupation__9	0.2406	0.1793	1.34	1.80	18.0%	-0.1108	0.5920
Female_Head_Education__1	0.5607	1.0897	0.51	0.26	60.7%	-1.5750	2.6964
Female_Head_Education__2	0.5030	0.4983	1.01	1.02	31.3%	-0.4737	1.4796
Female_Head_Education__3	-0.1926	0.1928	-1.00	1.00	31.8%	-0.5704	0.1852
Female_Head_Education__4	-0.0635	0.1807	-0.35	0.12	72.5%	-0.4177	0.2906
Female_Head_Education__5	-0.3741	0.1572	-2.38	5.66	1.7%	-0.6822	-0.0660
Female_Head_Education__6	-0.2604	0.1684	-1.55	2.39	12.2%	-0.5905	0.0697
Female_Head_Education__7	-0.2751	0.1588	-1.73	3.00	8.3%	-0.5863	0.0361
Female_Head_Education__8	-0.2740	0.1550	-1.77	3.12	7.7%	-0.5778	0.0298
Female_Head_Education__9	-0.1893	0.1487	-1.27	1.62	20.3%	-0.4807	0.1021
Female_Head_Education__10	-0.0834	0.1808	-0.46	0.21	64.5%	-0.4377	0.2709
Washing_Machine	0.3475	0.1973	1.76	3.10	7.8%	-0.0393	0.7342
Clothes_Dryer	-0.2189	0.2004	-1.09	1.19	27.5%	-0.6116	0.1738
Dishwasher__0	-0.0329	0.0651	-0.51	0.26	61.3%	-0.1604	0.0947
Dishwasher__1	0.1544	0.1051	1.47	2.16	14.2%	-0.0515	0.3604
Freezer__0	-0.1771	0.0696	-2.54	6.48	1.1%	-0.3135	-0.0407
Freezer__1	-0.2282	0.0997	-2.29	5.24	2.2%	-0.4236	-0.0329
Food_Processor__0	-0.1402	0.0587	-2.39	5.69	1.7%	-0.2553	-0.0250
Food_Processor__1	-0.2687	0.0921	-2.92	8.51	0.4%	-0.4493	-0.0882
Coffee_Maker__0	-0.1117	0.0677	-1.65	2.72	9.9%	-0.2444	0.0209
Coffee_Maker__1	0.0703	0.0932	0.75	0.57	45.1%	-0.1124	0.2530
Garbage_Disposal__0	0.0183	0.0660	0.28	0.08	78.1%	-0.1111	0.1478
Garbage_Disposal__1	-0.3125	0.1097	-2.85	8.12	0.4%	-0.5274	-0.0976
Curling_Iron__0	0.1108	0.0948	1.17	1.37	24.3%	-0.0750	0.2965
Curling_Iron__1	0.3255	0.1300	2.50	6.27	1.2%	0.0706	0.5804
Hair_Rollers__0	0.0540	0.0599	0.90	0.81	36.7%	-0.0633	0.1714

Atributo	Coef.	StDev	Z	Wald	p-value	Lower (*)	Upper (*)
Hair_Rollers__1	0.1793	0.0892	2.01	4.04	4.5 %	0.0044	0.3542
Vacuum_Cleaner	-0.2911	0.1658	-1.76	3.08	7.9 %	-0.6161	0.0339

(*) 95 % confidence interval

Tabla J.23: Set atributos B. Resultado Modelo (3): Competencia en precios, descuentos y marketing.

Estadístico	Valor
μ	-0.2
e	0.1
N	7117
Log-Likelihood	-4324.8
AIC	8840
BIC	9492
G statistic	3278.5
p-value(G statistic)	0
$\widetilde{R}^2$ (McFadden's)	9.27 %
$\widetilde{R}^2$ (McFadden's adjusted)	7.06 %
$\widetilde{R}^2$ (Cox & Snell)	10.48 %
$\widetilde{R}^2$ (Nagelkerke)	15.03 %

Tabla J.24: Set atributos B. Modelo (3), Estadísticos de Training.

Estadístico	Training	Test	Validation
N	7117	3008	3257
KS	31.58 %	26.47 %	40.90 %
AUC (ROC)	71.42 %	68.40 %	78.06 %
Cut-off (ROC)	0.3410	0.3410	0.3410
TP rate	64.77 %	59.48 %	75.59 %
FN rate	35.23 %	40.52 %	24.41 %
FP rate	33.10 %	34.02 %	35.09 %
TN rate	66.90 %	65.98 %	64.91 %
J de Youden	31.67 %	25.46 %	40.50 %
MSE	18.88 %	18.61 %	19.75 %

Tabla J.25: Set atributos B. Modelo (3), Estadísticos de Eficiencia por Base de datos.