

**Endogeneidad en modelos de precios: una aplicación desde la empresa**

Almacenes Éxito S.A.

**Juan Cristóbal López Orozco**

Trabajo de Grado

Asesorado por:

Jesús Alonso Botero

Juan Camilo Chaparro



Escuela de Economía y Finanzas

Departamento de Economía

Universidad EAFIT

Colombia

2020

# CONTENIDO

1. INTRODUCCIÓN . . . . .	3
1.1. PREGUNTA DE INVESTIGACIÓN . . . . .	4
2. JUSTIFICACIÓN . . . . .	5
3. MARCO TEÓRICO . . . . .	6
3.1. IDENTIFICACIÓN EN MODELOS DE ECUACIONES SIMULTÁNEAS . . . . .	6
3.2. ENDOGENEIDAD EN SISTEMAS DE PRECIOS . . . . .	7
3.3. VALIDEZ DEL INSTRUMENTO . . . . .	8
3.4. ANÁLISIS DE ELASTICIDADES EN RETAIL . . . . .	9
4. METODOLOGÍA . . . . .	10
4.1. DATOS . . . . .	10
4.2. MODELOS . . . . .	11
4.3. SELECCIÓN DEL INSTRUMENTO . . . . .	12
5. RESULTADOS . . . . .	15
6. CONCLUSIONES . . . . .	18
6.1. CONSIDERACIONES, LIMITACIONES Y PASOS FUTUROS . . . . .	19
7. REFERENCIAS . . . . .	20
8. APÉNDICE . . . . .	22

## **Resumen**

Se aplicó el método de variable instrumental para medir el posible sesgo por endogeneidad en un modelo de elasticidades-precio de la demanda aplicado para la cadena de almacenes *retail* más importante en Colombia. Se compararon los resultados por MCO con los obtenidos con la regresión en dos etapas utilizando como instrumento del precio el efecto sobre la movilidad laboral que tuvo la cuarentena obligatoria como medida preventiva al COVID-19. Se encontró gran variabilidad entre categorías pero en general se notó que los coeficientes por MCO tienden a subestimar la elasticidad de los productos.

Palabras clave: Sesgo por endogeneidad, elasticidad-precio demanda, retail, variable instrumental.

## **Abstract**

Using transactional data from the leading retail chain in Colombia, a two-stage model is used to correct for endogeneity bias in a pre-existing price-elasticity of demand model. OLS results are compared to those of Instrumental Variable regression for categories of products. Data related to work-related mobility restrictions is used as an instrument for the price time series with the intention of capturing the supply shock resulting from the first months of quarantine associated with COVID-19. Results were found to vary significantly between product categories but overall elasticity coefficients were being underestimated by the OLS model.

Keywords: Endogeneity bias, Price-elasticity of Demand, retail, instrumental variable.

# 1. INTRODUCCIÓN

En el análisis microeconómico clásico, se entiende que al equilibrio se llega a través de la interacción simultánea entre las fuerzas de demanda y las fuerzas de oferta. En el punto de intersección de estas dos funciones yace el equilibrio de ese mercado. Esta determinación simultánea, vacía el mercado y se llegan a valores específicos de precios y cantidades. Desde la perspectiva de la firma, en particular una cadena de almacenes *retail*, se tiene un mercado para cada producto (PLU) donde los precios pueden variar libremente dentro de unas reglas de negocio. La expresión del concepto de equilibrio en datos reales es el registro histórico de datos transaccionales, siendo cada venta uno de muchos equilibrios en uno de muchos mercados. Teniendo la nube de puntos de equilibrio de un bien particular es natural pensar, desde la perspectiva de la firma, en estimar la forma de la función de demanda de cada producto para mejorar su estrategia de precios, optimizar sus márgenes o para categorizar su portafolio de bienes ofrecidos de manera estratégica según el comportamiento de su función de demanda.

Una solución intuitiva al problema sería ajustar una regresión a los puntos para obtener el coeficiente de variación en las cantidades ante un cambio en el precio. Este método de estimación de la elasticidad-precio de la demanda (EPD) fue la solución que se utilizó para construir un modelo de elasticidades como proyecto de práctica empresarial, en Almacenes Éxito S.A. Se evaluaron algunas formas funcionales a las cuales se podría ajustar la función de demanda de cada producto. Dada la naturaleza del problema de ser un proyecto empresarial, se decidió por cuestión de efectividad en tiempo y por simplicidad interpretativa el modelo de elasticidad constante. Esta decisión tiene sentido práctico para el proyecto y en general ha dado buenos resultados para el proyecto pero tiene ciertas implicaciones teóricas. Es posible que el estimador de mínimos cuadrados ordinarios esté sesgado al tratarse de un problema de endogeneidad a causa de la simultaneidad.

Para sortear el problema de modelos de ecuaciones simultáneas (SEM), Wooldridge (2012) utiliza la regresión con variables instrumentales para identificar, a partir de choques a la función

de oferta una función de demanda. Este problema se remonta a Wright (1928), quien fue el primero en utilizar reportes climáticos de alta mar como instrumento a la oferta de su pescadería local. Si se tienen los equilibrios de un mercado ante cambios en la oferta y además se cuenta con información de dicho choque, se pueden utilizar las distintas funciones de oferta para estimar los parámetros de la función de demanda. La pregunta es entonces ¿cómo aislar un choque puramente de oferta que permita medir el efecto *ceteris paribus* sobre las unidades demandada?

En este caso se busca identificar la función de demanda de algún conjunto de bienes a partir de un instrumento para los choques de oferta de dichos bienes. Luego se puede utilizar esta identificación de la curva de oferta para estimar los parámetros de su función de demanda. Al tratarse de una firma productora, tiene mucho más importancia la deducción de la función de demanda del mercado que pensar en la función de oferta. Ésta tiende a ser más exógena a la empresa más allá de las decisiones de surtido de portafolio. Vale aclarar que hay alternativas al método de regresión por variable instrumental que podrán ser considerados. MacKay y Miller (2019) notan las dificultades en la selección de un instrumento válido y propone hacer un a identificación con instrumentos de demanda.

## **1.1. PREGUNTA DE INVESTIGACIÓN**

El núcleo de este trabajo es atacar el problema clásico de identificación en sistemas de demanda con dos particularidades. Por una parte, se utiliza como instrumento el choque más reciente y menos estudiado: Los efectos de una pandemia global, cuarentena obligatoria y el cierre obligatorio de varios sectores económicos. Por otra parte se trabajan con datos utilizados en un modelo preexistente cuyos resultados se busca corroborar o esclarecer los futuros pasos para su correcta estimación.

## 2. JUSTIFICACIÓN

Para entender la importancia de este trabajo es importante resaltar que se trata de un caso de estudio empresarial, utilizando data transaccional histórica de la cadena de almacenes retail líder del país. Esto es importante porque el trabajo busca cumplir dos metas importantes: Brindarle al mundo de la investigación, la teoría de la organización industrial y la academia en general la oportunidad de aplicar herramientas econométricas a conjuntos de datos reales, de gran tamaño.

Por otra parte está la oportunidad para la empresa de llevar a cabo una investigación que ayude a brindar información adicional al modelo de elasticidades existente en el Departamento de Analítica de Almacenes Éxito S.A y ayude en la explicación de ciertos resultados que van en contra de la intuición económica tradicional como lo son las elasticidades-precio de la demanda positivas. Desde sus inicios, se han obtenido resultados de este estilo para un porcentaje considerable de productos. Si bien nunca se ha pretendido interpretar estas elasticidades positivas como señales de que se podrían aumentar las ventas con aumentos de precios, siempre queda la incógnita de cómo interpretar de la mejor manera estos resultados o si es necesario corregir algún aspecto de la estimación actual.

Los resultados del modelo actual, de construcción propia, muestran las características de un problema de simultaneidad. El propósito del modelo de elasticidades que se construyó es servir como una de varias entradas en un modelo de optimización de precios. Para la empresa, la oportunidad de hacer una investigación a profundidad y poner a prueba una herramienta ya implementada tiene gran valor. Toda información adicional que se logre obtener sobre los patrones de consumo y el conocimiento del comportamiento de bienes de la canasta básica de consumo ante cambios en el precio son claves para el proyecto ya en desarrollo. Más aún, el intento de corregir los sesgos que puedan estar afectando la validez del modelo puede ayudar a su mejor comprensión o futuros pasos para mejorar el modelo.

Como efecto secundario está la posibilidad de explorar los efectos económicos de la pandemia desde datos de consumo real. Apenas se está empezando a estudiar las implicaciones a corto plazo del cierre de la economía mundial y las implicaciones hacia el futuro. Si bien no es el foco de este trabajo, es casi inevitable generar conclusiones acerca de la naturaleza del choque. Se hicieron algunos supuestos sobre su naturaleza, se cree que comenzó como un choque de oferta que luego afectó el poder adquisitivo de los consumidores y por consiguiente la demanda de bienes básicos de consumo. Si se lograra concluir acerca del impacto económico de los efectos del cierre de la economía como consecuencia de una pandemia de escala global, se podrían interpretar mejor los datos del 2020 que presentan un quiebre marcado a comparación de años anteriores.

### **3. MARCO TEÓRICO**

#### **3.1. IDENTIFICACIÓN EN MODELOS DE ECUACIONES SIMULTÁNEAS**

Uno de los pilares teóricos importante de este trabajo es el estudio de sistemas de determinación simultánea y las implicaciones de su estimación. El primer estudio donde se utilizó información adicional para identificar una de las dos ecuaciones de un sistema de demanda fue el desarrollo de Wright (1928). En este, propuso la entonces novedosa idea de capturar información sobre las funciones de oferta, midiendo su variación no observada en el tiempo para así poder estimar correctamente la relación entre el precio y las cantidades demandadas de bienes básicos de consumo. En su estudio, quiso estimar la elasticidad precio de la demanda y de la oferta de todo tipo de bienes, enfocándose en aceites de origen vegetal y en ciertos tipos de pescado. Lo que él llamó *shifters* se conocerían luego formalmente como variables instrumentales. En esencia son variables que contienen información sobre fenómenos que afecten directamente la función de demanda sin afectar la oferta o viceversa. Wright utilizó dos instrumentos principales: el precio de bienes sustitutos como instrumento a la demanda del bien en cuestión y para la medir los cambios sobre la oferta utilizó información meteorológica que

aproximara las condiciones de producción de los bienes que analizó. Este estudio seminal abrió el camino a muchos otros usos de instrumentos de distintas naturalezas y para distintos tipos de variables a instrumentar. Siguiendo la línea de estudios sobre sistemas de oferta está el desarrollo de Angrist y Evans (1998), donde a partir de datos sobre la composición del sexo de los hijos se pretende instrumentar la fertilidad que presenta problemas de endogeneidad con la oferta de trabajo. Un ejemplo de particular interés para el presente estudio es aquel de Imbens y Klaauw (1995). En este utilizan un conjunto de variables dummy como instrumento al haber sido llamados a prestar servicio militar en Holanda, con el fin de estimar el efecto que tiene la participación en el ejército sobre los ingresos de los individuos.

La clave para el éxito de la instrumentación de un proceso endógeno es precisamente la validez del instrumento. La gran barrera que deben sortear los investigadores es la selección de un buen instrumento que cumpla las condiciones de exogeneidad y relevancia para darle validez al método. Es por esto que común en la literatura que se recurra a variables puramente exógenas, como lo son los fenómenos meteorológicos.

### **3.2. ENDOGENEIDAD EN SISTEMAS DE PRECIOS**

Los sistemas de precios y la relación entre las decisiones de consumo y el precio son un caso específico de los modelos de ecuaciones simultáneas. Son quizás el ejemplo por excelencia de dicho tipo de sistemas. En la década de los 90's se presentó un gran número de estudios de organización industrial y estimación del efecto del precio en la concentración de mercado. Evans y col. (1993) notan que la estimación de los efectos sobre los precios por OLS en estos sistemas tiende a estar sesgada por la endogeneidad. En sus resultados encuentran que el sesgo es significativo y negativo, es decir que se estaba sobreestimando el efecto de la concentración de mercado en los precios. Para hacer la conexión con este estudio, se pretende hacer una estimación similar y ver si en efecto existe un sesgo y analizar su magnitud y dirección. En este caso se enfoca en el sesgo del efecto estimado del precio sobre las cantidades consumidas.

En los modelos estudiados por Evans y col. (1993), la gran mayoría utilizaba una función de demanda lineal. Es importante notar sobre los supuestos que se hicieron en el modelo de elasticidades actual. Por simplicidad y sobretodo por facilidad interpretativa y explicativa se optó por el modelo de elasticidad constante, derivado de una función logarítmica de demanda

$$Y = \beta_1 X_i^{-\beta_2} \quad (1)$$

$$\ln Y = \ln \beta_1 - \beta_2 \ln X_i \quad (2)$$

En este caso, en la ecuación (1) la variable dependiente son las cantidades vendidas de un producto específico y la variable regresora es el precio de dicho bien. Se utiliza la versión logarítmica para facilitar la interpretación del coeficiente  $\beta_2$  como un cambio porcentual en las cantidades ante un cambio porcentual en el precio.

### 3.3. VALIDEZ DEL INSTRUMENTO

Para poder sortear la endogeneidad en un sistema de determinación simultánea mediante la identificación de una ecuación con un instrumento, se tienen que cumplir los supuestos de MCO y reemplazar el supuesto de exogeneidad por los siguientes dos supuestos :

$$\text{Corr}(X_1, Z_1) \neq 0 \quad (3)$$

$$\text{Corr}(Y, Z_1) = 0 \quad (4)$$

La condición (3) es conocida como el supuesto de relevancia del instrumento. En este caso es que el hecho de que los puntos de equilibrio de las fechas entre el inicio de la cuarentena el 25 de marzo del 2020 y el 1ro de julio del mismo año, obedecen a una función de oferta desplazada hacia la izquierda a comparación de los meses anteriores. La ecuación (4) es la condición de exogeneidad, es decir que no hay correlación entre la variable que se va a usar como instrumento y la variable dependiente.

Brinca y col. (2010) hicieron la descomposición del choque sufrido entre marzo y abril cuando

empezó el efecto COVID-19, para diferentes sectores económicos de Estados Unidos. A partir de la relación entre los cambios de oferta laboral (medida en horas) y el salario, ante un choque, determinaron qué tanto se debía a fuerzas de oferta o demanda. Sus resultados apuntaban a que la mayor parte del choque se debe a una caída de la oferta laboral, transversal a todos los sectores. Para el retail encontraron que en su mayoría era un efecto de oferta laboral y el pequeño choque positivo de demanda que hallaron para el mes de marzo se ve anulado por el choque negativo del mes de abril. Basado en esto se puede asumir que esta variable es lo más cercano a un choque puramente exógeno sobre las condiciones de producción y abastecimiento, sin afectar la estructura de la demanda por estos bienes.

### **3.4. ANÁLISIS DE ELASTICIDADES EN RETAIL**

Con el desarrollo de técnicas de procesamiento de grandes volúmenes de datos, se ha popularizado el uso del valor de la elasticidad-precio de la demanda de los productos en fijación de precios en un entorno empresarial. Compañías consultoras como (Mosaic Data Science, 2020), han publicado casos de estudio donde realizan estimaciones de EPD para compañías con portafolios de 150 millones de códigos de *Price look-up* (PLU) únicos. Estos códigos corresponden a referencias específicas de productos que ofrece una cadena. En el caso de estudio notan la dificultad de medir efectos no observados como la posible estacionalidad de algunos productos y el efecto del inventario o productos agotados a través de regresiones directas. Otros estudios como Levy y col. (2004) concuerdan con las recomendaciones del caso de estudio mencionado y agregan la posibilidad de obtener elasticidades positivas. Esto lo explican como la relación positiva entre el precio de un bien y su percepción de calidad, sin embargo la relación encontrada entre el precio y la disposición a pagar es negativa.

## 4. METODOLOGÍA

### 4.1. DATOS

El conjunto de datos transaccionales utilizados son propiedad privada de Almacenes Éxito S.A y no puede revelarse información sensible acerca de las reglas de precios del negocio ni detalles de valores históricos de ventas. Por lo tanto, se optó por construir Índices de Fisher para las series de precios y cantidades de las categorías escogidas. Todas las series se trabajaron entre el 1ro de mayo de 2019 y el 31 de agosto del 2020

$$IPP = \frac{\sum p_{it}q_{it}}{\sum p_{i0}q_{it}} \quad (5)$$

$$IPL = \frac{\sum p_{it}q_{i0}}{\sum p_{i0}q_{i0}} \quad (6)$$

$$IQP = \frac{\sum p_{it}q_{it}}{\sum p_{it}q_{i0}} \quad (7)$$

$$IQL = \frac{\sum p_{i0}q_{it}}{\sum p_{i0}q_{i0}} \quad (8)$$

$$(9)$$

Estas son las ecuaciones para la construcción de los índices ponderados de Paasche y Laspeyres, tanto de precios como de cantidades. Estos utilizan los precios y cantidades del periodo base para compararlos con sus valores corrientes, midiendo la variación porcentual en el tiempo para estas variables. Ambas son utilizadas frecuentemente en el cálculo de la inflación.

$$IPF = \sqrt{IPP * IPL} \quad (10)$$

$$IQF = \sqrt{IPQ * IQL} \quad (11)$$

A partir de estas dos parejas de índices, se construyen los Índices de Fisher. En este caso, son utilizadas para que los resultados de regresión sean interpretables como las elasticidades del modelo de regresión lineal con variables en logaritmos. Así son comparables los resultados entre los dos modelos.

En el modelo actual de la compañía se trabaja con los datos más desagregados posibles. Es decir, se calculan valores de elasticidades para combinaciones únicas de PLU-punto de venta-días. Para este trabajo se agregaron los valores por categorías y se eliminó la dimensión geográfica. Cada categoría tiene un valor de  $P_t$  y  $Q_t$ , tomando la mediana del precio y la suma de unidades vendidas respectivamente. Para tener una base de datos sin elementos nulos fue necesario reducir las categorías a aquellos PLU's que tuviesen al menos una unidad vendida en cada día del periodo analizado. Esto implica que se está sesgando la muestra al incluir únicamente productos de muy alta rotación. Estos productos tienden a ser de necesidad básica por lo que se espera que tiendan a ser relativamente inelásticos a choques en precios y se espera que sobre ellos el choque de restricción de movilidad tenga un gran impacto. Esta reducción busca maximizar la cantidad de información mientras asegura que las bases de las series sean invariantes en el tiempo.

En cuanto a la selección de las categorías a analizar, se empleó una lógica similar. Se ordenaron las categorías de Productos de Gran Consumo según sus ventas en unidades durante los 16 meses analizados. De estos se descartaron los que fuesen productos a granel para evitar problemas con las unidades y se escogieron las 10 categorías más vendidas. La base de datos era de 25 millones de registros correspondiente a 17 millones de transacciones y 228 PLUs. Luego de aplicarle el proceso mencionado anteriormente se llegó al formato final de 10 series diarias de cantidades y precios respectivamente, cada una de 484 registros.

## **4.2. MODELOS**

En los modelos estudiados por Evans y col. (1993), la gran mayoría utilizaba una función de demanda lineal. Es importante notar sobre los supuestos que se hicieron en el modelo de elasticidades actual. Por simplicidad y sobretodo por facilidad interpretativa y explicativa se

optó por el modelo de elasticidad constante, derivado de una función logarítmica de demanda

$$Y = \beta_1 X_i^{-\beta_2} \quad (12)$$

$$\ln Y = \ln \beta_1 - \beta_2 \ln X_i \quad (13)$$

En este caso, en la ecuación (4.1) la variable dependiente son las cantidades vendidas de un producto específico y la variable regresora es el precio de dicho bien. Se utiliza la versión logarítmica para facilitar la interpretación del coeficiente  $\beta_2$  como un cambio porcentual en las cantidades ante un cambio porcentual en el precio. Nótese que bajo este esquema opera el modelo original de elasticidades que se quiere corregir.

### 4.3. SELECCIÓN DEL INSTRUMENTO

Partiendo del choque de oferta que se quería utilizar y de las condiciones para un buen instrumento, se buscó una variable que capturara las nuevas condiciones de oferta sin estar afectando las condiciones de demanda. Originalmente se pensó en utilizar una variable *dummy*, que tomase el valor de 1 para los días entre el 23 de Marzo y el 23 de mayo del 2020, correspondiente a los primeros dos meses de la cuarentena obligatoria decretada el 22 de Marzo («DECRETO NÚMERO 457 DE 2020», s.f.). Para este periodo se espera que las condiciones de consumo de alimentos de la población general se haya mantenido estable pero que debido a la incertidumbre se produjo un efecto de aprovisionamiento, visible en las series de cantidades de las categorías.

Finalmente se optó por utilizar datos del Reporte de movilidad de comunidades de Google («Google Mobility Report», 2020). Este consiste de 6 variables diarias que miden el cambio porcentual diario en la actividad o participación en un sector particular comparado con un periodo base. Dicha base corresponde a la mediana del periodo entre el 3 de enero y el 6 de febrero del 2020 (5 semanas). Al tratarse de restricciones a la movilidad se definió que esta variable tomaría el valor de 0 para los días previos al periodo base. Esto implica asumir que en

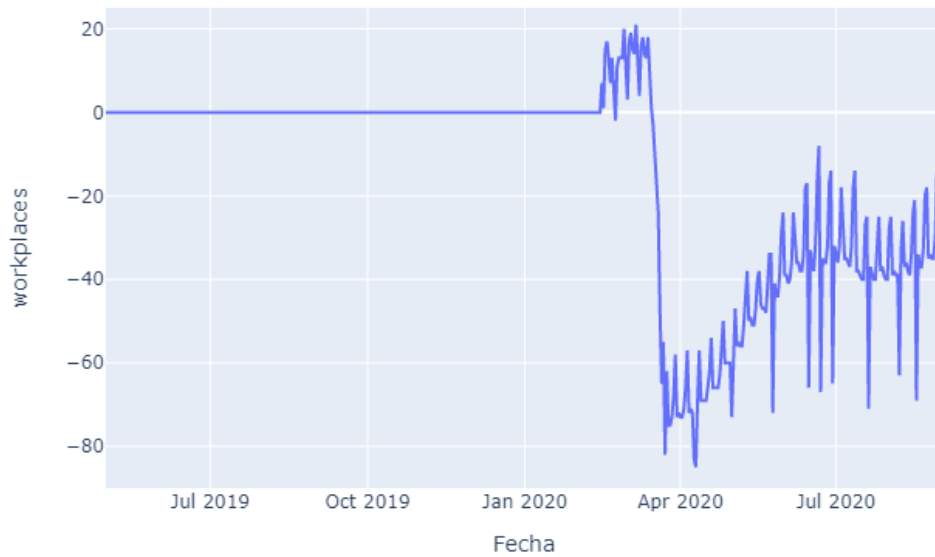


Figura 1: Serie del instrumento

promedio los datos del 2019 tuvieron las mismas condiciones del periodo base.

De las 6 variables, hay dos de particular interés: la participación en el sector *retail* y la movilidad hacia lugares de trabajo. La primera por tratarse del sector económico de la empresa y la segunda por ser la que mejor captura el choque de oferta que se quiere estudiar. Se optó por esta última al ser la medida más directa de la reducción a la oferta de bienes y servicios. La selección de esta variable además permite cuantificar la magnitud del choque y evita tomar decisiones arbitrarias en cuanto a las fechas límites.

El método de estimación de Mínimos Cuadrados en Dos Etapas (MC2E) fue desarrollado por Theil (1953) e independientemente por Basmann (1957) En este caso, se parte la regresión lineal del índice de cantidades explicado por el índice de precios. Donde  $X_1$  es la variable endógena (Precio) que se quiere instrumentar

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon \quad (14)$$

$$\hat{X}_1 = \gamma_0 + \gamma_1 Z_1 + v \quad (15)$$

Primero se lleva a cabo la regresión del instrumento sobre dicha variable endógena. Donde  $Z_1$  es la variable instrumental . Luego de capturar los valores estimados de  $X_1$  de la ecuación (15), se incluyen como regresor en la ecuación original (14) de tal forma que quede:

$$\ln Y = \beta_0 + \beta_1 \hat{X}_1 + v \quad (16)$$

Nótese que en la ecuación (16) el término de error ( $v$ ) proviene de la estimación de la primera etapa y no está correlacionado con  $\hat{X}_1$ . Lo que se busca entonces es comparar estos  $\hat{\beta}_{VI}$  con el ya existente  $\hat{\beta}_{MCO}$  para medir el sesgo por endogeneidad.

## 5. RESULTADOS

El Cuadro 1 muestra los resultados de regresión de las 10 categorías, tanto por MCO como por VI. Para ambos se reportan los valores de los coeficientes asociados a la variable de precio y su desviación estándar asociada.

Cuadro 1: Comparación con MCO

Categoría	$\hat{\beta}_{mco}$	$\sigma_{mco}$	$\hat{\beta}_{vi}$	$\sigma_{vi}$
Agua Normal	-3.1854***	0.2414	-8.3632***	0.6032
Crema Dental	-3.2873***	0.3105	-4.1574***	0.6647
Detergente	0.1741	0.4432	3.8230***	0.8866
Harina Maíz	1.2294**	0.4840	8.2339***	1.1147
Queso Mozzarella	0.5030*	0.2265	2.5075***	0.3471
Pastas Comunes	1.5492**	0.7022	16.0867***	2.5694
Papel Higiénico	-8.8533***	0.6266	-18.6258***	5.6162
Tortas y Ponqués	-1.3151***	0.2441	4.9008***	0.9232
Wafer	1.2783***	0.2401	2.8517	0.3909
Yogurt	1.2798***	0.2002	2.8740***	0.3190

\*\*\* p<0.01, \*\* p<0.05, \* p<0.1

Es interesante notar que todas las categorías a excepción de Detergentes son elásticas, es decir que su valor de elasticidad es mayor a uno. Para categorías como Detergentes, al estar el coeficiente tan cercano a cero, más probable que el intervalo de confianza incluya el cero y por consiguiente sea el coeficiente no significativo como se observa en este caso.

En el Apéndice se encuentran las series de de ambos índices para las 10 categorías junto con los resultados de regresión individuales. Las Figuras 2 y 3 muestran el ejemplo de las series de cantidades y precios respectivamente para la categoría Agua Normal. En este ejemplo se puede ver como el choque impactó negativamente la serie de cantidades con perfecta coincidencia y que el nivel de consumo se estabilizó en un nuevo nivel mucho más bajo que antes. Esto tiene

### AGUA NORMAL

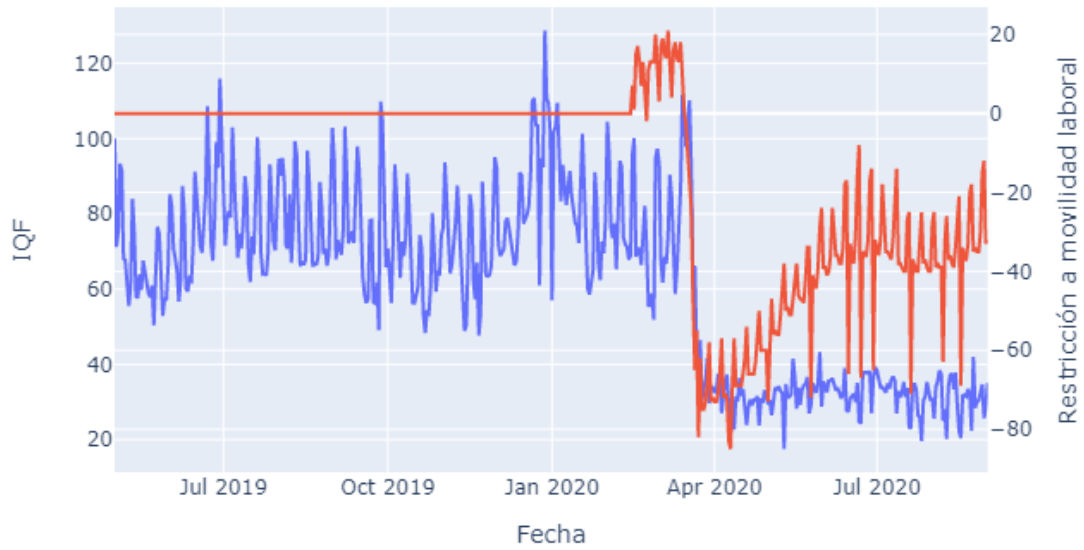


Figura 2: Agua Normal (Cantidades)

### AGUA NORMAL

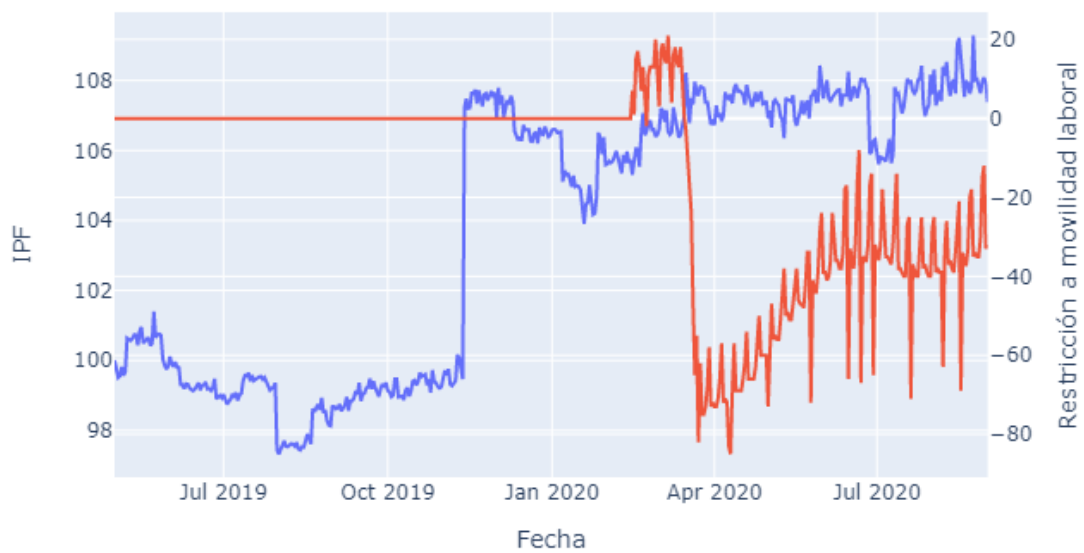


Figura 3: Agua Normal (Precios)

sentido si se supone que el consumo de agua embotellada tiende a disminuir drásticamente al pasar más tiempo en la casa. Se puede observar también que la serie de precios no reacciona al choque. Esto se debe a que los cambios en precios están sujetos a reglas del negocio que cambian en el tiempo y difieren entre categorías.

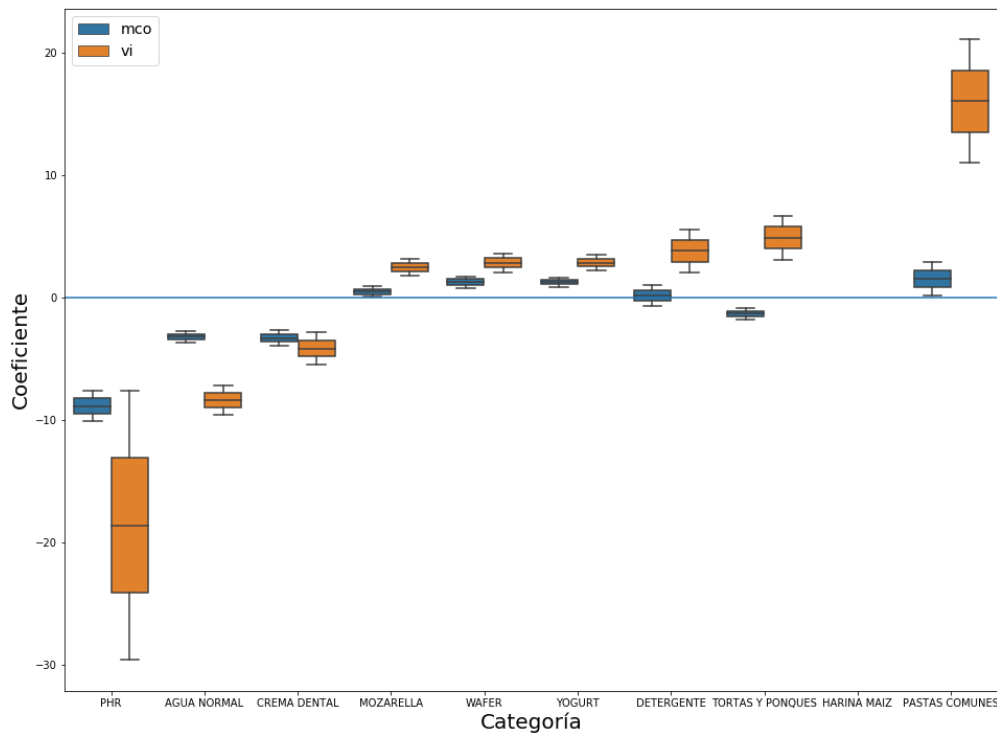


Figura 4: Resultados por Categoría

La Figura 4 permite visualizar los cambios en los coeficientes de estimación por categoría. Se organizó de menor a mayor según su valor de coeficiente por VI. Nótese que las bandas de confianza bajo este método son de mayor tamaño que sus contrapartes de MCO. Esto se debe a que por lo general, se tiene una mayor desviación estándar bajo la regresión en dos etapas que bajo MCO y esta diferencia en las desviaciones se acentúa dependiendo de la debilidad del instrumento.

## 6. CONCLUSIONES

Para poder entender el impacto de los resultados obtenidos, resulta útil recordar los objetivos planteados inicialmente. Principalmente, se buscaba ver si los resultados del modelo de elasticidades preexistente tenían claras señas de estar sesgados por problemas de endogeneidad. De ser este el caso, no estaría restándole validez al modelo actual pero sí ayudaría a determinar si el efecto de la incidencia del precio en las decisiones de compra está siendo subestimado o sobrestimado. Los resultados de la comparación entre la estimación por MCO y el método de VI muestra una clara diferencia entre los coeficientes en las 10 categorías analizadas. A excepción de la categoría "Tortas y Ponqués", que sufrió un cambio de signo, todas las categorías obtuvieron resultados más elásticos cuando se instrumentó su precio. Esto en sí mismo ya es un resultado interesante, más aún cuando se considera que una gran proporción de los resultados del modelo existente yacen dentro del círculo unitario, implicando que hay una abundancia de productos inelásticos. Es posible que para algún conjunto de estos resultados inelásticos, el modelo de regresión lineal no logre captar la verdadera relación entre las variables.

Otro aspecto interesante de los resultados es la diferencia entre categorías, tanto en el comportamiento de sus series en índices como en los resultados de regresión. Nótese por ejemplo, como las Figuras 6, 12, 14 y 16 del Apéndice muestran 4 periodos promocionales muy marcados en las respectivas categorías. Esto muestra la diferencia en estrategias comerciales a las cuales están sujetas las series de precios de los diferentes tipos de productos. El ejemplo se magnifica cuando se considera que solo se trabajó con un conjunto reducido de categorías, la mayoría pertenecientes a productos de alimentos.

En las series de cantidades también se puede observar gran heterogeneidad en los efectos del choque. La categoría Agua Normal es la única que presentó una gran reducción en sus ventas coincidente con la reducción en la movilidad. Esto ilustra las diferencias no observadas entre las condiciones de demanda de los distintos tipos de productos que se pretenden abarcar con una sola variable explicatoria.

## 6.1. CONSIDERACIONES, LIMITACIONES Y PASOS FUTUROS

Es importante reconocer que si bien se ha logrado tener más información sobre la relación entre precios y cantidades, aún no se han resuelto algunas dudas conceptuales grandes como lo son las elasticidades positivas. Ya se corrigió por endogeneidad pero aun queda el problema del sesgo por variable omitida. El modelo de elasticidades acepta de entrada esta posibilidad porque busca utilizar el mayor nivel de simpleza para sacar un valor aproximado con el cual se tomen decisiones de fijaciones de precio dentro de la compañía. Se sabe de antemano que el precio de un bien no es el único factor decisorio en las decisiones de los consumidores.

En este sentido, el presente trabajo muestra que aún corrigiendo por endogeneidad, hay categorías para las cuales se obtienen resultados contrarios a la intuición económica y que posiblemente se deban seguir los pasos de Angrist y Kreuger (2001) y utilizar variables adicionales en la estimación. Es común que se controle por efectos de ingreso y precios de productos sustitutos, que en este caso corresponde a agregar alguna medida del nivel de ingresos e información sobre los cambios de precios de cadenas de almacenes de la competencia. He aquí donde se encuentra el gran aporte de este trabajo investigativo para el proyecto empresarial: Se están justificando los pasos a seguir en caso de que se desee refinar las estimaciones del modelo actual, que si bien ha sido bien recibido, todavía cuenta con espacio para mejorar.

Otra posibilidad es seguir las líneas de trabajos como Cortés y Pérez (2010), donde se trata el problema como un sistema de ecuaciones de equilibrio general. Estiman a partir de datos de la Encuesta de Ingresos y Gastos, las elasticidades-precio y gasto de grandes sectores agregados como lo son el gasto en alimentos o el gasto en salud. La diferencia aquí sería señirse a un modelo completo de demanda y calcular en conjunto estos valores, contrario a las regresiones actuales que solo toman en cuenta choques a los precios propios de los bienes. Nuevamente cabe la aclaración de que son sugerencias en caso de que se busque refinar los resultados del modelo actual pero que no necesariamente resulten atractivos al tener que recurrir a fuentes

externas de datos y desarrollos futuros.

## 7. REFERENCIAS

- Angrist, J. & Evans, W. (1998). Children and Their Parents' Labor Supply: Evidence from Exogenous Variation in Family Size. *American Economic Review*.
- Angrist, J. & Kreuger, A. (2001). Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives*.
- Basman, R. L. (1957). A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica*.
- Brinca, P., Duarte, J. B. & Faria-e-Castro., M. (2010). Is the COVID-19 Pandemic a Supply or a Demand Shock? *Federal Reserve Bank of St. Louis*.
- Cortés, D. & Pérez, J. (2010). El consumo de los hogares colombianos, 2006-2007: Estimación de sistemas de demanda. *UNIVERSIDAD DEL ROSARIO - FACULTAD DE ECONOMÍA, DOCUMENTOS DE TRABAJO*.
- DECRETO NÚMERO 457 DE 2020 [Por el cual se imparten instrucciones en virtud de la emergencia sanitaria generada por la pandemia del Coronavirus COVID-19 y el mantenimiento del orden público.]. (s.f.).
- Evans, W., Froeb, L. & Werden, G. (1993). Endogeneity in the Concentration–Price Relationship: Causes, Consequences, and Cures. *The Journal of Industrial Economics*.
- Google Mobility Report*. (2020). <https://www.google.com/covid19/mobility/?hl=es-419>
- Imbens, G. & Klaauw, W. (1995). Evaluating the Cost of Conscription in The Netherlands. *Journal of Business Economic Statistics, American Statistical Association*.
- Levy, M., Grewal, D., Kopalle, P. K. & Hess, J. D. (2004). Emerging trends in retail pricing practice: implications for research. *Journal of Retailing*.
- MacKay, A. & Miller, N. H. (2019). Estimating Models of Supply and Demand: Instruments and Covariance Restrictions. *Harvard Business School*.

Mosaic Data Science. (2020). PRICE ELASTICITY OF DEMAND AUTOMATION CASE STUDY [<https://www.mosaicdatascience.com/2016/08/30/price-elasticity-of-demand-cs/>,].

Theil, H. (1953). Repeated least-squares applied to complete equation systems. *Central Planning Bureau, Memorandum*.

Wooldridge, J. (2012). *Introductory Econometrics: A Modern Approach* (5.<sup>a</sup> ed.). South Western Cengage Learning.

Wright, P. G. (1928). *The Tariff on Animal and Vegetable Oils*. *New York: Macmillan*.

## 8. APÉNDICE

Cuadro 2: Agua Normal

	<i>Variable dependiente:</i>	
	IQF	
	<i>OLS</i>	<i>instrumental variable</i>
	(1)	(2)
IPF	-3.185*** (0.241)	-8.363*** (0.604)
Constante	392.339*** (25.101)	930.234*** (62.811)
Observaciones	484	484
R <sup>2</sup>	0.265	0.806
Adjusted R <sup>2</sup>	0.264	-
Residual Std. Error (df = 482)	20.804	29.082
F Statistic	174.047*** (df = 1; 482)	

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Cuadro 3: Crema Dental

	<i>Variable dependiente:</i>	
	IQF	
	<i>OLS</i>	<i>instrumental variable</i>
	(1)	(2)
IPF	-3.287*** (0.311)	-4.157*** (0.666)
Constante	405.118*** (32.484)	496.068*** (69.644)
Observations	484	484
R <sup>2</sup>	0.189	0.810
Adjusted R <sup>2</sup>	0.187	-
Residual Std. Error (df = 482)	30.342	30.588
F Statistic	112.083*** (df = 1; 482)	

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Cuadro 4: Detergente

	<i>Dependent variable:</i>	
	IQF	
	<i>OLS</i>	<i>instrumental variable</i>
	(1)	(2)
IPF	0.174 (0.443)	3.823*** (0.889)
Constant	67.754 (44.700)	-299.036*** (89.394)
Observations	484	484
R <sup>2</sup>	0.0003	0.480
Adjusted R <sup>2</sup>	-0.002	-
Residual Std. Error (df = 482)	78.157	83.470
F Statistic	0.154 (df = 1; 482)	

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Cuadro 5: Harina Maíz

	<i>Dependent variable:</i>	
	IQF	
	<i>OLS</i>	<i>instrumental variable</i>
	(1)	(2)
IPF	1.229** (0.484)	8.234*** (1.117)
Constant	-26.826 (52.150)	-780.895*** (120.287)
Observations	484	484
R <sup>2</sup>	0.013	0.777
Adjusted R <sup>2</sup>	0.011	-
Residual Std. Error (df = 482)	45.405	54.381
F Statistic	6.452** (df = 1; 482)	

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Cuadro 6: Queso Mozzarella

	<i>Dependent variable:</i>	
	IQF	
	<i>OLS</i>	<i>instrumental variable</i>
	(1)	(2)
IPF	1.229** (0.484)	8.234*** (1.117)
Constant	-26.826 (52.150)	-780.895*** (120.287)
Observations	484	484
R <sup>2</sup>	0.013	0.929
Adjusted R <sup>2</sup>	0.011	-
Residual Std. Error (df = 482)	45.405	54.381
F Statistic	6.452** (df = 1; 482)	

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Cuadro 7: Pastas Comunes

	<i>Dependent variable:</i>	
	IQF	
	<i>OLS</i>	<i>instrumental variable</i>
	(1)	(2)
IPF	1.549** (0.702)	16.087*** (2.575)
Constant	-83.445 (72.130)	-1,576.041*** (264.373)
Observations	484	484
R <sup>2</sup>	0.010	0.563
Adjusted R <sup>2</sup>	0.008	-
Residual Std. Error (df = 482)	41.579	57.147
F Statistic	4.867** (df = 1; 482)	

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Cuadro 8: Papel Higiénico

	<i>Dependent variable:</i>	
	IQF	
	<i>OLS</i>	<i>instrumental variable</i>
	(1)	(2)
IPF	-8.853*** (0.627)	-18.626*** (5.628)
Constant	927.063*** (61.786)	1,889.654*** (554.354)
Observations	484	484
R <sup>2</sup>	0.293	0.329
Adjusted R <sup>2</sup>	0.291	-
Residual Std. Error (df = 482)	60.568	74.293
F Statistic	199.605*** (df = 1; 482)	

Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

Cuadro 9: Tortas y Ponqués

	<i>Dependent variable:</i>	
	IQF	
	<i>OLS</i>	<i>instrumental variable</i>
	(1)	(2)
IPF	-8.853*** (0.627)	4.901*** (0.925)
Constant	927.063*** (61.786)	-422.387*** (93.984)
Observations	484	484
R <sup>2</sup>	0.293	0.823
Adjusted R <sup>2</sup>	0.291	-
Residual Std. Error (df = 482)	60.568	33.113
F Statistic	199.605*** (df = 1; 482)	
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

Cuadro 10: Galletas Wafer

	<i>Dependent variable:</i>	
	IQF	
	<i>OLS</i>	<i>instrumental variable</i>
	(1)	(2)
IPF	1.278*** (0.240)	2.852*** (0.392)
Constant	-52.254** (25.172)	-217.031*** (41.045)
Observations	484	484
R <sup>2</sup>	0.056	0.922
Adjusted R <sup>2</sup>	0.054	-
Residual Std. Error (df = 482)	22.623	23.609
F Statistic	28.333*** (df = 1; 482)	
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

Cuadro 11: Yogurt

	<i>Dependent variable:</i>	
	IQF	
	<i>OLS</i>	<i>instrumental variable</i>
	(1)	(2)
IPF	1.278*** (0.240)	2.874*** (0.320)
Constant	-52.254** (25.172)	-234.785*** (32.812)
Observations	484	484
R <sup>2</sup>	0.056	0.916
Adjusted R <sup>2</sup>	0.054	-
Residual Std. Error (df = 482)	22.623	18.195
F Statistic	28.333*** (df = 1; 482)	
<i>Note:</i>	*p<0.1; **p<0.05; ***p<0.01	

CREMA DENTAL

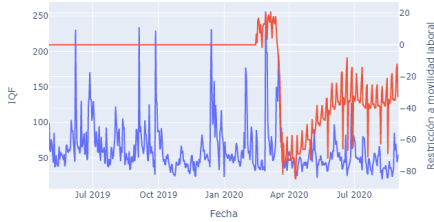


Figura 5: Crema Dental (Q)

CREMA DENTAL

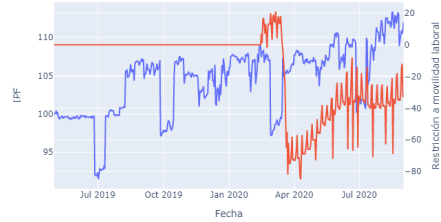


Figura 6: Crema Dental (P)

DETERGENTE

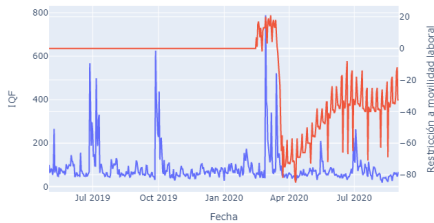


Figura 7: Detergentes (Q)

DETERGENTE

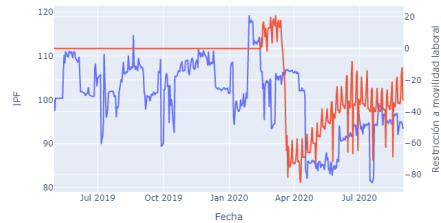


Figura 8: Detergentes (P)

HARINA MAÍZ

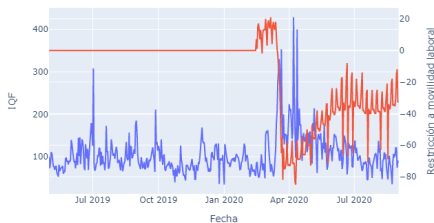


Figura 9: Harina Maíz (Q)

HARINA MAÍZ

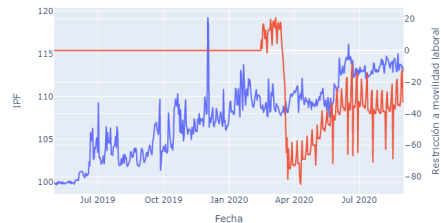


Figura 10: Harina Maíz (P)

QUESO MOZZARELLA

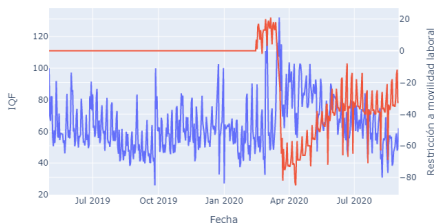


Figura 11: Queso Mozzarella (Q)

QUESO MOZZARELLA

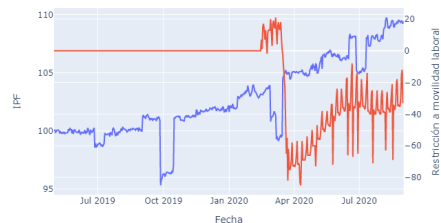


Figura 12: Queso Mozzarella (P)

PASTAS COMUNES

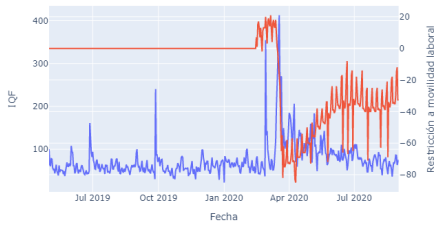


Figura 13: Pastas Comunes (Q)

PASTAS COMUNES

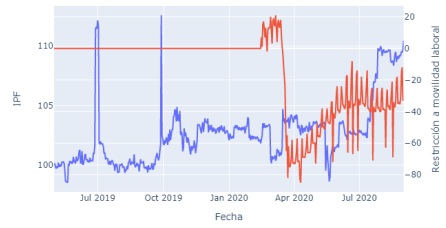


Figura 14: Pastas Comunes (P)

PAPEL HIGIÉNICO

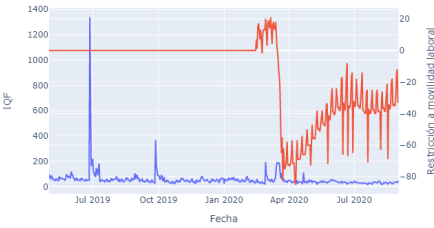


Figura 15: Papel Higiénico (Q)

PAPEL HIGIÉNICO

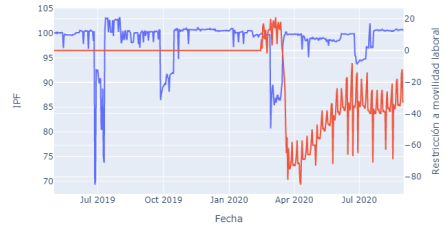


Figura 16: Papel Higiénico (P)

TORTAS Y PONQUÉS

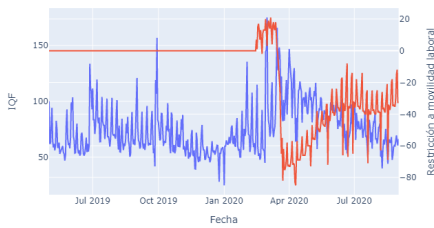


Figura 17: Tortas y Ponqués (Q)

TORTAS Y PONQUÉS

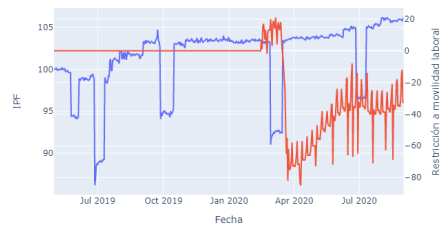


Figura 18: Tortas y Ponqués (P)

WAFER

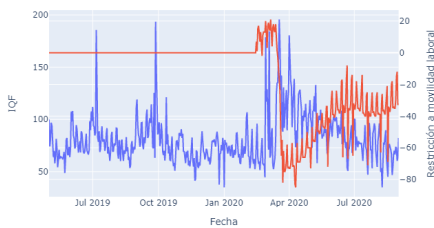


Figura 19: Galletas Wafer (Q)

WAFER

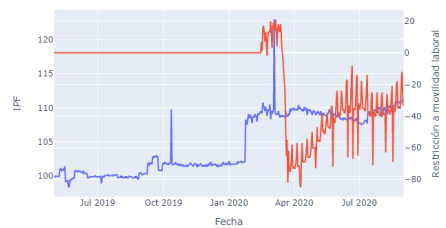


Figura 20: Galletas Wafer (P)

YOGURT

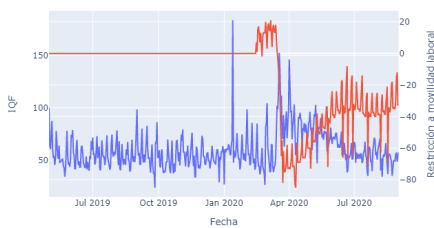


Figura 21: Yogurt (Q)

YOGURT

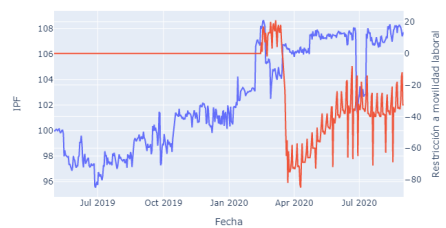


Figura 22: Yogurt (P)