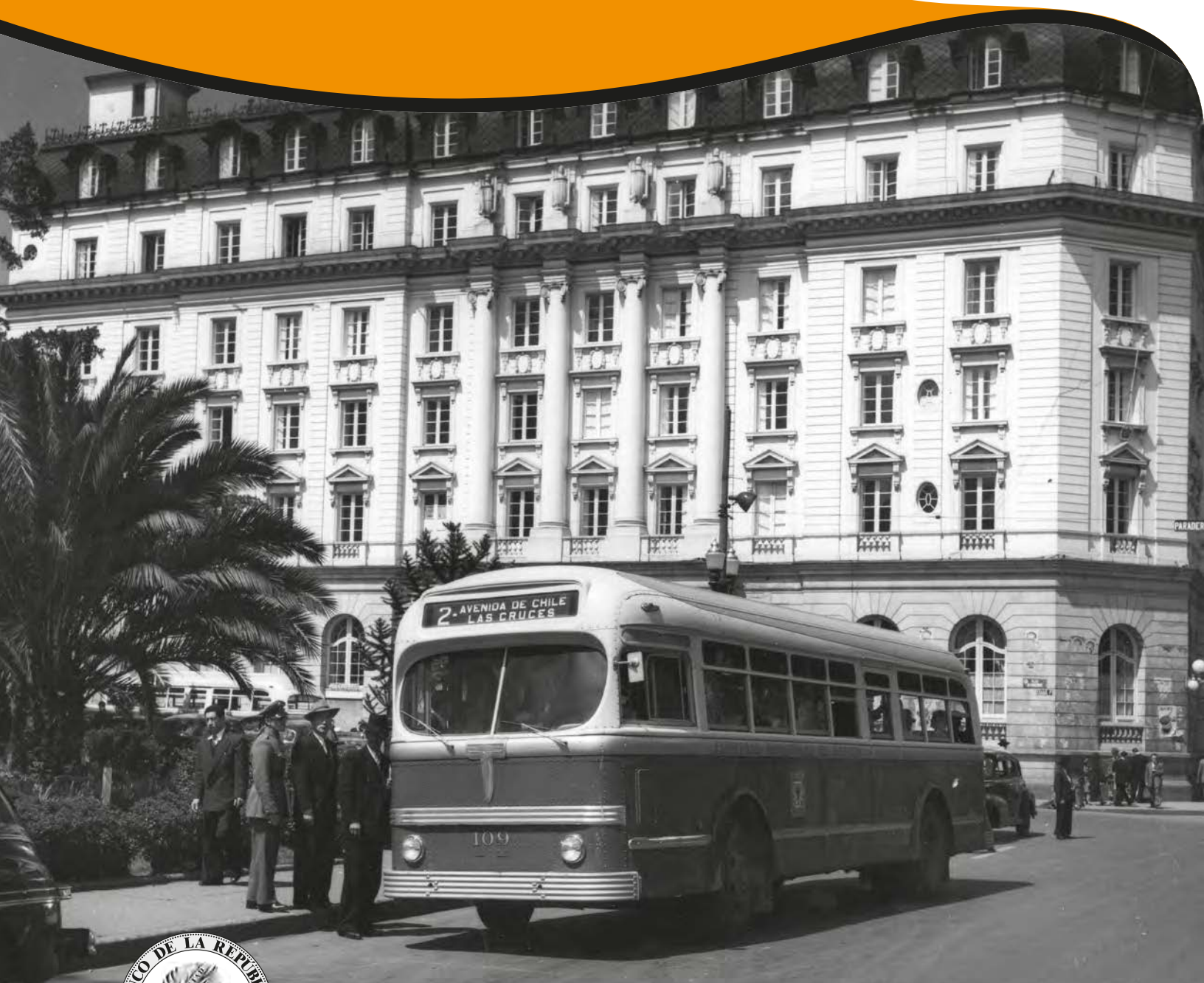


Borradores de ECONOMÍA

Estimación de la variación del precio
de los alimentos con modelos de
frecuencias mixtas

Por: Julián Alonso Cárdenas-Cárdenas
Edgar Caicedo-García
Eliana R. González Molano

Núm. 1109
2020



Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia

Estimación de la variación del precio de los alimentos con modelos de frecuencias mixtas

Julián Alonso Cárdenas-Cárdenas¹

Edgar Caicedo-García

Eliana R. González Molano

jcardeca@banrep.gov.co

ecaicega@banrep.gov.co

egonzamo@banrep.gov.co

Banco de la República

Las opiniones contenidas en el presente documento son responsabilidad exclusiva de los autores y no comprometen al Banco de la República ni a su Junta Directiva

Resumen:

El comportamiento de los precios de los alimentos en Colombia ha sido un factor que inquieta a la autoridad monetaria por su volatilidad, alta ponderación en la canasta de IPC y en ocasiones recurrentes altos niveles debido a su reacción a choques de oferta como el clima, lo cual dificulta la tarea de estabilizar la inflación alrededor de la meta. De lo anterior, se desprende la necesidad de tener pronósticos insesgados y más oportunos de los cambios en el precio de los alimentos en el corto plazo. En este documento se desarrolla una metodología que aprovecha la información disponible con alta frecuencia de precios y abastecimiento de alimentos y permite combinar información observada en varias frecuencias para generar pronósticos alternativos de la variación de los precios de los alimentos y sus diferentes componentes. Los resultados encontrados indican que los modelos propuestos de frecuencias mixtas, producen mejores pronósticos que los tradicionales que utilizan solamente información de precios del Sistema de Información de Precios del Sector Agropecuario (SIPSA-DANE).

Palabras claves: Inflación de alimentos, nowcasting, modelos de frecuencias mixtas, pronósticos, clima, inflación objetivo.

JEL: C32, C51, C53, E31, E52.

¹ En orden son: Profesional del Departamento Técnico y de Información Económica, profesional líder del Departamento de Programación e Inflación y Jefe de la Sección de Estadística. Banco de la República, Bogotá, Colombia.

Nowcasting of food price variation via mixed frequency models

Julián Alonso Cárdenas-Cárdenas²

Edgar Caicedo-García

Eliana R. González Molano

jcardeca@banrep.gov.co

ecaicega@banrep.gov.co

egonzamo@banrep.gov.co

Banco de la República

The opinions of this paper are the sole responsibility of it's authors and do not reflect the views of Banco de la República or its Junta Directiva

Abstract:

The behavior of food prices is a big issue for the monetary authority, due to the high volatility as well as the big weight it has in the CPI basket and because it reacts temporarily to supply shocks, such as climate conditions, what makes difficult the task of keeping total inflation around the target. Thus, it is needed to count with more accurate and timely forecasts of food inflation for the short run in order to guide the macroeconomic model for monetary policy and help the authority in the decision making process. For that purpose, in this document we apply a methodology that combines information of different frequencies (MIDAS) to produce forecasts for food inflation. In particular, information about food prices at a very disaggregate level and an indicator for food supply, which are available in a weekly basis, may help to generate a more accurate nowcast of total food inflation and its components: perishable and processed food. Compared to a naïve nowcast generated every week as the weighted average change of food prices taken by SIPSA, the results show an improvement in the nowcast, generated by the mixed frequency data models that includes not only high frequency variables as explanatory but also some other determinants of food price changes such as unemployment, climate conditions and international commodity prices. Thus, MIDAS models are a promising alternative to generate forecasts in the short run.

Key words: Food inflation, nowcasting, mixed frequency models, inflation targeting, climate conditions.

JEL codes: C32, C51, C53, E31, E52.

² In order are: Professional, Statistics Division, Technical and Economic Information Department, Leading Professional, Inflation and Programming Department and Chief Statistics Division, Technical and Economic Information Department, Banco de la República, Bogotá, Colombia.

1. Introducción

Desde comienzos del nuevo siglo se implementó en Colombia el esquema de inflación objetivo, con resultados exitosos en el control de la inflación y reduciendo la incertidumbre sobre la evolución cercana de los precios al consumidor, facilitando la adopción de políticas económicas y sociales a diferentes horizontes de planificación³. En este esquema, el uso de modelos de simulación de política monetaria ha tenido un rol relevante en la generación de pronósticos de largo plazo y como herramientas para ayudar a la toma de decisiones por parte de las autoridades pertinentes. No obstante, la evidencia empírica ha mostrado que, para mejorar la capacidad de pronósticos de largo plazo, estos modelos de simulación deben ser condicionados con estimaciones para el corto plazo de variables relevantes como el IPC de alimentos (también con proyecciones de corto plazo de otras canastas del IPC y del crecimiento económico). Para tal efecto se diseñan e implementan modelos satélites, lo cual hace parte del objeto de este documento.

Para capturar, evaluar y estimar adecuadamente la dinámica de la inflación al consumidor, los bancos centrales suelen dividir el IPC en varios componentes, por ejemplo: en bienes y servicios transables y no transables excluyendo por lo general los alimentos y algunos servicios que son regulados por el Estado. Esto lo hacen con el propósito de aislar los componentes del IPC que son afectados por choques de oferta que tienden a disiparse con relativa rapidez en el tiempo o aquellos grupos sujetos a cambios en las regulaciones, sobre los cuales la política monetaria no tiene ninguna injerencia. Esto permite tener una medida del IPC cuya tendencia podría tener mayor relación con el comportamiento de la demanda agregada.

Para capturar la naturaleza volátil de los precios de los alimentos, cuya dinámica suele estar afectada por choques con frecuencias e intensidades diferentes como perturbaciones climáticas (El Niño), paros de transporte o agrarios, precios internacionales y de la tasa de cambio, entre otros, el equipo técnico del Banco de la República ha elaborado un conjunto de modelos para pronosticar su comportamiento futuro. Sin embargo, ninguno de estos modelos se construye utilizando simultáneamente información de diferente frecuencia (diario, semanal, mensual), cuya ventaja radica en incorporar información disponible con diferente periodicidad a la del IPC para capturar la dinámica compleja de los precios de los alimentos.

De esta manera, incluir, aprovechar y adicionar nueva información de frecuencias mixtas, brinda la oportunidad de mejorar el conjunto de modelos y pronósticos para capturar el cambio en los precios de los alimentos. Además, desde el enfoque del nowcasting, tal oportunidad y disponibilidad de información reciente permite, en teoría, mejorar el pronóstico al reducir el error debido a la propia combinación de todos estos factores.

En consecuencia, el principal objetivo de este documento es presentar una herramienta que permita conocer oportunamente nowcasting (En Banbura et al. (2010) lo definen como una abreviación del inglés -now y forecasting- y se usa para predecir el pasado reciente, el presente y el futuro cercano) del ajuste mensual de precios de los alimentos introduciendo en el espectro de este tipo de problemas los modelos de frecuencias mixtas y evaluando su viabilidad por medio de la comparación con otro tipo de reportes y/o modelos de pronóstico. El aporte fundamental de este trabajo consiste en la introducción de variables de frecuencia semanal para predecir el cambio mensual de los precios de los alimentos, lo cual, ya sea por disponibilidad de datos o problemas en la aplicación, no ha sido muy implementado en la literatura.

³ Al respecto véase: López E., Vargas H., y Rodríguez N. (2016).

Haciendo la revisión bibliográfica correspondiente se encontraron casos en los cuales se tiene información financiera de frecuencia alta para predecir una amplia serie de variables objetivo, por ejemplo, Baumeister et al. (2015) recogen una diversidad amplia de información diaria y semanal con la cual intenta predecir los precios internacionales del petróleo dando resultados aceptables, en Alper et al. (2008) se utilizan modelos MIDAS con inclusión de frecuencia diaria para predecir los precios de ciertas acciones en economías emergentes y en Yang & Zhang (2014) se analiza por medios de modelos MIDAS de frecuencia semanal el impacto del sentimiento inversor en el retorno de las acciones, encontrando significancia en el efecto. Para el caso específico de intentar predecir valores futuros de la inflación, Breitung & Roling (2015) utilizan información diaria con el objetivo de predecir la inflación en Alemania usando modelos MIDAS de forma no paramétrica, Monteforte & Moretti (2013), encuentran como ciertas variables financieras tienen poder predictivo para predecir la inflación en la zona Euro haciendo uso de modelos MIDAS y comparando su desempeño con modelos VAR usuales. Por último, en Latinoamérica se pueden encontrar ejemplo como en Libonatti (2018) el cual utiliza índices de precios financieros obtenidos por medio web para evaluar su poder predictivo sobre la inflación Core en la Argentina bajo un marco de análisis de modelos MIDAS, encontrando que la inclusión de esta información diaria logra ayudar a explicar ciertos comportamientos en esta inflación.

Dentro de la literatura internacional cabe resaltar el artículo pionero de Ghysels et al. (2004), seguido de Ghysels et al. (2006), cuyo principal aporte fue introducir el uso de variables de diferente frecuencia para aproximar una mejor estimación de la inflación. Desde entonces, diversos autores utilizaron y adaptaron estos modelos de frecuencias mixtas para elaborar pronósticos de diferentes tópicos económicos, destacándose especialmente aquellos relacionados con el pronóstico de variables de crecimiento económico como el PIB (véase la amplia revisión de métodos que se hace en Camacho et al. (2013)).

Este documento contiene cinco secciones de las cuales la primera es esta introducción. A continuación, se encuentra una descripción de las variables analizadas y las candidatas a variables explicativas dentro de los modelos de frecuencias mixtas. En la siguiente sección, se describen los modelos que se utilizan y luego el procedimiento de selección y estimación de los modelos empleados para la evaluación de pronósticos. Posteriormente, se muestran los resultados de la evaluación de pronósticos y comparaciones entre los diferentes modelos. Por último, se presentan algunas conclusiones sobre la aplicación de modelos de frecuencias mixtas para generar nowcast de la variación mensual del precio de los alimentos.

2. Descripción de la información

Puntualmente, el objetivo de la metodología que se expone aquí es producir nowcasting de los cambios mensuales de los precios de los alimentos, excluyendo las comidas fuera del hogar. Este último segmento de alimentos no se tiene en cuenta por dos razones: la primera, porque ya no hace parte del grupo de alimentos en el nuevo IPC vigente desde el 2019; la segunda, porque los precios de las comidas fuera del hogar incorporan un alto componente de servicios (arriendos, salarios y servicios públicos), sin que los alimentos lideren la formación de precios de las comidas fuera del hogar. Adicionalmente, también se evaluarán las sub-canastas de alimentos perecederos y alimentos procesados con el fin de conocer si el pronóstico se refina al hacer esta segmentación. Definido el “objetivo” del procedimiento a continuación se hace una pequeña descripción de cada uno de los grupos de variables.

Desde el año 2001 el grupo de Estadística del Departamento Técnico y de Información Económica del Banco de la República (DTIE), viene presentado una estimación semanal de la variación mensual de alimentos utilizando la información publicada semanalmente por el Sistema de Información de Precios del Sector Agropecuario (SIPSA) del Departamento Administrativo Nacional de Estadística (DANE). Además de lo anterior, se recopila precios de otras fuentes de información para algunos productos que por su alta ponderación y volatilidad requieren de diversas fuentes de información (por ejemplo, papa y carne). Este seguimiento de fuentes y levantamiento de datos ha permitido disponer de información particular y oportuna para los diferentes componentes de la canasta de alimentos, permitiendo la elaboración de pronósticos por subcanastas, estimaciones que van actualizándose y cambiando a medida que se presenta nueva información disponible semanalmente.

Disponer de información de precios con alta frecuencia, como la señalada arriba, se aprovechó para actualizar cada semana los precios por componentes de alimentos y para construir estimaciones de los diferentes segmentos de este grupo. Dichos pronósticos que utilizan información de alta frecuencia resultaron ser un muy buen proxy del dato de cambio mensual de precio de los alimentos que reporta el propio DANE cada mes. Este dato de frecuencia semanal es el primero de los predictores, dado que es el dato usado actualmente como “nowcasting” de la variación mensual de los precios de los alimentos, a la vez sirve para contrastar sus resultados con otros modelos posteriormente.

Para la alimentación de nuestro modelo a evaluar se introducirá la información semanal de precios de los alimentos del SIPSA, tomando las variaciones semanales de un índice construido para cada una de estas series de precios de los alimentos.

El SIPSA también publica datos acerca del abastecimiento de alimentos en las principales centrales de abasto del país. Esta información, a priori, se concibe como una parte importante en la evolución de los precios de los alimentos, porque se espera que los precios de este segmento del IPC se muevan de acuerdo con la oferta de los productos agrícolas. Esta variable de oferta, también de frecuencia semanal, se convierte en una introducción interesante al desarrollo del modelo, debido a su naturaleza como variable de una frecuencia más alta al objetivo (inflación mensual) y la información importante que pueda contener. Esta variable no se había tenido en cuenta en un modelo de pronóstico de precios o de oferta en la literatura nacional.

Otra dimensión de variables que se espera que sean de alto impacto dentro del modelo son los precios de commodities, dentro de los cuales se seleccionó el petróleo WTI, debido a su estrecha relación con la producción agrícola y el transporte de los mismos. Además, se considera el índice de precios de los commodities de alimentos CRB (Commodity Research Bureau), el cual representa un indicador global de los mercados de commodities. Para el segmento de alimentos tiene en cuenta los precios internacionales de: Cacao, café, maíz, jugo de naranja, soya y azúcar. Por lo tanto, este índice se tuvo en cuenta como indicador importante de los precios agrícolas primarios en el exterior. Aunque estas variables tienen periodicidad diaria, por facilidad en el ajuste de los modelos y en el análisis se incluyeron con frecuencia mensual, tomando el valor promedio en el mes de análisis. Adicionalmente se introdujo la brecha de la tasa de desempleo, para tratar de emular el comportamiento de una curva de Phillips dentro de la modelación que se va a proponer. Esta

variable se puede segmentar como rural y nacional permitiendo encontrar relaciones más específicas, según la variable objetivo.

Finalmente, debido a la estrecha relación del clima con el precio de los alimentos, se introdujo como proxy de precipitaciones un indicador que contabiliza el nivel de los embalses a lo largo del país. Se entiende que este indicador es una buena aproximación del nivel de precipitaciones a nivel nacional, el cual influye positivamente sobre la productividad agropecuaria y la oferta de alimentos. Tal variable se obtiene por medio de un procedimiento de consolidación de información del Sistema de Información Energético Colombiano. Además, para complementar la información asociada al clima y para capturar el impacto de perturbaciones climáticas sobre los precios de los alimentos, que actualmente son más recurrentes e intensas, el modelo propuesto se alimentará con la información del Índice Oceánico del Niño (ONI). Este índice es utilizado por la Administración Nacional Oceánica y Atmosférica de Estados Unidos (NOAA por sus siglas en inglés) para identificar eventos cálidos (El Niño) y fríos (La Niña) en el océano Pacífico tropical. De esta manera, se puede capturar la presencia o ausencia de un evento El Niño⁴.

Un análisis de correlaciones cruzadas fue efectuado para conocer la estructura de la relación funcional y de rezagos que podían tener cada una de estas variables candidatas a explicativas dentro de los posibles modelos a construir. Lo anterior se resume en el siguiente cuadro.

Cuadro 1 – Análisis de correlaciones cruzadas entre la variable respuesta y las variables candidatas a explicativas

<i>Variables explicativas</i>	Correlación con Inflación Total Alimentos (sin CFH)
Precios SIPSA	Contemporánea MUY fuerte
GAP Desempleo Nacional	Sin relación aparente
GAP Desempleo Rural	2-5 rezago tienen relación
Índice oceánico del niño	1-3 rezago tienen relación
Nivel de embalses	Sin relación aparente
Abastecimiento SIPSA	Relación inversa contemporánea MUY fuerte
Índice alimentos Commodities	Relación contemporánea fuerte
Precios Petróleo WTI	4-5-6 rezago tiene una relación inversa
<i>Variables explicativas</i>	Correlación con Inflación de alimentos Perecederos
Precios SIPSA	Contemporánea MUY fuerte y a un rezago

⁴ Recientes estudios -Grothe et., al. (2019) y Wang et. al. (2019)- coinciden en señalar que los eventos climáticos como El Niño son más intensos y frecuentes que en el pasado. Este hecho estilizado aumenta la importancia de incorporar estas perturbaciones climáticas en los modelos de pronóstico de los precios de los alimentos.

GAP Desempleo Nacional	Sin relación aparente
GAP Desempleo Rural	Relación contemporánea y desde el sexto al décimo rezago
Índice oceánico del niño	Contemporánea y rezagos 1,2 tienen relación
Nivel de embalses	Contemporánea inversa parece haber relación
Abastecimiento SIPSA	Relación inversa contemporánea MUY fuerte
Índice alimentos Comoditas	Relación contemporánea fuerte
Precios Petróleo WTI	4-5-6 rezago tiene una relación inversa especialmente al 5
<i>Variables explicativas</i>	Correlación con Inflación de alimentos Procesados y otros
Precios SIPSA	Contemporánea MUY fuerte y a 1-2 rezago
GAP Desempleo Nacional	Sin relación aparente
GAP Desempleo Rural	Contemporánea además relación y desde el primero hasta el sexto parece fuerte
Índice oceánico del niño	Rezagos 3 al 8 tienen relación
Nivel de embalses	Sin relación aparente
Abastecimiento SIPSA	Relación inversa a un rezago fuerte
Índice alimentos Commodities	Sin relación aparente
Precios Petróleo WTI	Del 1 al 4 rezago tiene una relación inversa, significativamente fuerte en el 3 rezago

Teniendo en cuenta todas las relaciones funcionales descritas en el Cuadro 1 y siendo nuestro punto de partida para la especificación de los modelos, en la siguiente sección se explica la metodología que se va a utilizar para realizar la modelación.

3. Modelos teóricos

Para la especificación de los modelos, este documento se basa en la metodología de modelos de frecuencias mixtas introducida por Ghysels et al. (2004). Se tienen en cuenta los siguientes tres modelos:

3.1. Ecuación Bridge

Sea y_t la variable de baja frecuencia (variable objetivo, para nuestro caso la variación mensual de alimentos), y sea x_t las variables de alta frecuencia (semanales, de frecuencia mayor a la variable objetivo), las cuales son agregadas a baja frecuencia de acuerdo a su naturaleza mediante algún método. Sea t_m que indica el tiempo en meses y t_s que indica el tiempo en semanas. El modelo bridge tiene la siguiente estructura:

$$y_{t_m} = \alpha + \sum_{i=1}^j \beta_i(L)x_{it_m} + U_{t_m} \quad (1)$$

Donde $t_m = 1, \dots, T$. $\beta_i(L)$ son los polinomios asociados al operador de rezago L , tal que:

$$\beta_i(L) = \beta_{0i} + \beta_{1i}L + \dots + \beta_{pi}L^p \quad (2)$$

Uno para cada variable de alta frecuencia y finalmente U_{t_m} es un error i.i.d. bajo los supuestos usuales.

A pesar de que la frecuencia del modelo sea mensual el modelo se puede actualizar semanalmente y diversos pronósticos pueden ser construidos a partir de él. Esto dependerá de la forma de agregación de los datos semanales, la cual se define como:

$$x_{it_m} = x_{it_s} + 2x_{it_{s-1/4}} + 3x_{it_{s-2/4}} + 4x_{it_{s-3/4}} + 3x_{it_{s-1}} + 2x_{it_{s-5/4}} + x_{it_{s-6/4}} \quad (3)$$

Esta estructura hace depender a la información de un mes en particular de las 4 semanas que lo componen y las tres semanas del mes anterior, teniendo en cuenta lo que se evidenció en el análisis de correlaciones.

Debido a la estructura lineal de la ecuación del modelo y la agregación propuesta de las variables de alta frecuencia, el modelo es lineal, por lo tanto, este puede ser fácilmente estimado vía OLS bajo los supuestos usuales. Este tipo de estimación claramente facilita la interpretación de los resultados del modelo y permite hacer pronósticos de manera directa y sencilla.

3.2. Modelo MIDAS

Los modelos MIDAS (Ghysels et al. (2004)) es la alternativa ideal para trabajar con variables en diferentes frecuencias. El desarrollo de estos modelos está basado en los conocidos modelos de rezagos distribuidos (Almon (1965)), los cuales ayudan a prevenir una sobreparametrización del modelo soportado en un ajuste parsimonioso por medio de diversos métodos de restricción, como polinomios de menor grado o mínimos cuadrados no lineales. El modelo MIDAS en su versión general se compone por:

$$y_t - \alpha_1 y_{t-1} - \dots - \alpha_p y_{t-p} = \sum_{i=0}^k \sum_{j=0}^{l_i} \beta_j^{(i)} x_{tm_i-j}^{(i)} + \varepsilon_t \quad (4)$$

Donde tm_i define la indexación de las diferentes frecuencias para la variable i a través del tiempo. El principal inconveniente de esta ecuación radica en la cantidad elevada de parámetros, para lo cual Ghysels et al. (2004) propone una restricción paramétrica que asegura la reducción del espacio de parámetros a estimar de la siguiente forma

$$\beta_j^{(i)} = f_i(\gamma_i, j), j = 0, \dots, l_i \quad (5)$$

$$\gamma_i = (\gamma_1^{(i)}, \dots, \gamma_{q_i}^{(i)}) \quad (6)$$

Así, al reducir el número de parámetros por medio de la función f_i se llega a una forma similar a un modelo de rezagos distribuidos, por lo cual hablamos de una estimación/modelo no lineal. Detalles más específicos sobre el proceso de estimación y las formas que puede tomar la función f pueden encontrarse en Ghysels et al. (2004), pero para efectos prácticos se usa la descripción propuesta en Ghysels et al. (2016) donde se desarrollan los modelos MIDAS, para el software estadístico R mediante el paquete *midas_r*.

Para hacer un poco más de énfasis sobre la manera en que se reduce el número de parámetros, consideremos la parte derecha de la ecuación (4) como:

$$B(L)^T X_{t,0} = \sum_{i=0}^k \sum_{j=0}^{l_i} \beta_j^{(i)} x_{tm_i-j}^{(i)} = \sum_{i=0}^k \sum_{j=0}^q \lambda_j^{(i)} \tilde{x}_{tm_i-j}^{(i)} \quad (7)$$

Con $l > q$, algunos de los rezagos de las variables de baja frecuencia y sus parámetros son agregados como:

$$\tilde{x}_{t-r}^{(i)} := x_{t-r}^{(i)}(\delta_{i,r}) = \sum_{s=1}^{m_i} w_r^{(i)}(\delta_{i,r}; s) x_{(t-1-r)m_i+s}^{(i)} \quad (8)$$

Tal agregación depende de la función $w_r^{(i)}(\delta_{i,r}; s)$, la cual tiene como parámetros los diferentes $\delta_{i,r}$, los cuales varían en función de cada y cada frecuencia. Adicionalmente, para la identificación de los parámetros $\lambda_j^{(i)}$ se tiene que satisfacer que:

$$\sum_{s=1}^{m_i-1} w_r^{(i)}(\delta_{i,r}; s) = 1 \quad (9)$$

Por lo cual el tratamiento que se define para cumplir con esta restricción es imponer que para todo i, r :

$$w_r^{(i)}(\delta_{i,r}; s) = \frac{\psi_r^{(i)}(\delta_{i,r}; s)}{\sum_{j=1}^{m_i} \psi_r^{(i)}(\delta_{i,r}; j)} \quad (10)$$

Para que efectivamente se construya una función ponderadora, se tiene que cumplir que la función ψ sea no negativa. En el contexto de modelos de rezagos distribuidos, presentado por Almon (1965), se deriva la función exponencial de rezagos polinomiales, la cual se puede ver como:

$$\psi(\delta; j) = \exp\left(\sum_{j=1}^p \delta_j s^j\right) \quad (11)$$

Esta es la función de pesos que se va a usar en este procedimiento, junto con la versión que no es normalizada por medio de la ecuación (10). Una discusión más amplia de las diferentes funciones de ponderación puede ser encontrada en la bibliografía de este escrito.

3.3. Modelo U-MIDAS

Un caso particular del modelo anterior, motivado por la ausencia de restricciones al momento de estimar los parámetros es el modelo U-MIDAS. Esto teniendo en cuenta unas condiciones bajo las cuales el modelo pueda ser identificable. Tales condiciones se comentan en Marcellino and Schumacher (2015). La idea general del modelo mantiene la estructura del modelo MIDAS, pero dada la linealidad, explicada en la falta de restricciones, el modelo puede ser estimado directamente por OLS, esto teniendo en cuenta que los regresores de las variables de alta frecuencia serán estimados para cada una de estas subfrecuencias, es decir, un parámetro para cada número de semana si nos vamos al caso específico que se trabaja en este documento.

4. Procedimiento de selección de modelos

Debido a la disponibilidad de la información y de todas las series descritas en una de las secciones anteriores, se cuenta con información desde febrero de 2012 hasta julio de 2019.

La metodología de este documento está basada en las recomendaciones prácticas que se señalan en Camacho et al. (2013), con las modificaciones pertinentes para el caso de estudio, por lo tanto, la estrategia para la elección de los modelos y la evaluación de pronósticos es la siguiente, para el caso de la Ecuación Bridge y el Modelo U-MIDAS:

1. Se ajusta el primer modelo con la información desde febrero 2012 hasta junio 2018 haciendo una elección de las variables explicativas (y sus rezagos) dependiendo de la variable objetivo (variación mensual del precio de los alimentos, alimentos perecederos o alimentos procesados), pero siempre manteniendo dentro del modelo un componente autoregresivo de orden 1, además de las variables de alta frecuencia (Información de precios y abastecimiento del SIPSA).
2. Debido a que dentro de cada mes se tienen 4 semanas de información, se quiere también evaluar si el pronóstico mejora a medida que hay más información disponible (no solo el propio rendimiento del modelo con toda la información disponible), por lo cual, se predice un paso adelante con la información de solo 1 semana, luego 2, 3 y 4 semanas, dando como resultado 4 pronósticos por mes. Cuando se tiene menos de 4 semanas de información, habrá datos faltantes dentro de las variables de alta frecuencia, por lo tanto, estos se imputan a partir de pronósticos generados por modelos ARIMA de selección automática.

3. El paso 2 se repite mes a mes hasta julio 2019 manteniendo fija la forma funcional de los modelos, pero actualizando la información de sus parámetros con la nueva información agregada y prediciendo siempre un paso adelante.

4.1. Adecuación automática del MIDAS

El modelo MIDAS, debido a su forma funcional puede tener muchas formas de ajuste, teniendo en cuenta las diferentes funciones de restricción posibles sobre las variables de alta frecuencia. Con el fin de utilizar óptimamente estas bondades se sigue el siguiente procedimiento para la elección del modelo más adecuado para generar el nowcasting:

1. Se ajustan una serie de modelos con la información desde febrero 2012 hasta junio 2018, haciendo una elección fija de las variables explicativas de baja frecuencia (y sus rezagos) dependiendo de la variable objetivo (variación mensual del precio de los alimentos, perecederos o procesados), pero siempre manteniendo dentro del modelo un componente auto regresivo de orden 1 para la variable respuesta, además de las variables de alta frecuencia (Información de precios y abastecimiento del SIPSA).
 - a. La función de ponderación para cada variable de alta frecuencia se escoge entre la función exponencial de rezagos polinomiales descrita anteriormente y su versión no normalizada. Además, se tiene en cuenta varias opciones para los rezagos de las variables de alta frecuencia, esto se hace seleccionando el modelo con menor Error Cuadrado Medio (MSE) dentro de muestra para el conjunto de todos los modelos estimados bajo las condiciones dadas, es decir, teniendo en cuenta la combinación de número de rezagos para las variables de baja y alta frecuencia y la función de ponderación. Esto asegura, en cada paso (tanto de baja como alta frecuencia), tener el mejor modelo predictor dentro de muestra para hacer la estimación un paso adelante, por lo menos en consideración de estas dos funciones.
2. Finalmente se aplican los pasos 2 y 3 en la misma forma que se hace para la Ecuación Bridge y el U-MIDAS, pero con la diferencia de que en cada paso (semana) se hace la selección del mejor modelo según los criterios del punto anterior.

4.2. Formas funcionales

4.2.1. Ecuación Bridge y U-MIDAS

Los modelos, que presentaron un mejor ajuste dentro de muestra y finalmente estimados para las tres variables de interés están dados por:

$$\pi_t^A = \pi_{t-1}^A + s_{t_s}^A + a_{t_s}^A + a_{(t_s-1)}^A + oni_{t-1} \quad (7)$$

$$\pi_t^{Per} = \pi_{t-1}^{Per} + s_{t_s}^{Per} + a_{t_s}^{Per} + emb_{t-1} \quad (8)$$

$$\pi_t^{Pro} = \pi_{t-1}^{Pro} + s_{t_s}^{Pro} + a_{t_s}^{Pro} + a_{(t_s-2)}^{Pro} + wti_{t-1} \quad (9)$$

Tanto para el modelo para la variación mensual del precio de los alimentos (π_t^A) como para el de alimentos procesados (π_t^{Pro}) se utiliza un rezago adicional del abastecimiento a_{t_s} , lo que para la ecuación bridge implica tener en cuenta hasta 12 semanas hacia atrás de la información de abastecimiento y para el U-MIDAS 8 parámetros solo para esta variable explicativa. Las variables adicionales, de baja frecuencia, en estas formas funcionales se mantuvieron igual en cada tipo de

modelo para asegurar una comparación más homogénea. Debido a la cantidad elevada de parámetros en el U-MIDAS se dificulta anexar más variables explicativas de frecuencia baja.

4.2.2. MIDAS

Para el modelo MIDAS se seleccionaron las siguientes variables en cada modelo:

Alimentos Total

- Precios de petróleo WTI (Variaciones mensuales), wti_t
- ONI (Rezago a un mes) El Índice Niño Oceánico (ONI en inglés) es una medida de la condición de El Niño-Oscilación del Sur (ENSO) y sus fases cálidas (El Niño) y fría (La Niña) en el Pacífico ecuatorial central, oni_t .
- Precios semanales fuente SIPSA, con selección automática de rezago entre $s=0, \dots, 11$, donde s son los rezagos de la variable de alta frecuencia (semanal), s_{t_s} .
- Abastecimiento semanal con selección automática de rezago entre $s=0, \dots, 11$, donde s son los rezagos de la variable de alta frecuencia (semanal), a_{t_s} .

Alimentos Perecederos

- Índice de precios de alimentos de commodities (Variaciones mensuales)
- Tasa de embalses (Variaciones mensuales), emb_t en rezago 2
- Gap tasa de desempleo rural (Rezago a dos meses)
- Precios semanales fuente SIPSA de alimentos perecederos con selección automática de rezago entre $s=0, \dots, 7$, donde s son los rezagos de la variable de alta frecuencia (semanal), s_{t_s} .
- Abastecimiento semanal de alimentos perecederos con selección automática de rezago entre $s=0, \dots, 7$, donde s son los rezagos de la variable de alta frecuencia (semanal), a_{t_s} .

Alimentos Procesados

- Índice de precios de alimentos de commodities (Variaciones mensuales)
- Precios de petróleo WTI (Variaciones mensuales), wti_t
- Precios semanales fuente SIPSA de alimentos procesados con selección automática de rezago entre $s=3, \dots, 11$, donde s son los rezagos de la variable de alta frecuencia (semanal).
- Abastecimiento semanal de alimentos procesados con selección automática de rezago entre $s=3, \dots, 11$, donde s son los rezagos de la variable de alta frecuencia (semanal).

El modelo elegido para predecir será el mejor dentro de todas las combinaciones de los modelos, usando las dos opciones de funciones ponderadoras, con los rezagos de las variables de alta frecuencia que sea mejor *dentro de muestra* de acuerdo al criterio del MSE, así en cada paso.

5. Resultados

La referencia de comparación es el valor estimado semanalmente de acuerdo al informe de precios fuente SIPSA, que se ha venido utilizando como insumo de nowcasting de la variación mensual de los precios de los alimentos. Para hacer una comparación más objetiva, se extrajo el valor estimado de cada informe semanal y se comparó con la predicción mensual de cada semana obtenida de los modelos, para el periodo de agosto de 2018 a julio de 2019 para el total de alimentos, perecederos y procesados.

Para las tres variables de interés, se puede ver como sistemáticamente los nowcasting generados a partir de los modelos estimados son mejores que el valor estimado derivado del informe semanal de SIPSA. Adicionalmente, como los pronósticos mejoran a medida que hay disponible nueva información, en la mayoría de los casos, es decir, cada paso adelante semanalmente la predicción mejora, todo esto basado en el MSE de la predicción un paso adelante. Los resultados se presentan en cuadros donde se muestra el MSE para cada uno de los casos mencionados comenzando con la canasta total de alimentos, seguido por perecederos y terminando con procesados, marcando las principales características de la predicción en cada uno. En los anexos se encuentran gráficas donde se puede observar más detalladamente el desarrollo de tales pronósticos.

Alimentos canasta total (perecederos y procesados)

Los errores cuadrados medios (MSE) asociados a los modelos evaluados para el caso de la canasta total de alimentos son presentados en la Tabla 2, estableciendo que cada uno de los modelos mejoran sensiblemente los resultados del pronóstico del informe semanal derivado del SIPSA, lo que indica que las variables seleccionadas de uno u otro modo si aportan información relevante al momento de generar pronósticos. El modelo más apropiado para este caso es el modelo MIDAS, que sistemáticamente produce menor MSE.

Cuadro 2 – MSE modelos canasta total de alimentos

MSE	Semana 1	Semana 2	Semana 3	Semana 4
MIDAS	0,0876	0,0523	0,1177	0,0750
BRIDGE	0,0956	0,0832	0,1053	0,0903
UMIDAS	0,1029	0,0718	0,085	0,09
INFORME	0,2361	0,2017	0,2140	0,19

Alimentos Perecederos

Debido a la naturaleza volátil de los alimentos perecederos, por la cantidad de fenómenos que pueden llegar a afectarlo, los errores asociados al pronóstico de esta canasta en particular son más amplios que los presentados para los otros dos casos, aun así, respecto al informe semanal derivado de la información de SIPSA, los modelos producen mejores pronósticos. Aunque en este caso no se aprecian mejoras en el pronóstico por la inclusión de la información nueva semana a semana, el modelo con mejor desempeño es de nuevo el modelo MIDAS.

Cuadro 3 – MSE modelos para alimentos perecederos

MSE	Semana 1	Semana 2	Semana 3	Semana 4
MIDAS	0,9988	1,4735	1,7601	1,7089

BRIDGE	2,2058	2,1748	2,3429	2,2989
UMIDAS	1,6916	2,1644	2,0290	1,9131
INFORME	3,1945	2,9595	3,6357	2,3204

Alimentos Procesados

En este caso sucede lo contrario al caso de la canasta de perecederos, los alimentos procesados, por su propia definición, tienden a ser menos volátiles y, por lo tanto, la captura de sus movimientos parece ser más sencilla. Este hecho se confirma en los valores más bajos del MSE asociados a estos modelos. Nuevamente la metodología MIDAS es la más adecuada.

Cuadro 4 – MSE modelos para alimentos procesados

MSE	Semana 1	Semana 2	Semana 3	Semana 4
MIDAS	0,0607	0,0921	0,0717	0,0613
BRIDGE	0,0452	0,0458	0,046	0,0464
UMIDAS	0,0509	0,0494	0,0520	0,0563
INFORME	0,1274	0,1143	0,0880	0,1174

6. Conclusiones

Los modelos para capturar las variaciones mensuales de los precios de alimentos usando frecuencias mixtas, sistemáticamente mejoran el pronóstico o nowcast respecto a la estimación del cambio mensual de los precios de los alimentos derivada del informe semanal de SIPSA, esto indica que la selección de variables explicativas es adecuada para las tres variables analizadas (alimentos total, perecederos y procesados) y los tres tipos de modelos propuestos. Esto es un resultado importante porque dichos modelos proveen un mejor pronóstico para los tres agregados de alimentos a pesar de no llegar al nivel de desagregación o detalle (subclase de gasto) al que se puede llegar con el informe SIPSA.

También es importante recalcar que estos modelos pueden tener un margen de mejora desde varios puntos. Primero, a medida que se tenga nueva información disponible el pronóstico parece ir mejorando, sin embargo, dada la pequeña muestra usada para la evaluación de pronósticos todavía no es un resultado concluyente. Segundo, estos modelos permiten generar pronósticos para grandes agregados de alimentos, lo cual permite fácilmente la adecuación a los cambios de la nueva canasta del IPC 2018 para Colombia y sus nuevos componentes, mientras que con la estimación semanal a partir del informe de SIPSA, se debe buscar información para los nuevos rubros que conforman la canasta de alimentos. Adicionalmente, se hace necesario buscar más fuentes de información que ayuden a capturar los precios no solo de las centrales mayoristas sino de todas las fuentes posibles de comercialización de alimentos, un ejemplo de esto es la recolección de información por medio de web scrapping de precios online en grandes superficies. Tercero, al combinar información de diferentes frecuencias, estos modelos permiten contar con pronósticos oportunos que pueden irse actualizando a medida que se dispone de nueva información. Por todo lo anterior, los modelos de frecuencias mixtas son una alternativa para generar pronósticos y

algunas de las características que se aplicaron aquí, como la selección automática de modelos, son bastante prometedoras para el objetivo planteado.

Los modelos de frecuencias mixtas son relativamente nuevos en la literatura debido al auge en la disponibilidad de la información, como se señala en Camacho et al. (2013), por lo tanto, un campo de estudio interesante es desarrollar nuevas metodologías y aplicar las existentes en este tipo de problemas y evaluar sus resultados. Adicionalmente, al estimarse un modelo que incluye diferentes determinantes del cambio en los precios, se puede evaluar la contribución de cada uno de estos factores al error de pronóstico.

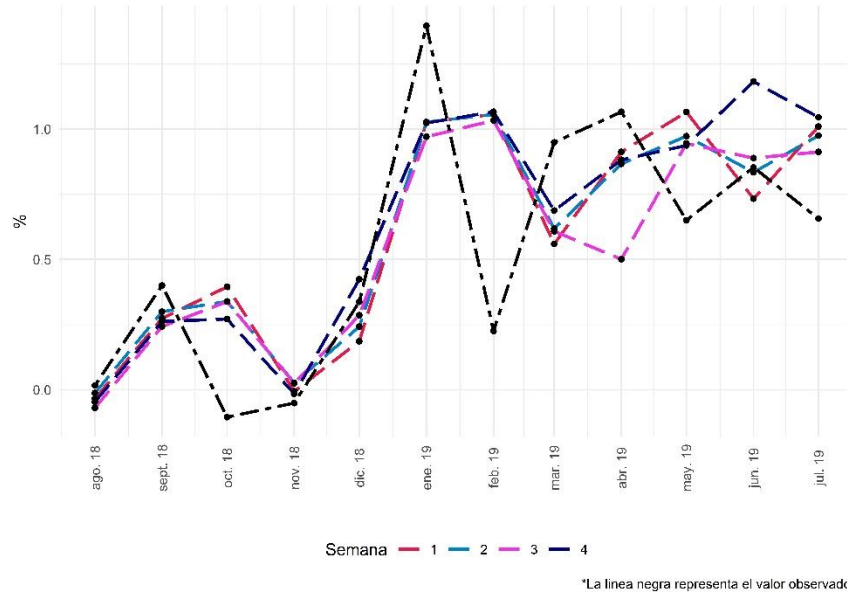
BIBLIOGRAFÍA

- Alper, C. E., Fendoglu, S., & Saltoglu, B. (2008). Forecasting stock market volatilities using MIDAS regressions: An application to the emerging markets.
- Almon, S. (1965). The distributed lag between capital appropriations and expenditures. *Econometrica: Journal of the Econometric Society*, 178-196.
- Banbura, M., Giannone, D., & Reichlin, L. (2010). Nowcasting.
- Baumeister, C., Guérin, P., & Kilian, L. (2015). Do high-frequency financial data help forecast oil prices? The MIDAS touch at work. *International Journal of Forecasting*, 31(2), 238-252.
- Breitung, J., & Roling, C. (2015). Forecasting inflation rates using daily data: A nonparametric MIDAS approach. *Journal of Forecasting*, 34(7), 588-603.
- Camacho, M., Pérez-Quirós, G., & Poncela, P. (2013). Short-term forecasting for empirical economists: A survey of the recently proposed algorithms. *Foundations and Trends® in Econometrics*, 6(2), 101-161.
- Foroni, C., Marcellino, M., and Schumacher, C. (2015). Unrestricted mixed data sampling (MIDAS): MIDAS regressions with unrestricted lag polynomials. *Journal of the Royal Statistical Society A*, 178(1):57-82.
- Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models.
- Ghysels, E., Kvedaras, V., & Zemlys, V. (2016). Mixed frequency data sampling regression models: the R package midasr. *Journal of statistical software*, 1-35.
- Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1), 53-90.
- Monteforte, L., & Moretti, G. (2013). Real-time forecasts of inflation: The role of financial variables. *Journal of Forecasting*, 32(1), 51-61.
- Libonatti, L. (2017). *MIDAS modeling for core inflation forecasting* (No. 72). Economic Research Working Papers.
- López, Enrique; Vargas, Hernando; Rodríguez, Norberto (2016), "La estrategia de la inflación objetivo en Colombia. Una visión histórica", Borradores de Economía No. 952, Banco de la República, Bogotá, Colombia.
- Grothe, Pamela R. et. al (2019), "Enhanced El Niño-Southern Oscillation variability in recent decades". *Geophysical Research Letter*. (<https://doi.org/10.1029/2019GL083906>)
- Wangt Bin et. al. (2019). "Historical change of El Niño properties sheds light on future changes of extreme El Niño". *Proceedings of the National Academy of Sciences of United States of America* (<https://doi.org/10.1029/2019GL083906>).
- Yang, C., & Zhang, R. (2014). Does mixed-frequency investor sentiment impact stock returns? Based on the empirical study of MIDAS regression model. *Applied Economics*, 46(9), 966-972.

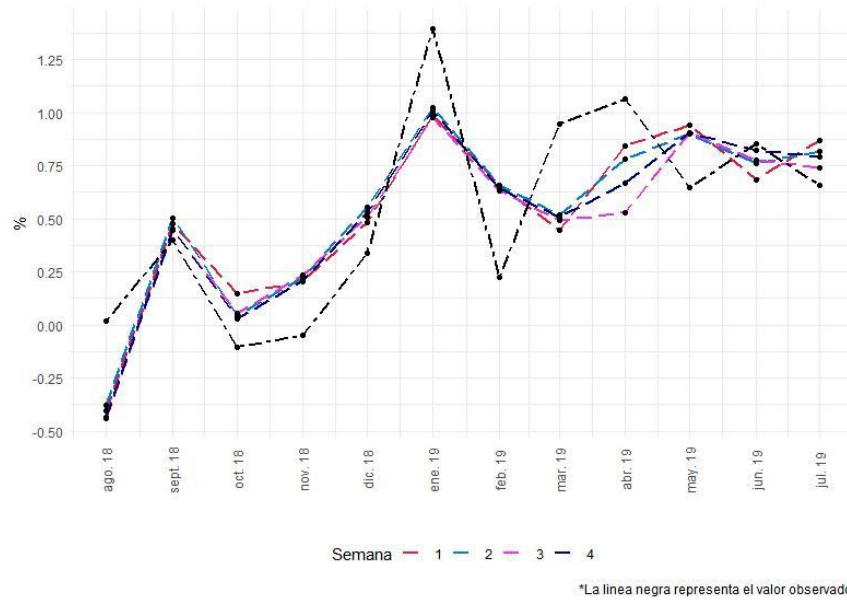
Anexos

A. Gráficos

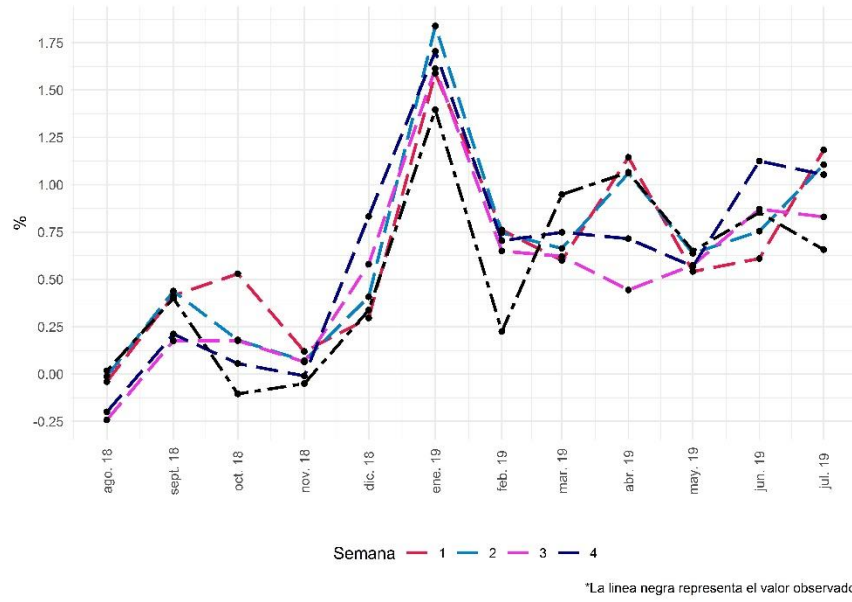
Gráfica 1 – Pronósticos modelo MIDAS – Total alimentos



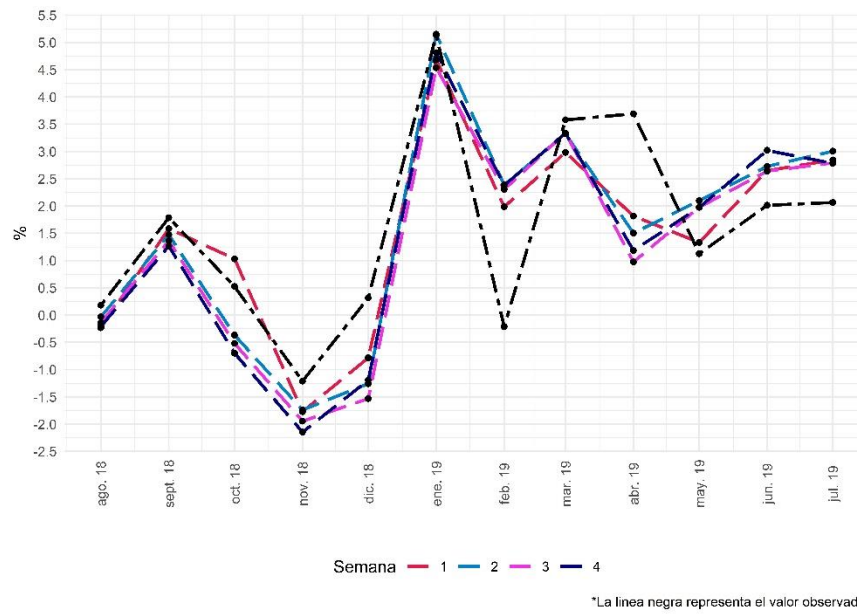
Gráfica 2 – Pronósticos modelo Bridge – Total alimentos



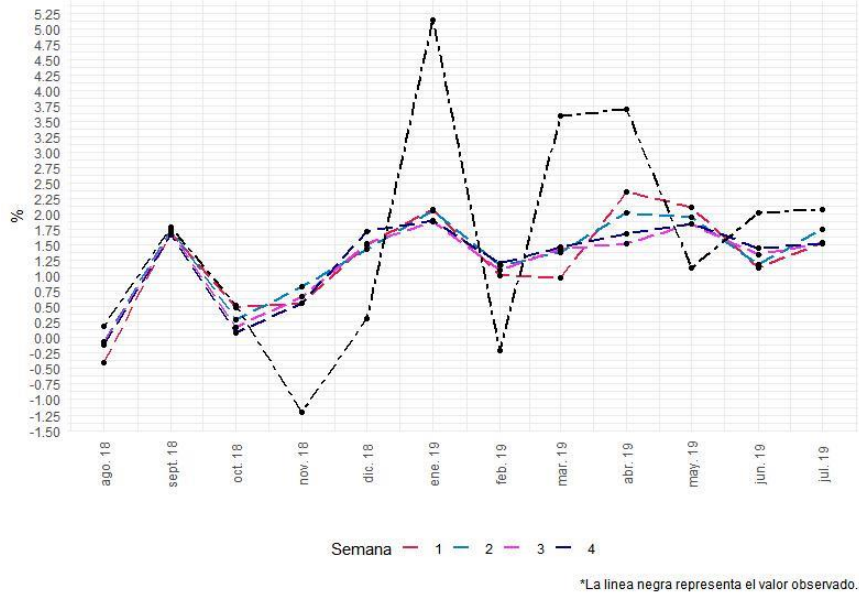
Gráfica 3 – Pronósticos modelo UMIDAS – Total alimentos



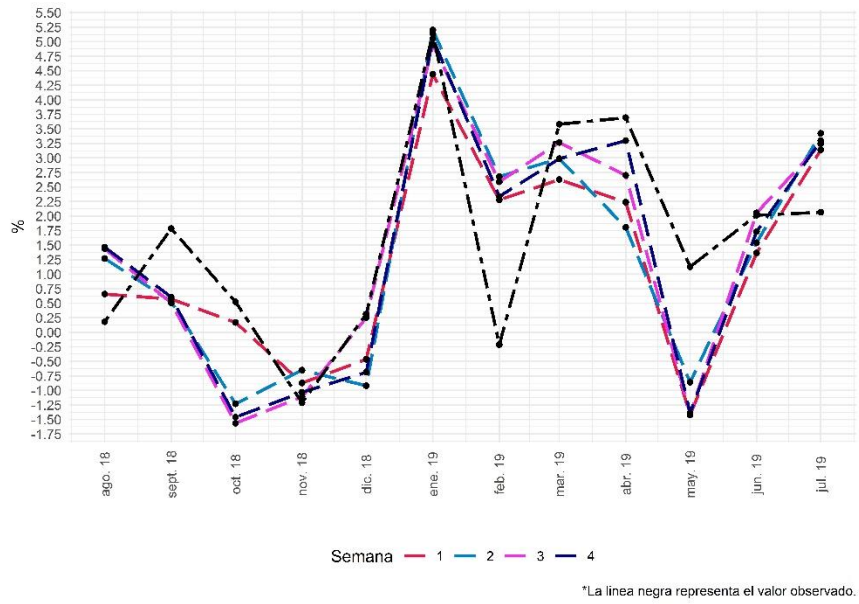
Gráfica 4 – Pronósticos modelo MIDAS – Alimentos perecederos



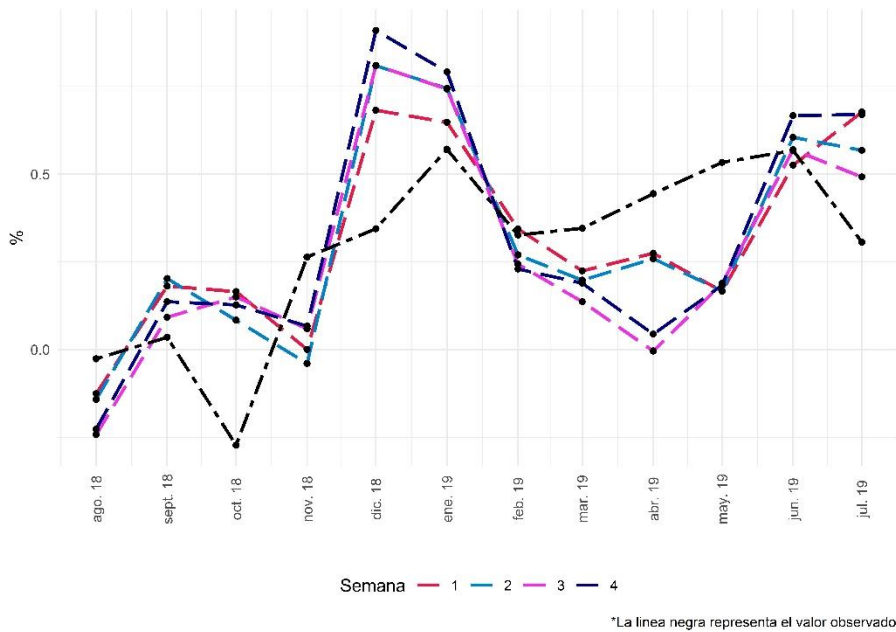
Gráfica 5 – Pronósticos modelo Bridge – Alimentos perecederos



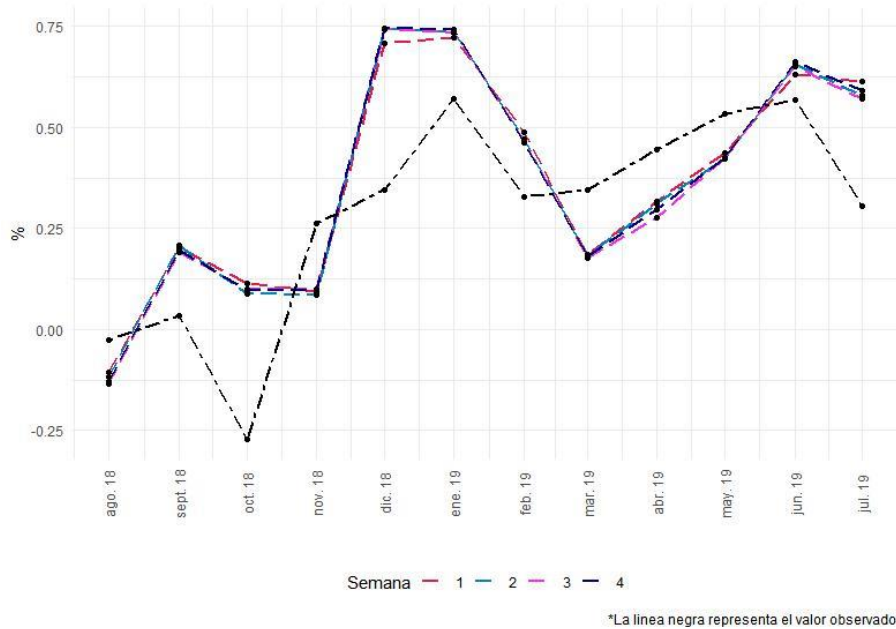
Gráfica 6 – Pronósticos modelo UMIDAS – Alimentos percederos



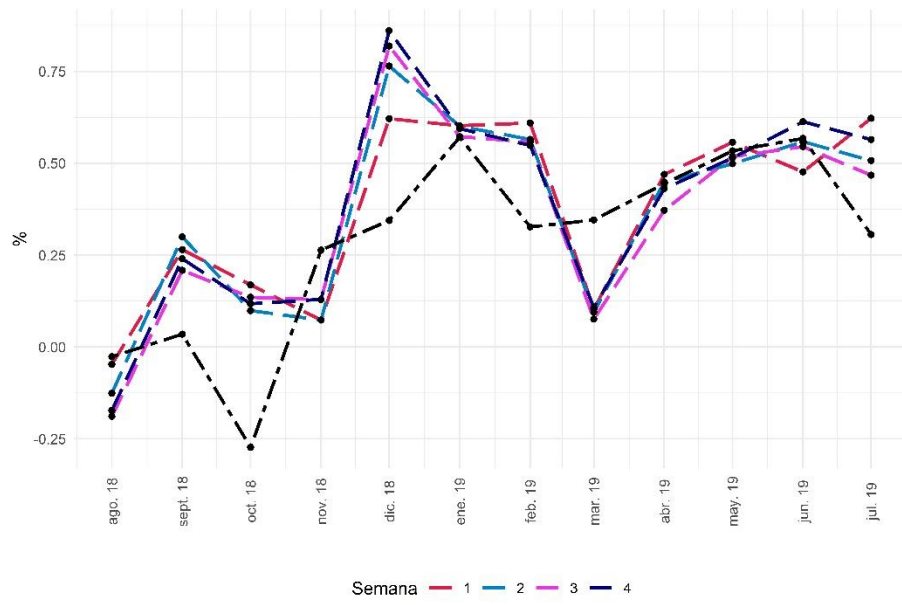
Gráfica 7 – Pronósticos modelo MIDAS – Procesados



Gráfica 8 – Pronósticos modelo Bridge - Procesados



Gráfica 9 – Pronósticos modelo UMIDAS - Procesados



*La línea negra representa el valor observado.

