



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

EFFECTO DE DISPONIBILIDAD DE VARIEDAD DE PRODUCTOS EN
GÓNDOLA EN EL COMPORTAMIENTO DE CLIENTES

TESIS PARA OPTAR AL GRADO DE MAGISTER EN GESTIÓN DE
OPERACIONES

JACQUELINE CECILIA ELIZABETH VON HAUSEN CÁRDENAS

PROFESOR GUÍA:

RICARDO MONTOYA MOREIRA

MIEMBROS DE LA COMISIÓN:

MARCELO OLIVARES ACUÑA

ANDRÉS MUSALEM SAID

ARIEL SCHILKRUT COGAN

SANTIAGO, CHILE

2014

RESUMEN DE LA TESIS
PARA OPTAR AL TITULO DE
INGENIERA CIVIL INDUSTRIAL
POR: JACQUELINE VON HAUSEN
FECHA: 22/05/2014
PROF. RICARDO MONTOYA MOREIRA

EFFECTO DE DISPONIBILIDAD DE VARIEDAD DE PRODUCTOS EN GÓNDOLA EN EL COMPORTAMIENTO DE CLIENTES

Tanto para los clientes como para las empresas la falta de productos en góndola (o quiebres de *stock*) produce costos indeseados muy elevados. Éstos, en su mayoría, se generan por falta de información, errores en el pronóstico de ventas y problemas en el sistema de reposición de productos. Para las empresas de retail es necesario entregar una buena experiencia de compra ofreciendo productos y servicios adecuados a las necesidades de los consumidores dada la alta competencia existente. Por esta razón es importante cuantificar y estudiar el efecto que los problemas en la disponibilidad de los productos generan sobre los clientes.

El objetivo principal de este trabajo es estudiar el efecto de la variedad de surtido en góndola en el comportamiento de los clientes mediante la incorporación de variables operacionales de *stock* a un modelo de incidencia de compra. Para ello, se utilizan datos transaccionales de los clientes y datos operacionales de quiebres de stock –medidos mediante fotografías en intervalos fijos de tiempo– de los distintos tipos de pan ofrecidos por una tienda de una cadena de supermercados norteamericana.

En primer lugar, se realiza un análisis agregado por clientes para los principales productos del supermercado para estudiar la relación entre las ventas de diferentes categorías de panes y los quiebres de *stock*, donde los resultados entregan evidencia preliminar del efecto agregado de la disponibilidad.

Por lo anterior, se realiza un análisis a nivel desagregado por clientes, para el cual se utiliza un modelo *mixed logit*, que captura la heterogeneidad de los clientes. En este modelo, se estudian los patrones de sustitución entre los productos y los efectos cruzados entre las categorías de productos. Los resultados muestran una disminución en la incidencia de compra general de panes con quiebres de *stock* de *rolls* de la categoría de un 0,91% y un efecto de sustitución entre los productos de un 0,19%.

Finalmente, con los resultados obtenidos, se estudia el efecto de los quiebres de *stock* de panes dentro de otros departamentos del supermercado –en particular, se estudia el departamento *Produce*–. Para esto, se utilizan variables instrumentales, que corresponden a las probabilidades de compra estimadas de los panes de manera agregada o de las categorías de panes. Los efectos encontrados para estos dos modelos son distintos, en donde los productos empaquetados toman mayor protagonismo sobre los de venta a granel en las compras.

.

.

.

.

.

.

.

.

.

.

..

....

A mis padres, Gladys y Hernán

Tabla de contenido

1. Introducción	1
1.1. Antecedentes Generales	1
1.2. Los quiebres de <i>stock</i>	3
1.2.1. Definición	3
1.2.2. Los contras (o costos) de los quiebres de stock	4
1.2.3. Causas y reacciones a los quiebres de stock	5
1.3. Descripción del problema	5
1.4. Motivación	6
1.5. Objetivos y Alcances	7
1.5.1. Objetivo General	8
1.5.2. Objetivos Específicos	8
1.5.3. Alcances	8
2. Hipótesis y Revisión bibliográfica	9
2.1. Efectos a capturar	9
2.2. Marco metodológico	11
2.2.1. Modelos de conteo	11
2.2.1.1. Modelos de regresión con datos de panel	12
2.2.2. Modelos de elección discreta	13
2.2.2.1. Modelos de utilidad aleatoria	14
2.2.2.2. Modelo Logit	15
2.2.2.3. Random Coefficient Logit	17
2.2.3. Modelos de panel	19
2.2.3.1. Modelos de panel con variables instrumentales	19
3. Análisis agregado de clientes	21
3.1. Preámbulo	21
3.2. Recopilación de información	25
3.3. Descripción de las variables del modelo	25
3.3.1. Variables dependientes	25
3.3.2. Variables de Stock	26
3.3.3. Variables de Estacionalidad	27

3.4. Modelos	28
3.5. Categoría 1: Los principales 2 productos	29
3.6. Categoría 2: Los principales 5 productos	31
3.7. Categoría 3: Rolls envasados v/s a granel	33
4. Modelo individual por cliente	36
4.1. Motivación	36
4.2. Conjunto de clientes a considerar	36
4.3. Definición de productos y categorías	38
4.4. Variables del modelo	42
4.5. Modelo utilizado	46
4.6. Resultados del Modelo	47
4.6.1. Consideraciones del modelo	49
4.6.2. Discusión de los resultados	49
4.7. Análisis de contrafactuales	52
4.7.1. Conjunto a elección de la tupla (p, oos, d, h)	52
4.7.2. Resultados del análisis	53
5. Análisis otras categorías	56
5.1. Segmentación de las otras categorías	56
5.2. Departamentos en estudio	59
5.3. Variables	60
5.4. Modelos utilizados	62
5.5. Resultados obtenidos	63
5.6. Análisis de contrafactuales	64
6. Conclusiones generales	67
6.1. Conclusiones	67
Bibliografía	70
7. Apéndice	72
7.1. Apéndice 1: Fluctuaciones de las ventas de productos y pan e incidencia de pan	72
7.2. Apéndice 2: Resultados obtenidos para los modelos del capítulo de datos agregados por clientes	73

Índice de figuras

1.1.	Importancia del retail en la industria estadounidense	1
1.2.	Causas de los quiebres de stock	5
1.3.	Respuesta de los consumidores a nivel mundial	5
1.4.	Capturas de canasta del pan a distintas horas del día	7
3.1.	Estacionalidades por hora del día	27
3.2.	Histogramas	29
4.1.	Histograma de la incidencia de compra de pan	41
4.2.	Análisis delta participaciones de mercado en el día miércoles en bloque horario H_1	54
5.1.	Para modelo (a)	65
5.2.	Análisis de contrafactuales para modelos estudiados	65
7.1.	Nº Productos vendidos y Cantidad de dólares recaudado v/s horas del día	72
7.2.	Nº de panes vendidos v/s horas del día	72
7.3.	Incidencia de productos v/s horas del día	73

Índice de tablas

3.1. Participación de mercado de los productos en la categoría de pan	22
3.2. Nombre y descripción de los productos agregados	23
3.3. Porcentaje de Stock por productos Total y por intervalos de tiempo	24
3.4. Stock desagregado por horas del último intervalo de tiempo	24
3.5. Definición de variables dependientes	26
3.6. Definición de variables de <i>stock</i>	26
3.7. Definición de variables de estacionalidad	28
3.8. Resultados modelos para variable de ventas y 2 productos	30
3.9. Variación porcentual de pasar de nivel de <i>stock</i> del 100 % a nivel 0 %	30
3.10. Análisis para modelo (2) con efectos cuadráticos	31
3.11. Resultados modelo para variable de ventas para la categoría 2	32
3.12. Variación porcentual al pasar de nivel de <i>stock</i> actual a nivel del 0 %	33
3.13. Resultados modelos para variable de ventas	34
3.14. Variación porcentual al pasar de nivel de <i>stock</i> actual al 0 %	34
4.1. Variables comportamiento de compra clientes	37
4.2. Máximo de panes comprados por boleta, agregado por cliente	38
4.3. Estadísticos diferenciado por quienes han o no comprado los <i>top</i> 10 productos	39
4.4. Participación de mercado de los productos más comprados	40
4.5. Percentiles para variable <i>recency</i>	45
4.6. Resultados de las variables de control del modelo <i>random coefficient logit</i>	47
4.7. Resultados de variables en estudio modelo <i>random coefficient logit</i>	48
5.1. Cantidad de PLU's por boleta para distintos conjuntos de compra de pan	58
5.2. Incidencia de compra por categoría para distintos conjuntos de compra de pan	59
5.3. Resultados primera regresión modelo (a)	63
5.4. Resultados segunda regresión modelo (a)	63
5.5. Resultados primeras regresiones modelo (b)	63
5.6. Resultados segunda regresión modelo (b)	64
7.1. Resultados modelos para variable de ventas y 2 productos	73
7.2. Resultados modelos para variable de incidencia y categoría 2 productos	74
7.3. Resultados modelos para variable de ventas y categoría 5 productos	75

7.4. Resultados modelos para variable de incidencia y categoría 5 productos	76
7.5. Resultados modelos para variable de ventas y categoría <i>rolls</i>	77
7.6. Resultados modelos para variable de incidencia y categoría <i>rolls</i>	77

Capítulo 1

Introducción

1.1. Antecedentes Generales

La definición de retail corresponde a toda actividad que está directamente relacionada con la venta de bienes y servicios para el consumidor final para uso personal, no comercial. (Lamb et al. (2009)). Éste es un eje importante dentro de la economía. Para el caso de la industria estadounidense, el retail forma parte importante de la generación de empleo por la gran proporción de empresas de este rubro; sin embargo, se debe destacar su importancia dentro del Producto Interno Bruto (PIB). Tal como menciona Lamb et al. (2009) las tiendas de retail registran casi \$4 mil millones de ventas anualmente, aproximadamente 40 % del PIB. Esto se puede ver en la Figura 1.1.

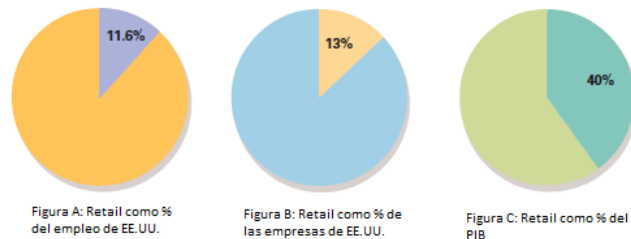


Figura 1.1: Importancia del retail en la industria estadounidense

Fuente: *Elaboración propia a partir de Lamb et al. (2009), pág. 360*

Los retailers se pueden clasificar de distintas maneras. La clasificación que se muestra a continuación busca abarcar en un limitado número de grupos la diversidad dentro de éste:

1. Tiendas de especialidad: Estas tiendas definen un método de operaciones de retail, especializadas en dar un tipo de mercadería. Tiene un surtido más estrecho, sin embargo, más profundo.
2. Supermercados: Son grandes, separados en departamentos, con sistema autoservicio y especializados en comida, con algunos items que no corresponden a comida. Por requerimiento de los consumidores de encontrar todo lo que buscan en una sola parada de compra, los supermercados tradicionales han sido reemplazados por grandes *superstores* que incluyen gran variedad de items que no corresponden a comida, además de algunos servicios utilizados por los consumidores.

3. Farmacias: Productos farmacéuticos y servicios como su principal atractivo. Cuenta además con una extensa selección de medicamentos sin receta, cosméticos, de salud y belleza, mercancía de temporada, artículos especializados tales como algunos alimentos de conveniencia no refrigerados.
4. Tiendas de conveniencia: Se define como un supermercado miniatura, con una variedad limitada de artículos de gran rotación, de servicio rápido. Típicamente se localizan en áreas residenciales. En general, los precios son mayores que en supermercados.
5. Tiendas de descuento: Es un retailer que compite en la base de precios bajos, alta rotación y gran volumen. Existen 3 tipos:
 - **De línea completa**: Ofrecen a los consumidores un servicio limitado y una amplia variedad de conocidos productos no perecederos. Los *supercenters* entran dentro de esta categoría.
 - **De especialidad**: Ofrecen una selección altamente completa de mercancía de un solo tipo y utiliza precio bajos, alto volumen y alta rotación.
 - **Off-price** : Venden a precios muy por debajo de los precios tradicionales de tiendas por departamento, ya que paga en efectivo por sus productos y, por lo general, no permite devolución. Un ejemplo particular de éstos son los *outlet* que son propiedad del fabricante. En general se encuentran en la periferia de las ciudades para no canibalizar sus productos.
6. Restaurant: Estos se sitúan en la línea entre los establecimientos minoristas y establecimientos de servicios. Venden productos tangibles -comidas y bebidas-; sin embargo, cuentan con un servicio en la forma en la cual los preparan y sirven. Es muy competitivo dadas las bajas barreras de entrada.

Existe una categoría de retail que se basa en compras sin visitar alguna tienda. Esta se puede clasificar de la siguiente manera:

1. Automatic Vending: Es el tipo de retail más generalizado de este tipo en los Estados Unidos
2. Direct Retailing: Venta de productos puerta a puerta, ejemplo: Avon.
3. Electronic Retailing - Venta por televisión: En pantalla se muestra la mercancía, con el precio de venta, a los televidentes. Éstos pueden comprar a través del teléfono y pagar con tarjeta de crédito.
4. Online Retailing: o e-tailing, es un tipo de compras disponible para los consumidores con acceso a internet. Éstos puede comprar sin salir de la casa, escoger una amplia gama discount retailersde mercancías, usar servicio de comparación de compra para buscar el mejor precio en la web, y luego recibir los items en una dirección indicada.

1.2. Los quiebres de *stock*

1.2.1. Definición

Según Gruen and Corsten (2007), existen dos conceptos previos que deben ser descritos antes de entrar en profundidad en la definición de quiebres de stock, para no entrar en ambigüedades:

- El evento de quiebre de *stock*. Este evento ocurre si el retailer pretende tener un item para la venta, pero no existe presencia física de una unidad vendible en la góndola. Este evento comienza cuando la última unidad de un SKU es removida de la góndola y empieza cuando alguna unidad vendible es repuesta en góndola.
- Ratio de quiebre de *stock*. Los atributos de una condición de quiebre de stock hacen alusión a aspectos de los eventos de quiebre de stock que pueden ser medidos y calculados como un ratio de quiebre de stock.
 1. Número de ocurrencias en cierto tiempo: Número de eventos de quiebre de stock para un item sobre una unidad de tiempo.
 2. Duración: Para un periodo fijo es tiempo total en el cual el item está “quebrado” dividido por el tiempo total de venta disponible. Esto mide el tiempo de ventas perdidas para el item en cuestión.
 3. Disponibilidad en góndola: Probabilidad que el cliente encuentre el item cuando entra a la tienda. Se calcula como $1 - (\text{ratio de duración del quiebre de } stock)$.
 4. Ventas perdidas: Ocurren cuando un cliente que quiere comprar un item que se encuentra a la venta, no encuentra en el lugar que debería encontrarse y, por lo tanto, no puede comprarlo. Se puede medir tanto las unidades de ventas perdidas como las ventas perdidas monetarias por periodo de tiempo.
 5. Número de clientes impactados: Se mide como $1 - (\% \text{ de canastas de compra en las cuales un item aparecería si éste estuviera estado disponible durante la medición})$.

Los quiebres de *stock* o *out-of-stock* (OOS) se miden de 3 formas típicas:

1. Auditoría de inventario físico
2. Análisis de los datos de Punto de Venta (PdV)
3. Uso de datos de inventario perpetuo (IP)

Éstos miden distintos aspectos y entregan distintos ratios de OOS. Las ambigüedades surgen debido a las diferencias en cómo son contadas:

- (a) Instancias
- (b) Ventas perdidas, en unidades
- (c) Ventas perdidas monetarias

En general, en las auditorías y en los análisis de inventario perpetuo cuentan instancias, observadas en un punto en el tiempo (para las primeras) y en un intervalo de tiempo (para las segundas). Los análisis de Punto de Venta (PdV) pueden ofrecer una amplia gama de medidas que incluyen (a), (b) y (c). Cuando ocurre un evento de quiebre de stock, no se incurre en ventas perdidas si el ítem es repuesto en la góndola antes de que afecte al próximo cliente que quiere comprar ese ítem.

1.2.2. Los contras (o costos) de los quiebres de stock

Para el retailer o los proveedores. El nivel promedio de *out-of-stock* mundial en la industria del retail para los bienes de consumo de rápido movimiento es 8 %. Gruen et al. (2002). En este mismo estudio, se puede observar que el costo de los quiebres de stock para los retailers alcanza un 4 % de las ventas, lo que se traduce en un 4 % menos de ganancias por ítem.

Algunas de las razones involuntarias por las cuales se sigue exponiendo a los clientes a niveles de quiebre de stock:

- Distorsionan la demanda real: disminuye la certeza de la predicción y orden de compra
- Proliferación de las promociones. Generalmente ante la insistencia de los proveedores.
- Se incrementan los costos operacionales debido al personal que busca los ítems con quiebre de stock en las bodegas, reposición no planeada de los ítems, etc.
- Presión en la reducción de costos de personal conlleva a oferta de mano de obra insuficiente.

Para los clientes. Algunos de los costos indirectos en los que incurre el cliente al encontrarse con quiebres de stock son los siguientes:

- Pérdida de tiempo incurrido tanto en buscar el producto por sí mismos y/o como por buscar personal para que lo ayude en esta tarea, que puede ser utilizado más productivamente en comprar (se estima que alcanza un 20 % del tiempo promedio de una ocasión de compra) Gruen and Corsten (2007).
- Aumento de requisitos en la toma de decisión.
- Costos psicológicos, como por ejemplo, menor confianza por tener que usar un sustituto no probado con anterioridad.
- Pérdida directa de lealtad a la tienda.
- Disminución de la satisfacción del consumidor.
- Fomenta compra en las tiendas de la competencia.
- Pérdida de los clientes por cambio a la competencia.

1.2.3. Causas y reacciones a los quiebres de stock

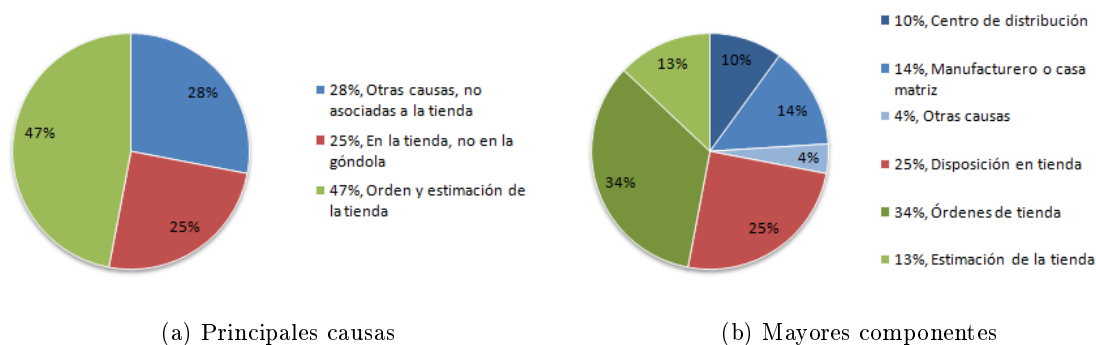


Figura 1.2: Causas de los quiebres de stock

Según entre el 70 % a 75 % de los quiebres de stock se producen directamente por prácticas directas de la tienda (ya sea subestimando la demanda o habiendo ordenando ciclos demasiado largos) y reposicionar productos en las góndolas (que están en tienda pero no en góndola).

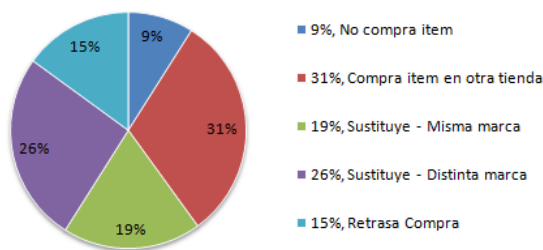


Figura 1.3: Respuesta de los consumidores a nivel mundial

En este mismo estudio se revisa el comportamiento de más de 71.000 clientes, donde se encuentra que los retailers están propensos a perder aproximadamente la mitad de las posibles compras cuando un cliente se encuentra frente a la falta de un producto. En esta pérdida no se encuentra el impacto de sustituir, que en general suele ser por un producto más barato.

1.3. Descripción del problema

A lo largo del tiempo se ha estudiado el comportamiento de compra de los individuos en las salas de supermercado utilizando datos transaccionales de venta, sobre los cuales se puede trabajar y encontrar resultados que conllevan a beneficios tanto para los clientes como para los *retailers*. Estos beneficios se pueden traducir tanto en ahorros monetarios para la empresa como aumento de calidad del servicio para los clientes.

Tal como menciona la Teoría del *Consumidor*, la competencia es cada vez más fuerte, por lo tanto, para estar en una constante renovación y prestar mejores servicios es imperioso conocer las necesidades de los clientes. Esto contribuye, a su vez, a aumentar la rentabilidad de ellos a través del tiempo.

En este contexto, el enfoque anterior puede ser enriquecido con información operacional que, relacionada con la ya disponible, genera una mejor calidad de resultados en el estudio de comportamiento de compra. La complejidad de esta adición de información se encuentra en determinar las variables necesarias y suficientes tanto operacionales como transaccionales para lograr un buen modelo que represente el comportamiento de los clientes.

Para llevar a cabo este estudio se requiere de información de calidad y suficiente, la cual es brindada por la empresa *Scopix Solutions* en conjunto con la cadena de supermercados norteamericana *Harris Teeter*.

Se cuenta con variables operacionales y transaccionales de los clientes; las primeras son extraídas de imágenes digitales, que contienen información relevante, obtenidas en intervalos periódicos de tiempo en distintas áreas de un supermercado en estudio, como reconocimiento de personas en cola, nivel de inventario de productos a través del día, flujo de entrada y salida de clientes, entre otras; en cambio, las segundas poseen datos de Punto de Venta (POS), descripción detallada de los SKU (Stock Keeping Unit), datos de la tarjeta de lealtad de los clientes (como datos socio demográficos de los clientes), entre otros.

Se puede mencionar que existen estudios preliminares sobre este tema. En el paper de Lu et al. (2011) se mide el efecto de las colas sobre el comportamiento de compra de los clientes añadiendo variables transaccionales como también operacionales (tamaño de la cola y servidores para intervalos fijos de tiempo). En base a lo anterior, y con la información disponible, se puede llevar a cabo la realización de un modelo de incidencia de compra sobre el comportamiento de los clientes, incorporando herramientas operacionales y transaccionales con la finalidad de aumentar los beneficios tanto para la empresa como para el cliente.

En este caso, se trabajará con los datos de las existencias de productos en el supermercado. En particular, se trabajará con los productos de la zona de los panes que se venden tanto a granel como empaquetados.

1.4. Motivación

La gran dinámica que se presenta dentro de un supermercado es un universo que tiene muchas aristas interesantes de explorar. Como fue mencionado recientemente, se poseen variables de *stock* para el estudio, lo que hace que su desarrollo sea novedoso. En un día normal en el supermercado, se puede observar algo similar a lo que se observa en la Figura 3.3, donde se muestran capturas reales realizadas en una de las tiendas de la cadena de supermercados en estudio.

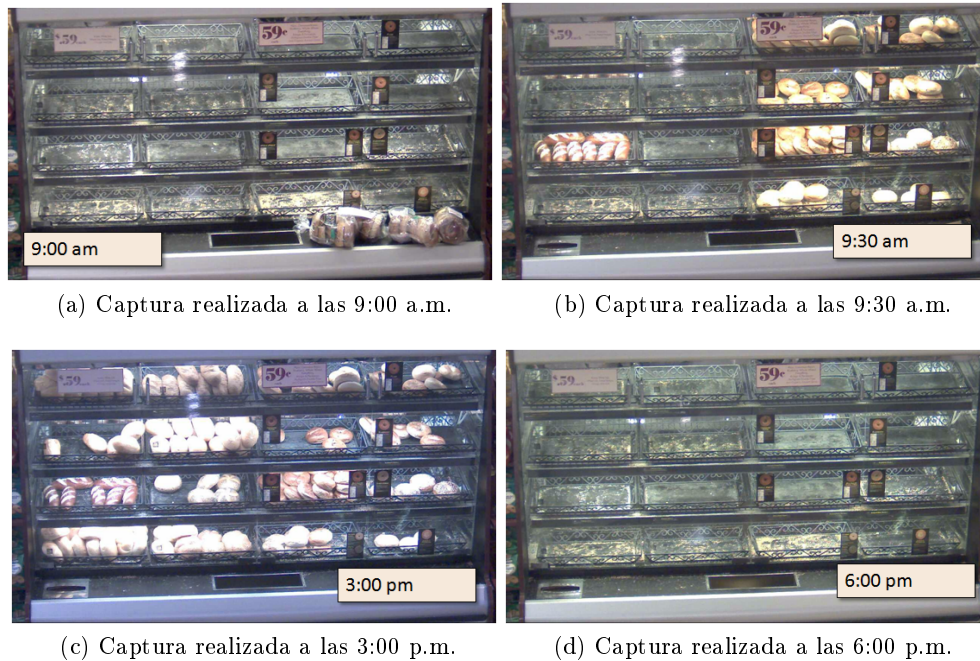


Figura 1.4: Capturas de canasta del pan a distintas horas del día

Se puede observar que existe una gran variación del *stock* observado en los distintos horarios del día. Esto genera muchas preguntas sobre las posibles causas de lo ocurrido en cada momento del tiempo y la relación y efecto que esto puede tener en las compras de los clientes, donde los fenómenos que no son observables directamente pueden ser deducidos a partir de la información disponible relacionada a la información transaccional de los clientes. Por ejemplo, un propósito válido podría ser querer determinar las causas por las cuales los clientes compraron (o no compraron) determinados productos. En ese caso, algunas de las alternativas que se pueden barajar como posibles respuestas (entre otras posibles) pueden estar relacionadas con ventas superiores al promedio en cierto horario, lo cual produjo que el *stock* del producto se haya acabado o disminuido drásticamente. Otra posibilidad es que, en ese día específico, las existencias no hayan sido repuestas después de cierto horario, lo cual afectaría de manera negativa las compras de los clientes venideros. Independientemente de las razones, puede ocurrir un efecto de interacciones entre el *stock* y las compras de los clientes, que puede ser desglosado y es de real interés estudiarlo y entenderlo.

Con todas las posibilidades posibles, se define a continuación cuáles son los objetivos y alcances de este trabajo.

1.5. Objetivos y Alcances

A continuación se hará referencia a los objetivos planteados para el trabajo de título y las limitaciones que éste tendrá.

1.5.1. Objetivo General

- Estudiar el efecto de la variedad de surtido en góndola en el comportamiento de los clientes mediante la incorporación de variables operacionales de *stock* a un modelo de incidencia de compra.

1.5.2. Objetivos Específicos

- Estudiar el efecto agregado de la variedad de surtido en góndola en el comportamiento de los clientes para realizar un diagnóstico previo.
- Estudiar el efecto de la variedad de productos en el comportamiento de los clientes mediante el desarrollo de un modelo que incluya heterogeneidad.
- Estudiar el efecto de la variedad de productos en el comportamiento de los clientes frente a otras categorías del supermercado.
- Determinar cuáles son los factores que inciden en el cambio del comportamiento de compra del consumidor.

1.5.3. Alcances

A continuación se presentan las limitaciones del trabajo de título:

- Las variables operacionales que serán estudiadas son solamente las correspondientes al *stock* de productos. Las variables correspondientes al reconocimiento de personas en cola y flujo de entrada y salida de clientes no serán consideradas en este estudio.
- Se estudiará para un tipo de producto en específico, que corresponde a los presentes en la categoría de panes.
- Se realizará un estudio del efecto a corto plazo de la variedad de productos en el comportamiento de los clientes. En la sección de extensiones se propone un modelo de mediano - largo plazo.

Capítulo 2

Hipótesis y Revisión bibliográfica

2.1. Efectos a capturar

Como fue mencionado en el Capítulo anterior, este trabajo busca estudiar el efecto de la variedad de surtido de panes en góndola en el comportamiento de los clientes. Para esto, se trabaja bajo la hipótesis de los clientes sustituyen entre productos Musalem et al. (2010), por lo tanto existen efectos de sustitución (o complementareidad) entre los productos y entre subcategorías de productos dentro de los mismos.

Para el estudio bajo la hipótesis descrita, se desglosa el trabajo en tres partes, cada uno separado por capítulos. La revisión bibliográfica de los modelos a utilizados se encuentra en el punto 2.2.

1. El efecto de las variables de disponibilidad de panes sobre la venta e incidencia de compra de panes agregadas por clientes.
2. El efecto de las variables de disponibilidad de panes sobre compra de productos de la categoría de panes para cada cliente.
3. El efecto de las variables de disponibilidad de panes dentro de la compra dentro de otros departamentos del supermercado.

El caso del primer capítulo se crea debido a la necesidad de reconocer, buscar y encontrar patrones dentro de los datos que sean de real ayuda a corroborar la hipótesis descrita. Para ello, se espera que ocurra alguno de los siguientes escenarios:

- Existencia de una correlación notoriamente negativa entre la venta de los productos (o su incidencia de compra) y el nivel de *stock* que éstos presenten.
- No sea observado un patrón claro entre la venta de productos y el nivel de *stock* de éstos.

Cualquiera sea el caso, se continuará con el análisis con los datos individuales. Para el primer escenario, si el patrón es observado con los datos agregados, entonces es esperable que se manifieste de manera más definida y precisa para el caso de los datos de clientes individuales.

Para el segundo escenario se espera que, si no se observa un patrón claro entre los datos – que puede ocurrir debido a que el “ruido” generado por la agregación de los datos sea tan significativo que encubra los patrones intrínsecos a los datos–, éste sea observado cuando se lleve a cabo el análisis de los datos desagregados a nivel de cliente, ya que los datos individuales tienen más poder para testear la hipótesis descrita.

Debido a la naturaleza de los datos, se utilizarán modelos de conteo. Los modelos más utilizados sobre métodos de conteo son el modelo de Poisson y el modelo de Binomial Negativa Cameron and Trivedi (1998).

Para el caso del segundo capítulo, se busca entender el efecto de la disponibilidad de los productos que le genera a cada cliente la compra de cada producto incluido dentro de la canasta de panes. Para esto, se necesita capturar y entender los efectos que se mencionan a continuación:

- **Patrones de sustitución:** Se desea conocer la importancia que tiene la compra de los distintos clientes del supermercado en cada subcategoría de panes; los que se venden a granel y aquellos que se venden empaquetados. Por esta razón, se desea estimar un efecto para cada cliente del supermercado de manera individual.

La captura de este efecto se logra a través del análisis de las variables de productos para las categorías existentes. Las variables creadas ayudan a capturar los cambios de preferencias para el conjunto de clientes en su totalidad. Sin embargo, para este análisis se desea además capturar el efecto de los cambios en las preferencias de los clientes dentro de las categorías de productos. Si existe varianza significativa de los datos, entonces se puede concluir que los clientes sustituyen por productos de la misma categoría, es decir, los clientes sustituyen entre *rolls* y *bagels*, que se encuentran dentro de la categoría de los productos frescos; y entre las 8 alternativas que se presentan para los productos no frescos.

- **Efectos cruzados:** Para este modelo los efectos cruzados existentes no están expuestos de manera explícita. Estos efectos se pueden ver a través de las subcategorías de productos cuyas ventas son a granel como de aquellos que se venden empaquetados, las cuales se encuentran expresadas a través de las variables binarias para cada una de éstas y, también, cuyos efectos se encuentran implícitos dentro de las variables de disponibilidad de variedad Lazari and Anderson (1994). Este efecto es capturado de la siguiente manera:

Al ocurrir un cambio en la preferencia (utilidad) de los clientes debido a una disminución (total o parcial) de alguno de los productos del conjunto de elección, éste es captado por los tres niveles jerárquicos en los cuales se descompone la variable de disponibilidad. Por lo tanto, lo anterior provoca que existan tanto efectos cruzados entre categorías (lo que acá se llaman efectos cruzados) como dentro de la misma categoría (los que acá se llaman efectos de sustitución).

- **Lealtad:** A pesar de que la lealtad de los clientes no es un efecto que se quiera estudiar en sí, las variables de lealtad ayudan a capturar ciertos efectos de sustitución y efectos cruzados entre los distintos productos (y categorías), debido a que éstas varían para cada cliente y cada producto según la ocasión de compra que se este observando.

El modelo a utilizar será un modelo de elección discreta, debido a que éstos se basan en la premisa de que un agente o tomador de decisiones, que en este caso corresponde a un cliente del supermercado, realiza una elección o una serie de elecciones a través del tiempo, sobre un conjunto de opciones. Train (2009). Dentro de éstos se encuentran los modelos de utilidad aleatoria (RUM, por sus siglas en inglés), donde el modelo logit se encuentra dentro de los más utilizados debido a su fácil implementación.

Para la realización de este capítulo se utiliza el modelo logit multinomial con coeficientes aleatorios (*random coefficient multinomial logit*), también llamado *mixed logit*, que es un modelo altamente flexible que se puede aproximar a cualquier modelo de utilidad aleatoria Gonul and Srinivasan (1993) Mc Fadden and Train (2000). Este modelo obvia las limitaciones que tiene el modelo logit estándar al permitir variación en los gustos aleatoria, patrones de sustitución irrestrictos y correlación entre los factores no observados durante el tiempo.

Para el caso del estudio dentro de los otros departamentos del supermercado, se estudiará cómo afectan los quiebres de *stock* de panes las compras en otras categorías del supermercado. Para esto se utilizará un modelo de panel con variables instrumentales. La necesidad de incorporar este modelo nace porque la compra en otros departamentos no debería explicarse directamente por la variable de disponibilidad de panes, sino que es más correcto asumir que la compra en éstos se ve afectada por las compras de pan que, a su vez, se ven afectadas por las variables de disponibilidad de panes.

2.2. Marco metodológico

2.2.1. Modelos de conteo

- Modelo de Poisson

Los modelos de regresión más simples para los datos de conteo son los modelos de regresión de Poisson, que son estimados mediante el estimador de máxima verosimilitud.

En muchas aplicaciones los datos de conteo se encuentran sobredispersados, donde la varianza condicional excede a la media condicional. Una solución a esto es el uso de regresión de Poisson, ya que produce estimaciones consistentes siempre que la media condicional se haya especificado correctamente. Es necesario, sin embargo, ajustar las estimaciones de error estándar. El supuesto más simple y manejable es asumir que la varianza es un múltiplo de la media. Por ejemplo, en el caso de las estimaciones de los coeficientes de máxima verosimilitud de Poisson y GLM (Modelo Lineal Generalizado o *Generalized Linear Model*, por su nombre en inglés) son idénticas, y los errores estándar de la estimación de Poisson usual y *t*-estadístico puede ser utilizados luego de reescalar apropiadamente los datos. Una alternativa a la regresión de Poisson es especificar una distribución más generalizada que la de Poisson, que no tenga que imponer equidispersión de los datos y se pueda realizar una inferencia mediante Máxima Verosimilitud Estándar.

- Binomial Negativa

La distribución estándar utilizada es la Binomial Negativa, donde se asume que la varianza es una función cuadrática de la media.

Los datos de conteo que se poseen en este caso en particular se encuentran sobredispersos. Esto produce que los resultados obtenidos luego de la implementación de un modelo de Poisson mediante Máxima Verosimilitud arrojen t -estadísticos que se encuentren considerablemente sobre-inflados. Esto puede llevar a conclusiones erróneas y, a la vez, muy optimistas sobre la significancia estadística de los regresores.

El modelo paramétrico estándar para dar cuenta de sobredispersión es la Binomial Negativa. Lo más común es que los datos utilizados se distribuyan Poisson, pero exista una heterogeneidad individual no observada que se distribuya mediante una distribución *Gamma*, donde se refleja que la media real no se puede observar perfectamente.

2.2.1.1. Modelos de regresión con datos de panel

Para este análisis se realizará una regresión. Como puede ser visto en Cameron and Trivedi (1998, 2009), se conocen distintos tipos de regresiones que pueden ser utilizadas para estos datos. Algunas de las posibles son la utilización de modelos de regresión con datos de panel. Éstos soportan dos tipos de formas de capturar efectos en los datos: efectos aleatorios (*random effects*) y efectos fijos (*fixed effects*).

- Efectos aleatorios (*random effects*)

Se utiliza cuando es altamente probable que se necesite controlar el carácter individual de cada producto, es decir, que el intercepto capture los efectos que posee cada grupo de productos en particular, por lo tanto, se supone que el intercepto es una variable aleatoria con cierta media y una desviación aleatoria sobre esa media. Existen tests como la *Prueba del Multiplicador de Lagrange para Efectos Aleatorios* T.S. and A.R. (1980) para comprobar si es preferible la utilización de modelos de efectos fijos o efectos aleatorios.

- Efectos Fijos (*fixed effects*)

Este modelo también permite modelar el carácter individual de cada producto. Se crea un intercepto para cada uno de los productos, es decir, una variable dicotómica para cada producto y se estima cada intercepto, obteniendo un valor fijo para cada estado. Este modelo no supone que las diferencias entre estados sean aleatorias, sino constantes o “fijas”.

- Autocorrelación

Los estimadores de Mínimos Cuadrados Ordinarios son los Mejores Estimadores Lineales Insesgados siempre y cuando los errores sean independientes entre sí y se distribuyan idénticamente con varianza constante σ^2 Cameron and Trivedi (2009). Con frecuencia estas condiciones son violadas en

datos de panel: la independencia se viola cuando los errores de diferentes unidades están correlacionados (correlación contemporánea), o cuando los errores dentro de cada unidad se correlacionan temporalmente (correlación serial), o ambos.

Existen varias maneras de corregir por correlación serial o autocorrelación. Una manera de hacerlo es a través de un modelo de efectos fijos con término autorregresivo de grado 1 –AR(1)–, que controla por la dependencia de t con respecto a $t - 1$ en término de los errores. El modelo AR(1) con efectos fijos se especifica como:

$$y_{2,t} = \alpha + \sum_{p'} \beta_1^{p'} x_t^{p'} + \sum_{p'} \beta_2^{p'} \left(x_t^{p'}\right)^2 + \sum_{p'} \beta_3^{p'} z_t^{p'} + \zeta_{d,t} + \varepsilon_t$$

con $\varepsilon_t = \rho\varepsilon_{t-1} + \nu_t$, donde $\rho \in [-1, 1]$ indica la correlación de primer grado que presentan los errores ε_t y ε_{t-1} , y ν_t es i.i.d. con media cero y varianza σ_ν .

Se puede notar que si $\rho \rightarrow 0$, entonces la función no posee términos autorregresivos y corresponde a una regresión común. Por lo tanto, el modelo por sí solo determinará, para cada caso, si es o no un modelo autorregresivo.

2.2.2. Modelos de elección discreta

La idea detrás de este modelo es entender el proceso de comportamiento que conduce al cliente tomar la decisión. Para esto se considera que existen diversos factores o causas, que determinan esta elección.

Se considera la siguiente notación:

- x = Factores observados por el investigador (factores deterministas)
- ε = Factores no observados por el investigador (factores aleatorios)

donde $f(\varepsilon)$ corresponde a la densidad de ε

Los factores están relacionados por una función, $y = h(x, \varepsilon)$, la cual es llamada también función del proceso de comportamiento. Como ε no es observable, entonces la elección del cliente no es determinista a ojos del investigador y no puede ser predecida exacta y precisamente. Sin embargo, se puede obtener la probabilidad de elección de cualquiera de las opciones del conjunto de elección.

De esta manera, la probabilidad de elegir la alternativa y dado que los factores observados están dados por x , se define de la siguiente manera:

$$\mathbb{P}(y/x) = \mathbb{P}(\varepsilon/h(x, \varepsilon) = y)$$

Para simplificar el problema, se define lo siguiente:

$$I(h(x, \varepsilon) = y) = \begin{cases} 1 & \text{si la afirmación es cierta} \\ 0 & \text{en caso contrario} \end{cases}$$

Por lo tanto, si se reescribe $\mathbb{P}(y/x)$ queda de la siguiente forma:

$$\begin{aligned}\mathbb{P}(y/x) &= \mathbb{P}(I[h(x, \varepsilon) = y] = 1) \\ &= \int I[h(x, \varepsilon) = y]f(\varepsilon)d\varepsilon\end{aligned}$$

Además, para ciertos modelos –como el modelo logit multinomial– se cumplen ciertas condiciones tanto sobre h como sobre f que provocan que la integral descrita anteriormente sea una integral cerrada, es decir, que puede ser calculada directamente sin intervención de simulaciones.

Propiedades de los modelos de elección discreta

El conjunto de elección corresponde al conjunto de alternativas que observan los clientes del supermercado. Este conjunto debe respetar las siguientes condiciones:

1. Las alternativas deben ser mutuamente excluyentes ante los ojos del cliente.
2. El conjunto de elección debe ser exhaustivo, es decir, todas las posibles alternativas deben estar incluidas.
3. El cliente debe elegir únicamente una alternativa.
4. El número de alternativas debe ser finito.

2.2.2.1. Modelos de utilidad aleatoria

Los modelos de elección discreta se derivan generalmente del supuesto de una maximización de utilidad por parte de los clientes, para decidir qué alternativa escoger. Los modelos de utilidad aleatoria – RUM, por su nombre en inglés (*random utility models*)–, consideran que un cliente i escoge entre J alternativas y obtiene cierto nivel de utilidad, es decir, recibe un beneficio de realizar cada acción, para cada una de las elecciones realizadas. La utilidad que le entrega al cliente i elegir la alternativa j es U_j^i , $j = 1, \dots, J$. Ésta, como medida de bienestar, no tiene un nivel natural o escala, lo cual tiene importantes implicaciones para la especificación y normalización de los modelos de elección discreta. Train (2009)

Más específicamente, la descomposición de la utilidad en su parte determinista y aleatoria se puede ver a continuación:

$$U_j^i = \underbrace{V_j^i}_{\text{parte determinista}} + \underbrace{\varepsilon_j^i}_{\text{parte aleatoria}}$$

La parte determinista queda descrita de la siguiente manera:

$V_j^i = V(\beta_j^i, x_j^i)$ corresponde a la porción de las utilidades que es observada por el investigador.

donde:

- $x^i = \langle x_{11}^i, \dots, x_{1j}^i, x_{21}^i, \dots, x_{nj}^i \rangle$ corresponde al vector de variables, donde n corresponde al número de variables incluidas en el modelo.
- $\beta^i = \langle \beta_{11}^i, \dots, \beta_{1j}^i, \beta_{21}^i, \dots, \beta_{nj}^i \rangle$ corresponde al vector de parámetros. Éstos en general son desconocidos y deben ser estimados estadísticamente.

La parte aleatoria ε_j^i , que es desconocida, corresponde a los factores que afectan la utilidad pero que no están incluidos en V_j^i . La densidad conjunta del vector de variables aleatorias $\varepsilon^i = \langle \varepsilon_1^i, \dots, \varepsilon_J^i \rangle$ es denominada $f(\varepsilon^i)$. Con esta densidad, se puede determinar probabilísticamente la elección del cliente.

Por su parte, el cliente elegirá la alternativa que le entregue una mayor utilidad. Por lo tanto, dentro del modelo esto queda definido de la siguiente manera:

Si un individuo i observa J alternativas distintas, donde elige únicamente una alternativa, digamos $j \in \{1, \dots, J\}$, entonces su función de utilidad queda escrita por:

$$\begin{aligned} \text{Como } \mathbb{P} &= \int I[\beta'x + \varepsilon > 0] f(\varepsilon) d\varepsilon = \int I[\varepsilon > -\beta'x] f(\varepsilon) d\varepsilon \\ &= \int_{-\beta'x}^{\infty} f(\varepsilon) d\varepsilon = 1 - F(-\beta'x) \\ &= 1 - \frac{1}{1 + e^{\beta'x}} = \frac{e^{\beta'x}}{1 + e^{\beta'x}} \end{aligned}$$

La probabilidad de que el cliente i elija la alternativa j está dado por:

$$\begin{aligned} P_j^i &= \mathbb{P}(U_j^i > U_k^i \forall k \neq j) \\ &= \mathbb{P}(V_j^i + \varepsilon_j^i > V_k^i + \varepsilon_k^i \forall i \neq k) \\ &= \mathbb{P}(\varepsilon_k^i - \varepsilon_j^i < V_j^i - V_k^i \forall k \neq j) \end{aligned}$$

2.2.2.2. Modelo Logit

Es el modelo de elección discreta más utilizado debido a su fácil implementación e interpretación Train (2009). Para el caso del modelo logit multinomial, los factores no observados ε siguen una distribución Gumbel, es decir, $f(\varepsilon_{ij}) = e^{-\varepsilon_{ij}} e^{-e^{-\varepsilon_{ij}}}$. Por lo tanto, la probabilidad de queda dada por la siguiente expresión cerrada:

$\mathbb{P}_{ij} = \frac{e^{V_{ij}}}{\sum_k e^{V_{ik}}}$. Si V_{ij} crece, con V_{ik} , $k \neq j$, \mathbb{P}_{ij} se aproxima a 1. Por otro lado, si V_{ij} decrece, se aproxima a cero cuando $V_{ij} \rightarrow -\infty$. La probabilidad de elección de una alternativa nunca será cero. También se cumple que la suma de las probabilidad para todas las alternativas es igual a 1.

Propiedades principales del modelo Logit

Existen 3 asuntos que demuestran tanto el poder de los modelos logit para representar el comportamiento de elección, como también exponen los límites de ese poder. Los asuntos son:

variación de gustos (*taste variation*), patrones de sustitución y las elecciones repetidas en el tiempo. La aplicabilidad de los modelos logit pueden ser resumidos de la siguiente manera:

1. El modelo logit puede representar variación de gustos relacionadas a características observables del cliente, pero no relacionadas a variaciones aleatorias o desconocidas (no observadas). Es decir, el modelo logit puede capturar variación de gustos pero con ciertos límites.
2. La utilización del modelo logit implica tener patrones de sustitución proporcionales a través de las alternativas, dada la especificación realizada sobre la utilidad representativa. Para capturar patrones de sustitución más flexibles, es necesario la utilización de otros modelos. El modelo Logit cumple la siguiente propiedad:
 - a) Propiedad de la Independencia de Alternativas Irrelevantes (IIA por sus siglas en inglés –*Independence from Irrelevant Alternatives*).
Para dos alternativas j y k , se cumple: $\frac{P_{ij}}{P_{ik}} = e^{V_{ij}-V_{ik}}$, es decir, esta división no depende de las otras alternativas. Por lo tanto, es independiente de alternativas irrelevantes. Este es un supuesto muy fuerte y restrictivo que, en caso de ser soportado por el modelo en estudio, ofrece ventajas en casos donde el número de alternativas es grande.
 - b) Sustitución Proporcional.
Una mejora en el atributo de una alternativa reduce la probabilidad de elección de las otras alternativas en el mismo porcentaje. Esto es una implicancia de IIA.
3. Si los factores no observados son independientes sobre el tiempo en repetidas situaciones de elección, entonces el modelo logit puede capturar las dinámicas de las elecciones repetidas, incluyendo la dependencia entre estados. Sin embargo, el modelo logit no puede capturar estos efectos cuando los factores no observados están correlacionados en el tiempo.

Estimación del modelo Logit

Se asume que las variables explicatorias son exógenas a la elección. Es decir, las variables que conforman la utilidad son independientes del componente no observado de la utilidad. Como las probabilidades de elección del modelo logit poseen una expresión cerrada, entonces se puede aplicar el supuesto de máxima verosimilitud.

Esta función queda expresada de la siguiente manera:

$$L(\beta) = \prod_i \prod_j (P_{ij})^{y_{ij}}, \text{ donde } \prod_j (P_{ij})^{y_{ij}} \text{ corresponde a la probabilidad de que el cliente } i \text{ elija}$$

la alternativa que realmente eligió, que es equivalente a la probabilidad de la alternativa elegida, debido a que $y_{ij} = 1$ si la alternativa j fue elegida por el cliente i ; 0 en otro caso. Si se asume que todas las personas son distintas, entonces $L(\beta)$ corresponde a la probabilidad de que cada cliente haya elegido la alternativa que realmente fue elegida.

La función de log-verosimilitud queda descrita como sigue:

$$LL(\beta) = \sum_i \sum_j y_{ij} \ln P_{ij} \text{ y el estimador es el valor de } \beta \text{ que maximiza la función. Cuando}$$

la función de utilidad es lineal en los parámetros, la función de log-verosimilitud es globalmente

cóncava y los valores pueden ser estimados directamente mediante las condiciones de primer orden, que corresponde a la siguiente expresión:

$$\sum_i \sum_j (y_{ij} - P_{ij}) x_{ij} = 0$$

2.2.2.3. Random Coefficient Logit

En el modelo *mixed logit*, las probabilidades son integrales de las probabilidades del modelo logit estándar sobre la densidad de los parámetros β . Esto queda definido de la siguiente manera:

$$P_{ni} = \int L_{ni}(\beta) f(\beta) d\beta$$

donde $L_{ni}(\beta)$ corresponde a la probabilidad del modelo logit evaluada en el parámetro β , la cual es equivalente a lo siguiente:

$$L_{ni}(\beta) = \frac{e^{V_{ni}(\beta)}}{\sum_{j=1}^J e^{V_{nj}(\beta)}}$$

y $f(\beta)$ es la función de densidad. Por lo tanto, la función se puede ver como un promedio ponderado de una gran cantidad de funciones, donde el peso está dado por la función de densidad. Los valores de β representan el gusto de cada individuo.

Existen varias formas de plantear los supuestos anteriormente mencionados. La forma más común de hacerlo es bajo la integración de coeficientes aleatorios, donde la función de utilidad puede ser vista como sigue:

$$U_{ij} = \beta_i x_{ij} + \varepsilon_{ij}$$

donde x_{ij} son las variables observadas, que están relacionadas a las alternativas y a los clientes, β_i es un coeficiente para cada cliente i , que representa el gusto particular de cada uno de ellos y el término aleatorio ε_{ij} es el valor extremo *iid*.

Otra forma de ver la función de utilidad es a través de los componentes de error que crean correlaciones entre las utilidades de las diferentes alternativas. Esto puede ser expresado de la siguiente manera:

$$U_{ij} = \alpha' x_{ij} + \mu'_i z_{ij} + \varepsilon_{ij}$$

x_{ij} y z_{ij} son los vectores de variables observadas relativas a la alternativa j , α es el vector de coeficientes fijos, μ es el vector de términos aleatorios de media cero y ε_{ij} es valor extremo *iid*. Los términos en z_{ij} son componentes de error que, en conjunto con ε_{ij} , definen la parte estocástica de la utilidad. Es decir, la porción de utilidad no observada queda representada por $\eta_{ij} = \mu' z_{ij} + \varepsilon_{ij}$. Para el modelo logit estándar, el término z_{ij} es cero, por lo que no hay correlación dentro de la utilidad

para cada alternativa. La falta de correlación es que lo hace que la propiedad IIA se cumpla y existan patrones de sustitución restrictivos.

Estas dos últimas definiciones son equivalentes si se considera que el coeficiente β_i puede ser descompuesto en una media α y una desviación estándar μ_i , donde en los dos modelos éste sigue la misma distribución. Cabe recordar que las distribuciones mixtas, ya sean por los parámetros aleatorios o por componentes de error, capturan varianza y correlaciones en los factores no observados.

Estimación del Modelo Random Coefficient Logit

Los coeficientes β_i se distribuyen con una densidad $f(\beta|\theta)$, donde θ representa los parámetros de la distribución, como la media y la varianza de β .

Las probabilidades son aproximadas por simulación para algún valor de θ .

1. Elegir un valor de β para $f(\beta|\theta)$ y llamarlo β^r , donde r representa los distintos pasos elegidos
2. Calcular mediante la fórmula del modelo logit $L_{ij}(\beta)$ para ese paso.
3. Repetir los pasos 1 y 2 cuantas veces sea necesario y promediar los resultados. Este resultado puede ser visto de la siguiente manera: $\check{P}_{ij} = \frac{1}{R} \sum_{r=1}^R L_{ij}(\beta^r)$, donde R es el número de pasos. \check{P}_{ij} es un estimador insesgado de P_{ij} por construcción. La varianza decrece si R crece.

Las probabilidades estan insertas en la función de log-verosimilitud para dar paso a la función de log-verosimilitud simulada:

$$SLL = \sum_i \sum_j y_{ij} \ln \check{P}_{ij}$$

donde $y_{ij} = 1$ si el cliente i escoge la alternativa j ; 0 en otro caso.

Por último, el estimador de máxima verosimilitud simulada (MSLE por sus siglas en ingles – *maximum simulated likelihood estimator*) es el valor de θ que maximiza SLL .

Para elegir los pasos r se utilizan las Secuencias de Halton (*Halton draws*). La secuencia de Halton está definida para el intervalo unitario, donde los elementos que contiene pueden ser considerados pasos bien situados para una función de densidad uniforme. Los pasos de Halton proveen una mejor cobertura que los pasos aleatorios en promedio, debido a que éstos son creados para ir llenando progresivamente el intervalo unitario uniformemente y de manera más densa cada vez. Los elementos en cada ciclo están equidistantemente ubicados, y cada ciclo cubre el intervalo unitario en áreas que no fueron cubiertas en los ciclos previos. Los pasos de una observación tienden a estar correlacionados negativamente con aquellos no cubiertos en los ciclos pasados. Esto hace que la varianza entre observaciones disminuya.

Las secuencias de Halton se definen de la siguiente manera:

$$s_0 = \{0\}$$

$$s_{t+1} = \{s_t, s_t + 1/k^t, s_t + 2/k^t, \dots, s_t + (k-1)/k^t\},$$

donde la secuencia contiene ciclos de largo k . Para la creación de los pasos de Halton para una muestra de observaciones, se crea una larga secuencia de Halton y, luego, parte de esa secuencia es utilizada para cada observación. Por conveniencia, se escogen valores de k primos debido a la potencial correlación. Se evita la posibilidad de que los ciclos coincidan a lo largo de la secuencia completa. El elemento 0 no se considera parte de la cadena, pero se incluye por simplicidad del modelo computacional.

Los elementos iniciales de dos secuencias cualquiera pueden estar altamente correlacionadas. Por esta razón, es común eliminar los valores iniciales de cada secuencia de Halton antes de asignar los elementos a las observaciones. Se recomienda que el número de secuencia de elementos a eliminar debe ser al menos tan grande como el número primo usado en la generación de las secuencias. Si hay un número de coeficientes aleatorios asociado al modelo igual a p , entonces el modelo random coefficient utiliza los primeros p primos para generar las secuencias de Halton para cubrir las p dimensiones del espacio.

Cuando el número de pasos utilizados para cada observación crece, entonces la cobertura para cada observación mejora. La covarianza negativa a través de las observaciones disminuye, debido a que hay menores brechas en la cobertura de cada observación para ser llenada por la siguiente observación.

2.2.3. Modelos de panel

El modelo general a utilizar puede ser descrito de la siguiente manera:

$$y_{it} = \alpha_i + x'_{it}\beta + \varepsilon_{it}$$

Dentro de este modelo general, se conocen distintas variaciones que dependen de los efectos que se desean capturar. Algunas de éstas son las siguientes: *Random effects* (RE), *Between effects* (BE), *Fixed Effects* (FE), *Population averaged* (PA), entre otros Cameron and Trivedi (2009).

2.2.3.1. Modelos de panel con variables instrumentales

En las regresiones lineales comunes se utiliza el método de Mínimos Cuadrados Ordinarios (MCO) para estimar los valores de los estimadores del modelo.

Para que estos estimadores sean consistentes, deben cumplir el supuesto de que los términos de error no estén correlacionados con los regresores, es decir, $E(u|x) = 0$. Si esto no ocurre, entonces los estimadores $\hat{\beta}_j$ no pueden ser interpretados como el efecto marginal sobre la variable y de un cambio exógeno de la variable regresora x_j . Para que esto se siga cumpliendo, se introducen las variables instrumentales.

Las variables instrumentales (IV, por sus siglas en inglés - *instrumental variables*) proveen estimadores consistentes bajo el fuerte supuesto de que los instrumentos z son variables que están correlacionadas con las variables regresoras, y satisfacen que $E(u|z) = 0$.

Las variable instrumentales z cumplen que los cambios en éstas están asociados con los cambios en las variables regresoras x , pero no con los cambios en la variable dependiente y (sólo

directamente a través de los regresores x).

$$y_{1i} = y_2'\beta_1 + x_{1i}'\beta_2 + u_i$$

donde y_2 corresponde a los regresores endógenos, x_1 corresponde a los regresores exógenos, que incluyen un intercepto.

Los errores, denominados por u_i , no están correlacionados con x_{1i} , pero sí están correlacionados con y_{2i} . Se asume que existen al menos m estimadores IV –denominados por x_2 – que satisfacen que $E(u_i|x_{2i}) = 0$. Los instrumentos x_2 están naturalmente correlacionados con y_2 , debido a que entregan información de las variables instrumentadas.

Capítulo 3

Análisis agregado de clientes

3.1. Preámbulo

Para realizar el análisis es preciso determinar las variables a incluir en el modelo. Las ventas del producto serán analizadas mediante dos variables diferentes:

- Ventas de pan.
- Porcentaje de ventas de pan, que corresponde a la incidencia de la categoría de pan.

Como se observa en la Tabla 3.1, el conjunto de productos es considerablemente numeroso como para analizar cada producto de manera individual. Por lo tanto, es menester realizar categorías que agrupen de manera coherente estos productos. Luego de observar el porcentaje de incidencia de cada producto dentro de la categoría de pan, se consideró analizar las siguientes categorías:

1. Los 2 principales productos (Los 2 productos con mayor porcentaje de ventas)
2. Los 5 principales productos (Los 5 productos con mayor porcentaje de ventas)
3. Comparación entre *rolls* envasados (*packaged rolls*) y *rolls* a granel (*bulk rolls*).

Cabe mencionar que no se consideró un análisis análogo al de la tercera categoría con los *bagels*, debido a que en la categoría de pan no existe un símil que se venda envasado, como puede ser observado en la Tabla 3.1.

Nombre	Producto	Porcentaje	Porc. Acum.
<i>Roll</i>	2	31,5 %	31,5 %
<i>Bagel</i>	1	17,2 %	48,7 %
<i>French Bread</i>	8	8,4 %	57,0 %
<i>Small French Bread</i>	11	7,4 %	64,4 %
<i>Kaiser Rolls</i>	10	7,3 %	71,8 %
<i>Take & Bake Small French Bread</i>	15	3,7 %	75,4 %
<i>White Sub</i>	25	2,7 %	78,2 %
<i>Take & Bake Small Petite and Wheat Petite</i>	16	2,4 %	80,5 %
<i>Challah</i>	4	2,2 %	82,7 %
<i>Take & Bake Club</i>	14	2,1 %	84,9 %
<i>Chicago Rolls</i>	5	1,7 %	86,6 %
<i>Small Wheat French Bread</i>	12	1,7 %	88,3 %
<i>Club Rolls</i>	6	1,7 %	90,0 %
<i>Wheat Sub</i>	21	1,4 %	91,4 %
<i>Brown and Serve Rolls/ Butter Flake Rolls</i>	3	1,4 %	92,8 %
<i>White Flatbreads</i>	23	1,1 %	93,9 %
<i>Yeast Rolls</i>	26	1,0 %	94,9 %
<i>English Toasting Bread/LaBrea Rustique Rolls</i>	7	1,0 %	95,9 %
<i>Wheat Pita Pocket</i>	20	0,9 %	96,9 %
<i>Tear and Share Garlic Bread</i>	19	0,8 %	97,6 %
<i>Wheat/ Multigrain Flatbread</i>	22	0,7 %	98,3 %
<i>White Pita Pocket</i>	24	0,5 %	98,9 %
<i>Tear and Share Cheese Bread</i>	18	0,4 %	99,3 %
<i>Southern Style Biscuits</i>	13	0,3 %	99,6 %
<i>Fresh Food Mini White Pita Pockets 2 rows</i>	9	0,3 %	99,9 %
<i>Take & Bake Small Wheat French Bread</i>	17	0,1 %	100,0 %

Tabla 3.1: Participación de mercado de los productos en la categoría de pan

Para el caso de los *rolls* y *bagels* que se venden agrupados bajo un mismo Código *PLU* (*Price look-up code*) del supermercado, se puede ver desagregado de la siguiente manera, como se muestra en la Tabla 3.2.

Producto	PLU	Producto
1	813	<i>Asiago/ Parmesan Bagel</i>
1	813	<i>Blueberry Bagel</i>
1	813	<i>Cinnamon Bagel</i>
1	813	<i>Everything Bagel</i>
1	813	<i>Plain Bagel</i>
1	813	<i>Poppy/Sesame Bagel</i>
1	813	<i>Whole Wheat Bagel</i>
2	818	<i>Chicago Rolls</i>
2	818	<i>Kaiser Rolls</i>
2	818	<i>Pretzel Rolls</i>
2	818	<i>White Rolls</i>
2	818	<i>Wheat Rolls</i>

Tabla 3.2: Nombre y descripción de los productos agregados

A continuación se muestra en la Tabla 3.3 el nivel de *stock* de todos los productos en análisis en total y por intervalo de horas. Adicionalmente, en la Tabla 3.4 se ilustra el último intervalo de tiempo descompuesto por cada hora del día. Esta última se creó al notar que existían muchos más quiebres de *stock* para el último intervalo de tiempo en los productos 1 y 2. Cabe mencionar que se incluyó también a los 3 productos siguientes que concentran más ventas, debido a su importancia en el análisis.

Producto	PLU	Stock	$St_{h \in [8A.M., 11A.M.]}$	$St_{h \in [11A.M., 6P.M.]}$	$St_{h \in [6P.M., 10P.M.]}$
1	813	80,84 %	97,13 %	99,33 %	40,20 %
2	818	80,66 %	93,68 %	99,33 %	42,19 %
3	7203695999	99,42 %	99,25 %	99,86 %	98,87 %
4	7203695005	95,87 %	94,12 %	97,94 %	93,98 %
5	7203695711	99,42 %	99,53 %	99,86 %	98,66 %
6	7203695712	96,90 %	95,61 %	99,09 %	94,48 %
7	7203695065	99,70 %	98,79 %	99,95 %	100,00 %
8	7203695175	99,85 %	99,91 %	100,00 %	99,58 %
9	7203695247	99,91 %	100,00 %	99,95 %	99,79 %
10	7203695174	99,89 %	100,00 %	99,95 %	99,72 %
11	7203695004	97,11 %	96,36 %	99,41 %	94,13 %
12	7203695050	98,76 %	97,67 %	99,95 %	97,74 %
13	7203695571	99,70 %	99,44 %	100,00 %	99,43 %
14	7203695713	96,23 %	95,42 %	98,58 %	93,21 %
15	7203695358	97,56 %	96,78 %	99,29 %	95,47 %
16	7203695360	98,74 %	97,76 %	99,63 %	98,09 %
17	7203695874	97,97 %	96,92 %	99,31 %	96,67 %
18	7203695905	99,00 %	98,83 %	99,73 %	98,02 %
19	7203695906	96,08 %	95,52 %	98,12 %	93,35 %
20	7203695251	99,16 %	99,07 %	99,77 %	98,30 %
21	7203695127	99,27 %	98,13 %	99,91 %	99,15 %
22	7203695416	99,98 %	100,00 %	99,95 %	100,00 %
23	7203695415	99,87 %	100,00 %	100,00 %	99,58 %
24	7203695248	99,66 %	99,16 %	99,91 %	99,65 %
25	7203695087	98,05 %	95,61 %	99,41 %	97,81 %
26	7203695242	97,73 %	94,93 %	99,23 %	97,54 %
Total ponderado (\bar{x})		97,35 %	97,70 %	99,52 %	93,73 %

Tabla 3.3: Porcentaje de Stock por productos Total y por intervalos de tiempo

Producto	6 P.M.	7 P.M.	8 P.M.	9 P.M.	10 P.M.
2	98,29 %	66,03 %	13,80 %	2,12 %	1,59 %
1	96,80 %	74,79 %	14,23 %	2,90 %	2,60 %
11	97,46 %	95,25 %	92,04 %	92,93 %	91,72 %
10	99,37 %	100,00 %	99,68 %	99,68 %	100,00 %
8	99,05 %	99,68 %	100,00 %	99,68 %	99,36 %

Tabla 3.4: Stock desagregado por horas del último intervalo de tiempo

Se puede ver en la Tabla 3.4 que para los productos 2 y 1, el nivel de *stock* va disminuyendo abruptamente a lo largo del día, no así para los otros productos. Esto se puede deber a que éstos son los únicos productos con una perecibilidad mayor, por lo tanto, poseer *stock* de dichos productos a final del día puede no ser tan rentable para el supermercado, aunque no se sabe *a priori* si esta medida es óptima para la reposición de productos.

3.2. Recopilación de información

Datos transaccionales Se poseen datos transaccionales para las fechas entre el 1 de agosto de 2010 al 30 de mayo de 2011, donde se deben excluir los días comprendidos entre el 18 de enero de 2011 al 26 de febrero de 2011, como además los días 25 de diciembre de 2010 y 29 de marzo de 2011, debido a que no se dispone de información en esas fechas. Por lo tanto, se tiene información para un total de 261 días.

Datos de Stock Con respecto a los datos de nivel de *stock* de los productos, la información se encuentra disponible para las fechas comprendidas entre el 22 de diciembre de 2010 y el 30 de mayo de 2011 –ambos inclusive– lo que corresponde a un total de 160 días.

Datos totales utilizados La muestra total está compuesta por los datos obtenidos entre las fechas del 22 de diciembre de 2010 al 30 de mayo de 2011, donde se excluyen los días comprendidos entre el 18 de enero de 2011 al 26 de febrero de 2011, como así también los días 25 de diciembre de 2010 y 29 de marzo de 2011. Consecuentemente, se puede observar que la muestra total considera un total de 118 días.

Con respecto a los horarios utilizados, se omitieron los datos de los horarios comprendidos entre las 9 P.M. a 10 P.M., debido a que se realizaban pocas compras en ellos, así también se desecharon todos los datos generados antes de las 8 A.M.; por lo tanto, se considera un rango desde las 8 A.M. a las 8.30 P.M.

Los datos fueron medidos cada media hora, es decir, cada periodo t equivale a un intervalo de 30 minutos.

3.3. Descripción de las variables del modelo

A continuación, se realiza la descripción de las variables que son utilizadas en los modelos en estudio. Para esta sección, se seleccionaron las variables más relevantes a juicio del experto y las variables de control más relevantes según la literatura existente.

3.3.1. Variables dependientes

Para el caso en estudio, se utilizarán las variables dependientes presentadas en la Tabla 3.5, que serán estudiadas de manera autónoma.

Variable	Descripción
$Ventas_{j,t}$	Cantidad de compras de pan del tipo j realizadas en el periodo t ; si se omite superíndice j corresponde al total de compras de pan en el periodo t
$Incidencia_{j,t}$	Incidencia del pan de tipo j en el periodo de tiempo t . Queda definida como el porcentaje de compras de pan de tipo j con respecto al total de compras de productos en el periodo t ; si se omite superíndice j corresponde porcentaje total de pan en el periodo t . En este caso la variable corresponde a la incidencia de compra de la categoría “pan” por cada periodo de tiempo t

Tabla 3.5: Definición de variables dependientes

3.3.2. Variables de Stock

Las variables de *stock* corresponden a las variables a través de las cuales se mide el nivel de *stock* de los productos en góndola. Las variables de *stock* consideran que, si un cliente realiza la visita y elección del producto en la categoría en estudio en un horario y fecha determinada igual a t' y, suponiendo que un cliente se demora un tiempo $\tau \sim U[0, 30 \text{ minutos}]$ en realizar la compra, entonces el cliente realizó su transacción en la fecha y horario $t = t' + \tau$.

El nivel de *stock* de un producto está construido por definición como continuo entre 0 y 1. Dentro de esto, existen productos en los cuales el nivel de *stock* es dicotómico y en otros corresponde a un continuo. Las variables que se encuentran descritas de manera continua consideran uno o más de los siguientes factores:

1. Caso de los *Bagels* y *Rolls* a granel: Éstos se venden mediante un único código-SKU (*Stock-keeping unit*) –1 y 2 respectivamente–, por lo tanto, cada tipo de *bagel* o *roll* no es distinguible como producto único al momento de ser enlazado con los datos transaccionales. Por lo tanto, para estos productos se saca un promedio del *stock* de cada SKU y, de esta manera, se trabaja como si fuera un único producto.
2. Múltiples canastas: Existen productos que se venden en más de una canasta (*basket*). El sistema captura el nivel de *stock* de cada canasta por separado para un cierto producto. Luego, para determinar el nivel de *stock* total de éste, se calcula el promedio del nivel de *stock* de las canastas en cuestión. Esto se puede ver ejemplificado en la Figura 1.1, donde se muestra el lugar físico donde se ubican los distintos panes y las diferentes canastas que los contienen.

Las variables utilizadas dentro de esta clasificación se presentan en la Tabla 3.6.

Variable	Descripción
$OOS_{j,t}^{cont}$	Grado de no-disponibilidad (<i>stock out</i>) del producto j en el periodo t , $OOS_{j,t}^{cont} \in [0, 1]$. Mientras más cercano a 1 sea el valor de la variable, mayores serán los quiebres de <i>stock</i> para el producto j observados en el periodo t
$OOS_{j,t}^{bin}$	1 si se encuentra disponible 50 % o menos de los panes del tipo j para el periodo t ; 0 en otro caso

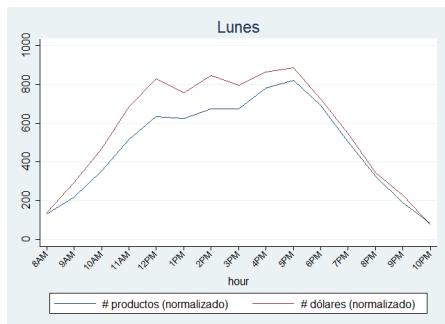
Tabla 3.6: Definición de variables de *stock*

3.3.3. Variables de Estacionalidad

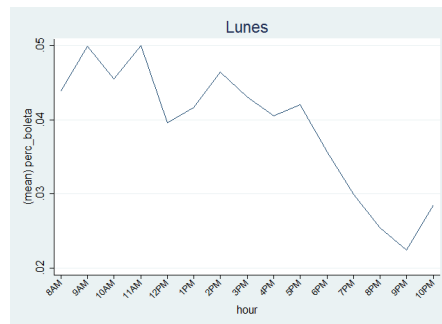
Para tener una idea más clara sobre cómo se comportan los datos, se puede observar la Figura 3.1, donde se muestran las estacionalidades de los datos por hora del día para el día lunes. El resto de las gráficas se encuentran en el Anexo 1. En (a), se muestra el número de productos totales vendidos y la cantidad de dólares recaudados normalizados.

Se observa que las ventas aumentan en el horario entre las 8 A.M. hasta las 11 A.M., se estabilizan entre las 11 A.M. y las 6 P.M. y decaen entre las 6 P.M. y las 10 P.M. Cabe destacar que en este periodo del día es donde ocurre la mayor cantidad de ventas por hora. Adicionalmente, se destaca que los días sábado y domingo presentan tanto una mayor venta de producto como mayor recaudación de dólares en las transacciones realizadas.

Para el caso de la incidencia de compra, se percibe que existen patrones distintos entre los días de la semana y el fin de semana. Para los primeros, la incidencia de compra se mantiene aproximadamente constante en las primeras horas del día, con un descenso de este porcentaje a partir de las 5 P.M. aproximadamente. Posteriormente, este descenso es relativamente constante si se consideran las ventas realizadas antes de las 9 P.M. Para el fin de semana, se ve un aumento las ventas de pan, aunque éstas descienden acorde transcurre el día.



(a) N^o productos vendidos y dólares recaudados



(b) Incidencia

Figura 3.1: Estacionalidades por hora del día

Mediante este análisis, las variables definidas son aquellas descritas en la Tabla 3.7, determinadas por el análisis de las variables de ventas e incidencia.

Variable	Descripción
W_d	1 si el día d corresponde a un fin de semana; 0 en otro caso
$H_{1,I^i(t)}$	Corresponde a: $\min\{t, \text{argmáx}\{I^i(t)\}\}$. Los intervalos considerados son los siguientes: $I^1 = [8 \text{ A.M.}, 11 \text{ A.M.}]$ $I^2 = (11 \text{ A.M.}, 6 \text{ P.M.}]$ $I^3 = (6 \text{ P.M.}, 10 \text{ P.M.}]$
$H_{2,I^i(t)}$	Corresponde a: $\min\{t, \text{argmáx}\{I^i(t)\}\}$. Los intervalos considerados son los siguientes: $I^1 = [8 \text{ A.M.}, 5 \text{ P.M.}]$ $I^2 = (5 \text{ P.M.}, 10 \text{ P.M.}]$
h_t	Corresponde a la hora en la cual se encuentra el periodo t .

Tabla 3.7: Definición de variables de estacionalidad

3.4. Modelos

Se muestra a continuación el modelo general empleado para el análisis, donde la función f queda determinada por la regresión que será utilizada.

$$E(y_t^j) = f \left(\alpha + \sum_{j'} \beta_1^{j'} OOS_{j',t}^{cont} + \sum_{j'} \beta_2^{j'} (OOS_{j',t}^{cont})^2 + CONTROLES_{d,t} \right)$$

$$E(y_t^j) = f \left(\alpha + \sum_{j'} \beta_3^{j'} OOS_{j',t}^{bin} + CONTROLES_{d,t} \right)$$

$$\text{Con } CONTROLES_{d,t} = \gamma W_d + \sum_j \sum_i \eta_{j,i} H_{j,I^i(t)} + \sum_t \psi_t h_t$$

Modelos relacionados a la variable de ventas Para determinar cuál es la distribución se ajusta mejor a los datos que se poseen, es necesario –en primera instancia– observar cuál es la distribución que siguen estos datos. A continuación –en la Figura 3.2a– se presenta la frecuencia de la variable $Ventas_{j,t}$, que corresponde a las unidades vendidas durante cada periodo de media hora.

Los datos utilizados corresponden a datos de conteo. Por la distribución de los datos, se puede dilucidar que éstos deben ajustar adecuadamente a una distribución binomial negativa.

Además de ser consideradas las características inherentes al modelo utilizado, éste presenta algunas ventajas absolutas sobre otros modelos que no son de conteo, entre las cuales se pueden mencionar las siguientes:

- La distancia considerada entre cada hora y día no es necesariamente lineal en la realidad. Dado que este modelo presentado tiene caracter multiplicativo puede capturar estos efectos de una manera distinta.
- La distribución presenta gran flexibilidad de ajuste a los datos.

Modelos relacionados a la variable de incidencia Para el análisis de la variable de incidencia, se construye el histograma de frecuencias –análogo al realizado para la variable de ventas –, el cual se presenta en la Figura 3.2b.

En ésta –y en muchas otras aplicaciones– los datos de conteo se encuentran sobredispersados, donde la varianza condicional excede a la media condicional. Una solución a esto es el uso de la Binomial Negativa.

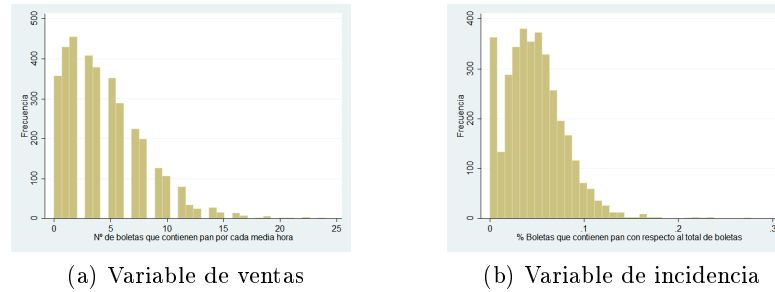


Figura 3.2: Histogramas

3.5. Categoría 1: Los principales 2 productos

Definición y resultados obtenidos

Los productos que se consideraron para este análisis son $j' = \{1, 2\}$. Es importante aclarar que éstos son los únicos productos que se venden a granel y corresponden a *bagels* y *rolls* respectivamente.

El análisis preliminar de las variables de *stock* a considerar es fundamental para la obtención de resultados congruentes, por lo tanto, se obtienen las correlaciones para $OOS_{j,t}^{cont}$ y $OOS_{j,t}^{bin}$ tanto para todo horario como también para los horarios en análisis. Para el caso del horario agregado en el día, ambas variables muestran una correlación muy alta (superior a 0,9). Si se desagrega por horario, la correlación para la variable continua baja a 0,87; en cambio, para la variable binaria se mantiene. Con esta información, se puede distinguir que el ajuste de β_1^{cont} y β_3^{bin} no será muy preciso, porque cuando la correlación entre regresores es alta, es difícil estimar con precisión los efectos individuales de cada uno de ellos.

Si el mismo análisis se realiza desagregado por hora, se ve que la correlación varía mucho dependiendo el horario. En las mañanas y en las tardes, las variables se comportan de manera muy similar, –con mínimos quiebres de *stock* en las primeras horas del día –y, en las tardes, con una gran cantidad de quiebres. En cambio, en el resto de los horarios analizados se puede ver baja correlación, donde los niveles de *stock* de cada producto siguen su propia tendencia.

Antes de presentar los resultados obtenidos, se debe mencionar que se probaron varios modelos, donde se consideró que el ajuste de las variables de *stock* sea razonable, que los controles añadidos puedan capturar los efectos estacionarios y que, al mismo tiempo, no se lleven parte de los efectos que las variables de *stock* deben explicar. Por lo tanto, los modelos que se consideraron en esta instancia fueron los siguientes:

	(1)	(2)	(3)	(4)	(5)
	<i>Ventas</i>	<i>Ventas</i>	<i>Ventas</i>	<i>Ventas prod. 1</i>	<i>Ventas prod. 2</i>
α	-2.736***	-2.759***	-2.720***	2.998***	-3.200***
	(-8.74)	(-8.80)	(-8.70)	(4.81)	(-5.66)
OOS_1^{cont}	-0.233*	0.508		-1.851***	-1.130***
	(-2.16)	(1.68)		(-4.16)	(-4.03)
OOS_2^{cont}	-0.0945	0.170		-0.345	-0.715**
	(-0.94)	(0.49)		(-1.08)	(-2.82)
$(OOS_1^{cont})^2$		-0.847*			
		(-2.50)			
$(OOS_2^{cont})^2$		-0.200			
		(-0.52)			
OOS_1^{bin}			-0.199*		
			(-2.01)		
OOS_2^{bin}			-0.140		
			(-1.49)		
Observations	2619	2619	2619	2619	2619
Log Lik.	-6136.6	-6132.8	-6135.7	-2176.2	-3392.0

t statistics in parentheses
 * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 3.8: Resultados modelos para variable de ventas y 2 productos

Los resultados de la variable dependiente de incidencia, se encuentran en el Anexo 2.

Análisis de los resultados

A continuación se realiza un análisis a los resultados para la variable dependiente de ventas obtenidos anteriormente. En la Tabla 3.9 se analizan los modelos (1), (3) y (4), que presentan variables lineales y en la Tabla 3.9 se analiza el modelo (2) que presenta variables cuadráticas con respecto a los niveles de *stock*.

	(1)	(3)	(4)	(5)
	Ventas	Ventas	Ventas producto 1	Ventas producto 2
OOS_1^{cont}	-20,78 %		-84,29 %	-67,70 %
OOS_2^{cont}	-9,0 %		-29,18 %	-51,08 %
OOS_1^{bin}		-18,05 %		
OOS_2^{bin}		-13,06 %		

Tabla 3.9: Variación porcentual de pasar de nivel de *stock* del 100 % a nivel 0 %

Para el caso del modelo (1), se puede observar en la Tabla 3.9 que los *bagels* reducen sus ventas cerca de un 20 % cuando no hay *stock* de ese producto; sin embargo, los *rolls* lo hacen en un 9 % solamente. Esta diferencia puede ser explicada por la existencia de una sustitución entre otros panes del mismo estilo, como puede ser *rolls* envasados. En cambio, los panes del tipo *bagel* no tienen un sustituto directo en la categoría. El modelo (4) muestra que, si hay un quiebre de *stock* de *bagel*, las ventas de éste caen un 84 % y cuando hay un quiebre en los *rolls*, las ventas de *bagel* bajan sólo

Nivel de stock	Reduc./Aum. Producto 1	Reduc./Aum. Producto 2
100 %	0,00 %	0,00 %
90 %	4,32 %	1,51 %
80 %	7,01 %	2,63 %
70 %	7,91 %	3,36 %
60 %	7,00 %	3,67 %
50 %	4,32 %	3,56 %
40 %	-0,01 %	3,05 %
30 %	-5,77 %	2,12 %
20 %	-12,69 %	0,80 %
10 %	-20,46 %	-0,90 %
0 %	-28,75 %	-2,96 %

Tabla 3.10: Análisis para modelo (2) con efectos cuadráticos

en 29 %. En el modelo (5) se da cuenta de que si no hay *stock* de *rolls*, las ventas de éste bajan un 51 %; sin embargo, si hay un quiebre de *stock* en los *bagels*, entonces las ventas de *rolls* disminuyen en 68 %. Este fenómeno parece contraproducente en un principio, pero puede suceder debido a que la gente, al no observar cierto producto en las canastas, decide no comprar pan a granel.

Con respecto al análisis de resultados del modelo (2), se puede notar que, desde el caso en que la mitad de los productos está disponible hasta el caso en que se tiene *stock* completo dentro de los *bagels*, se puede notar un aumento en las ventas. Esto se puede deber a efectos de complementariedad de diferentes tipos de *bagel*. Algo similar –aunque en mayor magnitud– sucede para el caso de los *rolls*, donde a un 20 % de disponibilidad de productos aún se ve aumento en las ventas. Luego, como es esperable, mientras menor es el grado de disponibilidad de los productos, mayor es la disminución de las ventas de los productos, debido a que el conjunto de productos comienza a ser menor, lo que causa que el conjunto de elección se reduzca y, por tanto, exista menor sustitución entre los productos.

3.6. Categoría 2: Los principales 5 productos

Definición y resultados obtenidos

El conjunto de productos está compuesto por $j = \{1, 2, 8, 10, 11\}$. Luego de la implementación de diversos modelos, se decidió utilizar los mismos descriptos para la categoría 1. De esta manera, los resultados obtenidos se presentan en la Tabla 3.11 para el caso de la variable dependiente de ventas y en la Tabla ?? para el caso de la variable de incidencia.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>Ventas</i>	<i>Ventas</i>	<i>Ventas p₁</i>	<i>Ventas p₂</i>	<i>Ventas p₃</i>	<i>Ventas p₄</i>	<i>Ventas p₅</i>
α	-2.743*** (-8.71)	-2.729*** (-8.68)	2.987*** (4.74)	-3.166*** (-5.53)	-9.492** (-2.98)	-4.513** (-2.62)	-4.042* (-2.07)
OOS_1^{cont}	-0.186 (-1.73)		-1.797*** (-4.03)	-1.074*** (-3.80)	-1.771* (-2.21)	-0.0667 (-0.11)	1.419** (2.84)
OOS_2^{cont}	-0.145 (-1.44)		-0.386 (-1.19)	-0.739** (-2.88)	0.882 (1.22)	-0.421 (-0.76)	-0.284 (-0.60)
OOS_3^{cont}	0.210 (0.81)		-0.258 (-0.26)	-0.538 (-0.76)	2.700*** (3.79)	0.681 (0.68)	-12.58 (-0.02)
OOS_4^{cont}	0.191*** (3.55)		0.0923 (0.59)	0.0815 (0.74)	0.369 (1.08)	-0.213 (-0.76)	-0.310 (-0.83)
OOS_5^{cont}	-0.584** (-2.75)		-0.470 (-0.66)	-0.561 (-1.11)	1.709* (2.31)	0.577 (0.80)	-12.76 (-0.03)
OOS_1^{bin}		-0.189 (-1.92)					
OOS_2^{bin}		-0.176 (-1.88)					
OOS_3^{bin}		0.205 (0.80)					
OOS_4^{bin}		0.197*** (3.68)					
OOS_5^{bin}		-0.625** (-2.95)					
Observation	2600	2600	2600	2600	2600	2600	2600
Log Lik.	-6082.2	-6080.1	-2164.4	-3361.3	-511.1	-1061.9	-840.4

t statistics in parentheses

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 3.11: Resultados modelo para variable de ventas para la categoría 2

Análisis de los resultados

Se procede a realizar un análisis similar al realizado para la categoría 1, donde los modelos (1) y (2) corresponden a los modelos con la variable dependiente que relaciona a todos los panes. Luego, los modelos (3), (4), (5), (6) y (7) hacen referencia a los productos 1, 2, 3 y 4 respectivamente, como es mostrado en la Tabla 3.2. Cabe destacar que no se incluyó un modelo donde el *stock* presentara efectos cuadráticos, debido a que su análisis para este caso es complejo porque el conjunto de productos a analizar es mayor.

	(1) Ventas	(2) Ventas	(3) Ventas p_1	(4) Ventas p_2	(5) Ventas p_3	(6) Ventas p_4	(7) Ventas p_5
OOS_1^{cont}	-16,97 %		-83,42 %	-65,84 %	-82,98 %	-6,45 %	313,30 %
OOS_2^{cont}	-13,50 %		-32,02 %	-52,24 %	141,57 %	-34,36 %	-24,72 %
OOS_3^{cont}	23,37 %		-22,74 %	-41,61 %	1387,97 %	97,59 %	-100,00 %
OOS_4^{cont}	21,05 %		9,67 %	8,49 %	44,63 %	-19,18 %	-26,66 %
OOS_5^{cont}	-44,23 %		-37,50 %	-42,94 %	452,34 %	78,07 %	-100,00 %
OOS_1^{bin}		-17,22 %					
OOS_2^{bin}		-16,14 %					
OOS_3^{bin}		22,75 %					
OOS_4^{bin}		21,77 %					
OOS_5^{bin}		-46,47 %					

Tabla 3.12: Variación porcentual al pasar de nivel de *stock* actual a nivel del 0 %

Al ocurrir quiebres de *stock* relacionados a los productos 1 y 2 se observa disminución en las ventas totales tanto para el modelo (1) como para el modelo (2). También se debe destacar que, para estos dos modelos, las ventas de los productos 3 y 4 aumentan. Se cree que esto ocurre debido al efecto de sustitución que existe dentro de estos panes, que son los más consumidos por los clientes.

Se puede ver que un quiebre del producto j , con $j = \{1, 2, 3, 4, 5\}$ afecta negativamente las ventas de ese mismo producto j para todos los productos, exceptuando el producto 3. El modelo asociado, que corresponde al modelo (5) en la Tabla 3.11, presenta mal ajuste de los datos, donde la *log*-verosimilitud entrega un valor mucho mayor a lo esperado. Por lo tanto, se omite análisis respecto a este modelo.

En general, se ve una disminución de las ventas de cierto producto j cuando existe quiebre de *stock* de cualquiera de los otros productos k , con $k \neq j$. Sin embargo, existen casos particulares que vale la pena mencionar. Al ocurrir un quiebre del producto 4, aumentan las ventas tanto del producto 1 como del producto 2. Sucede algo similar cuando ocurren quiebres de *stock* en el producto 3, con aumentos en las ventas del producto 4. Por último, un quiebre de *stock* del producto 1 causa que las ventas del producto 5 se cuadrupliquen.

3.7. Categoría 3: Rolls envasados v/s a granel

Definición y resultados obtenidos

Como se puede ver en la Tabla 3.2, los productos cuyo código es 1 corresponden a los *rolls* que se venden a granel; en cambio, los *rolls* que se venden envasados corresponden a los productos con código 3, 5, 6 y 10 (*Brown and Serve Rolls / Butter Flake Rolls*, *Chicago Rolls*, *Club Rolls* y *Kaiser Rolls* respectivamente). Por lo tanto, se crearon dos categorías: a granel (*bulk*) y empaquetado (*packaged*). La categoría *bulk* contiene un sólo producto, en cambio, la categoría *packaged*, que contiene un grupo de productos, considera las ventas de la categoría la suma de los productos asociadas a ésta y el *stock* se calcula mediante el promedio simple.

Los resultados obtenidos se presentan en la Tabla 3.13 para la variable de ventas y en la Tabla ?? para la variable de incidencia.

	(1)	(2)	(3)
	<i>Ventas</i>	<i>Ventas^{granel}</i>	<i>Ventas^{empaqa}</i>
α	-2.701*** (-8.55)	-3.138*** (-5.53)	-6.221*** (-7.53)
β_1^{granel}	-0.231** (-2.99)	-1.293*** (-5.76)	0.00355 (0.02)
β_1^{empaqa}	-0.0969 (-0.34)	-0.403 (-0.69)	0.263 (0.46)
Observations	2600	2600	2600
Log Lik.	-6095.1	-3371.1	-2952.1

t statistics in parentheses
 * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 3.13: Resultados modelos para variable de ventas

Análisis de los resultados

El análisis de los resultados se encuentra en la Tabla 3.14 a continuación:

	(1)	(2)	(3)
	<i>Ventas</i>	<i>Ventas^{granel}</i>	<i>Ventas^{empaquetado}</i>
OOS_{granel}^{cont}	-20,63 %	-72,56 %	0,36 %
OOS_{empaqa}^{cont}	-9,24 %	-33,17 %	30,08 %

Tabla 3.14: Variación porcentual al pasar de nivel de *stock* actual al 0 %

El modelo (1) refleja que, con un quiebre de *stock* del grupo de los *rolls* a granel, entonces las ventas totales de pan bajan en un 20 %; en cambio, con un quiebre de *stock* del grupo de los *rolls* empaquetados, éstas bajan solamente en un 9 %. Esto puede deberse a la mayor importancia que tiene para los consumidores el pan fresco sobre los envasados en la categoría de los *rolls*.

El modelo (2) muestra que, con un quiebre de *stock* de los *rolls* a granel, entonces las ventas de éste bajan en un 72 % aproximadamente y con un quiebre de *stock* del grupo de los empaquetados, las ventas a granel disminuyen cerca de un 33 %. Ésta última implicancia se cumple bajo el supuesto de que los productos son sustitutos.

Por último, el modelo (3) dice que si el *stock* de los *rolls* empaquetados se quiebra, entonces las ventas de ese mismo producto aumentan en en 30 %. Este resultado está fuera de todo lo esperado y se cree que ocurre debido a que, para ajustar los datos, la constante captura gran parte del efecto, entonces el ajuste de las variables de *stock* es inadecuado (las variables no son significativas a ninguno de los niveles de significancia aceptables). Es muy probable que si se trabajaran estos datos a nivel de cliente, los resultados serían muy distintos.

El análisis realizado en este capítulo deja abiertas muchas preguntas sobre el comportamiento de los clientes con respecto a las compras realizadas en relación a los quiebres de *stock* de

los productos. Se espera que un análisis a nivel individual ayude a despejar las dudas que quedaron abiertas debido a mal ajuste con los datos agregados y a corroborar los resultados significativos y coherentes que se obtuvieron. Para estudiar estos efectos, se da paso al próximo capítulo, que trata a cada cliente individualmente mediante un modelo de elección discreta.

Capítulo 4

Modelo individual por cliente

4.1. Motivación

El objetivo de este capítulo es entender el efecto de la disminución de variedad de los productos dentro de la utilidad con el modelo de elección discreta utilizado. Para esto se utiliza, la información y deducciones realizadas en el capítulo anterior, donde se muestra el estudio agregado de los clientes en conjunto con los propósitos particulares que se plantean para este capítulo. Estos corresponden principalmente al estudio de 3 efectos: Patrones de sustitución, efectos cruzados y lealtad de los clientes.

4.2. Conjunto de clientes a considerar

Como información preliminar, se debe mencionar que existen 46.709 clientes que han comprado en el supermercado en el periodo de compra estipulado, es decir, en el periodo de tiempo en donde se cruzan los datos transaccionales y los datos de *stock*. Estos periodos están descritos en detalle en el capítulo anterior. Del conjunto de clientes anteriormente mencionado, 6.485 de ellos han realizado compras de pan.

Es importante tener una idea clara del comportamiento de compra de estos clientes y saber qué tan importante es retenerlos y fidelizarlos. Por esta razón, se decide observar tanto la cantidad total de dólares gastados y la cantidad de productos en cada ocasión de compra, como también la cantidad de veces que visita el supermercado en el periodo en estudio. La información anteriormente descrita se presenta en la Tabla 4.1.

Variable	Quienes compran pan	Quienes no compran pan	Total
Dólares gastados por compra	56,4	37,5	40,1
Nº de productos adquiridos por compra	12,4	8,1	8,7
Nº de compras totales realizadas	16,0 (1 vez a la semana)	4,3 (1 vez al mes)	6,0 (2,5 veces al mes)
Nº de observaciones	6485 (13,9 %)	40224 (86,1 %)	46709 (100 %)

Tabla 4.1: Variables comportamiento de compra clientes

Se puede observar que los clientes que compran pan en su historial de compra (que corresponde al 13,9 %), compran en promedio más productos, gastan mayor cantidad de dólares por compra (cercano a un 50 % más en ambos casos) y la frecuencia con la que visita la sala de compras es cerca de 4 veces mayor. Por lo tanto, se puede decir que es un grupo de clientes rentable e interesante de estudiar. Debido a lo anterior, se trabajará con los clientes que han comprado alguna vez pan en su historial de compra.

Para continuar con el análisis, es menester considerar el siguiente supuesto:

“Si dos compras de un mismo cliente fueron realizadas con diferencia horaria menor o igual a 30 minutos, entonces éstas se consideran como una misma compra”.

Este supuesto se realiza debido a que se considera que, si un cliente vuelve a realizar una compra dentro de un intervalo tan acotado de tiempo, entonces es probable que haya vuelto por algún producto que no pudo adquirir la primera vez o, igualmente, que puede haber olvidado. Estos productos, bajo esa premisa, deberían ser considerados para los análisis como productos que fueron adquiridos en la misma ocasión de compra.

Desde ahora en adelante se trabaja bajo el supuesto anteriormente mencionado. Si se conserva a aquellos clientes cuya frecuencia de compra es, en promedio, 1 vez cada dos semanas - correspondiente a aproximadamente 8 compras en la historia de compra de un cliente, entonces el universo queda reducido a 3.758 clientes. Si se considera una cota superior en cuanto a frecuencia de compra de los clientes, una buena idea es trabajar con aquellos que han realizado a lo más 1 compra al día en promedio lo que equivale a 118 compras. Así, el universo muestral total es de 3.746 clientes.

Si se crea una cota inferior más restrictiva para la frecuencia de compra y se estudia a aquellos que han comprado al menos una vez a la semana en promedio, es decir, quienes han realizado un número igual o mayor a 16 compras, entonces el total es 2.397 clientes. Esta última restricción es más adecuada debido a que, a pesar de que el universo de clientes se reduce en un 36 %, se obtiene

un grupo de clientes que es más fiel al supermercado y, por lo tanto, será más fácil poder estimar el patrón de compras que presentan.

Adicionalmente, en la Tabla 4.2 se presenta la frecuencia de clientes remanente categorizada mediante el máximo de panes que han comprado en alguna ocasión de compra. Con esto se busca observar la diversidad de tipos de pan que estos clientes compran en cada ocasión de compra.

Máximo de PLU's comprados por boleta	Frecuencia Clientes	Porcentaje
1	2043	85,23 %
2	325	13,56 %
3	26	1,08 %
4	3	0,13 %
Total	2397	100 %

Tabla 4.2: Máximo de panes comprados por boleta, agregado por cliente

Al observar la frecuencia de clientes en cada categoría, se puede notar que los clientes que compran 3 y 4 productos de pan en la misma boleta como máximo con un porcentaje despreciable del total de clientes, correspondiente a un 1,21 % del total remanente. Por lo tanto, se decide eliminar a estos clientes del conjunto a utilizar.

Por otro lado, luego de analizar el grupo de clientes que ha comprado dos productos como máximo en una misma boleta, se pudo notar que no existe una combinación de dos productos que sean comprados en una boleta que sea predominante sobre las otras. Esto se puede deber a que la cantidad de combinaciones que se puede lograr sobre todos los productos disponibles es muy grande; por lo tanto, se decide eliminar del conjunto de clientes a estos últimos.

4.3. Definición de productos y categorías

El conjunto de productos que se posee para el análisis es inmenso, lo cual hace difícil el manejo de éstos. Adicionalmente, se poseen productos con muy baja participación de mercado. Por lo tanto, se decide considerar a aquellos clientes que han comprado alguno de los diez (10) productos más vendidos de la categoría dentro de su historia de compra. Esto considera que también pueden haber comprado alguno de los productos que no están dentro de esta categoría. La forma de manejar esta información será mencionada más adelante.

Para conocer más a los clientes que pertenecen a este grupo de productos, se desglosa la información de cantidad de productos comprados por boleta con respecto a quienes que se encuentran dentro de éste en comparación con aquellos que no se encuentran. Esta información está altamente correlacionada con la cantidad de dinero gastado en cada compra. La información obtenida se encuentra descrita en la Tabla a continuación.

	Promedio de compra por boleta	Frecuencia
No han comprado de los <i>top</i> 10 productos	5,61	359 (17,57%)
Han comprado de los <i>top</i> 10 productos	10,35	1684 (82,43%)
Total	9,51	2043 (100%)

Tabla 4.3: Estadísticos diferenciado por quienes han o no comprado los *top* 10 productos

Se puede ver que los clientes que compran dentro del grupo de los panes más consumidos, compran mayor cantidad de ítemes por boleta, correspondiente aproximadamente al doble de quienes no compran alguno de esos productos en sus boletas, lo cual sugiere que son clientes mucho más atractivos para el estudio en cuestión.

Los productos que se consideran corresponden al conjunto de alternativas $\{1, 2, 4, 8, 10, 11, 14, 15, 16, 25\}$. El desglose de éstos se muestra en la Tabla 4.4. Los productos que no están dentro del conjunto nombrado anteriormente que son comprados por el conjunto este grupo de clientes de igual manera, se encuentran agrupados en un producto que fue denominado como alternativa 99. El conjunto de las alternativas de compra queda descrito como conjunto C .

Es decir, formalmente:

$$C = \{1, 2, 4, 8, 10, 11, 14, 15, 16, 25, 99\}$$

La alternativa de no compra de un producto se nominó como alternativa 0, donde se define el conjunto NC como:

$$NC = \{0\}.$$

Por lo tanto, el conjunto total de productos considerado queda determinado por lo siguiente:

$$\mathcal{P} = C \cup NC$$

Código	Producto	Frecuencia	Participación de mercado con alternativa de no compra	Participación de mercado sin alternativa de no compra
0	No compra	58518	87,55 %	
1	<i>Bagels</i> (agrupados)	740	1,11 %	8,89 %
2	<i>Rolls</i> (agrupados)	2010	3,01 %	24,14 %
4	<i>Challah</i>	246	0,37 %	2,95 %
8	<i>French Bread</i>	950	1,42 %	11,41 %
10	<i>Kaiser Rolls</i>	985	1,47 %	11,83 %
11	<i>Small French Bread</i>	955	1,43 %	11,47 %
14	<i>Take & Bake Club</i>	280	0,42 %	3,36 %
15	<i>Take & Bake Small French Bread</i>	454	0,68 %	5,45 %
16	<i>Take & Bake Small Petite and Wheat Petite</i>	272	0,41 %	3,27 %
25	<i>White Sub</i>	305	0,46 %	3,66 %
99	Conjunto de otros productos	1128	1,69 %	13,55 %

Tabla 4.4: Participación de mercado de los productos más comprados

La incidencia de la categoría referente al total de compras realizadas luego de la segmentación efectuada (sin diferenciar por clientes), corresponde a 12,45 %, que es mucho mayor a la incidencia real de mercado de la categoría de panes. Se puede observar que el producto que tiene mayor participación de mercado es el producto que agrupa a los todos los panes del tipo *roll*. En segundo lugar, se encuentra la agrupación de los 16 productos restantes, que se denomina alternativa 99. Luego, se encuentran en similar proporción de participación los siguientes panes: *French Bread*, *Kaiser Rolls* y *Small French Bread*.

Cabe destacar que, luego de la segmentación, la cantidad de panes de tipo *bagel* disminuye notablemente, siendo, dentro de esta definición de alternativas, la sexta alternativa más comprada (cabe recordar que antes de los filtros utilizados ésta era la segunda categoría más comprada). Esta baja se explica debido a que los panes de tipo *bagel* pueden ser comprados por clientes más ocasionales.

A continuación, en la Figura 4.1, se encuentra el histograma de los clientes con respecto a la incidencia de compras de éstos en la categoría de pan.

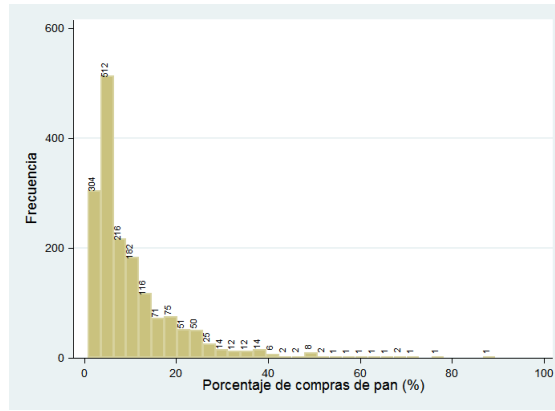


Figura 4.1: Histograma de la incidencia de compra de pan

Es importante notar que, dentro de la información que arroja el histograma, se puede notar que el mayor porcentaje de clientes (63,78 %) se concentra en el grupo en el cual el 10 % o menos de las compras corresponden a compras que contienen pan. Si se observa al grupo en el cual el 20 % o menos de las compras totales corresponden a compras de pan, entonces este porcentaje alcanza un 87,05 %. Lo cual quiere decir que los clientes están mayormente concentrados en el segmento de baja incidencia de pan. Sin embargo, se puede observar a clientes que alcanzan incidencias de compra alrededor del 80 % en la categoría.

Asimismo, se debe mencionar que la media de este grupo se encuentra en un 10,35 % de incidencia de compras en la categoría, con una desviación estándar de 9,63 %.

Para este estudio, se pondrá especial énfasis en la definición capturar los efectos mencionados en el capítulo anterior, según la siguiente distinción por niveles jerárquicos:

1. Productos dentro de la categoría $f \in F$, donde $F = \{granel, empaquetado\}$
2. Las categorías $f \in F$
3. La categoría pan, que incluye el conjunto total de productos.

Donde se considera que los categorías $f \in F$ quedan descritas de la siguiente manera:

$$granel = \{1, 2\}$$

$$empaquetado = \{4, 8, 10, 11, 14, 15, 16, 25\}$$

Se debe mencionar que la segmentación realizada corresponde a una partición \mathcal{P} , donde el conjunto es una colección de subconjuntos no vacíos de \mathcal{P} , disjuntos por pares, es decir, donde la unión de *granel*, *empaquetado* y *NC* corresponden al universo total y la intersección de estos elementos es vacía.

El modelo, por construcción, captura cuáles son los productos que pertenecen a cada subcategoría, es decir, distingue productos que se venden a granel de aquellos empaquetados. La categoría de panes intenta capturar el efecto dentro de la demanda total de los productos.

4.4. Variables del modelo

Las variables que se consideran para el modelo están basadas tanto en los efectos que quieren ser capturados como también están basadas en el análisis agregado que fue realizado en el capítulo anterior. A continuación se describirán la variable dependiente y las variables independientes que fueron consideradas en el modelo.

La variable dependiente que se considera es la siguiente:

$$y_{j,t}^i = \begin{cases} 1 & \text{si el cliente } i \text{ compra el producto } j \text{ en la ocasión de compra } t \\ 0 & \text{en otro caso} \end{cases}$$

Las variables independientes a considerar en el modelo de elección discreta se encuentran descritas en mayor detalle a continuación.

1. Variables binarias por alternativa

$$d_j = \begin{cases} 1 & \text{si el producto en análisis es el producto } j \\ 0 & \text{en otro caso} \end{cases}$$

La variable descrita anteriormente ayuda a capturar cualquier efecto que corresponda únicamente a la alternativa y que no sea capturado por alguna otra de las variables explicatorias. Esto ayuda a flexibilizar el modelo en cuanto al ajuste a las categorías Guadagni and Little (1983)

2. Variables binarias por categoría

$$d_F = \begin{cases} 1 & \text{si el producto } j \text{ en análisis pertenece a la categoría } F = F(j) \\ 0 & \text{en otro caso} \end{cases}$$

Al igual que para la variable anterior, se debe mencionar que esta variable aporta a la captura de cualquier efecto que corresponda únicamente a la categoría y que no sea capturado por alguna de las variables explicatorias

3. Variables de variedad / disponibilidad de productos

Se construyen 3 variables sobre el grado de no disponibilidad del producto para los clientes en cierta ocasión de compra t para las jerarquías definidas anteriormente:

- Para cada producto

OOS_{jt} = Grado de no-disponibilidad del producto j para el cliente i en la ocasión de compra t , donde $OOS_{jt} \in [0, 1]$

- Para cada categoría

$OOS_{f,t}$ = Grado de no-disponibilidad en la categoría f para el cliente i en la ocasión de compra t , donde $OOS_{f,t} \in [0, 1]$

Esto se calcula de la siguiente manera:

$$OOS_{f,t} = \frac{\sum_{j \in f} OOS_{jt}}{\text{card}(f)}, f \in F$$

donde la cardinalidad correspondiente a cada valor de f es:

$$\text{card}(\text{granel}) = 2$$

$$\text{card}(\text{empaquetado}) = 9$$

- Para el conjunto de productos

OOS_t = Grado de no-disponibilidad promedio para los productos para el cliente i en la ocasión de compra t , donde $OOS_t \in [0, 1]$

$$OOS_t = \frac{\sum_{j \in C} OOS_{jt}}{\text{card}(C)}, j \in C$$

donde la cardinalidad correspondiente a C es:

$$\text{card}(C) = 11$$

4. Variables de lealtad

- Variables de Lealtad a cada producto

$$LEAL_{j,t}^i = \alpha LEAL_{j,t-1}^i + (1 - \alpha) z_{j,t}^i$$

donde

$$z_{j,t}^i = \begin{cases} 1 & \text{si el consumidor } i \text{ compra producto } j \text{ en la ocasión de compra } t - 1 \\ 0 & \text{en otro caso} \end{cases}$$

La inicialización de la variable se realiza de la siguiente manera:

$$LEAL_{j,1}^i = \begin{cases} \alpha & \text{si el producto } j \text{ es el primer producto} \\ & \text{comprado en historia del cliente } i \\ \frac{1 - \alpha}{\text{card}(C) - 1} & \text{en el caso contrario} \end{cases}$$

donde $\text{card}(C) = 11$, que corresponde a la cardinalidad del conjunto sin la opción de no compra, para la cual no está asociada una lealtad. El valor de α se considera igual a 0,8, como está estipulado en la literatura Guadagni and Little (1983).

La razón por la cual se considera que, en el caso de la primera compra el valor estipulado sea $\frac{1 - \alpha}{\text{card}(C) - 1}$ para el caso de no comprar el producto j se debe a que, de esta manera, se asegura que la suma de las lealtades entre los productos sea igual a 1 para cada consumidor.

- Variables de Lealtad a la categoría de pan (no considera subíndice de producto)

$$LEAL_t^i = \alpha LEAL_{t-1}^i + (1 - \alpha)z_t^i$$

donde

$$z_t^i = \begin{cases} 1 & \text{si el consumidor } i \text{ compra algún producto de la categoría de pan} \\ & \text{en la ocasión de compra } t - 1 \\ 0 & \text{en otro caso} \end{cases}$$

La inicialización de la variable se realiza de manera análoga al caso del análisis de cada producto:

$$LEAL_1^i = \begin{cases} \alpha & \text{si algún producto de la categoría es el primer} \\ & \text{producto comprado en historia del cliente } i \\ \frac{1 - \alpha}{\text{card}\{CP, NC\} - 1} & \text{en el caso contrario} \end{cases}$$

En este caso se considera que la cardinalidad del conjunto es $\text{card}\{CP, NC\} = 2$, debido a que se evalúa entre dos alternativas, que corresponden a las alternativas de comprar algún tipo de pan (CP) o no comprar ninguno (NC). En este caso se considera igualmente que el valor de α es $\alpha = 0,8$.

Las variables de lealtad, por construcción, toman valores entre 0 y 1. Éstas son consideradas como el promedio exponencialmente ponderado de las compras pasadas del producto.

El cálculo de los valores de α según la literatura recientemente mencionada tanto para las variables de lealtad asociadas a los productos como a la variable de lealtad asociada a la categoría de panes se realiza mediante una calibración de los valores de éstos en la construcción del mismo modelo. Esto se realiza modificando los valores que puede tomar α y busca el valor máximo que toma la función de máxima verosimilitud con el valor óptimo de α . Por simplicidad, y como ha sido realizado y estudiado en varios otros trabajos (Guadagni and Little (1983)), este valor se fija en 0,8. Según la literatura mencionada anteriormente, este valor que se ajusta relativamente bien a los requerimientos.

Es importante destacar que las variables de lealtad capturan la probabilidad de volver a comprar cierto producto en particular (o la categoría), pero no captura el efecto que produce las

compras recientes dentro de la probabilidad de comprar. Por lo tanto, se hace imprescindible incluir el efecto de las compras recientes dentro del modelo. Esta variable se encuentra descrita a continuación.

5. Variables de diferencia de días

REC_t^i =Diferencia continua (en días) en la compra realizada entre las ocasiones de compra t y $t - 1$ para el cliente i de cualquier producto en el supermercado (*recency*). Para el caso de primera compra del cliente, esta variable toma valor cero.

Para ejemplificar el uso de esta variable se puede considerar que, si un cliente compra el 8 de agosto de 2010 a las 10:00 a.m. y la siguiente compra realizada ocurre el 10 de agosto de 2010 a las 12:30 a.m., entonces la variable toma el valor $REC_t^i = 2,104$.

En la Tabla 4.5 se puede observar los valores de la variable recientemente descrita para distintos percentiles. De esta información se puede observar que la mayoría de las compra son realizadas con poca diferencia de días. De hecho, cerca del 75 % de las compras son realizadas dentro de un rango de 5 días en comparación con la anterior ocasión de compra.

Percentiles		Percentiles	
25 %	1,07	90 %	8,90
50 %	2,71	95 %	14,05
75 %	5,05	99 %	47,88

Tabla 4.5: Percentiles para variable *recency*

6. Variables de control por estacionalidad

- Variables binarias de días de la semana. Se define de la siguiente manera:

$$DAY_{\delta(t)} = \begin{cases} 1 & \text{si la ocasión de compra } t \text{ ocurre en el día } D_{\delta}, \text{ con } \delta = \delta(t) \\ 0 & \text{en otro caso} \end{cases}$$

Donde $\{D_1 = \text{lunes}, D_2 = \text{martes}, D_3 = \text{miércoles}, D_4 = \text{jueves}, D_5 = \text{viernes}, D_6 = \text{sábado}, D_7 = \text{domingo}\}$

Esto ayuda a capturar las variaciones que ocurren dentro de los días de la semana, así como las tendencias entre éstos.

- Variables de control binarias sobre horas del día. La variable asociada se define de la siguiente manera:

$$H_{h(t)} = \begin{cases} 1 & \text{si la ocasión de compra } t \text{ ocurre en la hora } h = h(t) \\ 0 & \text{en otro caso} \end{cases}$$

donde $h \in [8 \text{ A.M.}, 10 \text{ P.M.}]$

Esto ayuda a capturar las variaciones que ocurren dentro de los días de la semanas, así como

las tendencias entre éstos.

4.5. Modelo utilizado

El modelo construido con las variables anteriormente descritas es el siguiente:

$$U_j^i = \sum_{j \in C} \alpha_{i,j} d_j + \sum_{f \in F'} \alpha_f d_f + \sum_j LOY_j \cdot LEAL_{jt}^i + LOY_{pan} LEAL_t^i + \sum_h H_h \cdot H_{h(t)} + \sum_d DAY_d DAY_{d(t)} + \kappa REC_t^i + ST_{jt}$$

donde ST_{jt} corresponde a las variables de disponibilidad de productos, las cuales se encuentran reflejadas en el modelo de la siguiente manera:

$$ST_{jt} = OOS_{1,g} \cdot OOS_{jt} \cdot \mathbb{1}_{(j \in g)} + OOS_{1,e} \cdot OOS_{jt} \cdot \mathbb{1}_{(j \in e)} + OS_{2,g} \cdot OOS_{g,t} \cdot \mathbb{1}_{(j \in g)} \\ + OOS_{2,e} \cdot OOS_{e,t} \cdot \mathbb{1}_{(j \in e)} + OOS_{3,g} \cdot OOS_t \cdot \mathbb{1}_{(j \in e)} + OS_{3,e} \cdot OOS_t \cdot \mathbb{1}_{(j \in e)}$$

Como forma de simplificar la notación, se decidió denominar g al conjunto de los productos a granel y e al conjunto de los productos empaquetados.

. Es importante mencionar que, debido a la colinealidad existente entre las variables binarias construidas para el modelo, no es posible incluir la totalidad de ellas, ya que el modelo no acepta variables que sean linealmente dependientes entre sí. Por lo tanto, la elección de las variables a incluir en el modelo se realiza de la siguiente manera:

En primer lugar, se incluyen las variables que son indispensables para el modelo. Como interesa capturar el efecto de las categorías sobre el comportamiento de los clientes, entonces se consideró menester adjuntar éstas al modelo utilizado, es decir, las variables binarias por categoría.

Luego, agregar de manera conveniente el máximo de variables posible sin que se retorne al problema de colinealidad de éstas. Después de un estudio de dependencia lineal de las variables, se pudo determinar que una combinación factible de variables únicas a cada alternativa que pueden ser incluidas son las referentes a las alternativas 8, 10, 11, 14, 15, 16 y 25.

Es de suma importancia mencionar que la variable $OOS_{g,t}$ fue omitida del estudio debido a que presenta una correlación mayor a 0,9 con la variable OOS_{jt} cuando $j \in g$. Se observó la correlación de estas variables en distintos bloques horarios y también era muy alta. Esto ocurre debido a que el movimiento de la disponibilidad de los dos productos que posee esta categoría es muy similar. Se decidió incluir OOS_{jt} en vez de $OOS_{g,t}$ porque es una variable más rica en cuanto a que posee mayor información.

El modelo que presenta coeficientes aleatorios difiere del otro modelo en la componente donde se estudia el gusto por los productos a granel o empaquetados. Este modelo incluye en los interceptos que se observan para las categorías coeficientes asociados a cada cliente, los cuales siguen una distribución normal $\alpha_{i,j} \sim N(\mu_j, \sigma_j)$ y rompe con los supuestos implícitos del logit tradicional.

4.6. Resultados del Modelo

Se muestran a continuación los resultados tanto para el modelo Logit tradicional y para el modelo *random coefficient* Logit.

A continuación se muestran los resultados del modelo *random coefficient* logit en las Tablas 4.6 y 4.7.

Los resultados obtenidos de los modelos descritos se encuentran a continuación:

	(1) <i>choice</i>	(2) <i>choice</i>	(3) <i>choice</i>
Mean			
H_8	0.619*** (3.97)	0.625*** (4.00)	0.769*** (4.32)
H_9	0.545*** (3.77)	0.549*** (3.80)	0.695*** (4.13)
H_{10}	0.438** (3.18)	0.455*** (3.29)	0.605*** (3.70)
H_{11}	0.395** (2.95)	0.418** (3.11)	0.567*** (3.55)
H_{12}	0.363** (2.78)	0.387** (2.94)	0.536*** (3.41)
H_{13}	0.353** (2.70)	0.374** (2.84)	0.523*** (3.32)
H_{14}	0.252 (1.92)	0.270* (2.05)	0.419** (2.66)
H_{15}	0.178 (1.36)	0.198 (1.51)	0.348* (2.21)
H_{16}	0.301* (2.35)	0.320* (2.49)	0.470** (3.02)
H_{17}	0.239 (1.86)	0.254* (1.98)	0.402** (2.60)
H_{18}	0.0975 (0.75)	0.111 (0.85)	0.257 (1.64)
H_{19}	0.0733 (0.54)	0.0880 (0.65)	0.196 (1.31)
H_{20}	-0.0883 (-0.59)	-0.0825 (-0.55)	-0.0542 (-0.36)
DAY_1	-0.232*** (-4.19)	-0.239*** (-4.32)	-0.238*** (-4.30)
DAY_2	-0.254*** (-4.17)	-0.257*** (-4.22)	-0.254*** (-4.18)
DAY_3	-0.363*** (-6.10)	-0.366*** (-6.14)	-0.363*** (-6.09)
DAY_4	-0.200*** (-3.61)	-0.199*** (-3.59)	-0.194*** (-3.50)
DAY_5	-0.228*** (-4.13)	-0.229*** (-4.16)	-0.226*** (-4.09)
DAY_6	-0.152** (-2.86)	-0.150** (-2.82)	-0.148** (-2.77)
REC	0.00641*** (3.70)	0.00659*** (3.79)	0.00661*** (3.81)
Observations	637608	637608	637608
Log Lik.	-24516.5	-24512.5	-24510.8

t statistics in parentheses

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 4.6: Resultados de las variables de control del modelo *random coefficient logit*

	(1)	(2)	(3)
	<i>choice</i>	<i>choice</i>	<i>choice</i>
Mean			
α_8	0.422*** (6.80)	0.419*** (6.76)	0.419*** (6.76)
α_{10}	0.133 (1.96)	0.131 (1.93)	0.131 (1.93)
α_{11}	0.435*** (7.02)	0.435*** (7.02)	0.435*** (7.02)
α_{14}	-0.794*** (-8.31)	-0.793*** (-8.29)	-0.793*** (-8.29)
α_{15}	-0.270*** (-3.48)	-0.270*** (-3.47)	-0.270*** (-3.48)
α_{16}	-0.629*** (-7.03)	-0.630*** (-7.05)	-0.630*** (-7.05)
α_{25}	-0.731*** (-7.85)	-0.732*** (-7.86)	-0.732*** (-7.87)
LOY_1	7.727*** (22.79)	7.746*** (22.86)	7.748*** (22.86)
LOY_2	7.768*** (30.02)	7.788*** (30.06)	7.793*** (30.07)
LOY_4	10.08*** (19.36)	10.08*** (19.36)	10.10*** (19.38)
LOY_8	8.572*** (30.16)	8.574*** (30.15)	8.577*** (30.16)
LOY_{10}	9.439*** (23.55)	9.439*** (23.52)	9.438*** (23.51)
LOY_{11}	8.623*** (28.38)	8.623*** (28.35)	8.630*** (28.35)
LOY_{14}	11.86*** (18.16)	11.88*** (18.15)	11.89*** (18.16)
LOY_{15}	9.767*** (25.56)	9.758*** (25.52)	9.768*** (25.54)
LOY_{16}	9.876*** (19.10)	9.853*** (19.03)	9.852*** (19.02)
LOY_{25}	10.65*** (19.44)	10.65*** (19.44)	10.66*** (19.44)
LOY_{99}	8.876*** (28.64)	8.873*** (28.61)	8.876*** (28.61)
LOY_{pan}	-3.303*** (-17.59)	-3.324*** (-17.66)	-3.339*** (-17.70)
$OOS_{1,g}$	-1.817*** (-7.17)	-1.062 (-1.63)	-1.043 (-1.61)
$OOS_{1,e}$	-0.625** (-2.60)	-0.779** (-3.14)	-0.781** (-3.15)
$OOS_{2,g}$		1.241* (2.55)	0.177 (0.22)
$OOS_{2,e}$		-0.827 (-1.22)	-0.872 (-1.22)
$OOS_{3,g}$			0.585 (0.64)
$OOS_{3,e}$			1.162 (1.75)
α_g	-4.613*** (-33.44)	-4.625*** (-33.41)	-4.777*** (-29.28)
α_e	-5.074*** (-37.99)	-5.101*** (-37.97)	-5.253*** (-32.65)
SD			
α_g	1.146*** (21.76)	1.146*** (21.76)	1.148*** (21.77)
α_e	0.911*** (28.49)	0.913*** (28.53)	0.914*** (28.55)
Observations	637608	637608	637608
Log Lik.	-24516.5	-24512.5	-24510.8

t statistics in parentheses

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 4.7: Resultados de variables en estudio modelo *random coefficient logit*

4.6.1. Consideraciones del modelo

Las simulaciones fueron realizadas mediante el Software Stata. Para el modelo logit de coeficientes aleatorios se debieron tomar las siguientes consideraciones:

El resultado fue obtenido mediante la utilización de Máxima Verosimilitud Simulada (MSL). Para la implementación de este método, donde se consideró un número de replicaciones igual a 50 y se eliminaron los primeros 15 elementos de la secuencia inicial.

Con respecto a las variables determinadas para el modelo final, se debe mencionar que se realizaron distintas pruebas con variables que no fueron incluidas en el modelo. La elección de las variables se realizó tanto mediante los efectos que se querían capturar a través de éstas como también de criterios estadísticos para la inclusión o exclusión de una variable dentro del modelo. Los dos criterios más comunes para realizar esta tarea son el *Akaike Information Criteria* (AIC) y *Bayesian Information Criteria* (BIC), que sirve para determinar la cantidad de variables a añadir al modelo sin caer en sobreajuste.

Se discutió la posibilidad de incluir variables de lealtad que, además de actualizarse mediante cada ocasión de compra del cliente, incluyeran de alguna manera el tiempo que ha transcurrido desde la última compra. Finalmente, luego de probar variables definidas para esta tarea, se determinó que la mejor opción era incluirlas como variables independientes.

Dentro de la discusión de las variables es inminentemente necesaria la inclusión de variables de grado de disponibilidad (o de variedad), donde las formas de incluirlas son tan amplias como sea la imaginación de los investigadores. Para este caso en particular, se veló por cumplir de forma esencial el propósito de la captura de los distintos efectos para cada categoría.

Calidad del ajuste

Si se observan los modelos presentados anteriormente, se puede ver que el ajuste logrado por el modelo de coeficientes aleatorios es mejor en cuanto a la verosimilitud obtenida que el modelo logit multinomial común. Esto quiere decir que el modelo de coeficientes aleatorios captura de mejor forma los distintos efectos y patrones existentes en los datos, debido a la flexibilidad que entrega la eliminación de ciertas restricciones y también por la heterogeneidad que agrega al modelo la presencia de coeficientes que capturan el gusto por cada cliente del supermercado. Adicionalmente, es importante mencionar que desviación estándar de los coeficientes aleatorios asociados al modelo son significativos, por lo que indica que el modelo ajusta bien en cuanto a capturar variaciones en gustos por parte de los clientes del supermercado.

4.6.2. Discusión de los resultados

VARIABLES DE NO-DISPONIBILIDAD Y EFECTOS CAPTURADOS

- Variables de no-disponibilidad

A continuación se encuentran el análisis de las variables de grado de no disponibilidad para los productos ($OOS_{1,g}$ y $OOS_{1,e}$), para las categorías ($OOS_{2,e}$) y para la categoría de pan ($OOS_{3,g}$ y

$OOS_{3,e}$).

Se puede observar que estas dos primeras – relacionadas a los productos – son muy significativas dentro del modelo y ambas poseen signo negativo para los 3 modelos revisados. Esto indica que, al enfrentar disminución de la variedad de cada producto (o un aumento del grado de no-disponibilidad de éstos), entonces disminuye la probabilidad de compra de algún producto de la *choice set*.

En segundo lugar, la variable de no-disponibilidad para la categoría de productos empaquetados entrega, a su vez, un coeficiente de signo negativo, lo que sigue en la línea de, que si disminuye la variedad dentro de la categoría, esto afecta a los productos en su totalidad. Sin embargo, este resultado no es significativo, por lo que no es determinante dentro del modelo (2) como el modelo (3)

Por último, dentro de las variables de no disponibilidad para las categorías de pan, donde se estudia la disponibilidad promedio total dentro de la categoría, los resultados obtenidos no son significativos y los signos que entrega el modelo, –que se incluyen dentro del modelo (3). Para el caso de los productos a granel va en dirección con lo anteriormente mencionado; sin embargo, para los productos empaquetados el grado de no disponibilidad agregado representa un aumento en las ventas de los productos. Se cree que por construcción esta variable presenta poca variabilidad dentro de cada ocasión de compra para los productos en estudio y, también, poca variabilidad entre compras para cada rango de fecha/hora controlada, por lo que no son determinantes para el modelo.

Como se mencionó recientemente, los coeficientes asociados tanto a las categorías de pan a granel y empaquetado como a la categoría de pan total no son significativos dentro del modelo. Este es un resultado relevante, ya que indica que el efecto de la disponibilidad dentro de las categorías creadas no es tan relevante para el caso en estudio, lo cual demuestra que los efectos cruzados entre las categorías son menores y el efecto total de la categoría sobre la elección de un producto también lo es.

Es importante notar que los coeficientes obtenidos tanto para $OOS_{2,e}$ como para $OOS_{3,g}$ y $OOS_{3,e}$ presentan valores positivos, aunque la magnitud de éstos es pequeña, lo cual genera una suerte de efecto sustitución entre categorías; sin embargo, como fue mencionado, es muy poco significativa esta forma de capturar el efecto.

- Efectos de sustitución y efectos cruzados

Se puede observar que las variables binarias asociadas a los productos (α_j), en su gran mayoría son significativas. Se debe mencionar que las variables de este tipo asociadas al modelo corresponden únicamente a variables pertenecientes a la categoría de productos empaquetados, por lo tanto, no son comparables con los productos de ventas a granel. Sin embargo, es interesante destacar que estas variables se ajustan de buena manera a las participaciones de mercado existentes, donde las variables asociadas a los productos 8, 10 y 11 -que son las que presentan mayores ventas dentro de la categoría de panes empaquetados- presentan valores positivos, y el resto de las variables presenta valores negativos.

Los efectos de sustitución se pueden ver implícitamente en estos coeficientes dado entre los productos que cuestión, ya que si no se encuentra alguno en stock, la presencia de los otros va a influenciar positivamente en la compra de pan ($\alpha_j > 0$) y, en otros casos, negativamente ($\alpha_j < 0$) en la compra de pan.

Sobre las variables binarias asociadas a las categorías (α_f) se observa que la media de los coeficientes es altamente significativa y presenta valores negativos. Esto se entiende debido a que están construidas en base al caso no compra, que es la categoría omitida en este caso para evitar efectos de colinealidad. No debe olvidarse que éstas son las variables que capturan la elección para cada cliente para las categorías, donde se obtiene una desviación estándar altamente significativa. Esto quiere decir que existe una alta varianza dentro de las elecciones realizadas por los clientes como también que existe varianza dentro de los productos que elige cada cliente, lo cual captura la elección por cada una de las categorías de panes a granel y empaquetados para cada cliente. Esto flexibiliza de gran manera el modelo y permite que la interacción dentro de cada categoría sea mucho mayor.

- Variables de Lealtad

Se procede la discusión de las variables de lealtad (LOY_j). Algo muy destacable observado en los coeficientes obtenidos para estas variables es que son muy significativos dentro del modelo como también son todos de carácter positivo. Esto significa que la lealtad a cierto pan influye positivamente en volver a comprarlo. También puede ser visto como que, al realizar una compra de un producto de cierto tipo en la ocasión de compra anterior, es más probable volver a comprar ese mismo producto. Otro tema a destacar en el contexto de estas variables es que en general no se observa una preferencia significativamente marcada por algún producto por los clientes del supermercado, a pesar de que las participaciones de mercado de estos productos varían harto entre éstos. También es importante notar que las preferencias de estos productos no están asociadas con las participaciones de mercado de los mismos. Lo anterior puede ser deducido al observar la magnitud de los coeficientes obtenidos para estas variables en el modelo.

En la misma línea, se puede observar que la lealtad a la categoría de pan (LOY_{pan}) altamente significativa y que presenta un coeficiente con valor negativo. Esto significa que la probabilidad de volver a comprar un producto cualquiera de la categoría disminuye para los clientes si la compra fue realizada en una ocasión de compra reciente. Cabe destacar que, de todas maneras, la magnitud de este coeficiente en comparación con la de los coeficientes de las variables de lealtad asociadas a cada producto es menor, aunque esta comparación no debe tomarse como algo absoluto debido a que las variables de lealtad para los productos por separado, por construcción, tienen media menor a la de la variable de lealtad de la categoría. Con lo anteriormente mencionado se quiere aclarar que el efecto de la categoría puede verse anulado por el efecto de cada producto, o viceversa, según el movimiento que tengan estas variables de lealtad.

Variables de control

Se puede mencionar que las variables de horario del día (γ_h) son significativas, excepto para los

horarios desde las 6 P.M. Esto se puede deber a que en estos horarios predomina más los efectos de otras variables sobre las ventas de los productos que el horario propiamente tal. La variable γ_9 fue omitida del modelo, por lo tanto, los coeficientes obtenidos en el modelo son relativos a este horario. Por último, es más probable que un cliente compre pan en los horarios de la mañana, lo cual puede ser visto por la magnitud de sus coeficientes.

Con respecto a las variables asociadas a los días de la semana (ω_d), se puede observar que todas son muy significativas para el modelo. Éstas están calculadas con respecto al día domingo (ω_7). Se observa que el día de mayores ventas es el domingo –seguido por el sábado. Por otro lado, como ha sido estudiado anteriormente, el día de menores ventas corresponde al v miércoles.

Por último, *recency* (κ), que indica la diferencia en días desde la última ocasión de compra del cliente, es también muy significativa para el modelo a pesar del pequeño valor que toma el coeficiente. Esto se ve explicado por la construcción de esta variable: Es una variable continua medida en días que toma valores mayores que la mayoría de las variables del modelo (con un promedio aproximado de 4). El carácter positivo de la variable indica que mientras más tiempo ha transcurrido desde la última ocasión de compra, es más probable que el cliente vuelva a realizar una compra en el periodo presente.

4.7. Análisis de contrafactuales

Se quiere entender y conocer el efecto de la disminución de variedad de los productos dentro de la función de utilidad de los clientes con el modelo realizado anteriormente. La problemática se aborda siguiendo los pasos que se detallan a continuación:

1. Determinar los productos a estudiar p
2. Determinar nivel de disminución de la variedad os para uno o varios de los productos en estudio
3. Definir los intervalos de tiempo en los que se llevará a cabo el análisis (día de la semana d y hora del día h)
4. Realizar el análisis de los resultados obtenidos para los escenarios elegidos

4.7.1. Conjunto a elección de la tupla (p, oos, d, h)

- Elección del producto p

Como fue observado anteriormente, existe un grupo selecto de productos que posee una mayor participación de mercado en comparación con el resto, es decir, existe un pequeño grupo que cubre un gran porcentaje del total de los productos de la categoría. Por lo tanto, debido a esta información, la elección de los productos a estudiar se realiza considerando los productos de mayores ventas en el supermercado tanto para la categoría de panes que se venden a granel como aquellos que se venden empaquetados. Los productos que poseen mayores participaciones de mercado corresponden al producto 2 (*Rolls* envasados), al producto 8 (*French Bread*) y al producto 11 (*Small French Bread*).

El primero de éstos corresponde a un producto de la categoría de productos que se venden a granel, mientras tanto los dos siguientes corresponden a productos que se venden empaquetados. Debido a lo anterior, el análisis se ve enriquecido porque se presentan productos de las dos subcategorías en análisis y, de esta manera, se puede hacer una comparación más detallada entre éstas.

- Elección del nivel de stock oos

Para el análisis en cuestión se desea observar cómo varía la participación de mercado de los distintos productos con respecto a los cambios de las variables de disponibilidad señaladas anteriormente y, de esta manera, observar la interacción que presentan estos productos. Para estudiar este fenómeno, se determinó el siguiente escenario comparativo con respecto a la disponibilidad de variedad a analizar: Cambio de la disponibilidad de variedad de los productos más comprado de un 80 % a un 20 %.

Este escenario fue elegido debido a que es de suma importancia observar cómo varía la disponibilidad de los productos desde una disponibilidad de variedad medianamente estándar de la categoría –como es un nivel de 80 %– a una mucho menor –20 %, la cual corresponde

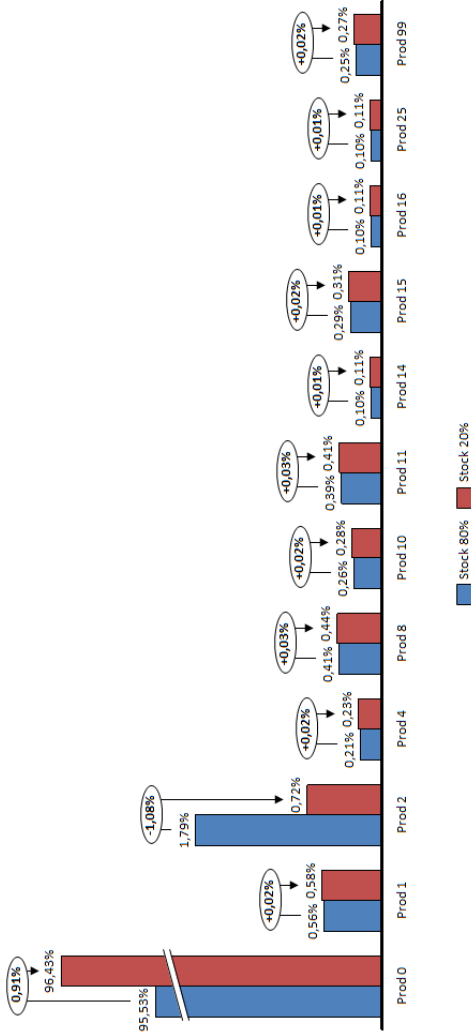
- Elección de días d y de horas h

Los días de la semana y horas del día bajo análisis se determinaron considerando el día en que hay menor demanda de los productos y el día en que hay mayor demanda de los productos. Esto se realiza de esta manera para poder contrastar los resultados obtenidos. El día de menor demanda corresponde al día miércoles y el día de mayor demanda corresponde al día domingo.

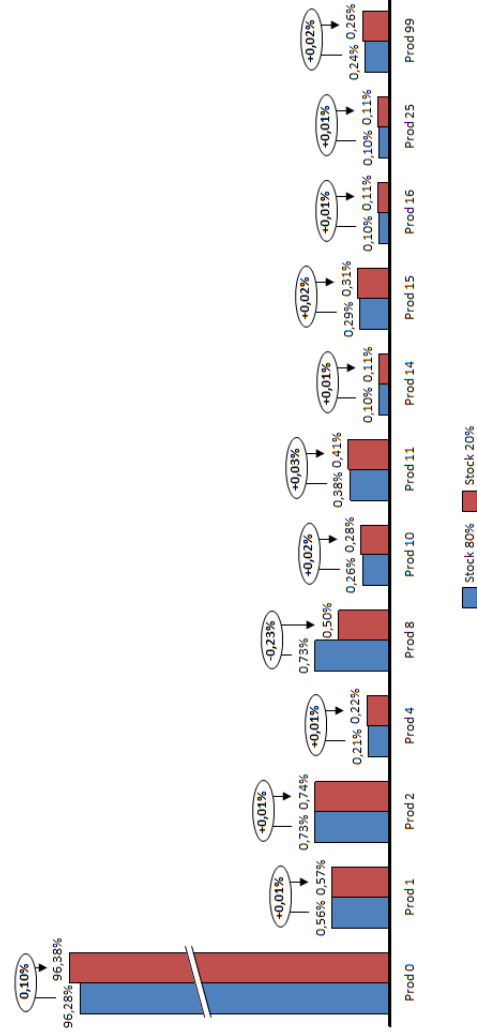
Luego, para cada uno de estos días se analiza en dos bloques horarios: $H_1 = [10 \text{ A.M.}, 11 \text{ A.M.}]$ y $H_2 = [6 \text{ P.M.}, 7 \text{ P.M.}]$. En el primero se producen mayores ventas de pan y, además, la disponibilidad de variedad de productos es menor. En cambio, en el segundo bloque horario, ocurren menores ventas de pan y la disponibilidad de variedad de pan en general es menor que en el primer bloque horario. Nuevamente es importante destacar que se eligen dos bloques horarios porque se busca ver los efectos de sustitución producidos en los distintos bloques horarios.

4.7.2. Resultados del análisis

Para este análisis se calcula la diferencia entre la participación de mercado de cada producto calculada con un nivel de 80 % de *stock* y la participación de mercado calculada para el caso de 20 % de *stock* de los productos en los días y bloques horarios estipulados. Los resultados del análisis se muestran en la Figura 4.2 para el producto 2 (a granel) y 8 (empaquetado) en el día miércoles en bloque horario H_1 .



(a) Producto 2



(b) Producto 8

Figura 4.2: Análisis delta participaciones de mercado en el día miércoles en bloque horario H_1

Se puede ver que el producto 2 presenta una mayor disminución de las ventas de pan que el resto de los productos cuando existe quiebre de éstos en los horarios estipulados. Esto se debe a que este producto, al ser líder en el mercado adicionado a una baja de *stock* de todos los productos asociados (*rolls* a granel), produce una gran baja en las ventas de éste. Para este producto, las variaciones de las participaciones de mercado son constantes con respecto a los días de la semana y a los horarios observados. Al ser este producto el más importante dentro de la categoría, una disminución de la disponibilidad de éste afecta de igual manera cualquiera sea la hora en la que no se encuentre disponible. De todas maneras, se puede notar que la mayor sustitución se obtiene con la opción de no compra (producto 0).

Para el producto 8, se puede observar que la variación de la participación de mercado disminuye en menor cantidad que para el producto anterior para el producto en cuestión y aumenta la participación de los otros productos de esta misma manera. Algo similar ocurre también para el producto 11.

No hay una tendencia clara en los productos estudiados sobre si la variación de participación es mayor dependiendo si el día en estudio es de alta o baja demanda, debido a que se observa que, para cada producto, este comportamiento es distinto y se cree que tiene que ver con el comportamiento del *stock* de éstos como también con el comportamiento de compra de los clientes más leales a cada uno.

Es importante recalcar que la diferencia de participación es más notoria en la mañana que en la tarde para todos los productos. Esto se puede deber al comportamiento de los clientes que visitan el supermercado en ciertas horas del día específicamente, aunque esto no es tan claro.

Por último, es claro ver que existe un efecto sustitución entre los productos. En particular, es importante destacar que, para los productos 8 y 11, la distribución de la participación de mercado que se pierde debido a la baja de *stock*, es repartida en su mayoría, –en más del 50 %– en el resto de los productos y su minoría, en la opción de no compra. Para el producto 2, líder del mercado, la sustitución se realiza en su mayoría con la opción de no compra. Esto se puede deber a muchas razones. Una de ella puede ser que los clientes que escogen productos frescos, optan por no comprar un producto empaquetado y, en cambio, optan por no comprar ninguno.

Es importante mencionar que los resultados obtenidos en este capítulo son coherentes con que se esperaba. La dinámica de interacción de los distintos productos mediante los distintos niveles de disponibilidad responden a lo esperado. Si se realiza una comparación con los resultados a nivel agregado, se puede notar que el estudio se condujo por el camino adecuado, donde se pudo observar y arreglar las deficiencias dejadas por este modelo, con variables más específicas sobre el gustos y comportamiento de clientes y productos.

Para mostrar cómo estos resultados pueden ser utilizados en otros estudios relacionados, a continuación se realiza un estudio donde se muestra la interacción existente entre los productos de la categoría de panes con las otras categorías existentes en el supermercado y cómo se relacionan las compras entre éstas, afectadas por los quiebres de *stock* de panes.

Capítulo 5

Análisis otras categorías

5.1. Segmentación de las otras categorías

Luego de haber estudiado la relación e interacción que presentan los productos de la categoría de pan entre ellos, se quiere extender este estudio más allá y poder entender la interacción que presentan los productos de la categoría de pan con otros productos del supermercado, viendo cuáles de ellos son sustitutos y cuáles de ellos son complementarios, y cuantificar esta información.

Se posee información que indica a qué departamento del supermercado pertenece cada producto. A continuación, en la Tabla 5.1, se observa la información sobre cada departamento, con la descripción de éste, y con la cantidad de PLU's correspondiente a cada uno y su respectivo porcentaje.

Para entender un poco más cómo se compone cada uno de estos departamentos, se desea introducir brevemente cuáles con algunas de las subcategorías que éstos presentan¹.

El Departamento 1, *Grocery*, es el más amplio en cuanto a tipo de productos de consumo, donde se pueden observar productos listos para cocinar, los bebestibles, productos enlatados y en conserva, cereales, fideos, arroz, sopas, té y café, condimentos, salsas y aderezos, galletas, insecticidas, productos de limpieza, productos plásticos, comida para animales, entre otros.

En el Departamento 2, *Meat*, se encuentran todos los distintos de tipos de carne -congeladas y no congeladas- y los distintos condimentos para éstas.

En el Departamento 3, *Dairy*, se encuentran principalmente los productos derivados de la leche, como mantequilla y margarina, queso, yogurt y algunos postres. También se encuentran los huevos frescos, papas y preparaciones prelistas.

En el Departamento 4, *Produce*, se encuentran los distintos productos frescos, como son las verduras y frutas que se venden en el Supermercado.

En el Departamento 5, *Frozen*, se encuentran la principal cantidad de productos que se venden congelados, donde se encuentran los productos que comidas preparadas, pizzas, sandwiches, distintas frutas, postres, helados, entre otros.

En el Departamento 6, *Deli*, se encuentran productos como comidas preparadas, como pizzas, sandwiches, sopas, pastas, carnes, *wraps*, ensaladas, entre otras.

¹Los Departamentos poseen números correlativos del 1 al 22. No aparecen en esta descripción los departamentos de los cuales no se poseen datos de cliente.

En el Departamento 7, *Commercial Bakery*, se encuentran distintos productos de panadería horneados, pan de molde, pitas y *wraps*, y productos de panadería dulces.

En el Departamento 9, *Floral*.

En el Departamento 12, *Seafood*, se encuentran todos los productos provenientes del mar, como distintos pescados y mariscos, tanto frescos como ahumados. Adicionalmente, se encuentran los distintos condimentos para éstos.

En el Departamento 13 y 16, *Wine* y *Beers*, se encuentran los distintos vinos y cervezas que posee el supermercado, respectivamente.

En el Departamento 14, *Bakery*, se encuentran los distintos productos vistos en la categoría de panes, como son los panes que se venden a granel como los que se venden empacados. Adicionalmente, en este departamento se venden *donuts*, galletas, queques, *muffins*, pasteles, entre otros.

En el Departamento 17, *HBC*, se encuentran los productos de baño e higiene personal, como también los productos de bebé y adultos. Adicionalmente se venden algunas vitaminas y remedios.

En el Departamento 18, *GM*, se encuentra Electrónica y electrodomésticos, libros, revistas y periódicos, discos, videos y fotografía, ropa de vestir, velas e inciensos, cristalería y artículos de cocina, lavandería, arriendo de películas, productos de escritorio, productos de fiesta, juguetes, entre otros.

En el Departamento 19, *Case Ready Meats*, se encuentran principalmente el tocino, embutidos y carnes listas para el consumir tanto listas como congeladas.

En el Departamento 21, *Tobacco*, se encuentran las bebidas alcohólicas, cigarrillos y accesorios.

En el Departamento 22, *Coffee Shop*, se encuentran productos vendidos por la tienda *Starbucks Coffee*.

A continuación, se pretende dar a conocer cómo varían las compras de los distintos departamentos descompuesto por boletas que contienen algún tipo de pan en sus compras, cuáles no contienen pan y el total de compras, como puede ser visto en la Tabla 5.1.

Depto.	Nombre	Frec.	Porc.	CPB ² (compras que incluyen pan)	CPB (compras que no incluyen pan)	CPB (total de compra)	Variación porcen- tual ³
1	<i>Grocery</i>	288706	38,22 %	5,385	3,499	3,719	44,81 %
2	<i>Meat</i>	31134	4,12 %	0,646	0,371	0,401	61,04 %
3	<i>Dairy</i>	12354	16,35 %	2,430	1,488	1,591	52,71 %
4	<i>Produce</i>	66919	8,86 %	1,296	0,769	0,862	50,41 %
5	<i>Frozen</i>	53534	7,09 %	0,994	0,651	0,690	44,14 %
6	<i>Deli</i>	47254	6,26 %	1,145	0,544	0,609	88,17 %
7	<i>Commercial Bakery</i>	21998	2,91 %	0,405	0,269	0,283	43,00 %
9	<i>Floral</i>	3051	0,4 %	0,035	0,039	0,039	-10,81 %
12	<i>Seafood</i>	8375	1,11 %	0,164	0,101	0,108	51,85 %
13	<i>Wine</i>	13814	1,83 %	0,187	0,174	0,178	4,85 %
14	<i>Bakery</i>	28798	3,81 %	1,209	0,270	0,371	226,00 %
16	<i>Beer</i>	8537	1,13 %	0,104	0,110	0,110	-5,25 %
17	<i>HBC</i>	20913	2,77 %	0,297	0,256	0,269	10,40 %
18	<i>GM</i>	13757	1,82 %	0,199	0,172	0,177	12,08 %
19	<i>Case Ready Meats</i>	22742	3,01 %	0,454	0,273	0,293	54,85 %
21	<i>Tobacco</i>	1482	0,2 %	0,007	0,009	0,019	-64,67 %
22	<i>Coffee Shop</i>	800	0,11 %	0,002	0,005	0,010	-78,96 %
Obs.		755372	100 %	8303	69336		

Tabla 5.1: Cantidad de PLU's por boleta para distintos conjuntos de compra de pan

Se puede observar que existe un aumento considerable de compras de los departamentos con más productos, como son *Grocery*, *Diary* y *Deli*. Es importante mencionar también que en la mayoría de los departamentos se ve un aumento en la compra. Se puede ver que los productos que poseen mayor porcentaje de ventas son aquellos relacionados o que son acompañamientos para los panes, como carnes, productos de *deli* y *case ready meats*. Por otro lado, se puede ver que existen disminuciones de ventas en algunos productos, como son cervezas –en un porcentaje pequeño del 5 % aproximadamente– y en el departamento de cigarrillos y alcoholes, que presenta una disminución significativa de un porcentaje cercano al 65 %. La misma correlación negativa existe con los productos del *Coffee Shop*. Esto se puede deber a que los clientes que acuden al café lo hacen para consumir productos de éste y, por lo tanto, no tienen interés en la compra de un productos complementarios.

Nombre	Depto.	Incidencia de compra (boletas que incluyen pan)	Incidencia de compra (boletas que no incluyen pan)	Incidencia de compra (boletas totales)	Variación porcentual (boletas pan v/s Total)
<i>Grocery</i>	1	0,806	0,717	0,727	10,94 %
<i>Meat</i>	2	0,415	0,256	0,273	52,12 %
<i>Dairy</i>	3	0,663	0,490	0,509	30,22 %
<i>Produce</i>	4	0,554	0,371	0,391	41,63 %
<i>Frozen</i>	5	0,406	0,293	0,305	32,75 %
<i>Deli</i>	6	0,518	0,303	0,326	58,70 %
<i>Commercial Bakery</i>	7	0,286	0,200	0,209	36,79 %
<i>Floral</i>	9	0,028	0,030	0,030	-5,61 %
<i>Seafood</i>	12	0,121	0,078	0,083	46,00 %
<i>Wine</i>	13	0,127	0,123	0,123	3,01 %
<i>Bakery</i>	14	0,780	0,195	0,257	203,19 %
<i>Beer</i>	16	0,088	0,091	0,091	-3,44 %
<i>HBC</i>	17	0,169	0,146	0,149	13,91 %
<i>GM</i>	18	0,131	0,119	0,120	8,82 %
<i>Case Ready Meats</i>	19	0,292	0,183	0,195	49,56 %
<i>Tobacco</i>	21	0,007	0,008	0,008	-18,62 %
<i>Coffee Shop</i>	22	0,002	0,004	0,004	-54,23 %
Observaciones		8303	69336		

Tabla 5.2: Incidencia de compra por categoría para distintos conjuntos de compra de pan

Se puede ver que los productos que poseen mayor incidencia cuando existen compras de pan en la boleta son *Grocery* (80,6 % de incidencia), *Bakery* (78 %), *Dairy* (66,3 %), *Produce* (55,4 %) y *Deli* (51,8 %). Esta información es muy útil como base para determinar cuáles departamentos se relacionan más con las compras de pan.

Adicionalmente, es importante mencionar que la información obtenida a partir del estudio de la incidencia de compra está orientada en la misma dirección que para la cantidad comprada. La diferencia se encuentra en la magnitud de la variación porcentual entre éstas. Esto se debe a que se sabe que las personas que compran pan, compran en promedio más productos en el supermercado, lo que hace que esta diferencia se vea magnificada.

5.2. Departamentos en estudio

Se quiere estudiar el efecto de las compras de pan, y la falta de variedad de éste, dentro de las compras realizadas por departamentos. Para esto se debe definir, en primera instancia, cuáles son los departamentos que son convenientes para realizar el estudio.

Al observar la frecuencia de compra se puede determinar cuáles son los departamentos más atractivos para el análisis. Hay 4 departamentos que son altamente atractivos –*Meat*, *Produce* y *Deli*–. Estos poseen una relación más directa con la categoría en estudio por el tipo de productos ofrecidos (en general, productos altamente perecibles) que otros departamentos. Adicionalmente,

estos departamentos poseen una variación porcentual significativa (sustentada por la frecuencia de compra) entre la incidencia de las boletas que contienen compras de pan y las boletas totales.

5.3. Variables

A continuación se detalla la información sobre las variables a considerar dentro del análisis.

Variable dependiente

$$D_{it}^d = \begin{cases} 1 & \text{si el cliente } i \text{ compra producto del departamento } d \text{ en la ocasión de compra } t \\ 0 & \text{en otro caso} \end{cases}$$

Variables independientes

- Compra de pan

- Para la compra de pan

$$x_{it} = \begin{cases} 1 & \text{si el cliente } i \text{ compra algún pan en la ocasión de compra } t \\ 0 & \text{en otro caso} \end{cases}$$

- Para la compra de categorías de pan

$$x_{it}^f = \begin{cases} 1 & \text{si el cliente } i \text{ compra algún pan de la categoría } f \in F \text{ en la ocasión de compra } t \\ 0 & \text{en otro caso} \end{cases}$$

- Variables binarias de días de la semana. Se define de la siguiente manera:

$$DAY_{\delta(t)} = \begin{cases} 1 & \text{si la ocasión de compra } t \text{ ocurre en el día } D_{\delta}, \text{ con } \delta = \delta(t) \\ 0 & \text{en otro caso} \end{cases}$$

Donde $\{D_1 = \text{lunes}, D_2 = \text{martes}, D_3 = \text{miércoles}, D_4 = \text{jueves}, D_5 = \text{viernes}, D_6 = \text{sábado}, D_7 = \text{domingo}\}$

Esto ayuda a capturar las variaciones que ocurren dentro de los días de la semana, así como las tendencias entre éstos.

- Variables de control binarias sobre horas del día. La variable asociada se define de la siguiente manera:

$$H_h = \begin{cases} 1 & \text{si la ocasión de compra } t \text{ ocurre en la hora } h = h(t) \\ 0 & \text{en otro caso} \end{cases}$$

donde $h \in [8 \text{ A.M.}, 10 \text{ P.M.}]$

Esto ayuda a capturar las variaciones que ocurren dentro de los días de la semanas, así como las tendencias entre éstos.

- Variables de Lealtad a la categoría de pan

$$LEAL_t^i = \alpha LEAL_{t-1}^i + (1 - \alpha)z_t^i$$

donde

$$z_t^i = \begin{cases} 1 & \text{si el consumidor } i \text{ compra algún producto de la categoría de pan} \\ & \text{en la ocasión de compra } t - 1 \\ 0 & \text{en otro caso} \end{cases}$$

La inicialización de la variable se realiza de manera análoga al caso del análisis de cada producto:

$$LEAL_1^i = \begin{cases} \alpha & \text{si algún producto de la categoría es el primer} \\ & \text{producto comprado en historia del cliente } i \\ \frac{1 - \alpha}{\text{card}\{CP, NC\} - 1} & \text{en el caso contrario} \end{cases}$$

Se considera que la cardinalidad del conjunto es $\text{card}\{CP, NC\} = 2$, debido a que se evalúa entre dos alternativas, que corresponden a las alternativas de comprar algún tipo de pan (CP) o no comprar ninguno (NC). En este caso se considera igualmente que el valor de α es $\alpha = 0,8$.

- Variables para primera regresión y variables instrumentales
 - Elección de la compra de pan

$$choice_{pan} = \sum_{j \in C} y_{jt}^i$$

Donde y_{jt}^i corresponde a la variable binaria utilizada en el capítulo anterior que indica 1 si el cliente i compró producto j en ocasión de compra t .

- Elección de la compra de pan en categorías

$$choice_f = \sum_{j \in f} y_{jt}^i, f \in F$$

Donde F corresponde a las categorías de pan granel y empaquetado, donde se cumple que $\text{card}(\text{granel}) = 2$ y $\text{card}(\text{empaquetado}) = 9$.

- Probabilidad estimada de compra de pan

$$IV_{pan} = \sum_{j \in C} \hat{y}_{jt}^i$$

Donde \hat{y}_{jt}^i corresponde a la estimación de la variable binaria utilizada en el capítulo anterior que indica 1 si el cliente i compró producto j en ocasión de compra t .

- Probabilidad estimada de compra de pan en categorías

$$IV_f = \sum_{j \in f} \hat{y}_{jt}^i, f \in F$$

5.4. Modelos utilizados

Los modelos considerados para este capítulo son los siguientes:

1. Modelo (a): Estudio de la categoría de pan agregada: Se realiza una regresión en dos etapas. En la primera etapa se estudia cómo afectan las compras de pan a la probabilidad estimada de comprar pan, la cual en la estimación considera las variables de disponibilidad de los productos. Luego, en la segunda etapa, esta probabilidad estimada de compra es utilizada para determinar si se compra en cierto departamento.

- Primera etapa: $choice_{pan} = C_{pan}IV_{pan} + CONTROLSES_t^i$
- Segunda etapa: $D_{d1t}^i = \nu_{pan}choice_{pan} + CONTROLSES_t^i$

2. Modelo (b): Estudio de la categoría pan dividida en productos de venta a granel y empaquetados: El procedimiento es análogo al anterior, con la particularidad de que se consideran dos regresiones en la primera etapa, donde se entrega la estimación de compra de comprar en cada una de las categorías. Luego, el resultado de ambas variables instrumentales se utiliza para determinar si se compra en cierto departamento.

- Primera etapa:
 - $choice_{granel} = \sum_{f \in F} IV_f + CONTROLSES_t^i$
 - $choice_{empa} = \sum_{f \in F} C_f IV_f + CONTROLSES_t^i$
- Segunda etapa: $D_{d2t}^i = \sum_{f \in F} \nu_f choice_f + CONTROLSES_t^i$

donde $CONTROLSES_t^i$ representa a las variables de control por estacionalidad, es decir, está definido de la siguiente manera:

$$CONTROLSES_t^i = \sum_h \gamma_h \cdot H_{h(t)} + \sum_d \omega_d DIA_{\delta(t)} + \phi_{pan} LOY_t^i$$

5.5. Resultados obtenidos

Los análisis que se muestran a continuación corresponden a los Departamentos 2 (*Meat*), 4 (*Produce*) y 6 (*Deli*). El comando utilizado en Stata para correr estos modelos son los siguientes:

(foreach p in 2 4 6)

(a) xtivreg D'p' (choice = IV) day1-day7 month1-month12 hour1-hour14

(b) xtivreg D'p' (choice_g choice_e = IV_e IV_g) day1-day7 month1-month12 hour1-hour14

Para cada una de éstas, se obtienen los siguientes resultados:

	(1)	(2)	(3)
	d_2	d_4	d_6
<i>choice</i>	0.0416	0.243***	0.199**
	(0.67)	(3.63)	(3.09)
Observations	52665	52665	52665
Log Lik.	-27742.6	-31858.6	-29795.8

t statistics in parentheses
 * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 5.3: Resultados primera regresión modelo (a)

	(1)
	<i>choice</i>
<i>IV</i>	0.643***
	(23.77)
Observations	52665
Log Lik.	-7328.3

t statistics in parentheses
 * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 5.4: Resultados segunda regresión modelo (a)

	(1)	(2)	(3)
	d_2	d_4	d_6
<i>choice_granel</i>	-0.123	0.0425	0.140
	(-1.16)	(0.37)	(1.27)
<i>choice_empaq</i>	0.143	0.367***	0.236**
	(1.84)	(4.37)	(2.93)
Observations	52665	52665	52665
Log Lik.	-27762.9	-31968.8	-29823.0

t statistics in parentheses
 * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 5.5: Resultados primeras regresiones modelo (b)

	(1)	(2)
	$choice_{granel}$	$choice_{empaqa}$
IV_{granel}	0.550*** (23.31)	0.0523 (1.62)
IV_{empaqa}	0.0324 (1.61)	0.640*** (23.24)
Observations	52665	52665
Log Lik.	19810.2	3261.1

t statistics in parentheses
* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 5.6: Resultados segunda regresión modelo (b)

Los resultados del análisis a nivel de pan agregados dan concordantes con lo esperado, es decir, las variables instrumentales nos indican que a más compras de pan, mayores son las compras en los departamentos estudiados, aunque éstas no son significativas para las 3 categorías analizadas por igual. Se puede observar que para el caso del modelo (a) y (b), el efecto no es significativo para el departamento de carnes, aunque sí lo es para *Produce* y *Delí*.

Para el modelo (a) se puede ver que tanto el aumento de *stock* en los panes aumenta las ventas de éste, como que las ventas de pan, incluyen positivamente en las compras en otros departamentos.

Para el modelo (b), desagregado por categorías de panes, se puede observar que para las dos primeras regresiones, la probabilidad de comprar productos a granel influye positivamente en comprar productos empaquetados, y viceversa, aunque hay que destacar que este resultado no es significativo. Por otro lado, la probabilidad de comprar productos de cierta categoría, influye positiva y muy significativamente en comprarlo.

Para la segunda regresión, tanto la compra de productos a granel como empaquetados influye positivamente en que las compras en los departamentos en estudio sean realizadas. Se debe notar que la magnitud de comprar un producto empaquetado es mucho mayor que comprar de uno a granel, por lo que una baja de *stock* de producto empaquetado afecta de gran manera la compra de los departamentos, lo cual no sería superado por el efecto de sustitución entre las dos categorías.

5.6. Análisis de contrafactuales

Como se espera que el comportamiento de baja de *stock* para las compras de productos empaquetados sea similar en ambos modelos revisados, se analiza cómo afecta la disminución de *stock* de actual a 0% en todos los productos a granel para un día en particular (domingos) en el bloque horario $H_1 = [10 \text{ A.M.}, 11 \text{ A.M.}]$ para el departamento *Produce*, que entregó los resultados más significativos. Para esto se observa la incidencia de compra estimada para cada caso.

Para esto se consideran ambos modelos:

- Para Modelo (a). Variable instrumental construida por el conjunto total de panes
- Para Modelo (b). Variables instrumentales construidas por el conjunto de panes dentro de cada categoría

Ambos se encuentran en la Figura mostrada a continuación:

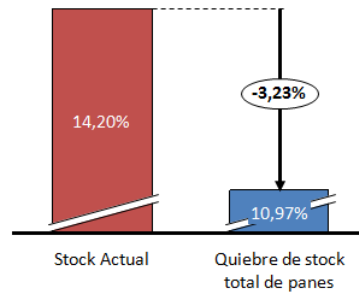
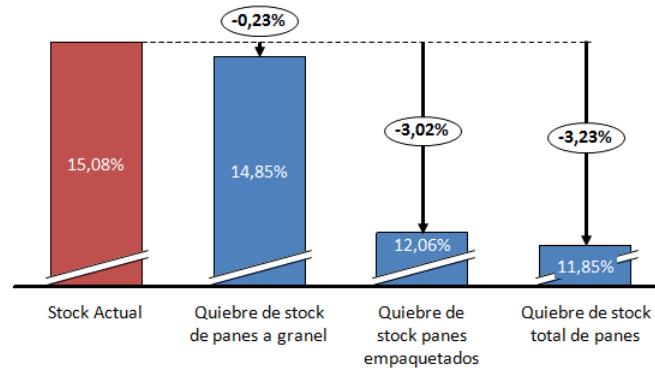


Figura 5.1: Para modelo (a)



(a) Para modelo (b)

Figura 5.2: Análisis de contrafactuals para modelos estudiados

Para el primer modelo, donde se ven las compras de pan de manera conjunta, la baja del *stock* de productos a granel genera una baja en la probabilidad de compra de los panes. Esto, a su vez, genera que las compras en el departamento *Produce* también se vean disminuidas, debido a que éstas variables están relacionadas positivamente. Es decir, un cliente que observa una baja de *stock* de panes de manera conjunta (en particular, de panes a granel), disminuye las compras de carnes.

Por otro lado, para el modelo por categorías de pan, se observa la variación entre el stock actual de la categoría y tres escenarios diferentes: quiebre de *stock* de los panes a granel, quiebre de *stock* de panes empaquetados y quiebre de *stock* total de panes para el día y horas señaladas.

Como se discutió en los resultados obtenidos, se puede ver en el análisis que el peso que tienen los productos empaquetados sobre los productos a granel es mucho mayor, por lo que los quiebres de *stock* por panes de venta a granel no es tan considerable como cuando se quiebran los panes empaquetados. Adicionalmente, se puede observar que los quiebres de *stock* total de panes disminuye aún más la probabilidad de comprar en el departamento *Produce*, lo cual está en línea con los resultados obtenidos: si hay quiebre de ambas categorías, la probabilidad de comprar en dicho departamento es aún menor que cuando se quiebra alguna de las categorías de pan por separado.

Para concluir, es importante mencionar que, para el caso en análisis, el porcentaje de compras de pan realizada en cada ocasión de compra es muy bajo (cercano a un 3%), por lo cual se tuvo especial cuidado en cada decisión tomada con respecto a la forma de medirlos. En particular, en la

elección del *choice set* utilizado finalmente, donde se consideró que el grupo de clientes comprara al menos una vez algún pan en su historial de compra, con una incidencia final es cercana a un 9%, y en las agrupaciones de los productos como también en las variables a crear para capturar los efectos relacionados a éstos. De cierta manera se están poniendo a prueba los modelos en casos “extremos” por la naturaleza de los datos, lo que beneficiaría a estudios posteriores con mayor incidencia de compra en los datos. A pesar de esto, los resultados obtenidos se encuentran satisfactorios y el espíritu de los efectos a estudiar fue capturado de manera favorable.

Capítulo 6

Conclusiones generales

6.1. Conclusiones

Para la realización de este trabajo se obtienen datos del nivel de *stock* de panes y las transacciones de clientes en un supermercado de Estados Unidos en un periodo comprendido entre el 22 de diciembre de 2010 al 30 de mayo de 2011. Las primeras obtenidas de manera novedosa mediante imágenes fotográficas analizadas por computador. Esta información es primordial para poder cumplir con el objetivo planteado en este trabajo, que corresponde al estudio del efecto de la variedad de surtido en góndola en el comportamiento de clientes mediante la incorporación de variables operacionales de *stock* a un modelo de incidencia de compra.

Para el caso en análisis, el porcentaje de compras de pan realizada en cada ocasión de compra es muy bajo (cerca a un 3%), por lo cual se tuvo especial cuidado en cada decisión tomada con respecto a la forma de medirlos.

En primer lugar se formula el modelo agregado por clientes, donde se estudia la relación entre las ventas totales y las ventas de cada uno de los productos en las siguientes categorías en relación entre el nivel de *stock* de los productos: los principales 2 productos, los principales 5 productos y *rolls* envasados v/s a granel.

Para la primera categoría, se estudian los modelos mencionados anteriormente y, adicionalmente, los efectos cuadráticos del nivel de *stock* en las ventas. Dentro de esta categoría, el producto 1 es significativo en los tres modelos y, para el modelo de las ventas del producto 2, el nivel de *stock* de ambos productos lo es también. Para los modelos lineales, al disminuir el nivel de *stock*, las ventas de todos los productos disminuyen, aunque en cantidades diferentes. Para el caso del modelo cuadrático, esta disminución repercute en un aumento en las ventas hasta cierto nivel de *stock*, donde finalmente decaen. Este aumento puede deberse a complementareidad entre ambos productos.

Para la segunda categoría, existen algunos resultados aisladamente significativos que, a pesar de no entregar resultados concluyentes, se observan patrones de sustitución y complementareidad entre los productos al observar las regresiones de ventas de cada producto de la categoría. Para el modelo de ventas totales no se observa ningún patrón claro.

Para la tercera categoría, para el modelo de venta agregada y para el de venta de *rolls* a

granel es sólo significativa la variable de *rolls* a granel, y para el modelo de *rolls* empaquetados, ninguna de ellas es significativa. Para los primeros modelos, al disminuir el nivel de *stock* de cada uno de los productos, las ventas disminuyen; en cambio, para el último, las ventas aumentan. Se cree que en este caso el ruido obtenido por los datos agregados oculta el efecto que se encuentra detrás de los datos.

De este modelo agregado por clientes, se obtuvieron patrones aclaradores sobre la relación entre los quiebres de *stock* y las compras de los clientes, aunque los resultados no siempre fueron significativos. Por esta razón, se hace imprescindible el análisis del modelo individual por cliente.

Para el modelo individual por cliente, se utilizó un modelo *logit* con coeficientes aleatorios, donde se incluyeron variables para capturar los efectos de sustitución entre productos y los efectos cruzados entre categorías, que corresponden a los panes que se venden a granel y los panes que se venden empaquetados. Para esto se crean variables de *out-of-stock* para los productos, categorías y total de productos (como promedio simple de los *out-of-stock*) que, en conjunto con las variables aleatorias de las categorías para cada cliente y las variables binarias por producto, se obtuvo resultados interesantes del comportamiento tanto de clientes de la manera en que los productos interactúan entre sí. Es importante mencionar que las variables de *out-of-stock* son significativas, excepto en el caso de la mirada de los *stock-out* como conjunto de panes. Desglosando estas variables, se puede que las variables de *out-of-stock* por producto son muy significativas y muestran una clara tendencia de a menor nivel de *stock*, mayores ventas de cada producto. Por otro lado, la variable de *out-of-stock* por categorías es levemente significativa y muestra tendencia al aumento como a la disminución de ventas dependiendo la categoría. Algo similar ocurre para las variables de *out-of-stock* por total de panes, donde se ve aumento de ventas con baja de *stock out*. Éstas últimas no son significativas y se cree que pierden peso por el promedio agregado de todos los productos.

Al realizar un análisis de contrafactuales en la variación del *stock* para el modelo en cuestión para distintos productos, se encuentra un claro patrón de sustitución entre los productos del supermercado y la opción de no compra, donde no se observa una tendencia clara del efecto entre categorías. Este efecto se encuentra en los ambos tramos horarios y días estudiados.

El modelo de compras en otros departamentos del supermercado fue realizado mediante un modelo de regresión con variables instrumentales, donde la variable dependiente corresponde a la compra dentro de los departamentos en estudio: *Produce*, *Meat* y *Deli*. Dentro de éste, se estudian dos modelos distintos: Uno que observa el conjunto de panes total y otro que estudia las categorías de panes: ventas a granel o empaquetado. Las variables instrumentales corresponden a la elección de pan total (o dentro de las categorías), instrumentada por la probabilidad de comprar pan total (o dentro de las categorías). Esta última variable incluye el efecto por *stock-outs*. Para el modelo agregado por panes, los resultados muestran que a mayor compra de pan, mayores compras en los departamentos en estudio, donde estos resultados son altamente significativos para el departamento *Produce*. Para el modelo por categorías, se observa la misma tendencia, aunque solamente la variable elegir productos empaquetados torna significativa.

Se realiza un análisis de contrafactuales en la variación del *stock* para las categorías y el *stock*

total de panes para observar el comportamiento de los clientes en las compras en el departamento *Produce*. Se observa una clara tendencia a la baja en las compras del departamento que, cuando se desglosa por categoría, se observa que los quiebres de *stock* en productos empaquetados hacen que las compras del departamento se reduzcan casi al igual que con un quiebre de *stock* total de panes.

Los resultados obtenidos en general son muy satisfactorios y ayudan a entender mejor la relación entre los quiebres de *stock* y el comportamiento de los clientes dentro del supermercado, con el foco en la elección de productos que realizan y las sustituciones que hacen entre ellos, a pesar de la baja incidencia de la categoría. Debido al comportamiento observado para estos modelos, éstos podrían ser utilizados para realizar otros estudios dentro del supermercado con similares características.

Bibliografía

- Cachon, G. P. and Olivares, M. (2013). Does adding inventory increase sales? evidence of a scarcity effect in u.s. automobile dealerships. *Management Science*, 56(7):1180–1197.
- Cameron, A. C. and Trivedi, P. K. (1998). *Regression Analysis of Count Data*. Cambridge University Press.
- Cameron, A. C. and Trivedi, P. K. (2009). *Microeconometrics using Stata*. Stata Press.
- Fader, P. S. and Hardie, B. G. S. (1996). Modeling consumer choice among skus. *American Marketing Association*, 33(4):442–452.
- Gonul, F. and Srinivasan, K. (1993). Modeling multiple sources of heterogeneity in multinomial logit models: Methodological and managerial issues. *American Marketing Association*, 12(3):213–229.
- Gruen, T. W. and Corsten, D. (2007). *A Comprehensive Guide: To Retail Out-of-Stock Reduction In the Fast-Moving Consumer Goods Industry*.
- Gruen, T. W., Corsten, D., and Bharadwaj (2002). Retail out of stocks: A worldwide examination of extent, causes, and consumer responses. *Marketing Science*.
- Guadagni, P. and Little, J. (1983). A logit model of brand choice calibrated on scanner data. *Marketing Science*, 2(3):203–38.
- Lamb, C. W., F. Hair, J., and McDaniel, C. (2009). *Essential of Marketing*. Neil Marquardt, OH, USA.
- Lazari, A. and Anderson, D. (1994). Designs of discrete choice set experiments for estimating both attribute and availability cross effects. *Journal of Marketing Research*, 31(3):375–383.
- Lu, Y., Olivares, M., Musalem, A., and Schilkrut, A. (2011). Measuring the effect of queues on customer purchases. *Management Science*.
- Manchanda, P., Ansari, A., and Gupta, S. (1999). The "shopping basket": A model for multicategory purchase incidence decisions. *Marketing Science*, 18(2):95–114.
- Mc Fadden, D. and Train, K. (2000). Mixed multinomial logit models for discrete response. *Marketing Science*.

Musalem, A., Olivares, M., and Corsten, D. (2010). Structural estimation of the effect of out-of-stocks. *Management Science*, 56(7):1180–1197.

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

T.S., B. and A.R., P. (1980). The lagrange multiplier test and its applications to model specification in econometrics. *the Review of Economic Studies*, 47.

Capítulo 7

Apéndice

7.1. Apéndice 1: Fluctuaciones de las ventas de productos y pan e incidencia de pan



Figura 7.1: N^o Productos vendidos y Cantidad de dólares recaudado v/s horas del día



Figura 7.2: N^o de panes vendidos v/s horas del día



Figura 7.3: Incidencia de productos v/s horas del día

7.2. Apéndice 2: Resultados obtenidos para los modelos del capítulo de datos agregados por clientes

	(1)	(2)	(3)	(4)	(5)
	$y_{1,t}$	Ventas	Ventas	Ventas p_1	Ventas p_2
α	-2.736*** (-8.74)	-2.759*** (-8.80)	-2.720*** (-8.70)	2.998*** (4.81)	-3.200*** (-5.66)
OOS_1^{cont}	-0.233* (-2.16)	0.508 (1.68)		1.851*** (4.16)	-1.130*** (-4.03)
OOS_2^{cont}	-0.0945 (-0.94)	0.170 (0.49)		-0.345 (-1.08)	-0.715** (-2.82)
$(OOS_1^{cont})^2$		-0.847* (-2.50)			
$(OOS_2^{cont})^2$		-0.200 (-0.52)			
OOS_1^{bin}			-0.199* (-2.01)		
OOS_2^{bin}			-0.140 (-1.49)		
γ	0.350*** (14.97)	0.344*** (14.59)	0.346*** (14.81)	0.425*** (6.57)	0.259*** (6.07)
$\eta_{1,1}$	0.462*** (14.48)	0.465*** (14.56)	0.461*** (14.46)	-0.184** (-3.09)	0.537*** (10.00)
$\eta_{1,2}$	0.0671*** (12.16)	0.0668*** (12.11)	0.0668*** (12.11)	0.0326* (2.06)	-0.00848 (-0.87)
ψ_{18}	-0.398*** (-8.64)	-0.402*** (-8.72)	-0.399*** (-8.66)	-0.419** (-2.89)	-0.341*** (-3.71)
ψ_{19}	-0.703*** (-12.21)	-0.707*** (-12.35)	-0.699*** (-12.12)	-0.604*** (-3.32)	-0.706*** (-5.74)
ψ_{20}	-1.086*** (-11.40)	-1.067*** (-11.23)	-1.069*** (-11.22)	-0.922* (-2.36)	-0.876*** (-3.68)
ξ_1	0.932*** (10.59)	0.916*** (10.37)	0.925*** (10.56)	0.874*** (3.31)	-0.199 (-0.71)
ξ_2	0.176 (1.56)	0.185 (1.64)	0.175 (1.56)	0.146 (0.45)	0.205 (1.04)
ξ_3	0.355*** (3.52)	0.361*** (3.58)	0.356*** (3.52)	-0.0510 (-0.14)	0.431* (2.50)
ξ_4	0.802*** (9.80)	0.792*** (9.68)	0.796*** (9.73)	0.576* (2.26)	0.214 (1.10)
Observations	2619	2619	2619	2619	2619
Log Lik.	-6136.6	-6132.8	-6135.7	-2176.2	-3392.0

t statistics in parentheses

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 7.1: Resultados modelos para variable de ventas y 2 productos

	(1)	(2)	(3)	(4)	(5)
	<i>Inc</i>	<i>Inc</i>	<i>Inc</i>	<i>Inc p₁</i>	<i>Inc p₂</i>
α	0.0680*** (22.30)	0.0677*** (22.22)	0.0680*** (22.28)	0.0198*** (18.40)	0.0264*** (19.27)
OOS_1^{cont}	-0.00925* (-1.96)	0.0228 (1.35)		-0.00465** (-2.80)	-0.00578** (-2.75)
OOS_2^{cont}	0.00936* (2.17)	0.0343 (1.80)		0.00167 (1.10)	0.0000493 (0.03)
$(OOS_1^{cont})^2$		-0.0335 (-1.90)			
$(OOS_1^{cont})^2$		-0.0239 (-1.20)			
OOS_1^{bin}			-0.00729 (-1.76)		
OOS_2^{bin}			0.00677 (1.78)		
γ	0.0116*** (7.51)	0.0112*** (7.23)	0.0116*** (7.51)	0.00206*** (3.88)	0.00222*** (3.52)
$\eta_{2,1}$	-0.00121*** (-4.25)	-0.00121*** (-4.27)	-0.00121*** (-4.27)	-0.00108*** (-11.10)	-0.00104*** (-9.18)
$\eta_{2,2}$	-0.00677*** (-5.52)	-0.00694*** (-5.65)	-0.00660*** (-5.40)	0.000482 (1.12)	-0.00107* (-2.02)
Observations	2618	2618	2618	2618	2618
Log Lik.	5517.2	5520.4	5516.6	8232.1	7555.4

t statistics in parentheses

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 7.2: Resultados modelos para variable de incidencia y categoría 2 productos

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>Ventas</i>	<i>Ventas</i>	<i>Ventas</i> p_1	<i>Ventas</i> p_2	<i>Ventas</i> p_3	<i>Ventas</i> p_4	<i>Ventas</i> p_5
main							
Const. (α)	-2.743*** (-8.71)	-2.729*** (-8.68)	2.987*** (4.74)	-3.166*** (-5.53)	-9.492** (-2.98)	-4.513** (-2.62)	-4.042* (-2.07)
OOS_1^{cont}	-0.186 (-1.73)		-1.797*** (-4.03)	-1.074*** (-3.80)	-1.771* -0.0667 (-2.21)	1.419** (-0.11)	(2.84)
OOS_2^{cont}	-0.145 (-1.44)		-0.386 (-1.19)	-0.739** (-2.88)	0.882 (1.22)	-0.421 (-0.76)	-0.284 (-0.60)
OOS_3^{cont}	0.210 (0.81)		-0.258 (-0.26)	-0.538 (-0.76)	2.700*** (3.79)	0.681 (0.68)	-12.58 (-0.02)
OOS_4^{cont}	0.191*** (3.55)		0.0923 (0.59)	0.0815 (0.74)	0.369 (1.08)	-0.213 (-0.76)	-0.310 (-0.83)
OOS_5^{cont}	-0.584** (-2.75)		-0.470 (-0.66)	-0.561 (-1.11)	1.709* (2.31)	0.577 (0.80)	-12.76 (-0.03)
OOS_1^{bin}		-0.189 (-1.92)					
OOS_2^{bin}		-0.176 (-1.88)					
OOS_3^{bin}		0.205 (0.80)					
OOS_4^{bin}		0.197*** (3.68)					
OOS_5^{bin}		-0.625** (-2.95)					
γ	0.356*** (15.23)	0.352*** (15.11)	0.421*** (6.50)	0.259*** (6.04)	0.201 (0.99)	-0.208 (-1.67)	0.602*** (4.60)
$\eta_{1,1}$	0.465*** (14.52)	0.464*** (14.52)	-0.178** (-2.96)	0.536*** (9.89)	0.772* (2.37)	0.589*** (3.83)	0.411* (2.14)
$\eta_{1,2}$	0.0679*** (12.35)	0.0676*** (12.30)	0.0327* (2.06)	-0.00826 (-0.85)	0.119** (2.68)	0.00978 (0.37)	0.109*** (3.37)
ψ_{18}	-0.405*** (-8.81)	-0.405*** (-8.83)	-0.425** (-2.93)	-0.339*** (-3.70)	-1.009* (-2.46)	-0.137 (-0.61)	-0.796** (-2.82)
ψ_{19}	-0.717*** (-12.43)	-0.706*** (-12.27)	-0.603*** (-3.32)	-0.709*** (-5.72)	-1.206* (-2.53)	-0.825* (-2.51)	-0.968** (-2.87)
ψ_{20}	-1.098*** (-11.48)	-1.064*** (-11.19)	-0.917* (-2.35)	-0.910*** (-3.75)	-0.646 (-1.01)	-1.031* (-2.02)	-2.543*** (-4.78)
ξ_1	0.851*** (9.40)	0.846*** (9.43)	0.841** (3.15)	-0.230 (-0.82)	2.923*** (8.70)	1.730*** (5.87)	0.734 (1.36)
ξ_2	0.182 (1.62)	0.182 (1.63)	0.143 (0.44)	0.208 (1.06)	-15.42 (-0.01)	0.0799 (0.15)	0.924* (2.02)
ξ_3	0.370*** (3.67)	0.371*** (3.67)	0.00569 (0.02)	0.454** (2.64)	1.106 (1.87)	-0.588 (-0.83)	0.476 (0.81)
ξ_4	0.810*** (9.95)	0.804*** (9.89)	0.572* (2.25)	0.218 (1.12)	1.905*** (4.64)	0.432 (0.99)	1.259** (3.19)
Observations	2600	2600	2600	2600	2600	2600	2600
Log Lik.	-6082.2	-6080.1	-2164.4	-3361.3	-511.1	-1061.9	-840.4

t statistics in parentheses

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 7.3: Resultados modelos para variable de ventas y categoría 5 productos

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	<i>Inc</i>	<i>Inc</i>	<i>Incp₁</i>	<i>Incp₂</i>	<i>Incp₃</i>	<i>Incp₄</i>	<i>Incp₅</i>
Const. (α)	0.0671*** (22.03)	0.0672*** (22.02)	0.0198*** (18.60)	0.0265*** (19.23)	0.000123 (0.44)	0.00284*** (6.09)	0.000552 (1.56)
OOS_1^{cont}	-0.00865 (-1.84)		-0.00453** (-2.76)	-0.00566** (-2.68)	-0.00113** (-2.65)	0.000389 (0.54)	0.00146** (2.68)
OOS_2^{cont}	0.00843 (1.94)		0.00147 (0.97)	0.000288 (0.15)	0.00113** (2.87)	0.0000913 (0.14)	-0.000351 (-0.70)
OOS_3^{cont}	0.00890 (0.60)		-0.00136 (-0.27)	-0.00416 (-0.65)	0.00419** (3.17)	0.00344 (1.57)	-0.00119 (-0.73)
OOS_4^{cont}	0.00745* (2.40)		0.000623 (0.59)	-0.00123 (-0.92)	0.000907** (3.27)	-0.000624 (-1.36)	-0.000156** (-0.47)
OOS_5^{cont}	-0.00651 (-0.71)		-0.00157 (-0.50)	-0.00328 (-0.84)	0.00297*** (3.64)	0.00274* (2.03)	-0.00141*** (-1.44)
OOS_1^{bin}		-0.00699 (-1.68)					
OOS_2^{bin}		0.00600 (1.56)					
OOS_3^{bin}		0.00853 (0.58)					
OOS_4^{bin}		0.00763* (2.46)					
OOS_5^{bin}		-0.00670 (-0.73)					
γ	0.0115*** (7.45)	0.0115*** (7.45)	0.00192*** (3.69)	0.00217*** (3.41)	-0.0000343 (-0.25)	-0.000526* (-2.37)	-0.000600*** (3.84)
$\eta_{2,1}$	-0.00116*** (-4.11)	-0.00117*** (-4.13)	-0.00108*** (-11.45)	-0.00105*** (-9.13)	0.0000268 (1.08)	-0.0000783 (-1.95)	0.0000260 (0.94)
$\eta_{1,1}$	-0.00688*** (-5.62)	-0.00667*** (-5.48)	0.000512 (1.21)	-0.00110* (-2.08)	-0.0000873 (-0.80)	-0.000366* (-2.01)	-0.0000391** (-2.93)
Observations	2600	2600	2600	2600	2600	2600	2600
Log Lik.	5494.8	5494.4	8207.8	7511.9	11724.0	10346.3	11001.7

t statistics in parentheses

* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 7.4: Resultados modelos para variable de incidencia y categoría 5 productos

	(1)	(2)	(3)
	<i>Ventas</i>	<i>Ventas_{granel}</i>	<i>Ventas_{empa}</i>
α	-2.701*** (-8.55)	-3.138*** (-5.53)	-6.221*** (-7.53)
OOS_{granel}^{cont}	-0.231** (-2.99)	-1.293*** (-5.76)	0.00355 (0.02)
OOS_{empa}^{cont}	-0.0969 (-0.34)	-0.403 (-0.69)	0.263 (0.46)
γ	0.348*** (14.80)	0.253*** (5.87)	0.598*** (11.91)
$\eta_{1,1}$	0.458*** (14.24)	0.526*** (9.67)	0.725*** (8.74)
$\eta_{1,2}$	0.0674*** (12.18)	-0.00793 (-0.81)	0.0633*** (5.24)
ψ_{18}	-0.400*** (-8.65)	-0.332*** (-3.61)	-0.555*** (-5.26)
ψ_{19}	-0.745*** (-13.39)	-0.844*** (-6.86)	-0.606*** (-5.27)
ψ_{20}	-1.165*** (-13.02)	-1.265*** (-5.32)	-1.164*** (-6.34)
ξ_1	0.908*** (10.34)	-0.259 (-0.93)	0.939*** (4.82)
ξ_2	0.169 (1.50)	0.194 (0.99)	0.155 (0.60)
ξ_3	0.365*** (3.60)	0.461** (2.67)	0.678*** (3.33)
ξ_4	0.802*** (9.81)	0.211 (1.08)	0.831*** (4.46)
Observations	2600	2600	2600
Log Lik.	-6095.1	-3371.1	-2952.1

t statistics in parentheses
* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 7.5: Resultados modelos para variable de ventas y categoría *rolls*

	(1)	(2)	(3)
	<i>Inc</i>	<i>Inc_{granel}</i>	<i>Inc_{empa}</i>
α	0.0676*** (22.17)	0.0264*** (19.20)	0.00658*** (6.03)
OOS_{granel}^{cont}	0.00382 (1.24)	-0.00354** (-2.61)	0.00191 (1.78)
OOS_{empa}^{cont}	0.000125 (0.01)	-0.0101 (-1.59)	0.00167 (0.33)
γ	0.0113*** (7.32)	0.00221*** (3.48)	0.00383*** (7.67)
$\eta_{2,1}$	-0.00117*** (-4.11)	-0.00120*** (-9.09)	0.0000333 (0.37)
$\eta_{2,2}$	-0.00779*** (-6.83)	-0.00159** (-3.23)	-0.00124** (-3.19)
Observations	2600	2600	2600
Log Lik.	5489.3	7508.5	8106.5

t statistics in parentheses
* $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 7.6: Resultados modelos para variable de incidencia y categoría *rolls*