



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

ESTUDIO DEL COMPORTAMIENTO DE LOS CLIENTES FRENTE A UN QUIEBRE
DE STOCK Y SU IMPACTO ECONÓMICO EN UN SUPERMERCADO

MEMORIA PARA OPTAR AL TÍTULO DE INGENIERA CIVIL INDUSTRIAL

LORETTO MONSERRAT GATICA QUINTANILLA

PROFESOR GUÍA:
RICARDO MONTOYA MOREIRA

MIEMBROS DE LA COMISIÓN:
RICARDO MONTOYA MORERIRA
MARCELO OLIVARES ACUÑA
ANDRÉS MUSALEM SAID

SANTIAGO DE CHILE
JUNIO 2014

ESTUDIO DEL COMPORTAMIENTO DE LOS CLIENTES FRENTE A UN QUIEBRE
DE STOCK Y SU IMPACTO ECONÓMICO EN UN SUPERMERCADO

La determinación de los niveles adecuados de servicio en góndola requiere equilibrar los costos de inventario y reposición con el beneficio de tener una mayor disponibilidad del producto para los clientes. Por lo general los costos de inventario son medibles, en cambio los relacionados con los quiebres de stock (Out Of Stock OOS por sus siglas en inglés) son menos visibles y más difíciles de medir.

El objetivo de este trabajo es cuantificar el impacto económico asociado al comportamiento del consumidor frente a los quiebres de stock. En específico, se estudia aquellos cambios de conducta de corto plazo que tiene el cliente en la misma compra y en la siguiente a la que se enfrentó al OOS, dentro de la categoría pan y en la categoría complementaria fiambrería.

Para ello se utiliza una base de datos de panel con 2.637 clientes, los que realizaron un total de 24.751 transacciones durante aproximadamente un mes y medio. Estos datos tienen información de punto de venta y de la disponibilidad de los productos que enfrentó el cliente al momento de realizar la compra. La información de quiebres de stock es capturada por cámaras que miden cada 30 minutos los niveles de disponibilidad de los distintos panes de la sección pan a granel. Para inferir si los clientes se enfrentaron a un quiebre de stock al momento de comprar, se realiza un cruce de información entre las mediciones de cámaras y las transacciones en el punto de venta utilizando la hora en la que el cliente realizó la transacción en caja. Con la información anterior, se desarrollan modelos de elección discreta, específicamente el modelo mixed logit, que incorporan heterogeneidad en las preferencias de clientes para estimar los efectos causados por el OOS dentro de la categoría pan y regresiones lineales para calcular los efectos cruzados en las ventas de otras categorías. Esta heterogeneidad es importante para que el modelo tenga flexibilidad suficiente para capturar patrones de sustitución entre productos.

El principal resultado obtenido es que los quiebres de stock en una categoría no sólo disminuyen las ventas de ésta, sino también las de las categorías complementarias. Produciéndose la mayor pérdida para el supermercado cuando la hallulla y la marraqueta no están disponibles, disminuyendo las ventas de la panadería en un 2,9 % y en un 0,31 % las de la fiambrería. La suma de esta pérdida asciende a los US\$ 313.480. Otro resultado interesante es la subestimación de elasticidad precio cuando se ignoran los OOS. Además, la lealtad de los clientes disminuye en importancia al incorporar la existencia de OOS intertemporal y contextual.

El conocer los efectos económicos que genera una baja en el nivel de disponibilidad de los productos le permite a la empresa determinar las estrategias de reposición óptimas que deben implementar. Como trabajo futuro se sugiere analizar los efectos de largo plazo que puede producir un OOS en el comportamiento del cliente a través de modelos que incorporen cambios estructurales en las preferencias del consumidor, tal como las cadenas de Markov ocultas (Hidden Markov Models).

A la vida por sus sorpresas y oportunidades...

Tabla de contenido

1. Introducción	1
1.1. Objetivo General	3
1.2. Objetivos Específicos	3
1.3. Alcances	4
2. Datos	5
2.1. Datos Transaccionales	5
2.2. Datos de Panadería	7
2.3. Datos Faltante Mercadería en Góndola	8
2.4. Datos de Panel	10
3. Efecto Dentro de la Categoría	13
3.1. Modelos de Elección Discreta	13
3.1.1. Características de un Modelo de Elección Discreta	13
3.1.2. Modelo Mixed Logit	14
3.1.3. Estimación del Modelo Mixed Logit	15
3.1.4. Selección del Modelo	16
3.2. Modelo	16
3.2.1. Conjunto de Alternativas	16
3.2.2. Variables	16
3.2.3. Definición de Modelos	20
3.2.4. Análisis de Resultados	21
4. Efecto en Categoría Complementaria	26
4.1. Variables Instrumentales	26
4.2. Reglas de Asociación	27
4.3. Selección de la Categoría Complementaria	28
4.4. Modelo	29
4.4.1. Variables	29
4.4.2. Definición de Modelos	30
4.4.3. Análisis de los Resultados	31
5. Efecto Económico	34
5.1. Efectos Económicos Dentro de la Categoría	34
5.2. Categoría Complementaria	38
6. Conclusiones	41

6.1. Conclusiones	41
6.2. Limitaciones	42
6.3. Recomendaciones y Trabajos Futuros	43
Bibliografía	44
Apéndice	46
A. Promedio de las Variables en Panadería	46
B. Promedio de las Variables en Fiambrería	46

Índice de tablas

2.1. Número de compras promedio por horario y tipo de día	6
2.2. Análisis de canasta en compras con y sin pan para todo el periodo de transacción	8
2.3. Características descriptivas del panel de datos	12
2.4. Análisis de canasta en compras con y sin pan en el panel de datos	12
3.1. Coeficientes Modelos MixedLogit en la Categoría Pan	22
4.1. Reglas de asociación panadería, fiambrería y electrodomésticos	29
4.2. Coeficientes Modelos Regresión RE con IV Fiambrería	32
5.1. Incidencia para distintos escenarios en la categoría pan	35
5.2. Pérdida económica mensual para distintos escenarios de quiebre de stock	38
5.3. Incidencia de la categoría fiambrería para distintos escenarios de quiebre de en la categoría pan	39
5.4. Pérdida económica mensual en la categoría fiambrería para distintos escenarios de quiebre de stock en la panadería	40
A.1. Promedio de las variables en los datos de panel para la categoría panadería	46
B.1. Promedio de las variables en los datos de panel para la categoría fiambrería	46

Índice de figuras

1.1.	Principales causas de los quiebres de stock	2
1.2.	Reacciones de los clientes frente a un quiebre de stock	2
2.1.	Número promedio de transacciones por hora del día	6
2.2.	Número promedio de transacciones por día de la semana	7
2.3.	Participación de los productos de la categoría pan en todo el periodo de transacción	8
2.4.	Ejemplo medición OOS	9
2.5.	Porcentaje promedio de OOS por hora del día en la categoría pan a granel .	9
2.6.	Porcentaje promedio de OOS por día de la semana en la categoría pan a granel	10
2.7.	Participación de los productos de la categoría pan en los datos de panel . . .	12
5.1.	Efecto de un quiebre de stock en la incidencia de la hallulla	36
5.2.	Efecto de un quiebre de stock en la incidencia de la marraqueta	36
5.3.	Efecto de un quiebre de stock en la incidencia de la hallulla y la marraqueta	36

Capítulo 1

Introducción

La determinación de los niveles adecuados de servicio en góndola requiere equilibrar los costos de inventario y reposición con el beneficio de tener una mayor disponibilidad del producto para los clientes. Por lo general los costos de inventario son medibles, en cambio los relacionados con los quiebres de stock (Out Of Stock OOS por sus siglas en inglés) son menos visibles y difíciles de medir.

Fisher y Raman (2010) determinaron que uno de los principales factores que influye en las ventas del retail es el nivel de disponibilidad de los productos, es decir, que el cliente encuentre lo que busca en el momento en que lo requiere. El índice que permite evaluar si se está cumpliendo con lo anterior es el quiebre de stock, el cual tiene diferentes definiciones, siendo una de las más aceptadas la siguiente, es el porcentaje de productos que no están en la góndola en un momento determinado (Gruen et al., 2002).

El promedio mundial de quiebres de stock es de un 8,3 % y varía según la región y las categorías que se estén evaluando (Gruen et al., 2002). En Chile Bosch et al. (2006) realizaron un estudio, en el que se muestra que el promedio de OOS es de un 15 %, muy por sobre el promedio mundial.

Los quiebres de stock son causados por una mala gestión del inventario en la cadena de suministro. El manejo del inventario es complejo debido a la gran cantidad de productos y proveedores. Como se observa en la Figura 1.1 las principales razones de la no disponibilidad de los productos en góndola son una subestimación de la demanda, un sistema lento de orden a los proveedores y mala reposición en góndola. A éstas se les llama causas dentro de la tienda y producen entre un 70 % y un 75 % de los quiebres de stock. El resto son causados por problemas río arriba, es decir, de los proveedores.

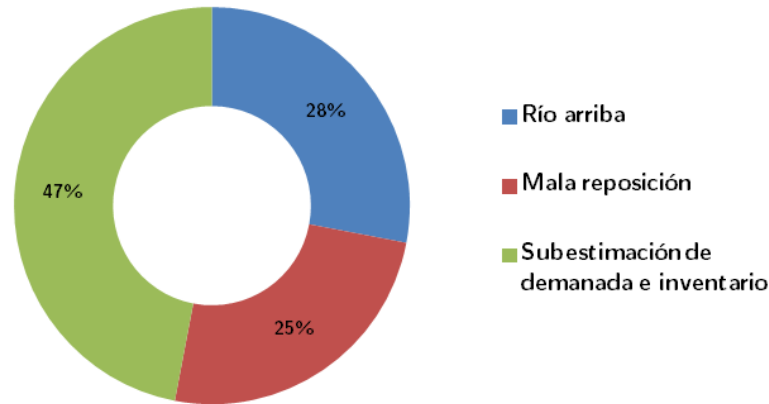


Figura 1.1: Principales causas de los quiebres de stock

La no disponibilidad de los productos en la góndola genera un cambio en el comportamiento de compra del cliente. Tal como se muestra en la Figura 1.2 las reacciones de un cliente frente a un quiebre de stock pueden ser: no comprar el producto en la tienda; sustituir el producto por uno de la misma marca pero de otro tamaño; sustituir el producto por uno de una marca diferente; retrasar la compra; comprar en otra tienda. Dentro de estas reacciones la más común es sustituir el producto por uno de otra marca (45 %).

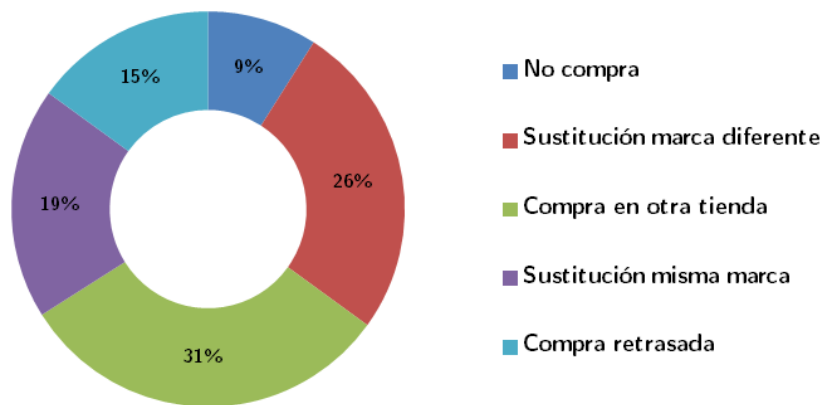


Figura 1.2: Reacciones de los clientes frente a un quiebre de stock

El cambio en el comportamiento del cliente producto de un quiebre de stock impacta no sólo al retail sino también a sus proveedores. Los efectos directos sobre el retail pueden ser una baja en las potenciales ventas del producto en OOS, ya que el consumidor puede decidir comprar en otra tienda e incluso simplemente no comprarlo, lo que afectaría al proveedor. Por otro lado, el cliente podría sustituir el producto generando pérdidas para el retail y los proveedores, porque esta sustitución se hace por un producto de menor tamaño y/o de una marca más económica según Gruen et al. (2002). Además el que el cliente enfrente un OOS genera una baja en su satisfacción, lo cual podría generar consecuencias indirectas como cambiar permanentemente la tienda de compra y el producto por el que se sustituyó.

Para el retail es fundamental conocer cuál es el impacto económico que tienen las diferentes reacciones del cliente frente a un OOS, sin embargo, no es directo cuantificarlo. La principal dificultad se encuentra en identificar el momento exacto en el que ocurre el quiebre de stock para luego definir cuáles fueron los clientes que estuvieron expuestos a él y así poder estudiar cuál fue su comportamiento.

Esta problemática ha sido abordada antes por otros estudios, los cuales utilizan distintas metodologías para definir el momento en que ocurre el quiebre de stock y cuáles fueron los clientes que lo enfrentaron. Che et al. (2012) cuentan con mediciones en un horario determinado de la disponibilidad de los productos y considera que aquellos clientes que pasaron por caja luego de esta medición fueron los que estuvieron expuestos en el caso de que hubiese sucedido el quiebre de stock, si es que existió. Si bien esta metodología funciona bien en casos con reposición diaria al fin del día puede no funcionar bien en el caso de reposición durante el día, como es el caso de productos de alta rotación. Si existe reposición posterior a la medición hay clientes que en realidad no se enfrentaron a una góndola sin producto. Además, antes de la medición también pudo existir productos en quiebres de stock que fueron repuestos. Musalem et al. (2010) también utilizan información de inventario proporcionada por mediciones al inicio y al final del día, para identificar el momento en que ocurre el OOS. Por otro lado, Campo et al. (2003) generan información de los OOS a través de los datos transaccionales, donde estiman la existencia de un quiebre de stock cuando las ventas de un producto son más bajas que las esperadas. Las medidas de quiebres de stock en este caso son más precisas, pero existen problemas cuando los productos son de baja rotación, ya que es normal observar cero ventas incluso si el producto está disponible.

Para disminuir los errores en la inferencia de los clientes que se encontraron con un quiebre de stock este trabajo utiliza datos de medición de góndola cada 30 minutos e información del punto de venta. De esta manera se pretende cuantificar de manera económica las distintas reacciones del cliente frente a un quiebre de stock. Así el retail podrá evaluar diferentes métodos para subsanarlos sin que la solución tenga un mayor costo que el efecto del OOS.

1.1. Objetivo General

Cuantificar el impacto económico asociado al cambio en el comportamiento del consumidor frente a los quiebres de stock.

1.2. Objetivos Específicos

- Identificar y medir los efectos en el corto plazo dentro de la categoría en donde se produce el quiebre de stock.
- Identificar y medir los efectos en el mediano plazo dentro de la categoría en donde se produce el quiebre de stock.
- Identificar y medir los efectos de corto plazo en una categoría relacionada a la que

presenta productos en quiebre de stock.

- Cuantificar cada uno de los efectos de un quiebre de stock identificados.

1.3. Alcances

- Se estudiarán sólo efectos de corto y mediano plazo, es decir, aquellos producidos en la misma compra en la que el cliente se enfrentó al quiebre de stock y la compra siguiente.
- Se utilizarán datos transaccionales de un supermercado y la información que proporcionen las cámaras dispuestas en las góndolas. Además de estos datos, no se consideran otras fuentes de información.
- Sólo se aplicará la metodología en un local de la cadena de supermercado.
- Las mediciones de disponibilidad en góndola se realizarán en una categoría focal, en este caso el pan.
- Se analizará sólo una categoría complementaria a la categoría focal, la cual corresponde a la fiambrería.

Capítulo 2

Datos

Para cumplir los objetivos de esta memoria, se necesita idealmente información que permita identificar a cada uno de los clientes que visitó un supermercado, los artículos comprados y la disponibilidad en góndola de los productos. En este caso se cuenta con dos bases de datos, una transaccional y la otra de medición de quiebres de stock. Se utilizarán estos datos para conformar un panel con las transacciones de los clientes en el periodo de medición de disponibilidad de los productos.

2.1. Datos Transaccionales

Para identificar los clientes y su comportamiento al comprar en un supermercado se utilizarán datos del punto de venta (Point Of Sale POS por sus siglas en inglés), los cuales entregan información detallada de las transacciones. Se cuenta con información de las visitas y montos comprados por los clientes en un local de supermercado durante un año. Adicionalmente, si la compra incluye algún producto de la categoría pan, fiambre o electrodomésticos se tiene información detallada de la canasta.

En general se tendrá la siguiente información en cada transacción:

- Identificador de cliente (ID).
- Fecha y hora de la compra.
- Productos que se llevaron en la compra con sus respectivos precios.
- Categoría a la que pertenece cada producto.

En esta base de datos se tiene que el 69,5% de los tickets poseen ID de cliente, teniendo así 95.840 clientes distintos. El 92,7% de estos clientes compraron en un local distinto al utilizado en esta memoria, con un promedio de 5 visitas externas por ID. Como para esta memoria es importante el comportamiento del cliente, se analizarán sólo aquellos datos de POS que posean ID para las estadísticas descriptivas que se mostrarán a continuación.

Como ya se mencionó hay 95.826 clientes distintos que visitan el supermercado en 12

meses, generando 1.636.204 transacciones con información de ID. En promedio estos clientes realizan 17,05 compras en el local. El número promedio de transacciones realizados por estos clientes son mayores en las tardes que en las mañanas y los fines de semana que los días de semana. Ambas diferencias son estadísticamente significativas ($p - value < 0,0001$ en ambos casos). En la Tabla 2.1 se puede apreciar el número de compras promedio por horario y tipo de día.

	Número de compras	
	Promedio	stdv
Mañana	238,77	169,94
Tarde	374,59	148,40
Día de semana	4.070,22	961,62
Fin de semana	5.769,53	915,15

Tabla 2.1: Número de compras promedio por horario y tipo de día

La Figura 2.1 y la Figura 2.2 muestran con mayor detalle lo mencionado anteriormente. Se observa que hay mayor cantidad de transacciones entre las 6pm y las 9pm al comparar por horario del día. En cuanto a los días de la semana, son los sábados y domingos los que reciben mayor cantidad de visitas. Esto podría indicar que existe estacionalidad en las ventas.

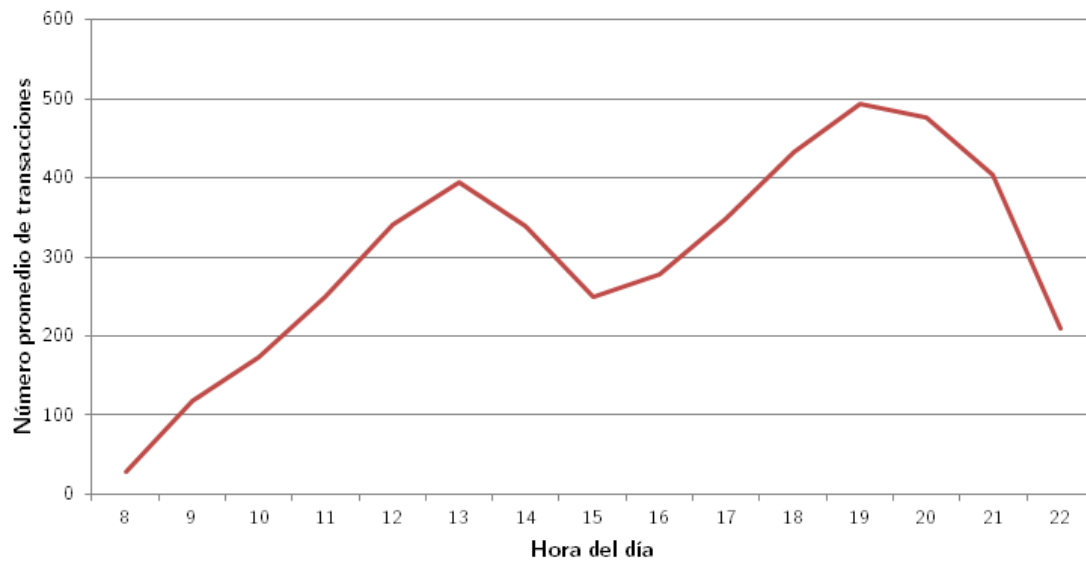


Figura 2.1: Número promedio de transacciones por hora del día

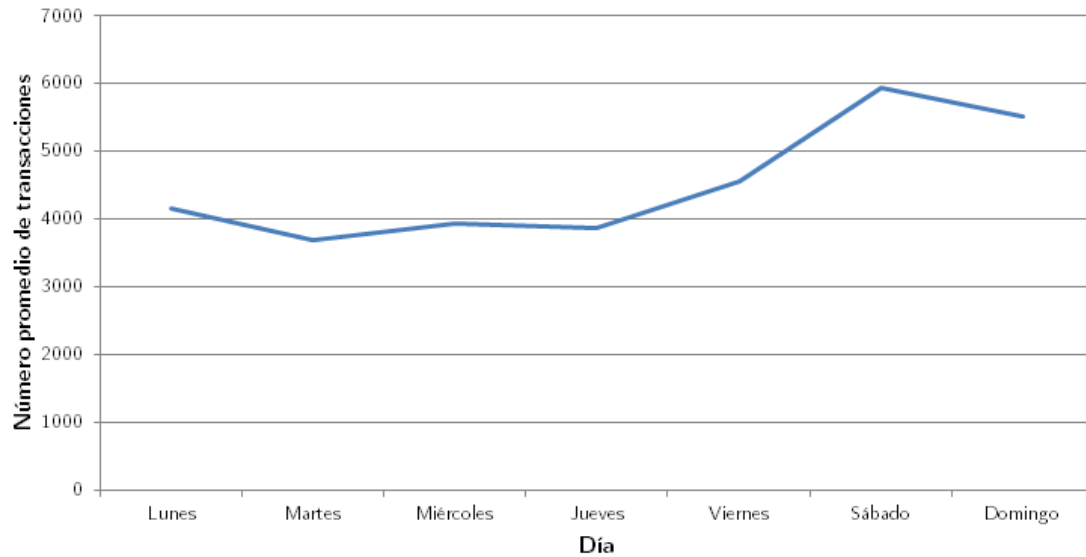


Figura 2.2: Número promedio de transacciones por día de la semana

2.2. Datos de Panadería

El supermercado en análisis tiene una gran variedad de panes a granel (34) y de panes envasados (72). La incidencia de la categoría pan en el total de compras que se realizaron en los 12 meses es del 47,22 %, es decir, de cada 100 compras que se observan en los datos, 47 de ellas llevan alguno de los productos de esta sección.

Dentro de la categoría pan hay dos productos que se encuentran en la mayoría de las compras donde se lleva pan. Al observar la Figura 2.3, se puede ver que la marraqueta es el pan más vendido (25 %), seguido por la hallulla (19 %). Ambos panes en conjunto representan el 44 % de las ventas de la categoría. Dentro de los otros panes a granel se encuentran los 32 productos restantes, teniendo cada uno de ellos en promedio un 0,97 %.

De manera de seguir analizando la importancia de esta categoría en las ventas del supermercado, se ha hecho una comparación entre las compras que incluyen algún producto de la categoría pan y aquellas que no los llevan. Los resultados se pueden ver en la Tabla 2.2.

Como se ve en la Tabla 2.2, las transacciones que incluyen algún producto de la categoría pan tienen en promedio un mayor monto y llevan una mayor cantidad de productos. Ambas diferencias son estadísticamente significativas con un $p - value < 0,0001$.

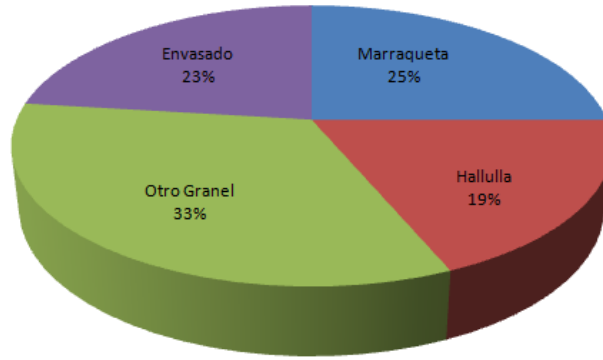


Figura 2.3: Participación de los productos de la categoría pan en todo el periodo de transacción

	¿Canasta incluye pan?			
	No		Si	
	<i>Promedio</i>	<i>stdv</i>	<i>Promedio</i>	<i>stdv</i>
Tamaño de la canasta (#SKU)	14,24	23,04	26,46	31,01
Monto de la canasta (\$)	18.505,40	26.810,97	29.862,89	34.084,10

Tabla 2.2: Análisis de canasta en compras con y sin pan para todo el periodo de transacción

Teniendo en cuenta esta información, se puede decir que la sección panadería es importante dentro de este supermercado.

2.3. Datos Faltante Mercadería en Góndola

Es necesario monitorear los niveles de disponibilidad de los productos en góndola para identificar el momento en que ocurren los quiebres de stock. El monitoreo en esta memoria se hace mediante cámaras que apuntan hacia las góndolas y registran la disponibilidad del producto cada 30 minutos. Estos datos son proporcionados por la empresa Scopix Solution. Específicamente, se cuenta con datos de OOS para la categoría pan a granel. Cada vez que se hace una observación se registra los SKUs dentro de la categoría que tienen menos del 20 % de su canasto asignado lleno, considerándose esto como un OOS¹. Estas mediciones se realizaron diariamente durante aproximadamente un mes y medio en un local de la cadena de supermercados (05-03-2009 al 19-04-2009).

¹En la categoría pan a granel esto se considera un quiebre de stock, dado que por lo general, el producto que queda al final del canasto está en malas condiciones, lo que hace que los clientes no lo compren. Esto podría ser diferente en otras categorías, como abarrotos.

En la Figura 2.4 se muestra un ejemplo de las mediciones de disponibilidad en góndola. Aquí se puede apreciar que si el producto no se encuentra disponible, en la base de datos se indica con un uno. En cambio, si el producto está en góndola se indica con un cero.



Figura 2.4: Ejemplo medición OOS

En el contexto del supermercado del que se tienen las mediciones, los quiebres de stock se producen en mayor medida en el horario de la tarde (entre las 16:00hrs. y las 22:00hrs.) y los fines de semana, tal como se muestra en la Figura 2.5 y la Figura 2.6. Las diferencias entre el promedio de quiebres de la mañana y la tarde son estadísticamente significativas ($p\text{-value} < 0,0001$), al igual que entre los días de semana y fin de semana ($p\text{-value} < 0,0001$). Estos aumentos en el porcentaje de OOS podrían estar correlacionados con el número de ventas que se registran en estos periodos, ya que tal como se mostró en la Sección 2.1 hay mayores compras en el horario de la tarde que en el de la mañana y los fines de semana que en los días de semana.

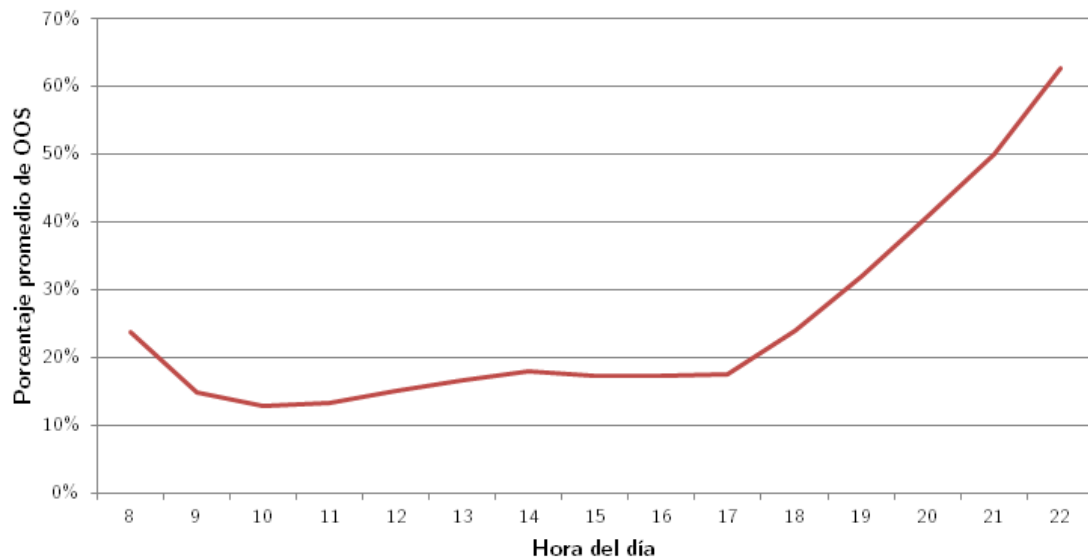


Figura 2.5: Porcentaje promedio de OOS por hora del día en la categoría pan a granel

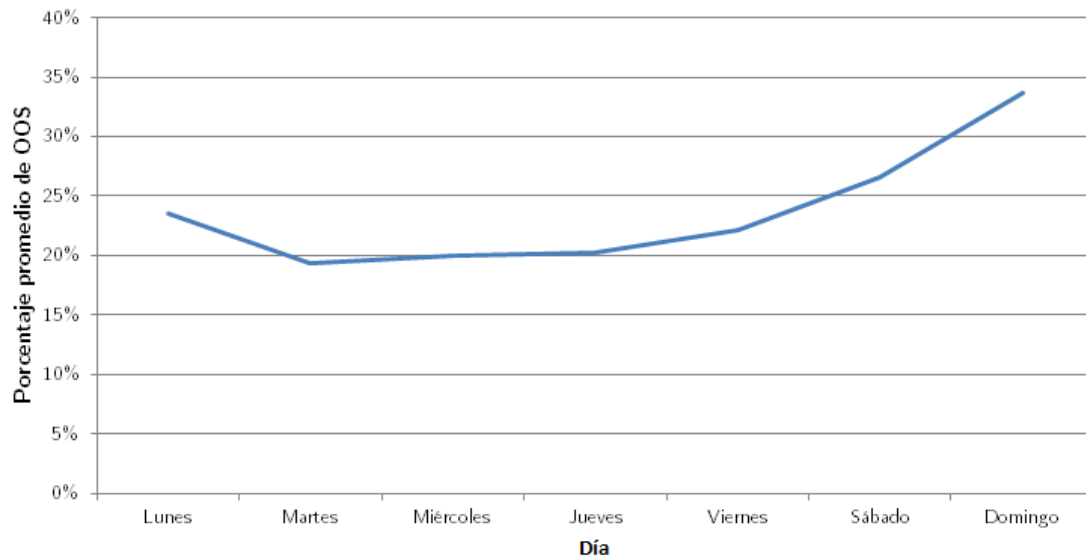


Figura 2.6: Porcentaje promedio de OOS por día de la semana en la categoría pan a granel

Con estos datos se puede identificar la disponibilidad de los productos al momento que el consumidor visitó el supermercado e inferir si enfrentó un quiebre de stock. El procedimiento para cruzar esta información se detallará en la Sección 2.4

2.4. Datos de Panel

Para analizar el comportamiento de los clientes frente a un OOS se crea una base con datos de panel, que proporciona un historial de compras, de manera que se puedan capturar las preferencias de los consumidores a lo largo del tiempo. Estos datos son cruzados con las mediciones de OOS, para analizar si esta variable tiene algún efecto en las decisiones que toma el cliente al momento de escoger un producto.

Al realizar el cruce de información entre las transacciones y las mediciones de OOS, se asume que el cliente pasa un tiempo en la sala del supermercado antes de pagar. Así cuando pasa por una caja se le conecta con aquella medición de disponibilidad anterior más cercana a la hora de esta transacción. Por ejemplo, si se tiene que el ID X compró a las 19:25 hrs. y se cuenta con las mediciones de disponibilidad en la categoría pan de las 19:00 hrs. y las 19:30 hrs, se selecciona la información de OOS correspondiente a las 19:00 hrs. Infiriendo que esa fue la disponibilidad que enfrentó el cliente X a la hora de pasar por la categoría pan.

Las transacciones donde no se incluye algún producto de la sección panadería, fiambrería o electrodomésticos sólo cuentan con la hora en que el cliente pasó por caja. Por lo que para realizar el cruce de información con las mediciones de disponibilidad es necesario simular el minuto en que se realizó la compra. Para esto se utiliza una simulación multinomial con los datos que presentan las horas y minutos de compra (Ross, 2006). Esta simulación puede tener un error al suponer que las personas que no compran pan, electrodomésticos o fiambres

tiene la misma distribución de visitas al supermercado que aquellas que si compran en estas categorías. Además, como los datos de OOS tienen algunas mediciones en blanco, se decide imputar. El criterio que se utilizó para esto es buscar si hubo una compra del producto en los 30 minutos posteriores a la medición de quiebre de stock, según corresponda.

Continuando con la construcción de los datos de panel, se realiza una limpieza de la base de datos, quedando así sólo los datos de POS que tienen el ID del cliente, ya que se necesita hacer un seguimiento de éste. Por otro lado, para efectos de este estudio el panel estará conformado por alternativas excluyentes para cada una de las visitas que realice el cliente, por lo que se eliminaron todas aquellos clientes que dentro del periodo de medición de OOS escogieron más de una de las alternativas, que a continuación se enumeran, en una de sus compras.

- Compró hallulla (CH).
- Compró marraqueta (CM).
- Compró algún pan a granel, excluyendo hallulla y marraqueta (CG).
- Compró algún pan envasado (CE).
- No compró ningún producto de la categoría pan (NC).

Además, sólo se trabajará con aquellos clientes que tuvieron un número de compras mayor o igual a 10 en la categoría pan a granel durante todo el periodo de transacciones (Che et al., 2012). En esta misma línea, los clientes deben tener al menos 6 compras antes de la etapa de medición de OOS y 4 dentro de este periodo, es decir contar con al menos una compra por mes antes de las mediciones de disponibilidad y una compra por semana dentro del periodo de medición, respectivamente.

Los filtros anteriores se aplicaron con tal de tener una mayor probabilidad de capturar los efectos que podría producir el OOS, ya que se considera que el comportamiento del cliente que compra de forma regular en la categoría se debiese ver más afectado que uno que no lo hace a menudo. En este mismo sentido, se decide trabajar con aquellos clientes donde las compras en la categoría pan a granel representen al menos el 30,0 % del total de sus compras, ya que esta categoría tiene una incidencia del 37,90 % en las compras totales del supermercado en el periodo de datos de POS.

Después de realizar los procedimientos anteriormente mencionados, se cuenta con un panel de 2.637 clientes los cuales realizaron un total de 24.751 compras durante aproximadamente un mes y medio (05-03-2009 al 19-04-2009). Un cliente compra en promedio 9.39 veces en este periodo, donde la menor cantidad de visitas que hizo uno de ellos fue 4 y la mayor fue 80. Dentro de las características de este panel, también se tiene que el 46,90 % de las transacciones lleva pan, es decir los clientes de cada 10 compras que hacen en 4 de ellas llevan algún producto de la categoría pan.

Todas las características descriptivas de los datos de panel mencionadas se encuentran resumidas en la Tabla 2.3.

En la Figura 2.7 se puede observar las preferencias de los consumidores en la sección panadería. La hallulla y la marraqueta son los productos más comprados por los clientes,

estando presentes en conjunto en un 50 % de las ventas de la categoría.

Datos de panel	
Periodo	05-03-2009 al 19-04-2009
Número de clientes	2.637,00
Número de compras	24,00
Promedio de compras	9,30
Mínimo de compras realizadas por un cliente	4,00
Máximo de compras realizadas por un cliente	80,00
Compras que incluyen pan	46,90 %

Tabla 2.3: Características descriptivas del panel de datos

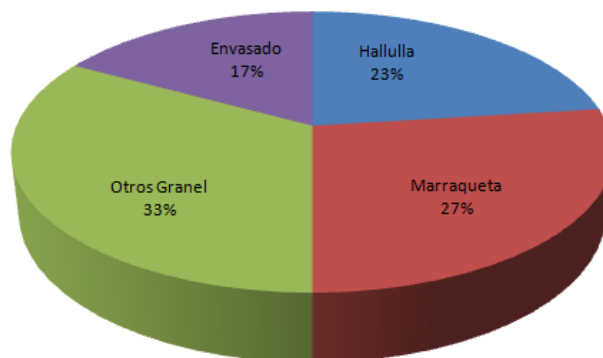


Figura 2.7: Participación de los productos de la categoría pan en los datos de panel

Otra característica interesante de mencionar son las diferencias entre las canastas de los consumidores que llevan algún producto de la sección panadería y las que no. En la Tabla 2.4 se observa que en promedio las compras que llevan pan tienen un monto de canasta más alto, así como también llevan una mayor cantidad de productos. Ambas diferencias son estadísticamente significativas (con ambos $p - value < 0,0001$). Lo cual es consistente con lo expuesto en la Sección 2.2.

	¿Compras incluyen pan?			
	No		Si	
	<i>Promedio</i>	<i>stdv</i>	<i>Promedio</i>	<i>stdv</i>
Tamaño de la canasta (#SKU)	14,97	22,94	22,71	31,01
Monto de la canasta (\$)	18.381,87	24.987,43	25.265,16	31.543,53

* $p < 0,01$

Tabla 2.4: Análisis de canasta en compras con y sin pan en el panel de datos

Capítulo 3

Efecto Dentro de la Categoría

En este capítulo se medirá el efecto que tiene un quiebre de stock en la categoría pan sobre el comportamiento de los clientes, en particular se estudiarán los cambios en la elección de los productos pertenecientes a la misma categoría según su disponibilidad. Para lo cual se utilizará un modelo de elección discreta, que permite calcular la utilidad que le genera a un tomador de decisión escoger una alternativa dentro de un conjunto de opciones. Se estudiarán los efectos de un quiebre temporal, intertemporal y contextual (quiebres de los demás productos en la categoría). Además de los efectos de la lealtad, precio y estacionalidad.

3.1. Modelos de Elección Discreta

Los modelos de elección discreta describen las decisiones realizadas por los tomadores de decisión al momento de escoger una alternativa dentro de un conjunto de opciones. Estos tomadores de decisión pueden ser personas, hogares, empresas, entre otras. Las alternativas pueden representar distintos productos en un mercado, los cursos de acción o cualquier otra opción o elementos sobre los que habrá que elegir.

3.1.1. Características de un Modelo de Elección Discreta

En el marco de un modelo de elección discreta el conjunto de alternativas dentro del conjunto de elección tiene que presentar tres características.

1. Las alternativas deben ser **mutuamente excluyentes** desde la perspectiva de la toma de decisiones, es decir, el tomador de decisiones escoge sólo una alternativa del conjunto de elección.
2. Se debe incluir todas las alternativas posibles, lo que se conoce con el nombre de un **conjunto exhaustivo**. Así, el tomador de decisiones elige necesariamente una de las alternativas.

3. El número de alternativas debe ser **finito**.

Es importante señalar que los modelos de elección discreta se derivan generalmente bajo el supuesto de un comportamiento maximizador de utilidad por parte de los tomadores de decisiones. A los modelos que utilizan este supuesto se le conocen como Modelos de Utilidad Aleatoria (Random Utility Models RUMs por sus siglas en inglés).

Los RUMs consideran que un tomador de decisión $i = 1, \dots, N$ escoge entre J alternativas y obtiene cierto nivel de utilidad, es decir, recibe un beneficio al realizar cada acción, para cada una de las elecciones realizadas. La utilidad que le entrega al tomador de decisiones i elegir la alternativa j es U_{ij} con $j = \{1, \dots, J\}$. Esta utilidad se descompone como se aprecia en la Ecuación (3.1). Donde V_{ij} es la parte determinista y ε_{ij} es la parte aleatoria. Además, V_{ij} queda descrita según la Ecuación (3.2), donde x_{ij} es un vector de variables observables y β_j corresponde al vector de parámetros de esas variables, que en general son desconocidos y deben ser estimados estadísticamente. La parte aleatoria ε_{ij} , que es desconocida para el investigador (el tomador de decisiones ve todo), corresponde a los factores que afectan la utilidad pero que no están incluidos en V_{ij} . La densidad conjunta del vector de variables aleatorias $\varepsilon_{ij} = \langle \varepsilon_{ij}, \dots, \varepsilon_{iJ} \rangle$ es denominada $f(\varepsilon_{ij})$. Con esta densidad, se puede determinar probabilísticamente la elección del tomador de decisión (Train, 2009).

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (3.1)$$

$$V_{ij} = V(\beta_j, x_{ij}) \quad (3.2)$$

Luego la probabilidad de que el cliente i escoja la alternativa j está dado por la Ecuación (3.3).

$$\begin{aligned} P &= P(U_{ij} > U_{ik} \quad \forall j \neq k) \\ &= P(V_{ij} + \varepsilon_{ij} > V_{ik} + \varepsilon_{ik} \quad \forall j \neq k) \\ &= P(\varepsilon_{ik} - \varepsilon_{ij} < V_{ij} - V_{ik} \quad \forall j \neq k) \end{aligned} \quad (3.3)$$

3.1.2. Modelo Mixed 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 por la Ecuación (3.4), donde $L_{ij}(\beta)$ corresponde a la probabilidad de que el individuo i escoja la alternativa j evaluada en el parámetro β lo cual es equivalente a la Ecuación (3.5), $f(\beta)$ es la función de densidad de los parámetros β y $V_{ij}(\beta_j)$ corresponde a la parte determinista de la utilidad que recibe el individuo i al escoger la alternativa j (Train, 2009).

$$P_{ij} = \int L_{ij}(\beta) f(\beta) d\beta \quad (3.4)$$

$$L_{ij}(\beta) = \frac{e^{V_{ij}(\beta_j)}}{\sum_{k=1}^J e^{V_{ik}(\beta_k)}} \quad (3.5)$$

Por lo tanto, la función de verosimilitud se puede ver como un promedio ponderado de distintas de funciones, el peso está dado por la función de densidad asociada a cada parámetro β . Donde los valores de β representan las preferencias de cada individuo.

El modelo mixed logit es altamente flexible, puede aproximarse a cualquier modelo de utilidad aleatoria y obvia las limitaciones que tiene el modelo logit estándar al permitir variación en los gustos aleatorios, patrones de sustitución irrestrictos y correlación entre los factores no observados durante el tiempo (Train, 2009).

3.1.3. Estimación del Modelo Mixed Logit

Según Train (2009), la estimación de un modelo Mixed Logit se ajusta bien con métodos de simulación. Si se tiene que la utilidad de un individuo i al escoger la alternativa j es $U_{ij} = \beta_i' x_{ij} + \varepsilon_{ij}$ y los coeficientes β_i se distribuyen con densidad $f(\beta | \theta)$, donde θ representa los parámetros de la distribución, como la media y la varianza de β . La probabilidad de elección se define como se muestra en la Ecuación (3.6).

$$P_{ij} = \int L_{ij}(\beta) f(\beta | \theta) d\beta \quad (3.6)$$

Luego, las probabilidades son aproximadas a través de simulación para cualquier valor de θ de la siguiente manera:

1. Elegir un valor de β para $f(\beta | \theta)$, al cual se le llama β^r , donde $r = 1$ representa el primer β escogido.
2. Calcular el valor de $L_{ij}(\beta^r)$.
3. Repetir el paso 1 y 2 R veces y promediar los resultados. Este promedio es la probabilidad simulada, tal como se muestra en la Ecuación (3.7).

$$\check{P}_{ij} = \frac{1}{R} \sum_{r=1}^R L_{ij}(\beta^r) \quad (3.7)$$

Algunas propiedades de \check{P}_{ij} son:

- Es un estimador insesgado.
- Su varianza decrece mientras incrementa R .
- Es estrictamente positivo, por lo que $\ln(\check{P}_{ij})$ está bien definido.
- La suma sobre todas las alternativas de \check{P}_{ij} es uno.

Luego estas probabilidades simuladas se utilizan en la función de verosimilitud dando paso así a la log-verosimilitud simulada, la cual se muestra en la Ecuación (3.8).

$$SLL = \sum_{n=1}^N \sum_{j=1}^J d_{ij} \ln(\check{P}_{ij}) \quad (3.8)$$

donde $d_{ij} = 1$ si el cliente i escoge la alternativa j y cero en otro caso. El estimador de máxima verosimilitud simulada (Maximum Simulated Likelihood Estimator MSLE por sus siglas en inglés) es el valor de β que maximiza la Ecuación (3.8).

3.1.4. Selección del Modelo

El test de hipótesis que se explicará a continuación se utiliza para comparar modelos anidados. Se tiene como hipótesis nula lo siguiente: muchos de los coeficientes de las variables explicatorias del modelo general son cero (Train, 2009). El estadístico que se usa para testear esta hipótesis se define como muestra la Ecuación (3.9).

$$LR = -2(LL(\hat{\beta}^H) - LL(\hat{\beta})) \quad (3.9)$$

Es así que para testear se necesita estimar dos modelos, uno donde se incluyen variables explicatorias que en el segundo no, con $LL(\hat{\beta}^H)$ y $LL(\hat{\beta})$ la máxima log-verosimilitud respectiva para cada modelo.

El estadístico LR se distribuye *chi-cuadrado* con grados de libertad igual a la diferencia entre la cantidad de variables explicatorias de los modelos a comparar. Si el valor del estadístico LR supera el valor crítico de la *chi-cuadrado* correspondiente se rechaza la hipótesis nula. La hipótesis nula corresponde a que los dos modelos entregan el mismo ajuste.

3.2. Modelo

Para estimar los distintos efectos de un quiebre de stock en la elección de los tomadores de decisión, que en este caso serán clientes, se utilizarán los datos de panel descritos en la Sección 2.4. Además, se decide trabajar con un modelo mixed logit multinomial ya que éste permite observar el comportamiento a nivel de cliente, teniendo en cuenta la heterogeneidad que éstos pueden presentar al momento de tomar una decisión.

3.2.1. Conjunto de Alternativas

El modelo mixed logit que se utilizará tendrá el conjunto de elección que se mencionó en la Sección 2.4, es decir, $j \in C = (\text{No Compra, Marraqueta, Hallulla, Granel, Envasado})$. El conjunto C cumplen con las tres características que debe tener un conjunto de alternativas de un modelo de elección discreta.

3.2.2. Variables

La manera de determinar el efecto del OOS en el comportamiento del cliente, será a través de la variación de demanda por las alternativas del conjunto C cuando hay un cambio en la

disponibilidad de estos productos. Por lo tanto, la variable independiente del modelo queda definida como sigue.

$$y_{ijt} = \begin{cases} 1 & \text{si el cliente } i \text{ escoge la alternativa } j \text{ en la ocasión de compra } t \\ 0 & \sim \end{cases}$$

La función de utilidad determinística se construye con las siguientes variables independientes que se creen están relacionadas con la variable dependiente y los efectos que se quieren capturar.

- **Variable de lealtad para cada producto**

Esta variable busca capturar la preferencia de un cliente por cada producto en una ocasión de compra determinada. Para esto, se utiliza su historial de lealtad al producto y la información sobre la ocasión de compra anterior, tal como se muestra en la Ecuación (4.3) (Guadagni and Little, 1983).

$$\text{Loyalty}_{ijt} = \alpha \text{Loyalty}_{ijt-1} + (1 - \alpha) z_{ijt-1} \quad (3.10)$$

donde

$$z_{ijt} = \begin{cases} 1 & \text{si el cliente } i \text{ escoge la alternativa } j \text{ en la ocasión de compra } t \\ 0 & \sim \end{cases}$$

La inicialización de la variable Loyalty_{ijt} se realiza de la siguiente forma:

$$\text{Loyalty}_{ijt} = \begin{cases} \alpha & \text{si la alternativa } j \text{ es la primera escogida por el cliente } i \\ \frac{1-\alpha}{\text{card}(C)-1} & \sim \end{cases}$$

donde $\text{card}(C)$ es la cardinalidad del conjunto de alternativas j , la cual es igual a 5 en este estudio.

Para la primera compra el valor es $\frac{(1-\alpha)}{(\text{card}(C)-1)}$ si el cliente no compra el producto j , ya que de esta forma se asegura que la suma de las lealtades entre los productos sea igual a 1.

Respecto a la constante α , ésta indica la rapidez con la que crece la lealtad del producto elegido y la disminución de los productos que no son comprados. Para efectos de esta memoria se utiliza un $\alpha = 0,7$ (Guadagni and Little, 1983).

- **Variables de estacionalidad**

Con estas variables se quiere modelar el efecto que podría tener el comprar en distintos días de la semana y el horario en que la transacción ocurre, ya que pueden existir variaciones dentro del día como entre los días, según lo mostrado en las Figuras 2.1 y 2.2 de la Sección 2.1.

Variables binarias de día de la semana

$$\text{week_day}_t = \begin{cases} 1 & \text{si la ocasión de compra } t \text{ ocurre un día de semana} \\ 0 & \sim \end{cases}$$

$$\text{weekend_day}_t = \begin{cases} 1 & \text{si la ocasión de compra } t \text{ ocurre un día de fin de semana} \\ 0 & \sim \end{cases}$$

Variables binarias de horario

$$\text{mañana}_t = \begin{cases} 1 & \text{si la ocasión de compra } t \text{ ocurre entre las 8am y las 11am} \\ 0 & \sim \end{cases}$$

$$\text{media_tarde}_t = \begin{cases} 1 & \text{si la ocasión de compra } t \text{ ocurre entre las 12pm y las 2pm} \\ 0 & \sim \end{cases}$$

$$\text{tarde}_t = \begin{cases} 1 & \text{si la ocasión de compra } t \text{ ocurre entre las 3pm y las 5pm} \\ 0 & \sim \end{cases}$$

$$\text{tarde_noche}_t = \begin{cases} 1 & \text{si la ocasión de compra } t \text{ ocurre entre las 6pm y las 9pm} \\ 0 & \sim \end{cases}$$

$$\text{noche}_t = \begin{cases} 1 & \text{si la ocasión de compra } t \text{ ocurre entre las 10pm y las 11pm} \\ 0 & \sim \end{cases}$$

• **Variable precio**

Se tiene conocimiento de que el precio afecta directamente la utilidad que le produce a un cliente la elección de un producto u otro.

$$\text{precio}_{jt} = \begin{cases} \text{precio} & \text{de la alternativa } j \text{ en la ocasión de compra } t \\ 0 & \text{en el caso que la opción sea no comprar} \end{cases}$$

$$\ln_precio_{jt} = \ln(\text{precio}_{jt} + 1)$$

Para el caso de la alternativa que no representa un producto, sino la compra en una categoría, esta variable será el promedio ponderado entre el precio de cada producto perteneciente a la categoría y su respectiva participación dentro de tal categoría.

$$\text{precio}_{gt} = \sum_{g=1}^G \text{market_share}_g * \text{precio}_{gt}$$

$$\text{precio}_{et} = \sum_{e=1}^E \text{market_share}_e * \text{precio}_{et}$$

donde G y E son el conjunto de productos a granel (sin la marraqueta y hallulla) y envasados, respectivamente.

- **Variables de disponibilidad**

a) *Disponibilidad de la alternativa*

$$OOS_{ijt} = \begin{cases} 1 & \text{si la alternativa } j \text{ se encuentra en estado de OOS en la} \\ & \text{ocasión de compra } t \text{ del cliente } i \\ 0 & \sim \end{cases}$$

Para el caso de la alternativa pan a granel, esta variable será el promedio ponderado entre el OOS de cada producto perteneciente a la categoría y su respectivo participación dentro de la sección pan a granel.

$$OOS_{iCGt} = \sum_{g=1}^G \text{market_share}_g * OOS_{igt}$$

En el caso de la alternativa pan envasado no se tienen mediciones de OOS, por lo que se considera que siempre está disponible.

b) *Disponibilidad Intertemporal*

Esta variable representa la disponibilidad que enfrentó el cliente en su ocasión de compra anterior y se denomina OOS_{ijt-1} .

c) *Disponibilidad Contextual*

Se construye una variable que indique cuál es la disponibilidad de los otros productos dentro de la categoría, lo cual se muestra a continuación.

$$OOS_{ijt}^C = \begin{cases} OOS_{iCMt} & \text{si la alternativa } j \text{ es comprar hallulla} \\ & \text{en la ocasión de compra } t \text{ del cliente } i \\ OOS_{iCHt} & \text{si la alternativa } j \text{ es comprar} \\ & \text{marraqueta en la ocasión de compra } t \\ & \text{del cliente } i \\ \min(OOS_{iCMt} + OOS_{iCHt}, 1) & \text{si la alternativa } j \text{ es comprar granel} \\ & \text{o comprar envasado en la ocasión de} \\ & \text{compra } t \text{ del cliente } i \\ 0 & \text{si la alternativa } j \text{ es no comprar pan} \\ & \text{en la ocasión de compra } t \text{ del cliente } i \end{cases}$$

donde CM y CH corresponden a comprar marraqueta y hallulla, respectivamente.

- **Variable de tiempo entre compras**

Si bien la variable lealtad permite inferir la probabilidad que tiene un cliente de comprar un producto en cierta ocasión, no captura el efecto temporal que producen las compras recientes, según la frecuencia de compra del consumidor. Es por esto que se incluye esta variable que indica cuántos días han pasado desde la última compra.

$$\ln_recency_{it} = \ln(\text{Fecha}_{it} - \text{Fecha}_{it-1})$$

donde Fecha_{it} es la fecha de la ocasión de compra t del cliente i .

3.2.3. Definición de Modelos

Utilizando las variables mencionadas en la sección anterior se construyen tres modelos, para comparar los efectos que tienen cada una de las covariables en la elección de compra del cliente.

Modelo sin Efecto de Disponibilidad

En general es difícil contar con la información del nivel de disponibilidad de los productos que se encuentran en el conjunto de alternativas, por lo que se asume que todos ellos están disponibles para el cliente al momento de realizar la compra. Esto es lo que se incluye en este modelo de demanda, por lo que será estimado sin los datos de OOS entregados por las cámaras. Así la función de utilidad determinística quedará definida por la Ecuación (3.11).

$$v_{ijt} = \beta_i^0 + \beta_i^p \ln(\text{price}_{jt} + 1) + \beta_i^L \text{Loyalty}_{ijt} + \sum_{r=1}^R \beta^r d_t^r \quad (3.11)$$

Donde d_t^r corresponde a las variables binarias día de la semana y de horario descritas en la Sección 3.2.2.

Luego la probabilidad de que el individuo i escoja la alternativa j evaluada en el parámetro β será, tal como se mostró en la Ecuación (3.5).

$$L_{ijt}(\beta_i) = \frac{e^{V_{ijt}(\beta_i)}}{\sum_{k=1}^J e^{V_{ikt}(\beta_i)}} \quad \forall i, j, k \quad (3.12)$$

donde $\beta_i = [\beta_i^0 \ \beta_i^p \ \beta_i^L \ \beta_i^1 \ \beta_i^2 \ \dots \ \beta_i^R]$

Modelo con Efecto de Disponibilidad

Para modelar de mejor forma la realidad con la que se enfrenta un cliente al momento de hacer la compra se utilizará la información de disponibilidad de las alternativas, ya que si un producto está en OOS en el momento que el cliente compra no debiese ser considerado en el conjunto de alternativas debido a que es imposible que pueda comprarlo.

De este modo se tiene la misma función de utilidad determinística mostrada en la Ecuación (3.11). Sin embargo, para considerar sólo el subconjunto de alternativas disponibles según la ocasión de compra, se realiza una modificación en la forma de calcular la probabilidad que se tiene en la Ecuación (3.13), ver por ejemplo Che et al. (2012) y Musalem et al. (2010).

$$L_{ijt}(\beta_i) = \frac{(1 - \text{OOS}_{ijt}) e^{V_{ijt}(\beta_i)}}{\sum_{k=1}^J (1 - \text{OOS}_{ikt}) e^{V_{ikt}(\beta_i)}} \quad (3.13)$$

Como se observa en la Ecuación (3.13) la probabilidad de comprar el producto j en la ocasión t es cero si se encuentra en quiebre de stock. De manera equivalente, esta probabilidad considera sólo los productos disponibles al momento de la compra.

Modelo Propuesto

En este modelo se incluirá en la función de utilidad descrita en la Ecuación (3.11) variables relacionadas con la disponibilidad del producto para capturar el efecto intertemporal y contextual del quiebre de stock.

El efecto intertemporal se modela a través de la variable OOS_{ijt-1} , lo cual indicará si es que la no disponibilidad del producto j en la compra anterior del cliente i afecta las preferencias de éste en la ocasión de compra actual, este efecto puede ser positivo o negativo (Fitzsimons, 2000).

Por otro lado, el efecto contextual está relacionado con el entorno del producto, es decir, corresponde al grado de disponibilidad de los otros productos que están dentro del conjunto de elección. El que un producto de la competencia esté en OOS puede generar cambios en la preferencias del consumidor al momento de tomar la decisión de comprar uno u otro producto. Este efecto es capturado por el coeficiente que acompaña a la variable OOS_{ijt}^C .

Además se incluyen unas interacciones entre la disponibilidad y la lealtad con el objetivo de estimar la variación en la importancia para el cliente de la variable lealtad, en el momento en que se encuentra con una situación de quiebre de stock. Esto queda reflejado en el cambio que se hace en el coeficiente β_i^L , tal como se muestra en la Ecuación (3.15).

Por lo tanto, la función de utilidad determinística queda descrita por la Ecuación (3.14).

$$v_{ijt} = \beta_i^0 + \beta_i^p \ln(\text{price}_{jt} + 1) + \beta_i^L \text{Loyalty}_{ijt} + \beta_i^{OOS_{t-1}} OOS_{ijt-1} + \beta_i^{OOS^C} OOS_{ijt}^C + \beta_i^{Re} \ln(\text{Recency}_{it} + 1) + \sum_{r=1}^R \beta^r d_t^r \quad (3.14)$$

donde

$$\beta_i^L = \beta_{0i}^L + \beta_{1i}^L OOS_{ijt-1} + \beta_{1i}^L OOS_{ijt}^C \quad (3.15)$$

Como se mencionó anteriormente se calculará $L_{ijt}(\beta_i)$ de la misma manera que en el modelo con efecto de disponibilidad.

3.2.4. Análisis de Resultados

En la Tabla 3.1 se presentan los resultados de las estimaciones para cada modelo descrito en la sección anterior. Cada modelo fue estimado utilizando el método de máxima verosimilitud simulada en el programa STATA/SE 12.0.

	(1)	(2)	(3)
	sin Efecto OOS	con Efecto OOS	Propuesto
Mean			
intercepto	-0,720 (0,640)	0,894 (0,656)	0,705 (0,664)
media_tarde	-0,0768 (0,0458)	-0,0670 (0,0463)	-0,0686 (0,0462)
tarde	0,135** (0,0480)	0,129** (0,0484)	0,126** (0,0484)
tarde_noche	0,230*** (0,0421)	0,244*** (0,0426)	0,233*** (0,0431)
noche	-0,276** (0,0859)	-0,0573 (0,0887)	-0,0842 (0,0910)
weekend_day	0,0248 (0,0282)	0,0385 (0,0288)	0,0307 (0,0293)
ln_price	-0,254** (0,0917)	-0,488*** (0,0940)	-0,465*** (0,0953)
ln_recency	0,135*** (0,0165)	0,138*** (0,0168)	0,130*** (0,0168)
Loyalty	3,618*** (0,0442)	3,684*** (0,0454)	3,781*** (0,0523)
OOS _{t-1}			0,317*** (0,0656)
OOS ^C			0,118* (0,0487)
Loyalty*OOS _{t-1}			-0,374+ (0,211)
Loyalty*OOS ^C			-0,433*** (0,127)
SD			
intercepto	0,148** (0,0493)	0,137* (0,0536)	0,108 (0,0755)
ln_price	0,000193 (0,0231)	0,000640 (0,0210)	0,00936 (0,0137)
Loyalty	0,00445 (0,0849)	0,00608 (0,0896)	0,00983 (0,0897)
ln_recency	0,185*** (0,0270)	0,191*** (0,0270)	0,185*** (0,0293)
OOS _{t-1}			0,0531 (0,480)
OOS ^C			0,114 (0,178)
Loyalty*OOS _{t-1}			0,387 (0,657)
Loyalty*OOS ^C			0,175 (0,412)
N	123.755	116.339	116.339
Log Lik.	-29.084,2	-27.503,1	-27.484,7

+ $p < 0,1$, * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$

Tabla 3.1: Coeficientes Modelos MixedLogit en la Categoría Pan

Comparando el modelo (1) y (2) se podrá entender el efecto que causa un OOS sobre el conjunto de alternativas. Por otro lado, contrastando el modelo (2) con el (3) se observarán los cambios en las preferencias del cliente causados por el OOS, más allá de la variación en el conjunto de alternativas disponibles.

Ajuste de los Modelos

En el caso de estos modelos no se puede hacer una comparación entre todos ellos, ya que no incluyen los mismos registros para estimarse. Por lo tanto sólo se comparará si tiene mejor ajuste el modelo (2) o el modelo (3). Para lo cual se utilizará el test de hipótesis descrito en la Sección 3.1.4. Se tiene que el estimador LR es igual a 36,8 y el valor crítico $\chi^2(0,05; 4)$ es igual 9,5. Así, se puede decir que el ajuste del modelo que incluye los efectos de OOS intertemporal y contextual es estadísticamente mejor al ajuste del modelo que sólo considera el OOS temporal.

Efecto de Estacionalidad

Comenzando con las variables de horario del día se tiene que hay diferencias estadísticamente significativas entre los horarios que van desde las 3pm a las 9pm (tarde y tarde-noche) y el horario de la mañana (8am-11am), con un $p - value < 0,01$ en todos los modelos analizados. Se puede apreciar que en el horario de la tarde y tarde-noche los clientes tienen una mayor preferencia por el pan que en la mañana. Por otro lado, el horario de la noche sólo es diferente al de la mañana en el modelo (1), indicando que los clientes son menos propensos a comprar alguno de estos productos al final de la jornada. En cambio, entre los días de la semana y fin de semana no se encuentran diferencias significativas, por lo que no se podría decir que los clientes compren más pan en uno que en otro día.

Efecto Elasticidad del Precio

Se observa que la elasticidad precio tiende a ser subestimada al considerar que todos los productos siempre están disponibles, es decir, cuando no se toma en cuenta el efecto temporal del OOS, tal como en el modelo (1). La elasticidad precio del modelo (2) es 1.92 veces mayor en magnitud que la estimada en el modelo (1).

Esta subestimación, tal y como lo expone Che et al. (2012), se da porque en el modelo (1) se observan las compras realizadas por los clientes considerando que pueden escoger entre cualquiera de los productos, incluso aquellos que no están disponibles. Entonces los datos pueden reflejar que los consumidores llevan aquellos productos con precios más altos, siendo que tienen alternativas más económicas. Lo que el modelo no captura es que probablemente aquellos productos que tienen un precio menor no se encuentran disponibles, por lo que realmente no son una alternativa viable, sesgando así la sensibilidad al precio que realmente tienen los clientes. Al considerar el OOS temporal se puede realizar una mejor estimación de la importancia que le entrega un consumidor al precio.

Este resultado es interesante a la hora de estimar la demanda de un producto del cual se quiere realizar una promoción. Si se hiciera con el modelo (1) se estimaría una menor cantidad de ventas de él, ya que se estimaría que los clientes son menos sensibles al precio de lo que realmente son.

Efecto del Tiempo entre Compras

La importancia del tiempo o el periodo entre compras es estadísticamente significativo al 1% y afecta de forma positiva las preferencias del consumidor. Se puede observar que el valor estimado se mantiene estable para los modelos desarrollados. Esto quiere decir, que la probabilidad de que un cliente compre pan aumenta a medida que aumenta el tiempo entre la última compra y la actual.

Efecto de la Lealtad

Para iniciar el análisis de esta variable se aprecia que en todos los modelos desarrollados es positiva, de una misma magnitud y estadísticamente significativa al 1%. El signo del coeficiente que acompaña a la lealtad señala que un consumidor tienen una mayor preferencia en su compra actual por aquel producto que llevó en la ocasión de compra anterior.

Al observar la magnitud del efecto en los distintos modelos pareciera que no hay diferencias entre considerar o no el efecto temporal del OOS sobre la lealtad, ya que el coeficiente se mantiene dentro de un rango constante. Pero si se observa en detalle el modelo (3) y más precisamente las interacciones entre la lealtad y las variables de OOS cruzado y el OOS enfrentado en la compra anterior, se tiene que hay un efecto negativo y significativo. Esto quiere decir que cuando el cliente enfrenta un OOS la lealtad es menos importante para él al momento de escoger una alternativa en esta y en la siguiente oportunidad. Esto se aprecia con mayor claridad en la Ecuación (3.16).

$$\beta_1^L = 3,781 - 0,374 \text{ OOS}_{ijt-1} - 0,433 \text{ OOS}_{ijt}^C \quad (3.16)$$

Luego, el efecto que genera un OOS sobre la lealtad es la disminución en la importancia de esta variable entre un 11,45% y un 21,34%. El valor máximo que podría tomar β_i^L es 3,781 y se da cuando las variables de interacción son cero, el mínimo es 2,974 y se tiene cuando existió un quiebre de stock en la ocasión de compra anterior y en los demás productos de la categoría. La menor variación que podría sufrir β_i^L es cuando no hay quiebre dentro de la sección panadería, tomando un valor de 3,407.

Efecto Intertemporal de la Disponibilidad

El coeficiente que acompaña a la variable OOS_{t-1} es positivo y significativo al 1%, lo cual indica que sí existe un efecto intertemporal de la disponibilidad en las preferencias del cliente.

Por lo que si el cliente se enfrentó a un OOS en la compra anterior, es más probable que en la compra actual lleve dicho pan. Esto da indicios de que existiría algún tipo de postergación o sustitución temporal en la compra de aquellos productos que el cliente no encontró.

Efecto Contextual de la Disponibilidad

Este efecto es positivo y estadísticamente significativo al 5%. Como la variable OOS^C mide la disponibilidad de los demás productos que están dentro del conjunto de elección, este resultado indica que si los demás productos de la categoría están en OOS es más probable que el cliente prefiera el producto que se está evaluando. Este efecto es por sobre el efecto directo del cambio (o disminución) en el conjunto de alternativas. Este último efecto es capturado directamente por el modelo mixed logit que a diferencia de un modelo logit permite patrones de sustitución no proporcionales a la participación de mercado de cada alternativa.

Capítulo 4

Efecto en Categoría Complementaria

La reducción en el nivel de disponibilidad no sólo puede afectar la demanda del producto que se encuentra en OOS, sino también la de aquellos productos complementarios en otras categorías. En este capítulo se mide el efecto que tiene un quiebre de stock presente en los productos de la categoría pan sobre las ventas de la sección fiambrería. Para ello se desarrolla un modelo de regresión lineal en el que se incluye el efecto de disponibilidad del pan, lo que requiere controlar por endogeneidad, ya que se utiliza el modelo de estimación de demanda para la categoría pan. Ésta proviene del modelo seleccionado en la Sección 3.2.3.

El incluir como covariable la demanda de la categoría pan en las ventas de fiambrería puede traer problemas de validez en la estimación de la regresión, ya que puede ser una variable endógena. El uso de variables instrumentales permite corregir este sesgo. Es por eso que a continuación se describe el uso de variables instrumentales. Para luego caracterizar la sección fiambrería dentro de los datos de panel, finalmente especificar el modelo y desarrollar los resultados.

4.1. Variables Instrumentales

Se tiene la siguiente regresión:

$$y = \beta X + u \tag{4.1}$$

donde y es la variable dependiente, x el vector de variables independientes, β el vector de coeficientes y u es un término de error. A través del método de estimación de mínimos cuadrados ordinarios (Ordinary Least Squares OLS por sus siglas en inglés), se tiene el estimador $\hat{\beta}$ de β . Este método asume que los regresores no están correlacionadas con los errores en el modelo, en el caso de la Ecuación (4.1) no existe asociación entre X y u . Entonces, el único efecto de X sobre y es un efecto directo a través de la expresión βX .

Sin embargo, en algunas situaciones hay asociación entre los regresores y los errores. Esto genera un efecto directo de βX sobre y , así como también un efecto indirecto a través de u en X y finalmente ésta sobre y . El estimador OLS genera un estimador $\hat{\beta}$ que combina ambos

efectos, siendo inconsistente.

La inconsistencia del estimador OLS se debe a la endogeneidad de X , lo que significa que cambios en X son asociados no sólo con cambios en y , sino también con cambios en el error u . Un método para generar sólo variaciones exógenas en X es utilizar variables instrumentales (Instrumental Variables IV por sus siglas en inglés).

Una variable z es llamada un instrumento o IV del regresor X de la Ecuación (4.1) si:

1. z no está correlacionado con el error u .
2. z está correlacionado con el regresor X .

La estimación de la regresión con IV se hace a través del método mínimos cuadrados de dos etapas (Two-Stage Least Squares 2SLS por sus siglas en inglés). El estimador 2SLS recibe este nombre porque es el resultado de realizar dos regresiones consecutivas por OLS (Stock and Watson, 2003):

1. Estimación mediante OLS de X sobre z .

$$X = \alpha z + \varepsilon \Rightarrow \hat{X} = \hat{\alpha} z$$

2. Reemplazar X por $\hat{X} = \hat{\alpha} z$ y estimar la siguiente regresión mediante OLS.

$$y = \beta \hat{X} + u \tag{4.2}$$

obteniendo $\widehat{\beta}_{2SLS}$. Donde ε y u son términos de error no correlacionados.

4.2. Reglas de Asociación

En el ámbito del marketing es interesante saber cuáles son los productos afines, es decir, aquellos que son complementarios, porque proveen información importante para diseñar variadas estrategias de marketing, como promociones y descuentos cruzados. Si bien, hay productos donde existe evidencia de su complementariedad, como el té y el azúcar, en un supermercado hay miles de SKUs por lo que determinar todas las asociaciones entre estos productos no se puede hacer a simple vista. Es así, que surgen las reglas de asociación, las cuales ayudarán a entender cuáles son los productos que se compran juntos y los que no.

Dentro de las principales reglas de asociación se encuentran las siguientes (Blattberg et al., 2008).

- **Support:** Es el porcentaje de transacciones que contienen una combinación particular de productos, lo que se puede ver como la probabilidad de llevar el producto A y B en la misma compra. ($P(AB)$). Una desventaja de esta medida de asociación se da cuando A o B son productos presentes en la mayoría de las transacciones, ya que la probabilidad de encontrarlo con cualquier otro producto es alta.

- **Confidence:** Mide la dependencia de un producto sobre otro, es decir, la probabilidad de que se lleve B si es que la canasta incluye A ($P(B | A)$). En este caso no es lo mismo $P(B | A)$ que $P(A | B)$. Cuando la $P(B | A) < P(B)$ se dice que no es una buena regla de asociación, ya que hay mayor probabilidad de encontrar B con una selección aleatoria que sólo mirando las transacciones donde se compró A.
- **Lift:** Mide la diferencia entre la confianza de una regla ($P(B | A)$) y su confianza esperada ($P(B)$). Por ejemplo, si se tiene que el lift de la regla de asociación 'si llevo helado entonces llevo cerveza' es 1.67, eso quiere decir que los clientes que compraron helado son 1.67 veces más propensos a comprar cervezas que si se escogiesen clientes al azar.

Esta regla de asociación puede resultar engañosa cuando el producto A tiene una baja confianza esperada, ya que esto podría generar un lift alto. Sin embargo, esta regla de asociación no sería interesante, debido a que pocos clientes compran A.

Los valores que puede tomar el lift se muestran a continuación.

- $Lift = 1$ significa que el evento B es independiente del evento A.
- $Lift < 1$ significa que el item A y B en conjunto son menos propensos a ser comprados juntos que si es que se hubiesen escogido al azar. Se puede decir que son productos sustitutos.
- $Lift > 1$ significa que el item A y B en conjunto son más propensos a ser comprados juntos que si es que se hubiesen escogido al azar. Se puede decir que son productos complementarios.

4.3. Selección de la Categoría Complementaria

Como se mencionó en la Sección 2.1 se tiene información detallada sobre la canasta de compra de aquellas transacciones que incluyeron algún producto de la categoría panadería, fiambrería o electrodomésticos. Es así que sólo se podría analizar la sección fiambrería o electrodoméstico para los propósitos de este capítulo.

En la categoría fiambrería hay cecinas, quesos y embutidos, tiene un total de 980 productos. Por otro lado, en la sección electrodomésticos hay productos para el hogar, de computación y de entretenimiento, con 3.071 SKUs.

Para determinar cuál de estas dos categorías analizar se utilizan las reglas de asociación. Éstas se calculan según la información que se tiene de los datos de panel y se muestran en la Tabla 4.1, donde A es la categoría pan y B puede ser fiambrería o electrodomésticos.

Como se observa en la Tabla 4.1 todas las reglas de asociación son mayores para la sección fiambrería que para la de electrodomésticos. Se tiene que en un 52,25% de las transacciones se lleva un producto de la fiambrería y que los clientes que compraron pan son 1.56 veces más propensos a comprar en la fiambrería que si se escogiese clientes al azar. Por lo tanto, se decide trabajar con esta categoría considerándola complementaria a la panadería.

Sección (B)	Confidence Esperado P(B) (%)	Confidence P(B A) (%)	Lift P(B A)/P(B)
Fiambrería	52,25	81,33	1,56
Electrodomésticos	4,11	4,68	1,14

Tabla 4.1: Reglas de asociación panadería, fiambrería y electrodomésticos

4.4. Modelo

4.4.1. Variables

Para capturar el efecto que tienen las compras de pan y sus respectivos niveles de disponibilidad en las compras del departamento fiambrería se definen las siguientes variables.

- *Variable dependiente*

$$y_{it} = \begin{cases} 1 & \text{si el cliente } i \text{ lleva un producto de la fiambrería en la ocasión de compra } t \\ 0 & \sim \end{cases}$$

- *Lealtad fiambrería*

$$\text{Loyalty}_{iFt} = \alpha \text{Loyalty}_{iFt-1} + (1 - \alpha) z_{iFt-1} \quad (4.3)$$

donde

$$z_{iFt} = \begin{cases} 1 & \text{si el cliente } i \text{ lleva un producto de fiambrería en la ocasión de compra } t \\ 0 & \sim \end{cases}$$

La inicialización de la variable Loyalty_{iFt} se realiza de la siguiente forma (Guadagni and Little, 1983).

$$\text{Loyalty}_{iFt} = \begin{cases} \alpha & \text{si el cliente } i \text{ lleva en su primera compra un producto de} \\ & \text{fiambrería} \\ \frac{1-\alpha}{\text{card}(C)-1} & \sim \end{cases}$$

En este caso la constante $\alpha = 0,7$ y $\text{card}(C) = 2$, ya que el conjunto de elección es comprar o no comprar un producto en la sección fiambrería.

- *Monto de la compra*

Se construye esta variable con el fin de capturar los distintos tipos de compras que un cliente puede hacer en el supermercado. Éstas pueden ser compras ocasionales o compras del mes. Se tiene que monto_compra_{it} es el monto que canceló el cliente i en su ocasión de compra t .

- *Compra en la categoría panadería*

Es una variable binaria que indica si el cliente compró algún producto de la categoría panadería.

$$\text{compra_pan}_{it} = \begin{cases} 1 & \text{si el cliente } i \text{ lleva un producto de panadería en la ocasión de compra } t \\ 0 & \sim \end{cases}$$

4.4.2. Definición de Modelos

Se realizarán distintos modelos de regresión incluyendo las variables mencionadas anteriormente, donde el modelo general será el que se muestra en la Ecuación (4.4).

$$\begin{aligned} \text{Fiambre}_{it} = & \mu_i^0 + \mu^{\text{pan}} \text{compra_pan}_{it} + \mu^L \text{Loyalty}_{iFt} + \mu^R \ln(\text{Recency}_{it} + 1) \\ & + \mu^M \ln(\text{monto_compra}_{it} + 1) + \sum_{r=1}^R \mu^r d_t^r + \xi_{it} \end{aligned} \quad (4.4)$$

donde ξ_{it} es el término de error.

En esta regresión se tiene que la variable que representa la demanda por pan del cliente i en la ocasión de compra t , compra_pan_{it} , puede estar correlacionada con el término de error, si este último incorpora por ejemplo un shock en la demanda. Si aumenta la demanda del supermercado se generaría un aumento en las compras de las distintas categorías, como el pan y la fiambrería y al estar incluyendo la variable compra_pan_{it} en la regresión se podría deducir incorrectamente que el efecto lo está causando la compra en panadería y no el shock en la demanda. Por esto, se instrumentaliza la variable de demanda de pan estimándola con el modelo (3) de la Sección 3.2.3, donde se tiene la función de utilidad determinística que se muestra a continuación.

$$\begin{aligned} v_{ijt} = & \beta_i^0 + \beta_1^p \ln(\text{price}_{jt} + 1) + \beta_1^L \text{Loyalty}_{ijt} + \beta_i^{\text{OOS}_1} \text{OOS}_{ijt-1} \\ & + \beta_i^{\text{OOS}_C} \text{OOS}_{ijt}^C + \beta_i^R \ln(\text{Recency}_{it} + 1) + \sum_{r=1}^R \beta^r d_t^r \end{aligned}$$

Como en el modelo (3) se calcula la probabilidad de compra del cliente i en la ocasión de compra t para cada una de las alternativas j del conjunto de elección (\widehat{p}_{ijt}), la probabilidad de compra de pan estimada $\widehat{\text{compra_pan}}_{it}$ se construye como se muestra en la Ecuación (4.5).

$$\widehat{\text{compra_pan}}_{it} = \sum_{j=1}^{J \setminus \{NC\}} \widehat{p}_{ijt} \quad (4.5)$$

Este instrumento cumple con las siguientes propiedades:

- $\widehat{\text{compra_pan}}_{it}$ es relevante, es decir $\text{corr}(\widehat{\text{compra_pan}}_{it}, \text{compra_pan}_{it}) \neq 0$, ya que se construye de manera de representar la incidencia de pan. En efecto, la correlación lineal es 0.20 ($p - \text{value} < 0,0001$).
- $\widehat{\text{compra_pan}}_{it}$ es exógeno, es decir $\text{corr}(\widehat{\text{compra_pan}}_{it}, \xi_{it}) = 0$, ya que no está correlacionada con factores que podrían afectar la compra de fiambre y que no están siendo capturados por el modelo. Estos factores incluyen cambios en el precio de productos de fiambrería y shock de demanda en visitas al supermercado. La probabilidad de compra de pan no debería afectar la fijación de precios de los productos de la sección fiambrería. Mientras que el shock de demanda que afecta a fiambrería y panadería (que está en los términos de error), no está correlacionado con el instrumento por construcción del mismo. Adicionalmente, cualquier patrón estacional en la demanda queda capturado por las variables de estacionalidad incluidas en los modelos.

Como el instrumento es relevante y exógeno se puede decir que es válido.

Luego de esto, el modelo general queda como se muestra en la Ecuación (4.6). Este modelo se denominará Modelo (d).

$$\begin{aligned} \text{Fiambre}_{it} = & \mu_i^0 + \mu^{\text{pan}} \widehat{\text{compra_pan}}_{it} + \mu^L \text{Loyalty}_{iFt} + \mu^R \ln(\text{Recency}_{it} + 1) \\ & + \mu^M \ln(\text{monto_compra}_{it} + 1) + \sum_{r=1}^R \mu^r d_t^r + \xi_{it} \end{aligned} \quad (4.6)$$

Del Modelo (d) se generan otros tres modelos anidados imponiendo restricciones en los distintos parámetros de μ , las que se detallan a continuación.

- **Modelo (a)** $\mu^L = \mu^R = \mu^M = \mu^{\text{Pan}} = 0$
- **Modelo (b)** $\mu^{\text{Pan}} = 0$
- **Modelo (c)** $\mu^L = \mu^R = \mu^M = 0$

4.4.3. Análisis de los Resultados

En la Tabla 4.2 se presentan los resultados de las estimaciones para cada modelo de regresión descrito en la sección anterior con *random effects*, mediante el método de máxima verosimilitud simulada en el programa STATA/SE 12.0.

	(a)	(b)	(c)	(d)
	incidencia fiambre	incidencia fiambre	incidencia fiambre	incidencia fiambre
_cons	0,338*** (0,0106)	-1,140*** (0,0318)	0,302*** (0,0172)	-1,185*** (0,0313)
media_tarde	-0,0254* (0,0111)	-0,0477*** (0,0116)	-0,0240* (0,0121)	-0,0461*** (0,0122)
tarde	-0,0176 (0,0142)	-0,00953 (0,0127)	-0,0198 (0,0126)	-0,0128 (0,0119)
tarde_noche	0,0393*** (0,0101)	0,0342** (0,0110)	0,0354** (0,0124)	0,0282** (0,0105)
noche	0,0329 (0,0183)	0,0105 (0,0182)	0,0353 (0,0198)	0,0126 (0,0200)
weekend_day	0,0393*** (0,00816)	-0,00779 (0,00566)	0,0395*** (0,00666)	-0,00822 (0,00584)
loyalty _F		0,0127 (0,0196)		-0,000117 (0,0211)
ln_recency		0,0249*** (0,00388)		0,0215*** (0,00272)
ln_monto_compra		0,157*** (0,00301)		0,157*** (0,00301)
$\widehat{\text{compra}}_{\text{pan}}$			0,0795* (0,0317)	0,109*** (0,0287)
N	24.751	24.751	24.751	24.751
Log Lik.	-16.329,7	-14.227,6	-16.326,6	-14.221,2

Standard errors in parentheses

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

Tabla 4.2: Coeficientes Modelos Regresión RE con IV Fiambrería

Selección del Modelo

En el caso de estos modelos se puede hacer una comparación entre todos ellos, ya que incluyen los mismos registros para estimarse. Para saber cuál es el modelo que ajusta mejor a los datos se utiliza el test de hipótesis de la Sección 3.1.4 y se selecciona aquél cuyo ajuste es significativamente mayor.

Los modelos con mayores verosimilitudes son el (b) y el (d), con -14.227,6 y -14.221,2 respectivamente. Por lo tanto, el estimador LR es igual a 12,8 y el valor crítico $\chi^2(0,05; 1)$ es igual a 3,8. Así, se puede concluir que el modelo que incluye la variable demanda de pan y por consecuencia los efectos del OOS es el seleccionado.

Los efectos que se analicen en las secciones siguientes corresponderán al modelo (d).

Efecto de las Ventas de la Sección Panadería

En el modelo (d) se puede observar que el coeficiente relacionado con la variable instrumental demanda de la sección panadería es positivo y significativo ($p - \text{valor} < 0,01$). Esto

quiere decir que un aumento en las ventas de pan aumenta las compras en la categoría fiambrería. Un aumento en las ventas de la panadería en un 1 % significa un aumento de ventas de 0,1 % en fiambrería.

En los resultados obtenidos del modelo (3) de la Sección 3.2.4, con el cual se estimó la variable instrumental, se puede observar que una disminución en el nivel de disponibilidad de los productos de la panadería disminuye la compra en esta categoría. Así, los quiebres de stock en la panadería no sólo afectan las ventas en la categoría a fin, sino también disminuyen las ventas de la sección fiambrería.

Efecto del Monto de Compra

El coeficiente que acompaña al monto de compra es positivo, significativo al 0,1 % y constante en los modelos en que se incluye. Este resultado indica que un mayor monto de compra significa una mayor probabilidad de comprar en la categoría fiambrería, lo cual está relacionado con las ocasiones de compra. Si un cliente realiza la compra del mes, llevará productos de variadas categorías de forma más o menos independiente entre ellas. En cambio, si un cliente realiza una compra ocasional, el monto será menor y existirá una menor posibilidad de llevar un producto fuera de la categoría.

Efecto de Estacionalidad

Se puede observar que existen diferencias estadísticamente significativas entre la media-tarde y tarde-noche respecto al horario de la mañana, con un $p - value < 0,01$. Esto indica que en la media-tarde se tiende a comprar menos en la categoría fiambrería que en la mañana, en cambio en la tarde noche existe una mayor probabilidad de observar compras que incluyan un producto de esta categoría.

Entre los días de la semana y fin de semana no se encuentran diferencias significativas, por lo que no se podría decir que los clientes compran más productos de la sección fiambrería en uno que en otro tipo de día.

Efecto del Tiempo entre Compras

Al igual que en los resultados para los efectos del OOS dentro de la categoría pan, el tiempo que pasa entre una compra y otra es estadísticamente significativo al 0,1 % y afecta de manera positiva las preferencias del consumidor. Además, su coeficiente se mantiene estable en los modelos que incluyen esta variable. Esto quiere decir, que la posibilidad de que un cliente compre algún producto de la fiambrería aumenta a medida que lo hace el tiempo que pasó entre la última compra y la actual.

Capítulo 5

Efecto Económico

En este capítulo se realizará un cálculo económico de la implicancia de los diferentes efectos estimados en la Sección 3.2.4 y la Sección 4.4.3. Estos efectos están relacionados con el cambio en el nivel de disponibilidad de los productos en la categoría pan, tanto dentro de la misma categoría como en la categoría complementaria fiambrería.

5.1. Efectos Económicos Dentro de la Categoría

Como se explicó en la Sección 3.2.4 el cambio en el nivel de disponibilidad de los productos en la categoría pan tiene efectos directos sobre las ventas de esta categoría. Como los productos con mayor participación dentro de la sección panadería son la marraqueta y la hallulla, se decide cuantificar los efectos económicos generados por un quiebre de stock en ellos a través de análisis de distintos escenarios.

- **Escenario base (Base):** Para construir este escenario se utilizan los coeficientes del modelo (3) mostrados en la Tabla ?? y el valor promedio de las variables correspondientes a estos coeficientes (ver Apéndice A), excepto para las variables relacionadas al quiebre de stock. Estas últimas variables son consideradas con un cero, que indica que el cliente no se está enfrentando a un OOS de los demás productos de la categoría (OOS_{ijt}^C) ni tampoco lo hizo en su ocasión de compra anterior (OOS_{ijt-1}), esto también implica que las variables de interacción son cero ($Loyalty * OOS_{ijt-1}$ y $Loyalty * OOS_{ijt}^C$).
- **Escenario marraqueta o hallulla en quiebre de stock (OOS Hallulla, OOS Marraqueta):** Al igual que en el caso anterior se consideran los mismos valores de los coeficientes, en este caso se modifica el valor de variable que representa la disponibilidad de la opción marraqueta o de la hallulla. En vez de cero ahora será un uno, que indica que se encuentra en OOS. Aquí se tienen tres escenarios, uno para la marraqueta, otro para la hallulla y uno donde ambos productos están en quiebre de stock.
- **Escenario marraqueta o hallulla en quiebre de stock en la compra anterior (OOS_{t-1} Hallulla, OOS_{t-1} Marraqueta):** Este escenario se utiliza para determinar el efecto que tiene un quiebre de stock en la siguiente ocasión de compra, de manera

de cuantificar el efecto intertemporal de un OOS. Al escenario base se le modifica las variables que representan el nivel de disponibilidad que enfrentó el cliente en la ocasión de compra anterior, en vez de cero ahora será uno (OOS_{ijt-1}). Esto produce además un cambio en la variable lealtad, ya que al no encontrar el producto en la compra anterior, no pudo haberlo comprado por lo que su lealtad a dicho producto disminuye según el factor $\alpha = 0,7$. Aquí también se tienen tres escenarios, uno para la marraqueta, otro para la hallulla y uno donde ambos productos están en quiebre de stock.

En la Tabla 5.1 se pueden observar las participaciones de mercado para cada alternativa de la categoría pan según los distintos escenarios.

	Escenarios						
		OOS	OOS _{t-1}	OOS	OOS _{t-1}	OOS	OOS _{t-1}
	Base	Hallulla	Hallulla	Marraqueta	Marraqueta	Ambos	Ambos
No Compra	56 %	61 %	54 %	62 %	55 %	72 %	53 %
Hallulla	11 %	0 %	13 %	13 %	11 %	0 %	13 %
Marraqueta	12 %	14 %	12 %	0 %	13 %	0 %	13 %
Granel	12 %	14 %	12 %	14 %	12 %	16 %	12 %
Envasado	9 %	11 %	9 %	11 %	9 %	12 %	9 %

Tabla 5.1: Incidencia para distintos escenarios en la categoría pan

En la tabla anterior se puede apreciar que si se tiene un quiebre de stock en:

1. Hallulla: la incidencia del pan disminuye en un 5 %, dado que la no compra pasa de un 56 % a un 61 %. Además, el cambio de participación de mercado de las alternativas que no se encuentran en OOS suma un 6 % (sustitución). También se aprecia que las ventas de la hallulla aumentan en un 3 % cuando estuvo en quiebre de stock en la ocasión de compra anterior por efecto de sustitución temporal. Por lo tanto, el efecto neto de un quiebre de stock en la hallulla es una disminución del 2 % en las ventas de este producto. Estos resultados se pueden ver en la Figura 5.1.
2. Marraqueta: la incidencia del pan disminuye en un 6 %, dado que la no compra pasa de un 56 % a un 62 %. Además, existe un cambio de participación de mercado de las alternativas que no se encuentran en OOS, lo que genera sustitución correspondiente al 6 %. Al igual que en el caso anterior, las ventas de la marraqueta aumentan en un 3 % si estuvo en quiebre de stock en la ocasión de compra anterior. Entonces, el efecto neto de un quiebre de stock en la marraqueta es una disminución del 3 % en las ventas de este producto. Para mayor claridad se puede observar la Figura 5.2.
3. Hallulla y Marraqueta: la no compra en la categoría pan aumenta de un 56 % a un 72 %, lo que genera una disminución del 16 % de la incidencia del pan. En este caso la sustitución llega al 7 % y tal como en los dos casos anteriores la venta de hallulla y marraqueta aumentan en la ocasión de compra actual cuando estuvieron en quiebre de stock en la ocasión de compra anterior. Este aumento es de un 2 % y un 1 %, respectivamente. Por lo tanto, el efecto neto de un quiebre de stock en la hallulla y la marraqueta a la vez es de una disminución del 13 % en las ventas de estos productos. Esto queda graficado en la Figura 5.3.

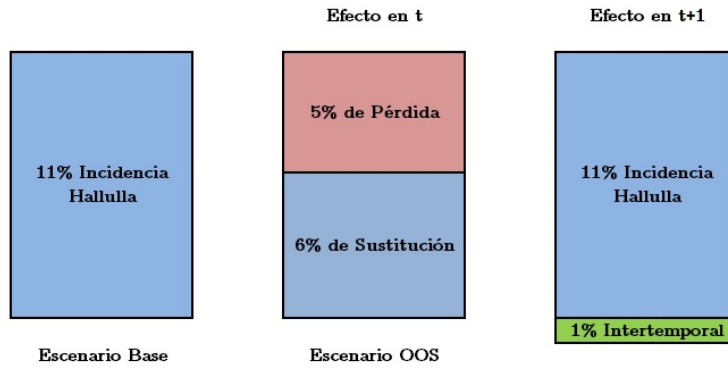


Figura 5.1: Efecto de un quiebre de stock en la incidencia de la hallulla

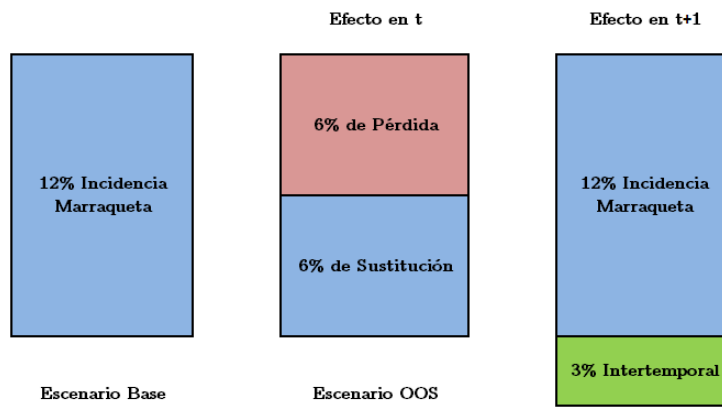


Figura 5.2: Efecto de un quiebre de stock en la incidencia de la marraqueta

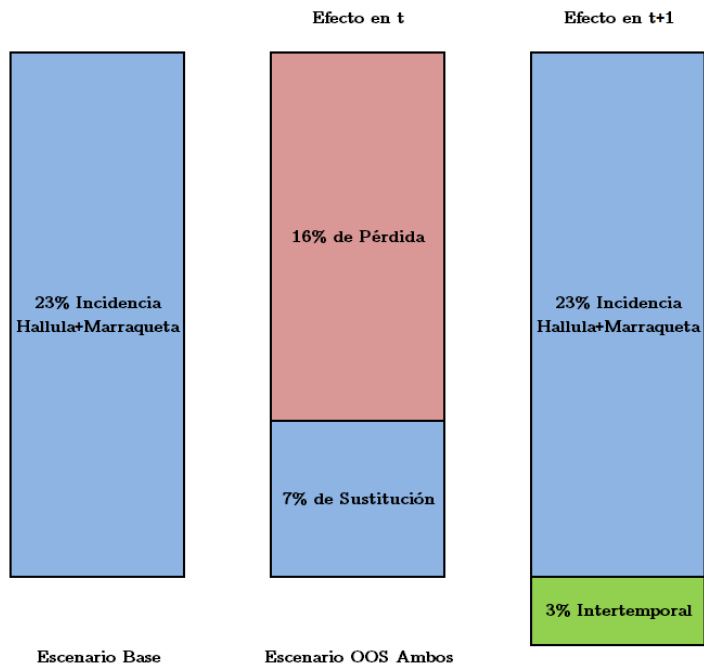


Figura 5.3: Efecto de un quiebre de stock en la incidencia de la hallulla y la marraqueta

Para estimar las pérdidas económicas de la disminución en las participaciones de mercado cuando la hallulla o la marraqueta están en quiebre de stock se considerará lo siguiente:

- Todos los clientes del supermercado se comportan como los clientes que se tienen en los datos de panel.
- Todos los locales de supermercados tienen la misma cantidad de boletas promedio mensuales.
- La proporción de ventas en donde un cliente se enfrenta a un quiebre de stock de un producto determinado durante un mes, se mantiene constante para todos los locales y meses.

Con estos supuestos se procede a realizar el cálculo en dinero de las ventas perdidas de manera mensual (\$ pérdida panadería) de la cadena que tiene 46 locales (n locales) y un promedio de 342.800 boletas mensual (boletas promedio), tal y como se muestra en la Ecuación (5.1).

$$\begin{aligned} \$ \text{pérdida_panadería}_j = & \text{incidencia_mensual}_j * \text{pérdida_neta}_j * \text{eventos_OOS}_j \\ & * \text{monto_promedio}_j * \text{boletas_promedio} * n_locales \end{aligned} \quad (5.1)$$

donde $j \in \{\text{Hallulla, Marraqueta, Ambos}\}$ y

$$\text{incidencia_mensual}_j = \begin{cases} 23\% & \text{si } j \text{ corresponde a Hallulla} \\ 27\% & \text{si } j \text{ corresponde a Marraqueta} \end{cases}$$

$$\text{pérdida_neta}_j = \begin{cases} 3\% & \text{si } j \text{ corresponde a Hallulla} \\ 5\% & \text{si } j \text{ corresponde a Marraqueta} \\ 13\% & \text{si } j \text{ corresponde a ambos productos} \end{cases}$$

$$\text{eventos_OOS}_j = \begin{cases} 7\% & \text{si } j \text{ corresponde a Hallulla} \\ 15\% & \text{si } j \text{ corresponde a Marraqueta} \\ 22\% & \text{si } j \text{ corresponde a ambos productos} \end{cases}$$

$$\text{monto_promedio}_j = \begin{cases} \$ 570 & \text{si } j \text{ corresponde a Hallulla} \\ \$ 662 & \text{si } j \text{ corresponde a Marraqueta} \\ \$ 616 & \text{si } j \text{ corresponde a ambos productos} \end{cases}$$

Los cálculo para cada uno de los 3 escenarios que se presentaron se pueden observar en la Tabla 5.2

	Hallulla	Marrqueta	Ambos
Venta sin OOS (US \$)*	3.546.092	4.834.695	8.331.018
Pérdida con OOS (US \$)*	7.840	36.966	238.267
Pérdida (%)	0,2 %	0,8 %	2,9 %

*Tipo de cambio al 21-04-2014

Tabla 5.2: Pérdida económica mensual para distintos escenarios de quiebre de stock

La mayor pérdida se produce cuando la hallulla y la marraqueta se encuentran en quiebre de stock con US\$ 238.267 mensuales menos para la recaudación de la cadena de supermercado, lo que equivale a un 2,9 % de las ventas si es que estos productos siempre se encontraran disponibles. Es importante observar que las pérdidas que se producen al estar simultáneamente la hallulla y la marraqueta en OOS es mayor (2,9 %) a la suma de las pérdidas individuales asociadas al quiebre de stock de estos productos en instantes distintos ($1,0\% = 0,2\% + 0,8\%$), debido a la alta sustitución entre estos panes. Por otro lado, la menor pérdida se da cuando la hallulla está en quiebre de stock, donde un 0,2 % de las ventas mensuales de este producto no se realizaría.

5.2. Categoría Complementaria

En la Sección 4.4.3 se concluyó que el cambio en el nivel de disponibilidad de los productos en la categoría pan tiene efectos indirectos sobre las ventas de la categoría complementaria fiambrería. Utilizando los resultados de los escenarios descritos en la sección anterior para los distintos niveles de disponibilidad de la marraqueta y la hallulla, se analizará los efectos económicos generados en la categoría fiambrería.

Para determinar las pérdidas generadas por un quiebre de stock en el pan en la fiambrería se construirá el escenario base con los coeficientes del modelo (f) mostrados en la Tabla 4.2 y el promedio de las variables que acompañan correspondientes a estos coeficientes (ver Apéndice B), excepto para las variable $\widehat{\text{compra}}_{\text{pan}}$. En este escenario base esta variable corresponde a la suma de la participación de mercado de las opciones hallulla, marraqueta, granel y envasado en el escenario base de la sección anterior. Luego, se sigue con los demás escenarios restándoles la pérdida neta calculada anteriormente según corresponda.

En la Tabla 5.3 se pueden ver las pérdidas para la categoría fiambrería según los resultados obtenidos en la sección anterior para las participaciones de mercado cuando se produce un quiebre de stock.

En la tabla anterior se puede apreciar que si se tiene un quiebre de stock en:

1. Hallulla: la incidencia de la fiambrería disminuye en un 0,3 %, dado que la incidencia pasa de un 37,0 % a un 36,7 %. Aquí se utiliza $\widehat{\text{compra}}_{\text{pan}} = 0,412$, que es la incidencia del pan cuando la hallulla está en quiebre.
2. Marraqueta: la incidencia de la fiambrería disminuye en un 0,5 %, es decir la no compra

	Base	OOS Hallulla	OOS Marraqueta	OOS Ambos
Estimación incidencia	37,0 %	36,7 %	36,5 %	35,6 %
Baja en la incidencia	-	0,3 %	0,5 %	1,4 %

Tabla 5.3: Incidencia de la categoría fiambrería para distintos escenarios de quiebre de en la categoría pan

aumenta de un 63,0 % a un 63,5 %. Aquí se utiliza $\widehat{\text{compra_pan}} = 0,392$, que corresponde a la incidencia del pan cuando la marraqueta está en quiebre.

- Hallulla y Marraqueta: la no compra en la categoría fiambrería pasa de un 63,0 % a un 64,4 %, dado que hay una disminución del 1,4 % de la incidencia de la fiambrería. Aquí se utiliza $\widehat{\text{compra_pan}} = 0,312$, que corresponde a la incidencia del pan cuando la hallulla y la marraqueta están en quiebre.

Para el cálculo económico se utilizan los mismos supuestos detallados en la Sección 5.1 y la Ecuación (5.2).

$$\begin{aligned}
 \$ \text{pérdida_fiambrería}_j &= \text{incidencia_mensual_fiambrería} * \text{pérdida_neta_fiambrería}_j \\
 &\quad * \text{eventos_OOS}_j * \text{monto_promedio_fiambrería} \\
 &\quad * \text{boletas_promedio} * n_locales
 \end{aligned} \tag{5.2}$$

donde $j \in \{\text{Hallulla, Marraqueta, Ambos}\}$, la incidencia mensual de la fiambrería es del 53 % y el monto promedio fiambrería es de \$1.687 en los datos de panel. El número de boletas y de locales son los mismos de la sección anterior (342.800 y 46, respectivamente).

$$\text{pérdida_neta_fiambrería}_j = \begin{cases} 0,3\% & \text{si } j \text{ corresponde a hallulla} \\ 0,5\% & \text{si } j \text{ corresponde a marraqueta} \\ 1,4\% & \text{si } j \text{ corresponde a ambos productos} \end{cases}$$

$$\text{eventos_OOS}_j = \begin{cases} 7\% & \text{si } j \text{ corresponde a hallulla} \\ 15\% & \text{si } j \text{ corresponde a marraqueta} \\ 22\% & \text{si } j \text{ corresponde a ambos productos} \end{cases}$$

Como se puede observar en la Tabla 5.4 la mayor pérdida en la categoría fiambrería se produce cuando la hallulla y la marraqueta se encuentran en quiebre de stock, disminuyendo sus ventas mensuales en un 0,31 %, lo cual equivale a US \$75.213. La menor pérdida se produce cuando sólo la hallulla está en quiebre de stock disminuyendo sus ventas mensuales en un 0,02 %.

	OOS	OOS	OOS
	Hallulla	Marraqueta	Ambos
Venta sin OOS (US \$)*	24.184.567	24.184.567	24.184.567
Pérdida con OOS (US \$)*	5.814	9.690	75.213
Pérdida (%)	0,02 %	0,04 %	0,31 %

*Tipo de cambio al 21-04-2014

Tabla 5.4: Pérdida económica mensual en la categoría fiambrería para distintos escenarios de quiebre de stock en la panadería

Capítulo 6

Conclusiones

6.1. Conclusiones

La determinación de los niveles adecuados de servicio en góndola requiere equilibrar los costos de inventario y reposición con el beneficio de tener una mayor disponibilidad del producto para los clientes. Por lo general los costos de inventario son medibles, en cambio los relacionados con los quiebres de stock son menos visibles y difíciles de medir, ya que se necesita determinar cuándo ocurrió el OOS, qué clientes estuvieron expuestos a él y cuál fue su comportamiento frente a éste.

Se han realizado diversos estudios que analizan el comportamiento de los clientes al enfrentarse a una disminución en el nivel de disponibilidad de un producto. Para detectar los quiebres de stock y los clientes que estuvieron expuestos a él, los trabajos siguen principalmente dos líneas. La primera mide físicamente la góndola en un momento determinado del día, asumiendo que los clientes que pasaron por caja después de esta medición se enfrentaron al OOS. Esto puede conducir a errores si es que existe reposición posterior al quiebre. La segunda lo hace a través de los sistemas de control de inventario, lo cual es impreciso ya que si bien el producto está en stock, puede estar en la bodega y no necesariamente visible para el cliente (en góndola). Además, es difícil que estos sistemas estén actualizados minuto a minuto, ya que los productos que pasan por caja no son los únicos que salen del supermercado, están los que se pierden por mermas de distinto tipo. En este estudio se utilizan cámaras que apuntan hacia las góndolas, midiendo cada 30 minutos la disponibilidad de los productos de la sección panadería, por lo que el error al identificar qué clientes se enfrentaron al quiebre de stock es menor.

Teniendo la información del nivel de stock en las góndolas de la panadería y las transacciones realizadas en el local de supermercado se estiman los efectos de un quiebre de stock en el comportamiento de los clientes dentro de la panadería y en una categoría complementaria.

El principal resultado obtenido es que los quiebres de stock en una categoría no sólo disminuyen las ventas de ésta, sino también las de las categorías complementarias. Para la sección panadería la mayor pérdida se produce cuando la hallulla y la marraqueta se

encuentran en quiebre de stock con US\$ 238.267 mensuales menos para la recaudación de la cadena de supermercado, lo que equivale a un 2,9 % de las ventas si es que estos productos siempre se encontraran disponibles. En el caso de la fiambrería (categoría complementaria) la mayor pérdida también se produce al tener la hallulla y la marraqueta en OOS, disminuyendo las ventas mensuales de la categoría en un 0,31 % equivalente a US \$75.213.

Al estimar los modelos mixed logit incorporando la disponibilidad del producto se encontraron resultados interesantes respecto a la elasticidad precio, la lealtad y los quiebres de stock intertemporal y contextual. Se observa que la elasticidad precio se subestima cuando no se consideran los OOS temporales; la lealtad disminuye su importancia para el cliente entre un 11,45 % y un 21,34 % cuando se tiene un quiebre de stock en la ocasión de compra anterior (OOS intertemporal) o en los demás productos de la categoría (OOS contextual); el OOS intertemporal aumenta la compra del producto en la próxima compra, por lo tanto existe un retraso en la compra del producto que se encuentra en quiebre de stock; el OOS contextual indica que la decisión del cliente se ve afectada por el nivel de disponibilidad de los demás productos de la categoría. El hecho de que exista sustitución temporal (retraso) hace que aumente la demanda en el siguiente periodo, lo que puede llevar a un mayor nivel de quiebre.

Finalmente, el conocer los efectos económicos que genera una baja en el nivel de disponibilidad de los productos le permite al retail determinar las estrategias de reposición que deben tomar, de manera que éstas no tengan un costo mayor al costo de tener quiebres de stock.

6.2. Limitaciones

El presente trabajo presenta limitaciones principalmente desde el punto de vista de la información disponible. Entre éstas, las visitas de los clientes en el supermercado tienen una mayor frecuencia que las mediciones de nivel de disponibilidad de los productos. Si bien tener mediciones de stock cada 30 minutos es un importante avance respecto a la literatura en la forma de identificar tanto la ocurrencia como los clientes que se enfrentaron al OOS, los errores en esto último aún pueden ocurrir si se considera que el pan es un producto que se repone varias veces al día.

Otra limitación tiene relación con la identificación de los clientes dado que puede que éstos no siempre den su ID al momento de pagar. Por lo que en los datos de panel se podría no estar considerando realmente todas las compras que los clientes realizaron.

Una tercera limitación del trabajo está dada por los supuestos que se hicieron al cuantificar los efectos económicos para la cadena de supermercados, porque en ellos se dice que los clientes que no están en el panel se comportarían de la misma manera que los que sí están. Además, puede que los clientes que dan su ID tengan un comportamiento distinto al de aquellos que no lo dan.

Por último, como sólo se tiene el detalle de la canasta para aquellas compras que incluyen

algún producto de la categoría pan, fiambre o electrodomésticos no se pudo analizar otra categoría complementaria a la panadería. Lo que podría generar variaciones en las pérdidas económicas para el supermercado.

6.3. Recomendaciones y Trabajos Futuros

Dentro de las recomendaciones que se pueden dar a la administración de los supermercados, dado los resultados de este trabajo, está el hecho de priorizar el control de disponibilidad en aquellos productos cuyos quiebres de stock generan mayores pérdidas económicas. Tomando en cuenta que no sólo se producen pérdidas en la venta del producto en OOS, sino también en aquellos relacionados a este producto.

Para tener más precisión en la determinación de cuáles fueron los clientes que se enfrentaron a un quiebre de stock se sugiere tener mediciones de disponibilidad de los productos con más frecuencia y algún sistema que permita saber cuáles son los pasillos en los que el cliente estuvo durante la visita. Por ejemplo, se podría implementar tecnología de reconocimiento de imagen, de modo que cada 5 minutos se tuviera una medición y utilizar tecnología de RFID en los carros y canastos de compra de manera de tener la localización del cliente en el local. Si bien, esto ayuda a disminuir el error de inferencia de los clientes que estuvieron expuestos a un OOS, es costoso tanto la implementación como el procesamiento la información que los sistemas recogen.

Con el fin de complementar lo hecho en este trabajo, se pueden realizar mediciones de disponibilidad en categorías que tengan menor cantidad de reposición en el día, ya que en ellas se podría encontrar una mayor proporción de quiebres, por consiguiente una mayor cantidad de clientes expuestos a éstos. Por lo que el impacto económico podría ser aún mayor para la cadena de supermercados.

Como en este trabajo se estimaron los efectos en la misma compra y en la siguiente, se propone como trabajo futuro investigar cuáles son los efectos a largo plazo que podría generar un quiebre de stock en las preferencias de los consumidores, lo que generaría un cambio en el comportamiento de éste. Para esto se podrían utilizar modelos que consideren cambios estructurales, como cadenas de Markov ocultas (Hidden Markov Models) incorporando la disponibilidad de un producto como una covariable en la la matriz de transición entre estados latentes (Netzer et al., 2008; Montoya et al., 2010).

Bibliografía

- Anderson, E. T., Fitzsimons, G. J., and Simester, D. (2006). Measuring and mitigating the costs of stockouts. *Management Science*, 52(11):1751–1763.
- Blattberg, R., Kim, B., and Neslin, S. (2008). *Database Marketing Analysing and Managing Customers*. International Series in Quantitative Marketing. Springer, New York.
- Bosch, M., Hilger, R., and Schilkrut, A. (2006). Medición de faltante en góndola. *Working Paper, Universidad de Chile*, pages 1–11.
- Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: methods and applications*. Cambridge university press.
- Campo, K., Gijsbrechts, E., and Nisol, P. (2003). The impact of retailer stockouts on whether, how much, and what to buy. *International Journal of Research in Marketing*, 20(3):273–286.
- Che, H., Chen, X., and Chen, Y. (2012). Investigating effects of out-of-stock on consumer stockkeeping unit choice. *Journal of Marketing Research*, 49(4):502–513.
- Fisher, M. and Raman, A. (2010). *The new science of retailing: how analytics are transforming the supply chain and improving performance*. Harvard Business Review Press.
- Fitzsimons, G. J. (2000). Consumer response to stockouts. *Journal of Consumer Research*, 27(2):249–266.
- Gruen, T. W., Corsten, D. S., and Bharadwaj, S. (2002). *Retail out-of-stocks: A worldwide examination of extent, causes and consumer responses*. Grocery Manufacturers of America Washington, DC.
- Guadagni, P. M. and Little, J. D. (1983). A logit model of brand choice calibrated on scanner data. *Marketing science*, 2(3):203–238.
- Montoya, R., Netzer, O., and Jedidi, K. (2010). Dynamic allocation of pharmaceutical detailing and sampling for long-term profitability. *Marketing Science*, 29(5):909–924.
- Musalem, A., Olivares, M., Bradlow, E. T., Terwiesch, C., and Corsten, D. (2010). Structural

estimation of the effect of out-of-stocks. *Management Science*, 56(7):1180–1197.

Netzer, O., Lattin, J. M., and Srinivasan, V. (2008). A hidden markov model of customer relationship dynamics. *Marketing Science*, 27(2):185–204.

Ross, S. (2006). *Simulation*. Academic press.

Stock, J. H. and Watson, M. W. (2003). *Introduction to econometrics*, volume 104. Addison Wesley Boston.

Train, K. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Apéndice A

Promedio de las Variables en Panadería

	No Compra	Hallulla	Marraqueta	Granel	Envasado
ln_price	0,000	6,832	6,874	7,021	7,114
Loyalty	0,000	0,118	0,140	0,164	0,091
OOS_{t-1}	0,000	0,061	0,127	0,177	0,000
OOS^C	0,000	0,153	0,074	0,194	0,194
ln_recency	0,000	1,208	1,208	1,208	1,208
Loyalty*OOS_{t-1}	0,000	0,007	0,018	0,029	0,000
Loyalty*OOS^C	0,000	0,018	0,010	0,032	0,018

Tabla A.1: Promedio de las variables en los datos de panel para la categoría panadería

Apéndice B

Promedio de las Variables en Fiambrería

	$\widehat{\text{compra_pan}}$	loyalty_F	ln_recency	ln_monto_compra
Fiambre	0,474	0,364	1,208	9,280

Tabla B.1: Promedio de las variables en los datos de panel para la categoría fiambrería