
Propuesta de un índice SIPSA y su relación con la inflación de alimentos para el caso colombiano: evidencia empírica

Proposal for a SIPSA index and its relation to food inflation for the Colombian case: empirical evidence

Sébastien Lozano Forero^a
sebastien.lozano@libertadores.edu.co

Juan Camilo Santana Contreras^b
jcsantanac@libertadores.edu.co

Resumen

La inflación de alimentos es un agente macroeconómico de notable impacto en el costo de vida de los ciudadanos y en el acercamiento a la toma de decisiones de las entidades que inciden en política monetaria. Por tanto, resulta de interés tener la capacidad de, con razonable grado de certeza, obtener información relevante de la variación en los precios de los alimentos. Por otro lado, el SPSA (dependencia del DANE) es el ente gubernamental encargado de recopilar los precios con los que se comercializan los productos en las centrales de abasto del país. En el presente trabajo, se aborda el problema de aportar a la predicción de la inflación de alimentos a partir de un índice SPSA propuesto por los autores. Se abordan tres metodologías para tal fin. A saber: modelos auto regresivos integrados de medias móviles estacionales con variables exógenas (SARIMAX), modelos vectoriales estructurales auto regresivos (SVAR) y modelos vectoriales de corrección de errores (VECM). Nuestros resultados indican que las variaciones del SPSA explican alrededor del 40% de la variabilidad de la inflación de alimentos, además de la relación en el corto y largo plazo. La creación y el uso de tales indicadores, pueden convertirse en herramientas ideales para la planeación y la gestión de estrategias de inversión de portafolio con miras a la tecnificación de las propuestas de inversión en el mercado de capitales.

Palabras clave: Índice SPSA; inflación de alimentos; SARIMAX; VAR; VEC.

Abstract

Food inflation is a macroeconomic agent of notable impact on the cost of living of citizens and in the approach to decision-making of entities that have an impact on monetary policy. Therefore, it is of interest to have the ability, with some degree of certainty, to obtain relevant information on the variation in food prices. On the other hand, The SPSA (DANE unit) is the government entity responsible for collecting the prices with which the products are sold in the country's supply centers. In this paper, the problem of contributing to the prediction of food inflation based on a SPSA index proposed by the authors, is addressed. Three methodologies are used for this purpose. Namely: integrated auto-regressive average mean with exogenous variables models (SARIMAX), vectorial auto-regressive structural models (SVAR) and vector error correction model (VECM). Our results indicate the variations of SPSA explain about 40 % of the variability of food inflation, in addition to the relationship in the short and long term. The creation and use of indicator stories can become ideal tools for the planning and management of portfolio investment strategies with a view to the technology of investment proposals in the capital market.

Keywords: SPSA Index; food inflation; SARIMAX; VAR; VEC.

^aFundación Universitaria los Libertadores

^bFundación Universitaria los Libertadores

1. Introducción

Según el Banco de la República de Colombia, "la inflación se define como el aumento sustancial, persistente del nivel general de precios a través del tiempo. La inflación puede promover el crecimiento económico, básicamente, en aquellas economías que muestran bajos niveles de ingreso y, por tanto, una baja capacidad para generar ahorro que permita financiar inversiones, incrementar el empleo y, en general, la actividad productiva". Por lo mismo, el entendimiento de la inflación resulta de interés para los agentes económicos que constantemente interactúan en una sociedad.

Para el caso de Colombia, el Departamento Administrativo Nacional de Estadística (DANE) propone un índice de inflación que se desagrega en 9 grupos de gasto principal que representa las dinámicas de gasto de un hogar promedio. Esta canasta de consumo representada en estos grupos, se divide en otros subgrupos y unidades de gasto que detalla los precios de los productos básicos que adquieren en promedio los hogares en Colombia. Es importante destacar que el DANE actualiza esta canasta cada 10 años. Dichos grupos en los que se desagrega el índice de inflación son los siguientes: alimentos (que supone una contribución del 28.22% en el índice global de inflación), vivienda (30.09%), vestuario (5.17%), salud (2.44%), educación (6.41%), diversión (2.43%), transporte (15.19%), comunicaciones (3.72%) y otros gastos (6.35%). Es decir, se cuenta con un índice particular de inflación (índice de variación de precios) para cada uno de los agentes anteriormente mencionados. Dichos índices pueden ser consultados en SIPSA (2018).

Por lo anterior y debido al interés de los agentes del mercado financiero por conocer la trayectoria de corto y mediano plazo de la inflación; y por ende de la inflación de alimentos, el interés de este artículo es proponer un índice líder, basado en la información dispuesta por la agencia SIPSA (a dicho índice, en adelante, será llamado de *índice SIPSA*) que permita una explicación aceptable de la inflación de alimentos oficial del DANE y al mismo tiempo proporcione los escenarios ideales para el pronóstico de la inflación total. La importancia de tal propuesta radica en lo sensible que resulta la variación de los precios de los alimentos para la población colombiana por tratarse de un elemento de primera necesidad. Adicionalmente, resultaría de interés poder generar pronósticos razonablemente acertados para la variación de los precios de los alimentos, para que el mercado pueda tomar sus propias acciones regulatorias.

El DANE reporta, semanalmente, cuatro niveles de inflación de alimentos. éstas son: ingreso bajo, ingreso medio, ingreso alto e ingreso total; para este trabajo consideramos la variación en los ingresos totales que, a su vez, puede ser desagregada en varios tipos de gasto. Los agentes contributivos a la inflación de alimentos son: Verdura/Hortaliza (con una contribución a la inflación de alimentos de 6%), fruta (4%), tubérculos (3%), grano/cereal (12%), huevos/lácteos (15%), carnes (17%), pescado (2%), procesados (11%) y comidas fuera de casa (30%). Dicha ponderación puede ser calculada basada en la información consignada en IPC (2018).

Las siguientes son las partes compositoras del trabajo: En la sección 2 se presentan las metodologías estadísticas que serán utilizadas para encontrar una relación de dependencia y en el tiempo entre el índice de inflación de alimentos y el índice SIPSA. Serán presentados los modelos tradicionales de la metodología Box-Jenkins y los modelos tradicionales de series de tiempo multivariadas. En la sección 3 se presenta la construcción del índice SIPSA que será utilizado en el presente trabajo. En la sección 4 se presentan los resultados obtenidos. Se presentarán resultados de la metodología univariada, un modelo SARIMAX que utiliza el índice SIPSA para predecir la inflación de alimentos; y de la metodología multivariada, un modelo SVAR para establecer la relación en el tiempo entre las variables consideradas y un modelo VEC para encontrar la relación de equilibrio en las mismas. Finalmente, en la sección 5 se presentan las principales conclusiones de este trabajo y algunas ideas para futuras investigaciones utilizando las ideas aquí expuestas.

Soporte computacional

El análisis estadístico y todos los gráficos presentados fueron realizados en el ambiente de programación libre *R* en su versión 3.5.0 para la plataforma Windows, con ayuda del editor *Rstudio* en su versión 1.1.442. Este lenguaje de programación fue creado por Ross Ihaka y Robert Gentleman en la Universidad de Auckland.

La elaboración de este documento fue hecho integralmente en el sistema tipográfico \LaTeX que, básicamente, es una serie de macros y rutinas basadas en el sistema \TeX . Este lenguaje fue creado por Donald Knuth en 1984 en la Universidad de Stanford, USA. Como plataforma fue usado el dominio www.sharelatex.com

2. Marco metodológico

De forma general, es de resaltar que serán empleados tres modelos para dar respuesta a tres preguntas centrales: ¿Cómo incide el índice SIPSA propuesto para el pronóstico de la variación en los precios de los alimentos?, ¿Cómo es la relación en el corto plazo entre el índice SIPSA propuesto y la variación en los precios de los alimentos? y ¿Cómo es la relación en el largo plazo entre el índice SIPSA propuesto y la variación en los precios de los alimentos?. A continuación se presentan los modelos empleados en el abordaje de los anteriores interrogantes.

2.1. Modelo SARIMAX

Los modelos SARIMAX (auto regresivos integrados de medias móviles estacional con variables explicativas) son una estructura matemática que contiene todas las bondades de un modelo SARIMA (véase Guerrero (2003)) y adicionalmente, pueden capturar información sobre variables exógenas que permiten a entender y pronosticar la variable de interés. Para una única variable de interés y , y una única variable independiente x . El modelo que describe su relación es:

$$y_t = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \beta_0 x_{t-1} + \beta_1 x_{t-1} + \dots + \beta_r x_{t-r} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (1)$$

donde α (media general del proceso), β_i (parámetros que indexan los valores contemporáneos y rezagados de la variable exógena), ϕ_j (parámetros que indexan el componente auto-regresivo del modelo) y θ_k (parámetros que indexan el componente de medias móviles del modelo); con $i = 0, \dots, r$, $j = 1, \dots, p$, $k = 1, \dots, q$ son parámetros fijados y $\epsilon_t \sim RB(0, \sigma^2)$.

Finalmente, el modelo (1) puede ser escrito como:

$$\phi(B)\Phi(B^s)(1-B)^d(1-B^s)^D y_t = \theta(B)\Theta(B^s)\epsilon_t + \sum_{j=1}^r \beta_j x_{t-j}$$

2.2. Modelo VAR

Los modelos vectoriales auto regresivos, conocidos como modelos VAR, son modelos en los que se asume estacionariedad en la serie de tiempo. Su principal aplicación se da en los análisis de series multivariadas y en análisis macroeconómicos y fueron desarrollados por Sims hacia la década de los ochenta. Estos modelos permiten definir a todas las variables como variables endógenas dado que las consideran funciones lineales de sus propios valores rezagados y también de los valores rezagados de las variables adicionales que componen el modelo. Este modelo tiene ciertas limitaciones al no tener

en cuenta problemas como la heterocedasticidad, cambios estructurales en las variables estimadas o relacionales no lineales entre las mismas variable. En general, los detalles del uso de este tipo de modelos, pueden ser consultados en Pfaff (2008), Tsay (2013) y (Lutkepohl2005).

El modelo vectorial autor regresivo de orden p , VAR(p) es una estructura que da cuenta del comportamiento de varias series de tiempo en una estructura unificada y que generaliza a los modelos auto regresivos (AR(p)) univariados. Puede escribirse, en su forma más general como:

$$\mathbf{y}_t = \mathbf{v} + A_1 \mathbf{y}_{t-1} + \dots + A_p \mathbf{y}_{t-p} + \mathbf{u}_t \quad (2)$$

donde $\mathbf{y}_t = (y_{1t}, \dots, y_{Kt})^\top$ es un vector aleatorio de dimensión ($K \times 1$), A_i son matrices fijas de coeficientes de dimensión ($K \times K$), $\mathbf{v} = (v_1, \dots, v_K)^\top$ es el vector de interceptos de dimensión ($K \times 1$) (que permite la posibilidad de tener un media diferente a cero). Finalmente, $\mathbf{u}_t = (u_{1t}, \dots, u_{Kt})$ es ruido blanco K -dimensional o proceso de innovación. Es decir, $E(\mathbf{u}_t) = \mathbf{0}$, $E(\mathbf{u}_t^\top \mathbf{u}_t) = \Sigma$ $E(\mathbf{u}_t^\top \mathbf{u}_s) = 0$ para $s \neq t$. La matriz de covarianzas Σ se asume no singular.

Definiendo el operador de rezago polinomial $A(L) = I_K - BA_1 - \dots - B^p A_p$, el proceso dado en la ecuación 6 puede definirse como

$$A(L)\mathbf{y}_t = \mathbf{v} + \mathbf{u}_t \quad (3)$$

2.3. Modelo SVAR

El modelo Vectorial Auto Regresivo Estructural es una generalización del modelo VAR dado en la ecuación (3). Consiste en la introducción de las matrices A y/o B en el modelo VAR y recoge la relación contemporánea existente entre el conjunto de variables.

$$A\mathbf{y}_t = A\mathbf{v} + A_1^* \mathbf{y}_{t-1} + \dots + A_p^* \mathbf{y}_{t-p} + B\mathbf{u}_t \quad (4)$$

donde $A_i^* = A \times A_i$ para $i = 1, \dots, p$. Los modelos SVAR se utilizan para identificar los choques y rastrearlos a través de herramientas como el análisis de impulso a la respuesta o análisis de descomposición de la varianza del pronóstico imponiendo ciertas restricciones en las matrices A y/o B . En este caso, los residuos del modelo están dados por $\mathbf{u}_t = A^{-1}B\mathbf{u}_t$ y su matriz de varianzas y covarianzas está dada por $\Sigma_u = A^{-1}BB^\top A^{-1\top}$. Más detalles sobre estos modelos pueden ser encontrados en Holtz (1988) y Asteriou (2011)

2.4. Modelo VEC

Los modelos de Vectores de Corrección de Error (VEC) son modelos VAR restringidos que dentro de su especificación incluyen las relaciones de cointegración, es decir, su aplicación está centrada en series no estacionarias cointegradas, que guardan relaciones a largo plazo entre ellas.

Considerando el modelo VAR(p),

$$\mathbf{y}_t = +A_1 \mathbf{y}_{t-1} + \dots + A_p \mathbf{y}_{t-p} + \boldsymbol{\epsilon}_t.$$

Se define el modelo de corrección de errores como

$$\Delta \mathbf{y}_t = \boldsymbol{\alpha} \boldsymbol{\beta}^\top \mathbf{y}_{t-p} + \boldsymbol{\Gamma}_1 \Delta \mathbf{y}_{t-1} + \dots + \boldsymbol{\Gamma}_{p-1} \Delta \mathbf{y}_{t-p+1} + \mathbf{u}_t,$$

donde

$$\boldsymbol{\Gamma}_i = -(I - A_1 - \dots - A_i), \quad i = 1, \dots, p-1.$$

y

$$\Gamma = \alpha \beta^T = -(I - A_1 - \dots - A_p).$$

Las matrices Γ_i contienen los impactos acumulados de largo plazo. Es así que la especificación del VEC es sinónimo de de una relación a largo plazo. Más detalles sobre estos modelos pueden ser consultados en Rodríguez, et. al. (2019), Fernandez-Corugedo (2003) y Wei (2006).

2.5. Función de Impulso Respuesta

La función impulso respuesta (FIR) analiza la respuesta de las variables explicadas en el modelo para sus valores actuales y futuros, ante innovaciones (cambios) en una de ellas, asumiendo que dicho cambio tiende a desaparecer con el tiempo y que las demás variables permanecen constantes Cavaliere (2003). El cambio que se presenta en una variable perteneciente a un modelo VAR/VEC en un momento t , afecta directamente a la variable y a las demás variables que la acompañan debido a la estructura dinámica de estos modelos. Los vectores de las innovaciones por lo general están correlacionados entre sí, por lo que se presenta un componente común que al no poder atribuírsele a una variable en específico se atribuye a la variable que se considera más importante dentro del modelo VAR/VEC. Para más detalles, ver Lutkepohl 2018.

2.6. Análisis de Descomposición de Varianza

El análisis de descomposición de varianza (ADV) es un método para determinar cómo interactúan los sistemas de ecuaciones del modelo, permitiendo visualizar el impacto en el tiempo de las innovaciones de las variables del sistema sobre una de ellas (explicada), separando el porcentaje de variación del error de cada variable que es explicado por el cambio en cada ecuación. Si una variable explica en gran parte su varianza con sus mismas innovaciones, esta variable será de un nivel más exógeno que las demás, por lo que el ADV contribuye a establecer si el orden con el que se genera la FIR sea correcto. Para más detalles, ver Lutkepohl (2005) y Shahbaz (2012).

2.7. Evaluación de pronósticos

Para evaluar los pronósticos generados por los modelos propuestos se procederá de la siguiente manera. Al considerar una serie de tiempo con T observaciones, serán retiradas de ésta serie, las últimas n observaciones. Se genera el modelo con las $T - n$ observaciones y al generar los pronósticos, se evaluará la diferencia de los valores pronosticados y los valores verdaderos contenidos en la serie (los n puntos que fueron retirados). Considere $e_j = y_j - \hat{y}_j$, donde \hat{y}_j denota el pronóstico y_j .

La capacidad predictiva del modelo estimado puede ser evaluada utilizando las siguientes medidas de error, donde $k = T - n$

1. Error medio (ME):

$$ME = \frac{1}{n-1} \sum_{j=k+1}^T e_j$$

2. Raíz de error cuadrático medio (RMSE):

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=k+1}^T e_j^2}$$

3. Error absoluto medio (MAE):

$$MAE = \frac{1}{n} \sum_{j=k+1}^T |e_j|$$

4. Error medio porcentual (MPE)

$$MPE = \left(\frac{1}{n} \sum_{j=k+1}^T \frac{e_j}{y_j} \right)$$

5. Error absoluto medio porcentual (MAPE)

$$MAPE = \left(\frac{1}{n} \sum_{j=k+1}^T \frac{|e_j|}{y_j} \right)$$

6. Autocorrelación de los errores con rezago 1. (ACF1)
7. índice U de Theil de desigualdad.

3. Propuesta de índice SIPSA

Para la consolidación del índice SIPSA se procedió a la siguiente manera:

En SIPSA (2018), el SIPSA, dependencia del DANE, ha dispuesto para descarga del público las bases de datos que dan cuenta de los precios reportados por las centrales mayoritarias de abasto. Dichas bases de datos contienen la información semanal de los precios reportados para la venta de productos agrícolas en las principales centrales de abasto del país, desagregada por ciudad o departamento. Dichos precios se reportan a lo largo de ocho grupos alimentarios. A saber: verduras y hortalizas, frutas frescas, tubérculos y raíces y plátanos, granos y cereales, pescados, huevos y lácteos, carnes y finalmente, productos procesados. Las bases de datos contienen en promedio 4500 registros en donde se especifica, producto a producto, el precio de venta y su tendencia frente a la última semana. El sitio web disponibiliza un reporte (formato pdf, llamado *Boletín*) y la respectiva base de datos (formato xlsx, llamado *Anexo*). Las mencionadas bases de datos se encuentran desde el año 2012 hasta 2018 (año en que se realizó esta investigación), se debe descargar el archivo *Anexo* disponible para cada semana en el mencionado periodo. Se tienen en total 280 archivos que registran la actividad mencionada desde enero de 2013 hasta abril de 2018. Es de aclarar que, si bien se tienen reportes del SIPSA desde junio de 2012, solamente desde noviembre de dicho año se comienza la publicación de la base de datos de la respectiva semana.

Posteriormente, para cada producto se calculó el precio mediano (por ser una medida más robusta que la media) de venta que fue reportado en las principales centrales de abasto de todo el país. Una vez establecido el precio de venta mediano de cada producto, se calculó el valor mediano de estos precios para obtener un valor que da cuenta de cada una de las categorías alimentarias previamente descritas (es decir, verduras, frutas, etc.). Finalmente se obtuvo un consolidado semanal a partir de la ponderación, de acuerdo a la participación en la inflación de alimentos de cada uno de los valores calculados. Dichas ponderaciones es del 9.4% para verduras y hortalizas, 5% para frutas, 4.1% para tubérculos, raíces y plátanos, 18.6% para granos y cereales, 21.4% para huevos y lácteos, 23.7% para carnes, 3% para pescados y 15.87% para productos procesados.

De esta manera, se plantea el valor CS_{ij} como el consolidado de la semana i del mes j que se construye con los precios reportados por SIPSA, con el procedimiento indicado anteriormente, y las respectivas ponderaciones por grupo de alimentos de la siguiente manera:

$$\begin{aligned}
CS_{ij} = & 0.094 \times \text{mediana}(\text{VerduraYHortaliza}_i) + 0.050 \times \text{mediana}(\text{Fruta}_i) \\
& + 0.041 \times \text{mediana}(\text{Tubérculos}_i) + 0.186 \times \text{mediana}(\text{GranoYCereal}_i) \\
& + 0.214 \times \text{mediana}(\text{HuevosYLácteos}_i) + 0.237 \times \text{mediana}(\text{Carnes}_i) \\
& + 0.030 \times \text{mediana}(\text{Pescado}_i) + 0.148 \times \text{mediana}(\text{Procesados}_i).
\end{aligned} \tag{5}$$

Aquí, $\text{mediana}(\text{Categoría}_i)$ hace referencia a la mediana de los precios reportados para cada categoría (Verduras, fruta, etc), en todas las centrales de abasto del país de las que se tenga información, en la i -ésima semana de observación. Es decir, corresponde al i -ésimo informe semanal dado por la agencia SIPSA. Así, CS_{ij} contiene la información consolidada de lo reportado en una de las 280 bases de datos que fueron consultadas al momento de esta investigación.

Posteriormente, se define $CM_t = \text{Mediana}(CS_{1j}, \dots, CS_{kj})$ donde k es el número de semanas que componen el mes j . A su vez, CS_{ij} es el consolidado de la semana i del mes j que se construye con los precios reportados por la agencia SIPSA y las respectivas ponderaciones por grupo de alimentos. Finalmente, el índice SIPSA propuesto se obtiene como la variación de los resultados obtenidos semana a semana, con la metodología anteriormente descrita y que se calcula mediante la ecuación:

$$\text{SIPSA}_t = 100 \times \left(\frac{CM_t - CM_{t-1}}{CM_{t-1}} \right),$$

Finalmente, cabe aclarar que la construcción del índice SIPSA propuesto omite el rubro de 'Comidas fuera del Hogar' que incorpora el grupo de inflación de alimentos del DANE, y cuya ponderación representa el 30% del gasto que destinan los hogares promedio en alimentos. La investigación en busca de encontrar un proxy de este instrumento no fue posible debido a la ausencia de indicadores o informaciones de gasto relativos a este ítem, motivo por el cual se omitió dentro del análisis.

4. Resultados: Relación entre el índice propuesto y la inflación de alimentos

En la presente sección serán presentados los modelos en series de tiempo que fueron obtenidos a partir del análisis de los datos disponibles. A saber, variación en los precios de los alimentos y el índice SIPSA, propuesto en la sección 3.

El desempeño de los modelos propuestos será evaluado a través de una partición de la muestra en dos partes: una dentro de la muestra (o *in-sample*) que permitirá estimar la capacidad de ajuste del modelo y otra fuera de la muestra (o *out-of-sample*) que permitirá medir el desempeño en pronóstico de los modelos. Es decir, los datos serán divididos en dos partes: La primera parte de los mismos, que serán llamados de *in-sample*, contendrá los datos en el periodo de enero de 2013 hasta Junio de 2018. La segunda parte de los mismos, que serán llamados de *out-sample*, corresponden al periodo de Julio de 2018 hasta diciembre de 2018. Los modelos que se presentarán fueron calculados solamente con los datos *in-sample*. La idea general es que, en el entendido que el modelo resulta ser bueno, los pronósticos generados por estos modelos, deberán ser razonablemente parecidos a los valores verdaderos presentes en el *out-sample*.

4.1. Modelo SARIMAX

Para comenzar el análisis de los datos, lo primero es entrar a fase de identificación e inspección visual de las series. En la figura 1 se presentan las figuras de la variación en los precios de los alimentos y el índice

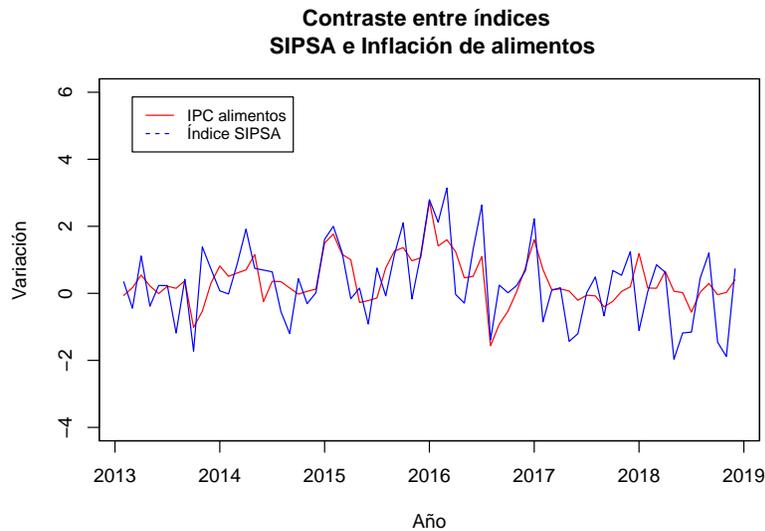


Figura 1: Interacción en el tiempo entre los índices considerados

SIPSA propuesto anteriormente. En general, el índice SIPSA propuesto parece dar buena cuenta de la variable de interés toda vez que “la persigue” en algunos de los picos y en la tendencia que presenta. De la misma forma, la correlación entre las variables de estudio es de 0.68.

Para la etapa de identificación, de acuerdo al esquema de Box-Jenkins, es necesario considerar los autocorrelogramas y autocorrelogramas parciales muestrales de los datos. La figura 2 presenta los datos con dos diferencias: en su componente estacional y en su componente no estacional. La prueba de Dickey-Fuller aumentada aporta un p -valor del 0.01 para estos datos. Por tanto procederemos con los datos con las diferenciaciones ya mencionadas ($d = 1, D = 1$).

El proceso metodológico de estimación paramétrica de Box & Jenkins llevó a la propuesta de 7 modelos de los que se revisaron varios criterios: Resultados en la prueba de autocorrelación serial, resultados en la prueba de normalidad de los residuos, comportamiento visual de los residuos, BIC y menor cantidad

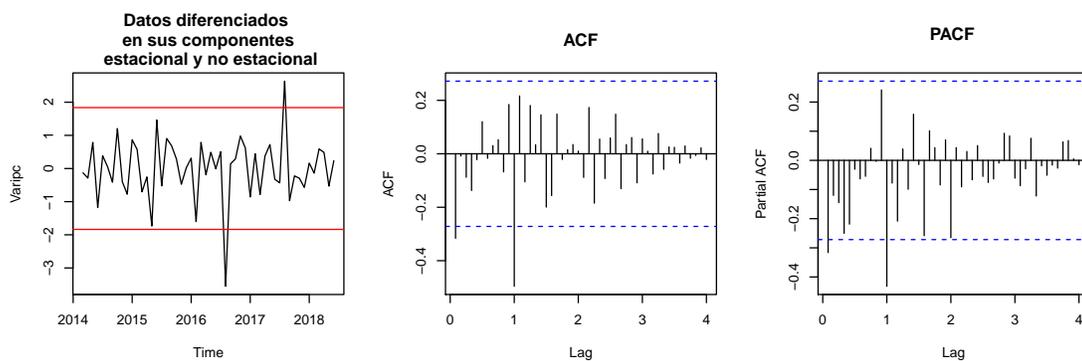


Figura 2: ACP y ACF para la serie IPC con diferencia en sus componentes estacional y no estacional

Parámetro	ϕ_1	ϕ_{10}	θ_1	θ_{13}	Φ_1	SIPSA _t	SIPSA _{t-1}
Estimación	-0.466	-0.261	-0.492	0.323	-0.457	0.300	0.157
Error Estándar	0.1147	0.1115	0.1357	0.1246	0.1130	0.0425	0.0434
t-valor	-4.06	-2.34	-3.62	2.60	-4.04	7.06	3.63

Tabla 1: Estimativas para el modelo SARIMAX(10, 1, 13) × (1, 1, 0)₁₂

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's
0.285	0.515	0.411	849	990	-0.558	1.24

Tabla 2: Medidas de error del modelo SARIMAX con out-sample

de parámetros (principio de parsimonia); se llegó a que el mejor modelo para describir y pronosticar la variación en los precios de los alimentos es un modelo SARIMAX(10, 1, 13) × (1, 1, 0)₁₂ que incorpora dos valores del índice SIPSA, a saber, el valor contemporáneo y un valor rezagado. Los parámetros del modelo seleccionado se dan en la tabla 1. La varianza estimada del modelo es 0.422 y el BIC calculado es 120.

En la figura 3 se presentan las figuras para el diagnóstico del modelo. Se presentan los datos originales vs. los datos ajustados y en general, se presenta un buen ajuste. Según el ACF y PACF de los residuos, no es aparente ningún problema de autocorrelación serial. Adicionalmente, el p -valor de la prueba de Ljung-Box es de 0.1686, lo que sugiere que con una significancia estadística del 5% no hay evidencia para rechazar la hipótesis de no correlación serial en los residuos. Finalmente, los residuos se configuran como ruido blanco, pero no como ruido blanco gaussiano pues el p -valor en la prueba de Jarque-Bera < 1%, sugiriendo, con una significancia del 5% que los residuos no se distribuyen normalmente.

Finalmente, para evaluar el modelo, de acuerdo a lo descrito anteriormente, se evaluarán los pronósticos del modelo para ver si son razonablemente cercanos a los datos contenidos en la muestra out-sample, en el sentido de la inclusión de los valores observados dentro de los intervalos de confianza del 95% de los valores pronosticados. En la figura 4 se presentan en rojo los pronósticos generados con el modelo dado en la tabla 1. Los datos originales observados se encuentren en rojo. Se observa que los valores estimados se encuentran razonablemente cerca a los valores originales de los datos. Adicionalmente, la tabla 2 presenta las medidas numéricas para evaluar los pronósticos.

De esta forma, el modelo propuesto puede ser utilizado para pronosticar la variación en los precios de los alimentos en un determinado mes a partir de la observación del índice SIPSA, propuesto en la sección 3, del valor contemporáneo (mismo mes) y el primer rezago (mes inmediatamente anterior).

4.2. Modelo SVAR

Para obtener el modelo SVAR respectivo, primero es necesario obtener un modelo VAR. Para tal fin, se comienza con la identificación de la relación que tienen las series consideradas. éstas son: variación en los precios de los alimentos (que notaremos como *Inf* e índice SIPSA propuesto (que notaremos como *SIPSA*). En la figura 5 se presenta el correlación cruzada de la variación en los precios de los alimentos y el índice SIPSA propuesto. Se puede observar que en los rezagos $k = 0$ y $k = -1$, el valor de la correlación es notablemente alto. Es decir, la variación en los precios de los alimentos está altamente correlacionada con los valores del índice SIPSA contemporáneo y un rezago hacia atrás. Lo anterior tiene sentido pues el modelo SARIMAX propuesto en la sección 4.1 incluía la misma dinámica.

Por lo tanto, visualmente hay una relación entre las variables de estudio en el tiempo. Se procedió a ajustar un modelo VAR(1) que incorpora la estructura estacional, la tendencia y el intercepto presente en los datos, con el fin de modelar tal comportamiento conjunto. El modelo VAR(1) ajustado está dado por:

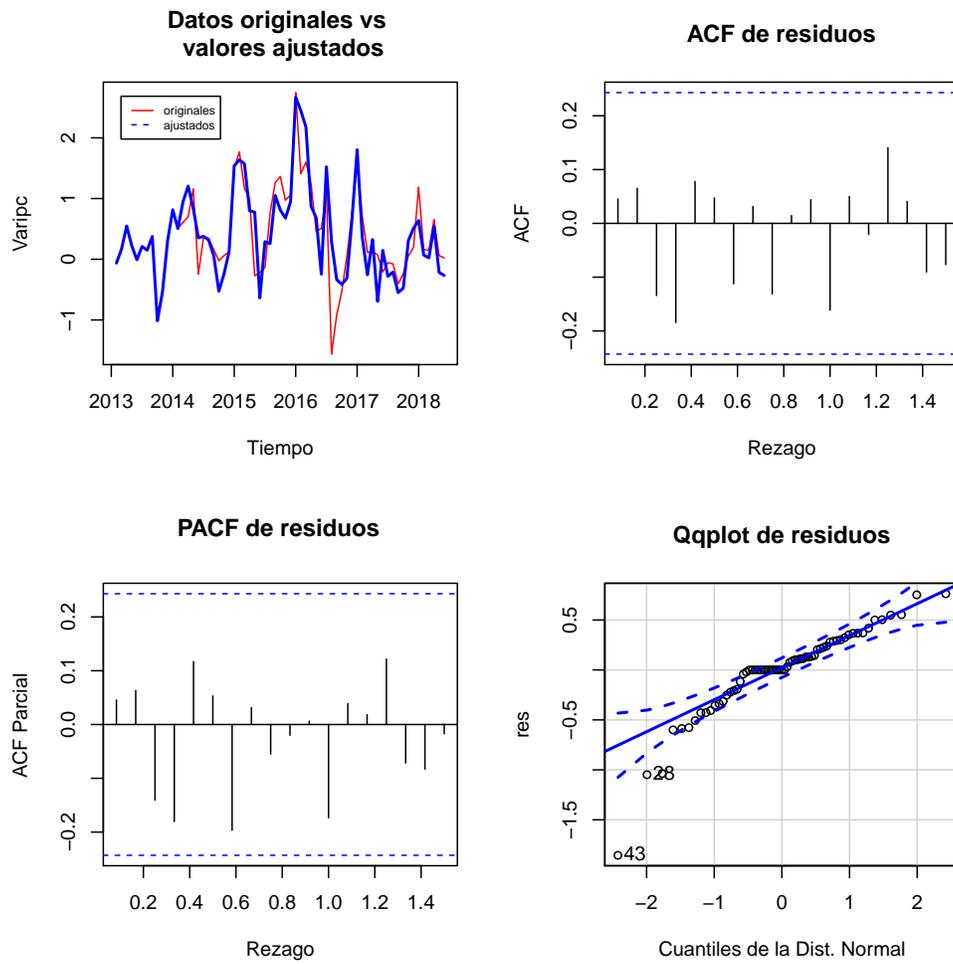


Figura 3: Validación del modelo dado en la tabla 1

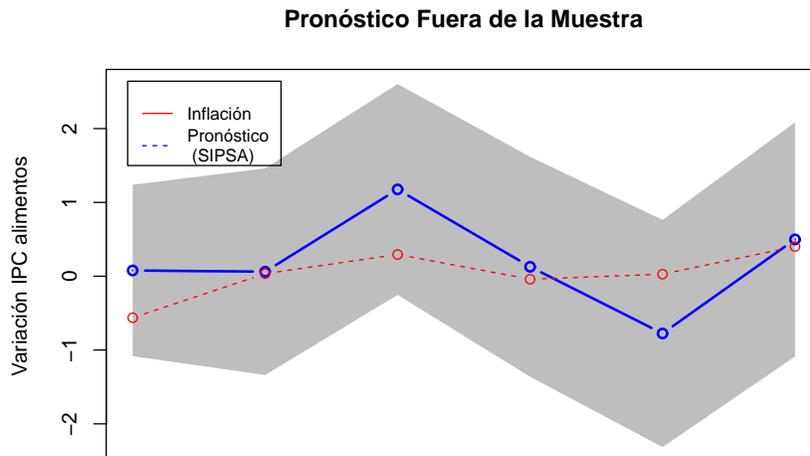


Figura 4: Pronóstico in-sample del modelo dado en la tabla 1. Fancharts con un 95% de confianza

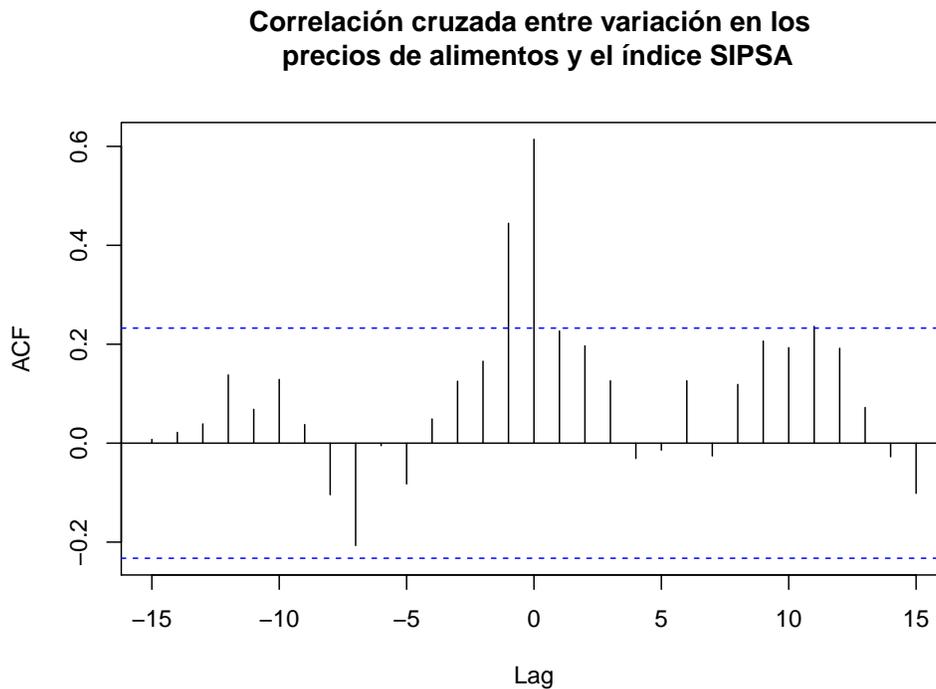


Figura 5: Correlación cruzada entre la variación en los precios de los alimentos y el índice SIPSA

i	1	2	3	4	5	6	7	8	9	10	11
SDinf $_i$	-1.1881	-1.055	-0.941	-1.384	-1.417	-0.991	-1.567	-1.131	-1.428	-1.147	-0.877
SDSIPSA $_i$	-0.930	-0.293	-0.590	-1.755	-1.210	-0.060	-1.597	-0.895	-0.655	-0.535	-0.200

Tabla 3: valores para las variables dummies que explican la estacionalidad de los datos

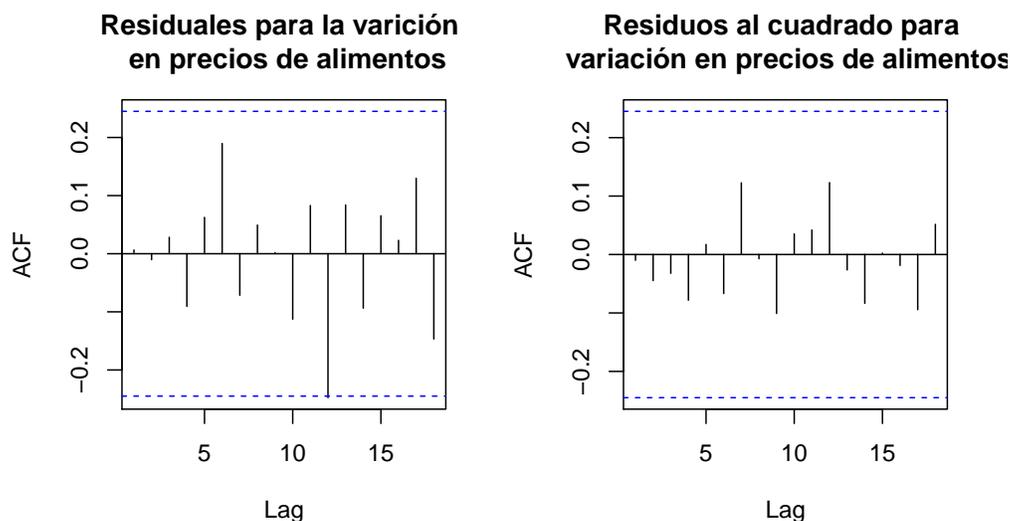


Figura 6: Correlaciones cruzadas para el modelo ajustado

$$\begin{cases} \text{Inf}_t &= 0.26475 - 0.00195t + 0.47304\text{Inf}_{t-1} + 0.06235\text{SISPA}_{t-1} + \sum_{i=1}^{11} \text{SDinf}_i \mathbf{1}_i, \\ \text{SIPSA}_t &= 0.34206 - 0.00464t + 0.44231\text{Inf}_{t-1} + -0.0003\text{SISPA}_{t-1} + \sum_{i=1}^{11} \text{SDSIPSA}_i \mathbf{1}_i \end{cases} \quad (6)$$

donde $\mathbf{1}_i$ es una variable dummy que toma el valor 1 cuando i representa un determinado mes (Enero, febrero, etc.) y 0 en otro caso. Las constantes dadas en la tabla 3 contiene la información del aporte de cada mes a la inflación y al SIPSA. Finalmente, la matriz varianzas y covarianzas de los residuos es de

$$\hat{\Sigma} = \begin{bmatrix} 0.280 & 0.307 \\ 0.307 & 1.118 \end{bmatrix}$$

Una vez establecido este modelo, se procede a validarlo. En las figuras 6 y 7 se presentan las funciones de correlación cruzadas. En general, el ajuste parece bueno, sin embargo llama la atención el pico que presenta el ACF cruzado de los residuos. Se aplicaron las pruebas de Portmanteau (H_0 : Residuos no correlacionados), Jarque-Bera multivariado (H_0 : Dist. Normal en los Residuos) y ARCH (H_0 : residuos homoscedásticos) para determinar que los residuos tengan las propiedades deseadas. Para las pruebas citadas, se obtuvieron p -valores de 0.8, < 0.01, 0.3, con lo cual, con un 5% de significancia, hay evidencia estadísticamente significativa para concluir que los residuos no presentan correlacionados serial, son homoscedásticos pero no son normales. Con lo cual se concluye que los residuos son ruido blanco no gaussiano. En la figura 8 se presentan los valores observados contra los valores ajustados por el modelo.

Una vez establecido el modelo VAR, se procede a hallar su forma estructural. ésta será determinada con

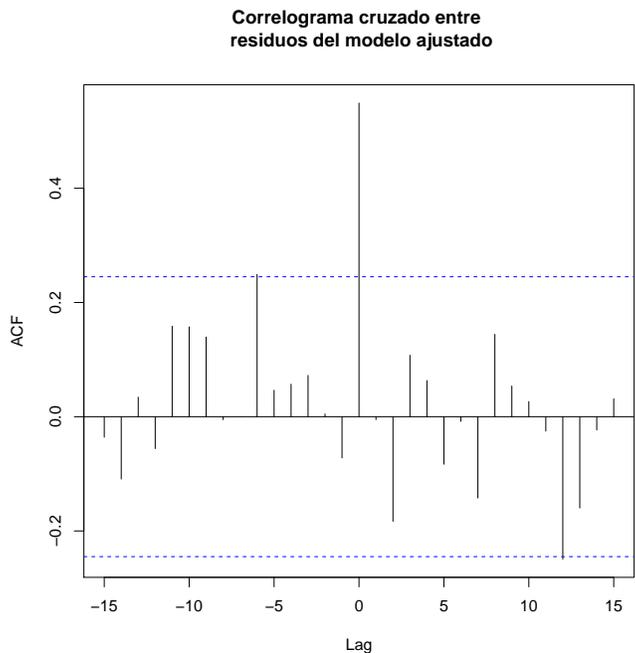


Figura 7: Correlograma cruzado entre los residuos del modelo ajustado

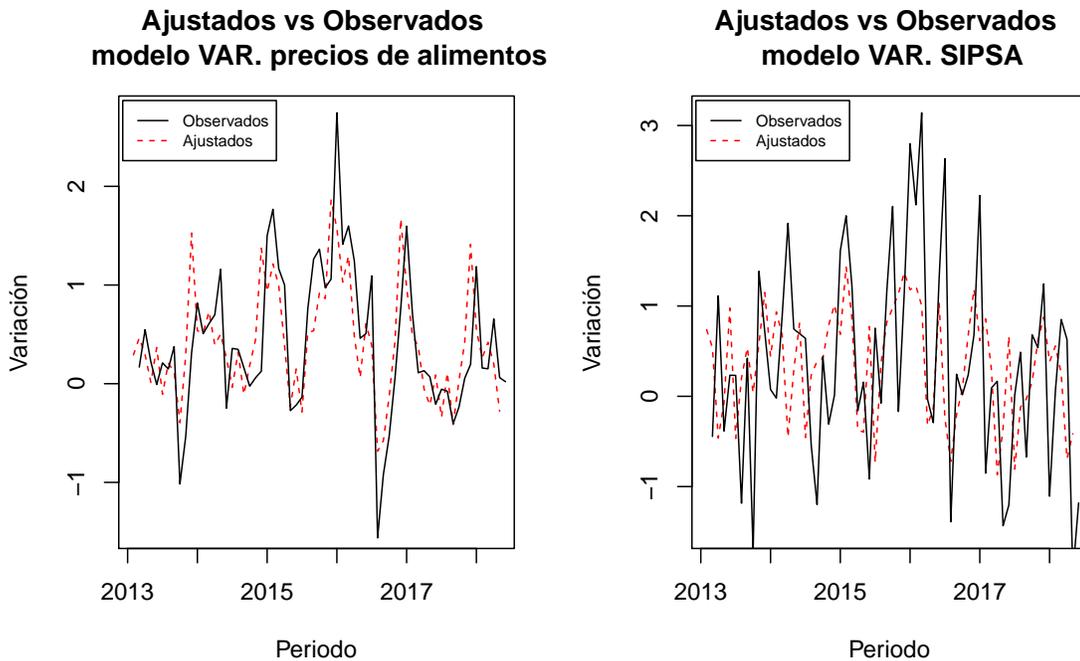


Figura 8: Estabilidad estructural del modelo propuesto

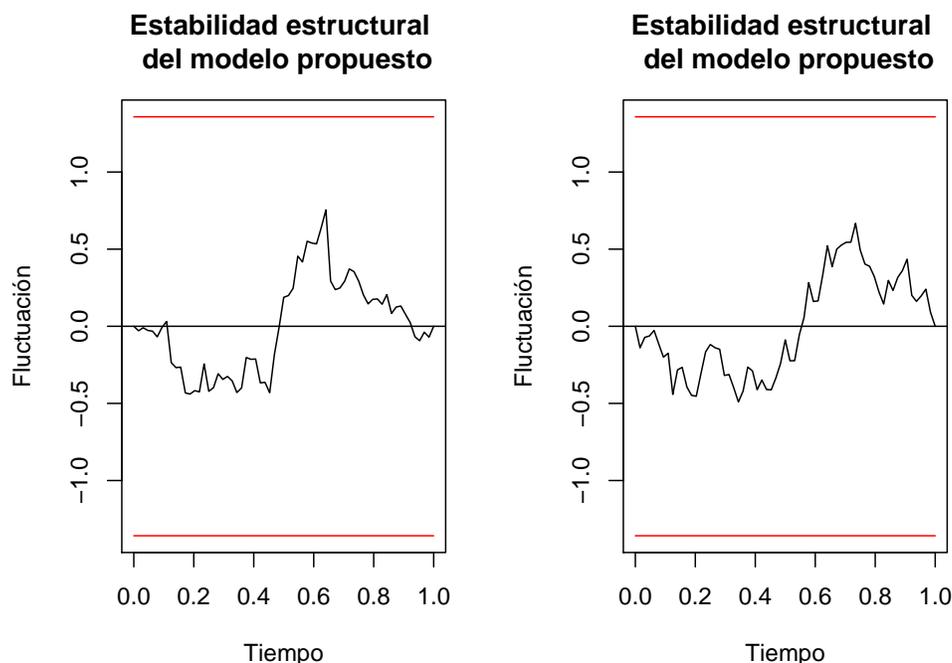


Figura 9: Estabilidad estructural del modelo propuesto

las matrices de restricciones A y B dadas por

$$A = \begin{bmatrix} NA & NA \\ 0 & NA \end{bmatrix} \quad B = \begin{bmatrix} NA & 0 \\ 0 & NA \end{bmatrix} \quad (7)$$

La matriz A incorpora las relaciones de largo plazo entre el índice SIPSA y la inflación de alimentos, previendo que el índice SIPSA incide contemporáneamente en el comportamiento de la inflación de alimentos; y no el caso contrario. En el corto plazo reflejado sobre la matriz B , las fluctuaciones de cada una de las variables es independiente. El modelo SVAR está determinado por la ecuación (6), la forma estructural dada en la ecuación (7) y las matrices estimadas A y B dadas por:

$$A = \begin{bmatrix} 1 & -0.2598 \\ 0 & 1 \end{bmatrix} \quad B = \begin{bmatrix} 0.4488 & 0 \\ 0 & 1.027 \end{bmatrix}$$

Con el modelo SVAR planteado, se estudia la dinámica entre las variables en estudio. En la figura 9 se presentan las fluctuaciones para ambas variables (izquierda para la variación en precios de los alimentos y a la derecha para el índice SIPSA propuesto). Hay evidencia de la estabilidad paramétrica del modelo y, por tanto, válido para establecer las relaciones que se presentan a continuación, ver Zeileis, et. al. (2002).

En la figura 10 se presentan los impulsos asociados al modelo SVAR propuesto. En la figura (a) se tiene el impulso de la variación en los precios de los alimentos en sí mismo. Se observa que tiene un efecto con caída suave, lo cual hace sentido toda vez que tal variación depende cada vez menos de sus valores pasados. Ahora, la figura (b) presenta el impulso asociado al índice SIPSA. Es interesante notar que tiene un efecto alto momentáneo. El índice SIPSA refleja la dinámica en periodos razonablemente inmediatos de la variación en los precios de los alimentos. Es decir, la variación en los precios cambio en función del

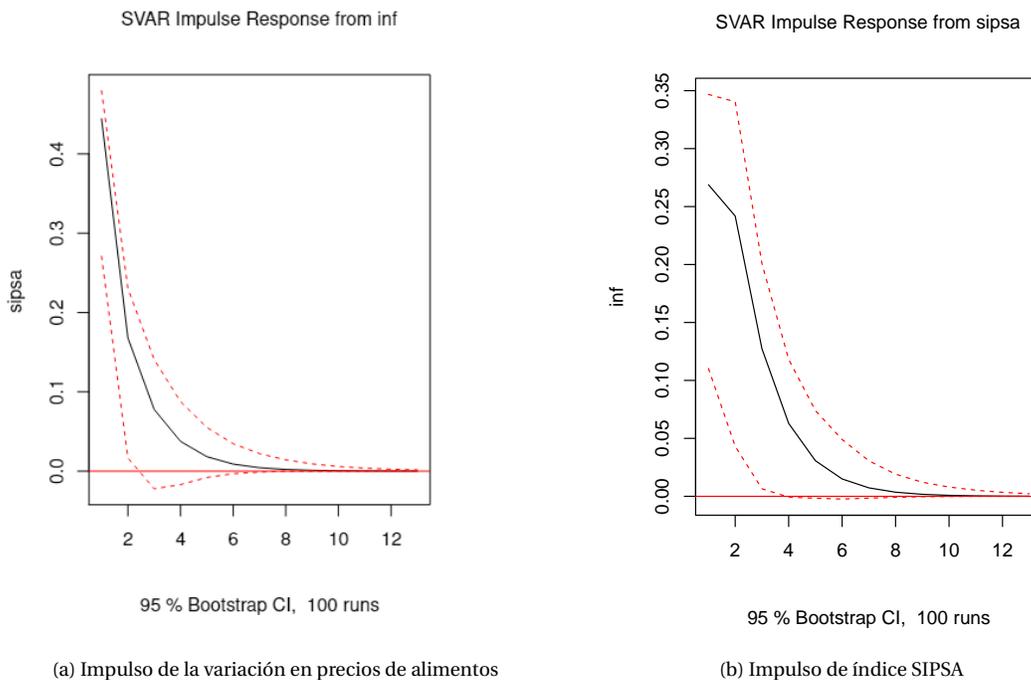


Figura 10: Impulsos respuesta asociados al modelo

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's
-0.282	0.426	0.321	-95.9	189	0.0348	0.382

Tabla 4: Medidas de error del modelo VAR con out-sample para variación en los precios de los alimentos

índice SIPSA. Adicionalmente, tales impactos son positivos lo que significa que ambos índices se mueven en la misma dirección en el tiempo.

En la figura 11 se presentan las descomposiciones de la varianza en los pronósticos, para la variación en los precios de los alimentos y el índice SIPSA, respectivamente. Se puede apreciar que el índice SIPSA propuesto explica entre un 30% y 40% de la varianza del pronóstico. Adicionalmente, el SIPSA no presenta dependencia externa puesto que la contribución de la variación en los precios es apenas marginal (menor al 5%). En conclusión, el índice SIPSA propuesto es útil para explicar la variación en los pronósticos en precios de alimentos.

Finalmente, como fue comentado al inicio del capítulo, la evaluación de los modelos se realizará a través del esquema in-sample y out-sample. Por lo tanto, se realizará el cálculo de los pronósticos in-sample y se contrastará la información pronosticada con aquella que fue observada. Serán generados pronósticos con el modelo VAR propuesto al inicio de esta sección. En la figura 12 se presentan los pronósticos para la variación en los precios de los alimentos. El ajuste es bueno toda vez que los puntos se encuentran dentro de los intervalos de confianza considerados. Adicionalmente, la tabla 4 presenta las medidas numéricas para evaluar los pronósticos para la variación en los precios de los alimentos.

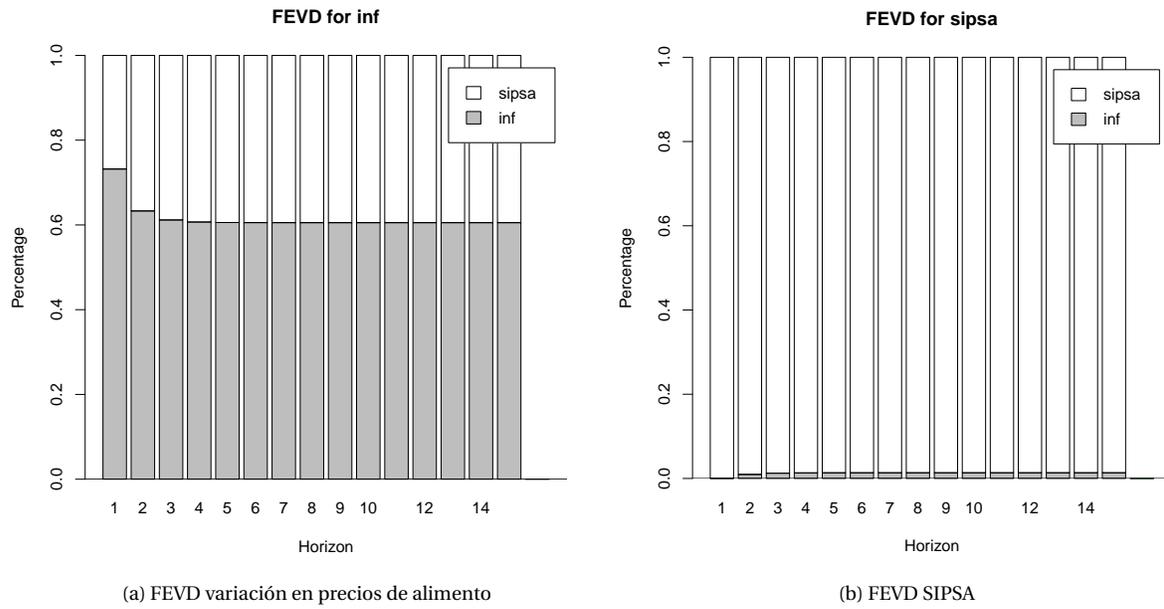


Figura 11: Descomposición en la varianza del error pronóstico para la variación en los precios de los alimentos y SIPSA

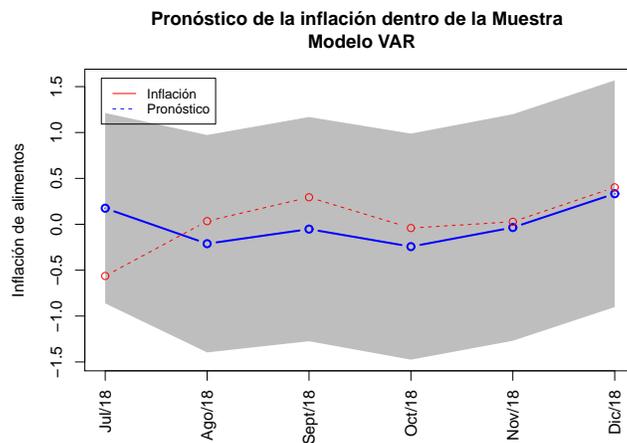


Figura 12: Pronóstico conjunto para la variación de los precios de los alimentos usando el modelo VAR

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's
-0.276	0.422	0.318	-93.4	189	0.0328	0.372

Tabla 5: Medidas de error del modelo VEC con out-sample para variación en los precios de los alimentos

4.3. Modelo VEC

Finalmente, para establecer la relación a largo plazo en las variables consideradas, se aplicó la prueba de cointegración por método de Johansen, ver Johansen (1991). Para determinar la existencia de cointegración y por tanto de la eficiencia de usar un modelo VEC, se tienen en cuenta las pruebas *Traza* y *Max eigenvector* que genera el test de cointegración de Johansen (ver Cheung, et. al. (1993)). En las dos pruebas deben aparecer vectores de cointegración para sustentar el uso del modelo.

Los resultados computaciones muestran el rechazo de la hipótesis nula de no relación de cointegración (Estadístico de prueba 26.46 y un valor crítico de 14.90 al 5% de significancia). Lo anterior significa que, en efecto la relación entre estas dos variables puede ser analizada a través de un modelo VEC.

El modelo VEC estimado para este punto está descrito en la tabla 6. Tal modelo presenta valores de 0.7796 y 0.0909 en las raíces del polinomio característico y por tanto, es un modelo estable. Se aplicaron las pruebas de Portmanteau (H_0 : Residuos no correlacionados), Jarque-Bera multivariado (H_0 : Dist. Normal en los Residuos) y ARCH (H_0 : residuos homoscedásticos) para determinar que los residuos tengan las propiedades deseadas. Se hizo un análisis de intervención determinando que la observación correspondiente a mayo de 2016 era atípica y podría influenciar desmedidamente en la verificación de supuestos. Por lo anterior, se realizó la verificación sin dicha observación. De esta forma, para las pruebas citadas, se obtuvieron p -valores de 0.6, 0.5 e 0.5, con lo cual, con un 5% de significancia, hay evidencia estadísticamente significativa para concluir que los residuos no están correlacionados serialmente, son homoscedásticos y son normales. Con lo cual se concluye que los residuos son ruido blanco gaussiano.

En la figura 13 se presentan los impulsos asociados al modelo VEC propuesto. En la figura (a) se tiene el impulso de la variación en los precios de los alimentos en sí mismo. Se observa que tiene un efecto constante y un intervalo de confianza que que no contiene al cero. Ahora, la figura (b) presenta el impulso asociado al índice SIPSA. Es interesante notar que tiene un efecto momentáneo bajo y que en el tiempo se muestra constante con un intervalo de confianza que contiene al cero.

Una vez especificado que el modelo, se obtiene la estimación del vector de corrección o cointegración dado por $\beta = (1, -1.2218, -0.0007)$ donde el primer coeficiente ha sido normalizado para que tenga como valor 1. De acuerdo a Engle y Granger y Granger (1987) este vector presenta equilibrio en el largo plazo por lo tanto se puede representar de la siguiente manera:

$$y_t = \text{Variación en precios de alimentos}_t - 1.2218\text{SIPSA}_t - 0.0007t \sim I(0) \quad (8)$$

La ecuación (8) da cuenta de la relación en el tiempo en el largo plazo entre la variación de los precios de los alimentos y el índice SIPSA propuesto.

La figura 14 presenta el pronóstico para la variación en los precios de los alimentos. Los pronósticos se encuentran por dentro del intervalo de confianza del 95% de confianza. El ajuste es bueno toda vez que los puntos se encuentran dentro de los intervalos de confianza considerados. Debe ser destacado que los pronósticos de los modelos VAR y VEC son razonablemente similares, con lo cuál podemos decir que el análisis presentado en esta sección es correcto pues se llegó a conclusiones similares de pronóstico con dos modelos diferentes. Adicionalmente, la tabla 5 presenta las medidas numéricas para evaluar los pronósticos.

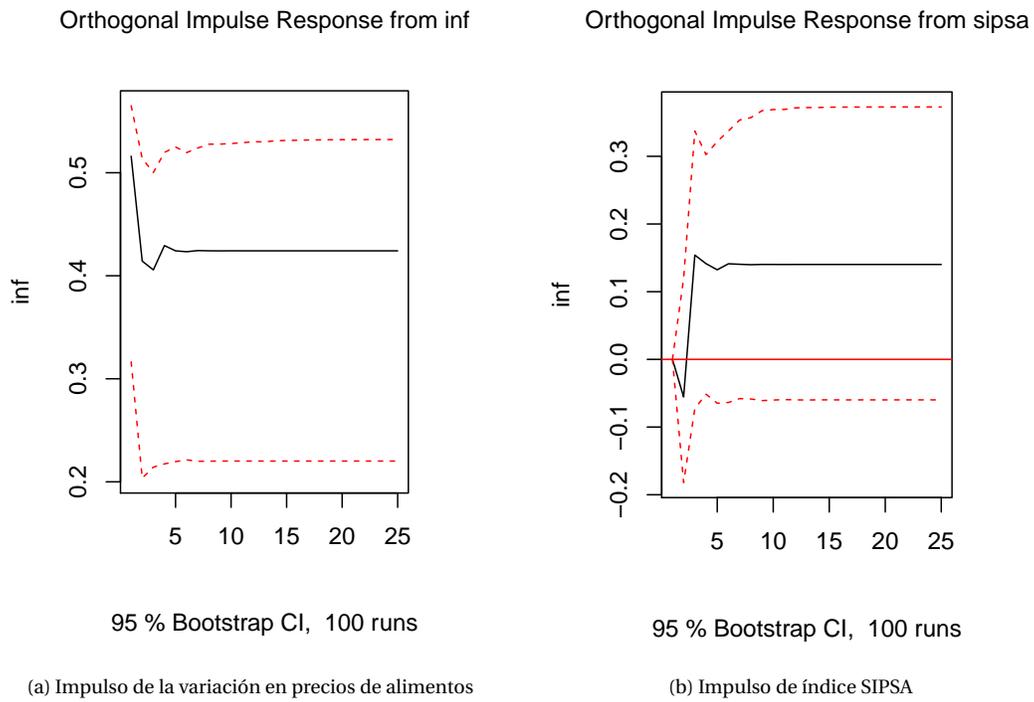


Figura 13: Impulsos respuesta asociados al modelo

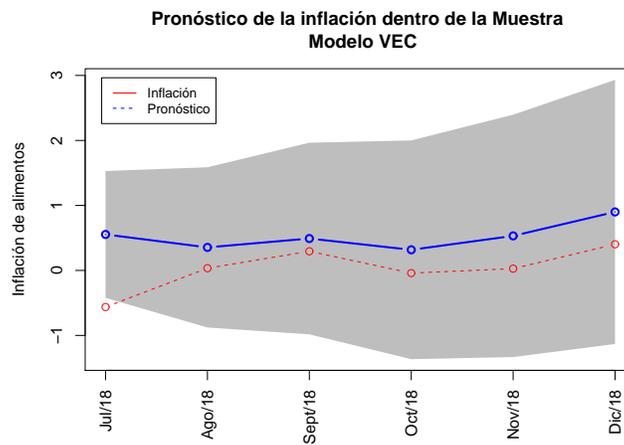


Figura 14: Pronóstico conjunto para la variación de los precios de los alimentos usando el modelo VEC

5. Conclusiones y recomendaciones

De acuerdo con los objetivos establecidos en este documento, se llegaron a las siguientes conclusiones:

- Fue creado un índice SIPSA que consolida la información de los precios de venta de los alimentos a lo largo de las principales centrales de abasto del país. Dicho índice, de forma visual, presenta un comportamiento razonable para entender la variable de interés, que en este trabajo, es la variación en los precios de los alimentos.
- Se consolidó un modelo SARIMAX para la variación en los precios de los alimentos que utiliza el índice SIPSA como variable exógena. Los pronósticos de dicho modelo fueron evaluados a través del esquema in-sample y out-sample dando resultados notables en términos de medidas de evaluación de error ME, MAE, RMSE y MAPE.
- Se estableció un modelo VAR estructural para entender la relación en el corto plazo entre las variables consideradas cuando se incorporan matrices de restricciones. A partir de las herramientas recomendadas en la literatura considerada tales como la generación de impulsos y descomposición de la varianza en el pronóstico. En general, se ve una buena relación a corto plazo entre la variación en los precios de los alimentos y el índice SIPSA.
- Finalmente, se establece a través de un modelo VEC la relación a largo plazo entre las variables consideradas en el presente trabajo. Se observa en general que los índices en el largo plazo presentarán un comportamiento razonablemente similar.

En general, el índice SIPSA propuesto se propone como una herramienta adicional para el entendimiento o generación rápida de pronósticos para la variación en los precios de los alimentos a través de metodologías en series de tiempo univariadas. Con metodologías de series de tiempo multivariadas se muestra el buen desempeño del índice propuesto en términos de su buena relación temporal (en el corto plazo y en el largo plazo) con la variable de interés.

De la misma manera, el trabajo presentado deja varias tareas pendientes en el marco de establecer información relevante para la obtención de buenas herramientas para estudiar la variación en los precios de los alimentos. Entre ellas, podemos formular las siguientes:

- La formulación de un índice SIPSA basado en otra metodología (diferente al cálculo de medianas) para la obtención de un índice más robusto y que ayude a capturar mejor la información presente en los reportes dados por el SIPSA.
- En el entendido que un índice SIPSA puede ser de interés para la variación en precios de alimentos, se podrían establecer variables que puedan ocasionar variación en el índice usado, tales como: precio de combustibles, clima y demás.
- Formulación de otros índices, no necesariamente, basados en SIPSA, para ver su relación en el tiempo con la variación y con el índice aquí propuesto.
- Basándose en el índice SIPSA presentado, indagar por una metodología complementaria que incorpore la dinámica del consumo de alimentos fuera del hogar (dicho consumo contribuye un 30% en la inflación de alimentos).

Recibido: Julio de 2019
Aceptado: Enero de 2020

Referencias

- Asteriou, D. and Hall, S.-G. (2011) Vector Autoregressive (VAR) models and causality tests. Applied Econometrics. London: Palgrave MacMillan.
- Cavaliere, G. (2003). Asymptotics for unit root tests under markov regime-switching. The Econometrics Journal, 6(1), 193-216.
- Cheung, Yin-Wong and Lai, Kon S. (1993) Finite-sample sizes of Johansen's likelihood ratio tests for cointegration. Oxford Bulletin of Economics and statistics, 55(3), pp. 313-328.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. Econometrica, 55, pp. 251-276.
- Fernandez-Corugedo, E (2003) Exercise on unit roots (including structural breaks), estimating a VECM and the implications of the VECM. Curso publicado por Center for Central Banking Studies (CCBS).
- Guerrero, V. (2003). Análisis Estadístico de Series de tiempo económicas. Ed. 2, International Thomson Editores.
- Holtz-Eakin, D., Newey, W. and Rosen, H.-S. (1988) Estimating vector autoregressions with panel data. Econometrica, pp. 1371-1395.
- IPC (2018) Subgrupos para consolidación de ipc colombiano. <https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc/grupos-ipc-2012>. Último Acceso: Julio 10, 2018.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. Econometrica, 59, pp. 1551-1580.
- Lütkepohl, H. (2005). New Introduction To Multiple Time Series Analysis. Springer.
- Lütkepohl, H. (2018) Impulse response function. The new Palgrave dictionary of economics. Springer. pp. 6141-6145.
- Pfaff, B. (2008). Analysis of Integrated and Cointegrated Time Series with R (Use R). Ed. 2, Springer.
- Rodríguez, Y. Agudelo, S. Rincón, L. Montañez, K. (2019) Inflación importada vía tipos de cambio y su aplicación al caso colombiano mediante el uso de modelos VEC para el periodo enero 2010 - diciembre 2018. Comunicaciones en Estadística, 12(1), pp. 33-52.
- SIPSA (2018) Reportes semanales. agencia sipsa. <https://www.dane.gov.co/index.php/estadisticas-por-tema/agropecuario/sistema-de-informacion-de-precios-sipsa/mayoristas-boletin-semanal-1>. Último Acceso: Julio 10, 2018.
- Tsay, R. S. (2013). Multivariate Time Series Analysis: With R and Financial Applications. Ed. 1, Wiley.
- Wei, W. (2006) Time series analysis. The Oxford Handbook of Quantitative Methods in Psychology. Ed. 2. Oxford.
- Zeileis, A., Leisch, F, Hornik, K., and Kleiber, C. (2002). strucchange: An r package for testing for structural change in linear regression models. Journal of Statistical Software. 7(2), pp. 1-38.
- Shahbaz, M. (2012) Does trade openness affect long run growth? Cointegration, causality and forecast error variance decomposition tests for Pakistan. Economic Modelling, 29(6), pp. 2325-2339.

Anexos

	Modelo Inf	Modelo SIPSA
ect1	0.09 (0.11)	0.85*** (0.21)
sd1	-1.39** (0.41)	-1.33 (0.74)
sd2	-1.46** (0.42)	-1.00 (0.76)
sd3	-1.22** (0.38)	-1.62* (0.70)
sd4	-1.37*** (0.38)	-1.93** (0.70)
sd5	-1.53*** (0.39)	-1.72* (0.72)
sd6	-1.08** (0.39)	-0.64 (0.71)
sd7	-1.41*** (0.38)	-2.00** (0.69)
sd8	-0.66 (0.43)	-1.24 (0.77)
sd9	-1.57*** (0.41)	-1.11 (0.74)
sd10	-1.16** (0.41)	-1.35 (0.75)
sd11	-0.61 (0.40)	-0.60 (0.72)
inf.dl1	-0.41* (0.16)	-0.58 (0.29)
sipsa.dl1	0.24* (0.10)	0.25 (0.19)
R ²	0.50	0.57
Adj. R ²	0.31	0.40
Num. obs.	53	53
RMSE	0.56	1.02

*** p<0.001, ** p<0.01, * p<0.05

Tabla 6: Modelo VEC ajustado