

FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS

ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

PROPUESTA DE UN ÍNDICE SIPSA PARA EL
PRONÓSTICO DE LA INFLACIÓN DE ALIMENTOS.
EVIDENCIA EMPÍRICA

SÉBASTIEN LOZANO FORERO

MATEMÁTICO Y MG. EN ESTADÍSTICA



Bogotá
2018

FUNDACIÓN UNIVERSITARIA LOS LIBERTADORES
FACULTAD DE INGENIERÍA Y CIENCIAS BÁSICAS
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

SÉBASTIEN LOZANO FORERO

PROPUESTA DE UN ÍNDICE SIPSA PARA EL
PRONÓSTICO DE LA INFLACIÓN DE ALIMENTOS.
EVIDENCIA EMPÍRICA

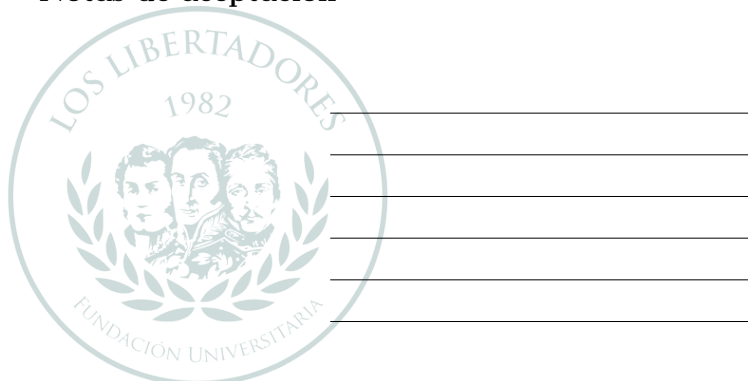
ORIENTADOR: PROF. JUAN CAMILO SANTANA CONTRERAS

Área de Concentración: Series de tiempo

Trabajo presentando al programa de Especialización en Estadística Aplicada, de la Facultad de Ingeniería y Ciencias Básicas de la Fundación Universitaria los Libertadores para optar al título de **Especialista en Estadística Aplicada**.

Bogotá
2018

Notas de aceptación



LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Firma del presidente del jurado

Firma del jurado

Firma del jurado

Bogotá DC, Mayo de 2018.



LOS LIBERTADORES

FUNDACIÓN UNIVERSITARIA

Las directivas de la Fundación Universitaria Los Libertadores, los jurados calificadores y el cuerpo docente no son responsables por los criterios e ideas expuestas en el presente documento. Estos corresponden únicamente a los autores y a los resultados de su trabajo.

A mi familia

Agradecimientos

En esta pequeña pero sustanciosa sección, deseo agradecer a todas las personas que hicieron posible este proceso posgradual. A mi familia por todo su paciencia y soporte. A Mariapaula por su apoyo constante. A Vladimir Ballesteros por su confianza. Al profesor Juan Camilo Santana por sus enseñanzas y visión experta en el tema. A la Fundación Universitaria los Libertadores y al cuerpo docente de la especialización por la excelente formación recibida.

Índice general

Agradecimientos

Resumen

1. Introducción	12
1.1. Objetivos	13
1.1.1. Objetivo General	14
1.1.2. Objetivos Específicos	14
1.2. Organización del trabajo	14
1.3. Soporte computacional	15
2. Marco metodológico	16
3. Marco teórico	18
3.1. Series de Tiempo Univariadas	18
3.1.1. Metodología Box-Jenkins	18
3.1.2. Criterio BIC	20
3.1.3. Evaluación de pronósticos	20
3.1.4. Prueba de Dickey-Fuller	21
3.1.5. Modelo ARIMA	22
3.1.6. Modelo SARIMA	23

3.1.7. Modelo SARIMAX	24
3.2. Series de Tiempo Multivariadas	24
3.2.1. Modelo VAR	25
3.2.2. Modelo SVAR	25
3.2.3. Modelo VEC	26
3.2.4. Función de impulso respuesta	26
3.2.5. Análisis de descomposición de varianza	27
4. Inflación e índice SISPA	28
4.1. Inflación de alimentos	28
4.2. Construcción del índice SIPSA	29
5. Resultados	32
5.1. Modelo SARIMAX	33
5.2. Modelo SVAR	37
5.3. Modelo VEC	47
6. Conclusión y recomendaciones	51
7. Trabajo Futuro	53
Referencias bibliográficas	55

Índice de figuras

3.1. Metodología de Box-Jenkins	19
5.1. SIPSA y variación de alimentos	33
5.2. Identificación para la serie IPC	34
5.3. Validación del modelo SARIMAX	35
5.4. Pronóstico del modelo SARIMAX in-sample	36
5.5. Autocorrelograma cruzado entre variables	37
5.6. Autocorrelograma cruzado entre variables	39
5.7. Autocorrelograma cruzado entre residuos	40
5.8. Estabilidad estructural del modelo propuesto	40
5.9. Estabilidad estructural del modelo propuesto	42
5.10. Impulsos asociados al modelo	42
5.11. FEVD variación en precios de alimentos	43
5.12. FEVD SIPSA	44
5.13. Pronóstico conjunto. SIPSA. Modelo VAR	45
5.14. Pronóstico conjunto. Variación en precios de alimentos. Modelo VAR	46
5.15. Pronóstico conjunto. SIPSA. Modelo VEC	49
5.16. Pronóstico conjunto. Variación en precios de alimentos. Modelo VEC	50

Índice de cuadros

5.1. Estimativas para el modelo SARIMAX(10, 1, 1) × (0, 1, 1) ₁₂	34
5.2. Medidas de error del modelo SARIMAX con out-sample	36
5.3. valores para las variables dummies que explican la estacionalidad de los datos	38
5.4. Medidas de error del modelo VAR con out-sample para variación en los precios de los alimentos	46
5.5. Medidas de error del modelo VAR con out-sample para SIPSA	46
5.6. Modelo VEC ajustado	48
5.7. Medidas de error del modelo VEC con out-sample para variación en los precios de los alimentos	49
5.8. Medidas de error del modelo VEC con out-sample para SIPSA	50

Resumen

La inflación es un agente macroeconómico delicado de lidiar dado que afecta directamente la vida los ciudadanos. En el caso del índice de variación en los precios de alimentos, tal afectación resulta mucho mas sensible, tanto para empresas del sector alimentario como para la población en general. Es así que resulta de interés poder, con algún grado de certeza, obtener información relevante del índice de variación en los precios de los alimentos. A su vez, el SIPSA (dependencia del DANE) es el ente gubernamental encargado de dictar a las centrales mayoritarias de abasto del país; los precios con los que deben comerciar los productos. En el presente trabajo, se aborda el problema de predecir la inflación de alimentos a partir de un índice SIPSA propuesto por el autor. Se abordan dos 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).

CAPÍTULO 1

Introducción

Según el banco de la República, 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. Es así que el entendimiento de la inflación puede ser de interés para los agentes económicos que constantemente interactúan en una sociedad.

Para el caso de la economía colombiana, el departamento administrativo nacional de estadística (DANE) desagrega el índice de inflación en nueve agentes contributivos. A saber: alimentos (que supone una contribución del 27.09 % en el índice global de inflación), vivienda (30.42 %), vestuario (5.11 %), salud (2.41 %), educación (5.99 %), diversión (3.19 %), transporte (15.32 %), comunicaciones (3.98 %) y otros gastos (6.47 %). Es decir, se cuenta con un índice particular de inflación (índice de variación de precios) para cada uno de los agentes anteriormente mencionados.

En este trabajo, nos enfocaremos en la inflación de alimentos como un agente importante puede regular la economía nacional en el sector alimentario. Lo anterior debido a que este frente es, por razones naturales, el más sensible de todos los que componen el índice de inflación dada su gran impacto en la población. 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. Por ejemplo, si la inflación sube hay dos opciones globales: o la demanda de alimentos baja (lo cuál no es del todo razonable dado la urgencia por la adquisición de los alimentos) o algún otro sector de la economía se verá afectado.

Dicha inflación de alimentos puede ser vista en tres grupos sociales: de ingreso bajo, de ingreso medio y de ingreso alto. Para este trabajo consideramos el caso de ingreso medio. Para dicho grupo social, la inflación de alimentos, a su vez puede ser desagregada en varios tipos de gasto. Consideramos, para este trabajo los agentes: 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 %).

Por otro lado, el sistema de información de precios (SIPSA) es una dependencia del DANE responsable de informar los precios mayoristas de los productos agroalimentarios que se comercializan en el país, así como la información de insumos y factores asociados a la producción agrícola y a nivel de abastecimiento de alimentos en las ciudades. En pocas palabras, el SIPSA es responsable de fijar los precios de los alimentos en las centrales mayoritarias de abasto en todo el país.

El SIPSA reporta, semanalmente, a través de reportes técnicos y comunicados de prensa, los precios a las centrales de abasto. Dichos comunicados están desagregados en las categorías: Verdura/Hortaliza, fruta, tubérculos, grano/cereal , huevos/lácteos, carnes, pescado y procesados. Dichas categorías se encuentran desagregadas por producto y central mayoritaria de abasto.

1.1. Objetivos

Una vez planteados los dos elementos fundamentales de este trabajo, índice de inflación y el índice SIPSA, planteamos los objetivos fundamentales

1.1.1. Objetivo General

- Contribuir al entendimiento de la inflación de alimentos, para el grupo social de ingresos medios de la población colombiana, a partir de un índice SIPSA.

1.1.2. Objetivos Específicos

- Consolidar la información disponible en los reportes semanales dados por el SIPSA (dependencia del DANE) en un índice mensual.
- Obtener un modelo de pronóstico SARIMAX para la variación de precios de alimentos utilizando el índice SIPSA como variables regresora.
- Establecer la relación en el tiempo entre la variación en el precio de los alimentos y el índice SIPSA a partir de un modelo SVAR
- Definir la relación a largo plazo entre las dos variables consideradas y su cointegración a partir de un modelo VEC.

1.2. Organización del trabajo

El presente trabajo se organiza de la siguiente forma:

En el capítulo 3 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 (AR, MA, ARMA, ARIMA, SARIMA, SARIMAX) y los modelos tradicionales de series de tiempo multivariadas (VAR, SVAR, VEC).

En el capítulo 4 se presenta información más general respecto al índice de inflación, y por tanto de la inflación de alimentos. Adicionalmente, se presenta la construcción del índice SIPSA que será utilizado en el presente trabajo.

En el capítulo 5 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.

Finalmente, en los capítulos 6 y 7 se presentan las principales conclusiones de este trabajo y algunas ideas para futuras investigaciones utilizando las ideas aquí expuestas, respectivamente.

1.3. 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 ([Ihaka, 1996](#)).

La elaboración de este documento fue hecho integralmente en el sistema tipográfico L^AT_EX que, básicamente, es una serie de macros y rutinas basadas en el sistema T_EX. Este lenguaje fue creado por Donald Knuth en 1984 en la Universidad de Stanford, USA ([Knuth, 1986](#)). Como plataforma fue usado el dominio www.sharelatex.com

CAPÍTULO 2

Marco metodológico

El estudio que aquí se propone es de tipo correlacional toda vez que se pretende proponer un índice SIPSA y ver su interacción en el tiempo con la variación mensual en los precios de los alimentos. Ninguna relación de este tipo está referenciada en la literatura consultada para este trabajo.

La información en su totalidad fue conseguida a través del portal web del Departamento Administrativo Nacional de Estadística (DANE), consultado entre marzo y abril de 2018. Los datos históricos corresponden a lo reportado entre enero de 2013 y abril de 2018.

A partir de la literatura consultada se determinó que los mejores tipos de modelos para hallar relaciones entre variables son los de tipo SARIMAX(Auto Regresivo Integrado de Medias Móviles Estacional con variables exógenas) y VAR/VEC (Vectorial Auto Regresivo y Vectorial con Corrección de Errores, respectivamente). Lo anterior dado que este tipo de modelos permiten analizar las interacciones de forma causal y simultánea, respectivamente; entre las variables.

El trabajo a continuación tiene tres grandes momentos. El primero de ellos es la construcción del índice SIPSA que resume, en buena parte, la información suministrada por el departamento SIPSA (Sistema de Información de Precios), dependencia del DANE; sobre los precios de venta de productos alimenticios reportados en las mayores centrales de abasto del país.

El segundo momento es la consolidación de un modelo SARIMAX (metodología univariada) para la variación en los precios de los alimentos, que incorpora valores del índice SIPSA propuesto. La idea general es que información sobre el índice SIPSA es de utilidad para pronosticar valores para la variación en los precios y por tanto, los pronósticos basados en utilizando SIPSA son más precisos que aquellos que sólo se generan con valores rezagados de la variable de interés.

El tercer momento consiste en estudiar la interacción entre las variables de interés en este trabajo (índice SIPSA y variación en los precios de los alimentos), esta vez, desde una perspectiva multivariada. Serán propuestos dos grandes modelos para este escenario: un modelo SVAR (VAR estructural) a fin de poder generar pronósticos conjuntos dado el histórico de interacción entre las variables. El segundo modelo será un modelo VEC para determinar la interacción en el largo plazo entre las variables y así, formular su relación de cointegración.

3.1. Series de Tiempo Univariadas

En esta sección, se presentan de forma muy general los modelos estadísticos tradicionales en series de tiempo univariadas. En general, se presenta la teoría general. Los detalles pueden ser consultados en ([Metcalfé, 2009](#)), ([Tsay, 2012](#)) o ([Guerrero, 2003](#))

3.1.1. Metodología Box-Jenkins

El modelamiento que se introducirá en esta sección, consiste en ajustar modelos al a estructura general $SARIMA(p, d, q) \times (P, D, Q)_s$ y alguna de sus variaciones. La estrategia para la construcción del modelo se basa en un ciclo iterativo y utiliza los propios datos para la elección de la estructura de este modelo. El modelado de Box-Jenkins, inicialmente presentado en ([Box, 1970](#)) se puede escribir como:

1. Postule una clase de modelos (AR, MA, ARMA, ARIMA, SARIMA, etc)

2. Identifique el modelo perteneciente a la clase postulada que mejor se ajusta a los datos. (Etapa de identificación)
3. Estime los parámetros del modelo seleccionado
4. Evalúe la calidad del ajuste.

Gráficamente, la metodología de Box-Jenkins se puede entender como el flujograma dado en la figura 3.1. Si las evaluaciones son correctas, entonces se pueden hacer las previsiones.

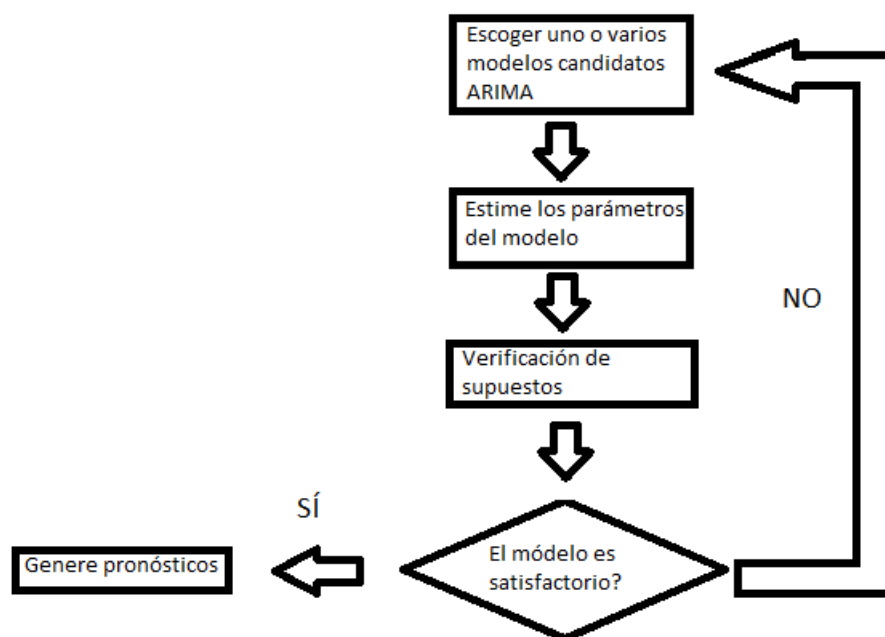


Figura 3.1: Metodología de Box-Jenkins

En el caso contrario, comience nuevamente. En general, los modelos más convenientes son aquellos que tienen pocos parámetros ya que son mejores para predecir lo que está por ocurrir.

En general, será utilizado el siguiente criterio para decidir cual entre varios modelos, de la misma clase, es mejor para ajustar los datos en cuestión.

3.1.2. Criterio BIC

El criterio de información Bayesiano (BIC), introducido por Gideon Schwartz en (Schwarz, 1978), es una medida alternativa para el AIC para obtener información sobre el mejor modelo que se puede utilizar. El BIC se define como

$$BIC = -2 \log(L) + k \log(T),$$

Donde k es el número de parámetros que varían ($k = p + q$ en los modelos ARMA) y T es la cantidad de datos disponibles. (En inglés) Algunas generalidades sobre estas medidas

1. El AICc es una versión corregida del AIC, pero que es asintóticamente equivalente.
2. El AIC no es consistente. En la clase ARIMA, por ejemplo, proporciona \hat{p} y \hat{q} tales que $p \lim(\hat{p}) \neq p$ y $q \lim(\hat{q}) \neq q$. Proporciona modelos superparametrizados
3. El BIC penaliza más fuertemente que el AIC y es consistente.

La parte sistemática del modelado consiste en estimar varios modelos y elegir aquel que minimiza el criterio utilizado.

3.1.3. 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.

3.1.4. Prueba de Dickey-Fuller

De una forma muy general, se notará como $I(0)$ a las *Series integradas de orden 0*, es decir, la serie presenta media y varianza constantes. De dichas series se dirá que como que no presentan problemas de estacionariedad. Análogamente, se nota como $I(1)$ a las series no estacionarias. Formalmente, son aquellas series que tienen una raíz unitaria, es decir, que tienen media o varianza no constantes.

Así, en el caso que sea de nuestro interés saber cuándo una serie es “bien comportada” en el sentido de estacionariedad, se tiene la prueba aumentada de Dickey-Fuller, originalmente presentada en (Dickey, 1979) que establecerá si la serie es integrada de orden 1. Es decir, formalmente

$$\begin{cases} H_0 : y_t \sim I(1) \\ H_1 : y_t \sim I(0) \end{cases}$$

donde $y_t = \alpha y_{t-1} + \epsilon_t$ y $\epsilon_t \sim \text{ARMA}(p, q)$ estacionario. Aquí se consideran varios casos,

R1 $\Delta y_t = \rho y_{t-1} + \sum_{j=1}^m \gamma_j \Delta y_{t-j} + \text{error}_t$ (menos usado)

R2 $\Delta y_t = \mu + \rho y_{t-1} + \sum_{j=1}^m \gamma_j \Delta y_{t-j} + \text{error}_t$ (más usado)

R3 $\Delta y_t = \mu + \delta t \rho y_{t-1} + \sum_{j=1}^m \gamma_j \Delta y_{t-j} + \text{error}_t$ para series con una tendencia fuerte.

Estime $[Ri]$ con $i = 1, 2, 3$ y calcule $\tau = \frac{\hat{\rho}}{\sqrt{\widehat{\text{Var}}(\hat{\rho})}}$. Dickey-Fuller, en 1979 realizaron simulaciones de Monte-Carloy estimaron los valores críticos de la estadística τ . La regla de rechazo es: Rechace H_0 si τ es menor que el valor crítico. Se puede demostrar que bajo H_0 cumple con

$$\tau \rightarrow \frac{1/2(W(1))^2 - 1}{\sqrt{\int_0^1 W(r)^2 dr}}$$

donde la convergencia es en distribución y $W(\cdot)$ representa un movimiento Browniano.

3.1.5. Modelo ARIMA

La estacionariedad se define, en forma sencilla, como la constancia en media, varianza y covarianza de una serie de tiempo y es, típicamente, uno de los primeros supuestos que deben verificarse en la etapa de identificación.

Para poder abordar escenarios más generales que involucren que tal supuesto de estacionariedad no se cumple, es necesario introducir el operador diferencia.

Sea $\{y_t\}$ una serie de tiempo. Se definen los operadores diferencia y desfase como

[Op. Diferencia] Se define la n -ésima diferencia de una serie de tiempo como

$$\Delta^n y_t = \Delta^{n-1} y_t - \Delta^{n-1} y_{t-1} = \sum_{r=0}^n (-1)^r \binom{n}{r} y_{t-r}.$$

[Op. desfase] Se define el operador desfase B de tal forma que $B y_t = y_{t-1}$, mas generalmente, $B^n y_t = y_{t-n}$ con $n = 0, 1, 2, \dots$.

Ahora, debemos enfrentarnos a la situación de si la serie no es estacionaria, el proceso puede ser integrado de orden d , es decir, y_t es no estacionario, pero $\Delta^d y_t = (1 - B)^d y_t$ es estacionario. Se tiene entonces un proceso ARIMA (p, d, q) definido como

$$\phi(B)(1 - B)^d y_t = \theta(B)\epsilon_t, \tag{3.1}$$

donde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$ es llamado de *polinomio auto-regresivo* y $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ es un polinomio de *medias de móviles*

Es necesario comentar que los modelos puros Auto Regresivos de orden p , denotado por $AR(p)$ y de medias móviles de orden q , denotado por $MA(q)$ son casos particulares del modelo ARIMA que se acaba de presentar. Basta con escribir $AR(p)=ARIMA(p, 0, 0)$ y $MA(q)=ARIMA(0, 0, q)$.

Con el objetivo de tener en cuenta los patrones de fluctuación estacional, se amplía la clase de modelos ARIMA, dados en (3.1). Dada una serie de tiempo $\{y_t\}$, muchas veces no es posible transformar la variable para remover estacionalidad periódica, es decir, dicha estacionalidad puede presentar un patrón dinámico. Luego, es necesario considerar un modelo ARIMA con componente estacional (SARIMA).

3.1.6. Modelo SARIMA

Sea y_t la serie de interés observada con periodo de estacionalidad s . Defina

$$\Phi(B^s) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{sP},$$

el operador auto regresivo estacional de orden P , estacionario, y

$$\Theta(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ}$$

el operador de medias móviles estacional de orden Q , invertible y $\Delta_s^D = (1 - B^s)^D$ con D indicando el número 'diferencias estacionales'. La clase de modelos estacionales multiplicativos $(p, d, q) \times (P, D, Q)$ está dada por

$$(1 - \phi_1 B - \dots - \phi_p B^p)(1 - \Phi_1 B^s - \dots - \Phi_P B^{sP})[(1 - B)^d(1 - B^s)^D - \mu]y_t = \\ (1 - \theta_1 B - \dots - \theta_q B^q)(1 - \Theta_1 B^s - \dots - \Theta_Q B^{sQ})\epsilon_t$$

donde $\epsilon_t \sim RB(0, \sigma^2)$. El proceso puede ser simplificado escribiendo

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D y_t = \theta(B)\Theta(B^s)\epsilon_t$$

3.1.7. Modelo SARIMAX

Los modelos SARIMAX (auto regresivos integrados de medias móviles estacional con variables explicativas) es una estructura de modelamiento que contiene todas las bondades de un modelo SARIMA y adicionalmente, puede capturar información sobre variables exógenas que puedan ayudar a entender y pronosticar la variable de interés. Para el caso en que se tiene una única variable de interés y , y una única variable independiente x . El modelo que describe la relación entre y en un tiempo t en términos de r observaciones anteriores de x está dado por

$$y_t = \alpha + \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + \beta_0 x_{t-1} + \beta_1 x_{t-1} + \dots + \beta_r x_{t-r} + \epsilon_t + \theta_1 \epsilon_{t-1} + \dots + \theta_q \epsilon_{t-q} \quad (3.2)$$

donde $\alpha, \beta_i, \phi_j, \theta$ 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 (3.2) puede ser escrito como

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

3.2. Series de Tiempo Multivariadas

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. Los pronósticos obtenidos con estos modelos son muchos más ajustados que los obtenidos con modelos Box Jenkis en el corto plazo, sin embargo, 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) o (Lütkepohl, 2005)

3.2.1. Modelo VAR

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:

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

donde $\mathbf{y}_t = (y_{1t}, \dots, y_{Kt})$ es un vector aleatorio de dimensión ($K \times 1$), las A_i son matrices de coeficientes de dimensión ($K \times K$), $\boldsymbol{\nu} = (\nu_1, \dots, \nu_K)$ 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 $A(L) = I_K - A_1 L - \cdots - A_p L^p$, el proceso anterior puede definirse como

$$A(L)\mathbf{y}_t = \boldsymbol{\nu} + \boldsymbol{\epsilon}_t \quad (3.3)$$

3.2.2. Modelo SVAR

El modelo Vectorial Auto Regresivo Estructural es una generalización del modelo VAR dado en la ecuación (5.1). 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.

$$A\mathbf{y}_t = A\boldsymbol{\nu} + A_1^* \mathbf{y}_{t-1} + \cdots + A_p^* \mathbf{y}_{t-p} + B\boldsymbol{\epsilon}_t \quad (3.4)$$

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 $\mathbf{u}_t = A^{-1}B\boldsymbol{\epsilon}_t$ y su matriz de varianzas y covarianzas está dada por $\Sigma_{\mathbf{u}} = A^{-1}BB^\top A^{-1\top}$

3.2.3. 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)

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

Se define el modelo de corrección de errores como

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

donde

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

y

$$\boldsymbol{\Gamma} = \boldsymbol{\alpha}\boldsymbol{\beta}^\top = -(I - A_1 - \cdots - A_p)$$

Las matrices $\boldsymbol{\Gamma}_i$ contienen los impactos acumulados de largo plazo. Es así que la especificación del VEC es sinónimo de de una relación a largo plazo.

3.2.4. 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 este cambio desaparece con el tiempo y que las demás variables permanecen *ceteris paribus* (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 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. Este proceso se denomina ortogonalización por descomposición de Cholesky, donde la matriz de innovaciones y covarianzas es diagonal. Esto causa también que el cambio de orden en las variables a la hora de generar la FIR influya de manera significativa en el resultado, motivo por el que hay que tener claridad que variable busca ser explicada con el modelo.

3.2.5. 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 él se genera la FIR sea correcto.

CAPÍTULO 4

Inflación e índice SISPA

En el presente capítulo se mostrará la metodología general para la obtención de los índices de inflación y SIPSA. Ambos índices fueron obtenidos a partir de la información disponible en la página web del Departamento Administrativo Nacional de Estadísticas (DANE).

4.1. Inflación de alimentos

La obtención de la variación de los precios de los alimentos fue consolidada a partir de la información disponible en la página web <https://www.dane.gov.co/index.php/estadisticas-por-tema/precios-y-costos/indice-de-precios-al-consumidor-ipc/grupos-ipc-2012>, en el grupo de gastos **Ingresos medios** en donde se presentan las contribuciones de los diferentes actores económicos necesarios para el cálculo de la inflación general como y sus respectivos valores mensuales. La consolidación de la variación de precios con la que se realizó el presente trabajo consiste en el cálculo de la diferencia relativa entre observaciones sucesivas reportadas en el la categoría **alimentos**. Es decir, si por ejemplo en un determinado mes el valor reportado es de 102 y en el mes inmediatamente

siguiente el valor reportado desde 107, se considera la variación como

$$100 \times \left(\frac{107 - 102}{102} \right)$$

Al consolidado del índice así calculado, nos referiremos como **la variación en los precios de alimentos** y será el punto de partida del presente trabajo.

Adicionalmente, se consideran los pesos, por grupo alimentario, para el cálculo de la variación en los precios de los alimentos. Dichos pesos son de 6 % para verduras y hortalizas, 4 % para frutas, 3 % para tubérculos, raíces y plátanos, 12 % para granos y cereales, 15 % para huevos y lácteos, 17 % para carnes, 2 % para pescados, 11 % para productos procesados y 30 % para productos alimentarios que se consumen "fuera de casa". Esta ponderación será capital para el cálculo del índice SISPA que se propone en la sección 4.2.

4.2. Construcción del índice SISPA

El primer gran objetivo del presente trabajo es proponer un índice consolidado que resuma la información que SISPA (como agencia) facilita a través de sus reportes semanales. Para la consolidación del índice SISPA se procedió a la siguiente manera:

Se descargaron todas las bases de datos que se encuentran en el sitio web <https://www.dane.gov.co/index.php/estadisticas-por-tema/agropecuario/sistema-de-informacion-de-precios-sipsa/mayoristas-boletin-semanal-1>. Dichas bases de datos contiene la información semanal de los precios reportados para la venta de productos agrícolas en las principales centrales de abasto del país. 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 contiene en promedio 4500 registros en donde se especifica, producto a producto, cual fue precio de venta en las principales centrales de abasto del país. Se tienen en total 280 archivos que registran la actividad mencionada desde enero de 2013 hasta abril de 2018.

Luego de esto, se utilizó la siguiente metodología en cada una de las bases de datos: para cada producto se calculó el precio medio de venta que fue reportado en las principales centrales de abasto de todo el país. Luego, se calculó la mediana de estos precios para cada una de las categorías alimentarias arriba reportadas. 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. Dicha ponderación es del 6% para verduras y hortalizas, 4% para frutas, 3% para tubérculos, raíces y plátanos, 12% para granos y cereales, 15% para huevos y lácteos, 17% para carnes, 2% para pescados y 11% para productos procesados.

Es de aclarar que en las ponderaciones anteriores se tiene el 70% de la información presente en la variación de precios de los alimentos. El 30% restante se refiere a las comidas fuera del hogar y, por tanto, que no se presentan en los reportes que SIPSA realiza toda vez que no son productos que se venden en plazas de abasto.

Finalmente, a partir de los consolidadas semanales se llegó a un valor único mensual. Esto se realizó mediante la suma de los valores unificados semanales que se obtuvieron teniendo en cuenta la participación de cada semana en el mes respectivo. Es decir, si por ejemplo un reporte contenía tres días de un mes y dos de otro mes, el consolidado semanal era reportado al mes que tenía mayor representación de días. El índice SIPSA que aquí se propone es el resultado de multiplicar por 100 la diferencia relativa de los consolidados mensuales construidos. Por ejemplo si el consolidado en un determinado mes es de 2700 y el consolidado del siguiente mes es 2710, el índice en este caso se calcula como,

$$100 \times \left(\frac{2700 - 2710}{2710} \right)$$

Así, el índice SIPSA propuesto recoge la información en la variación de los precios consolidados mensuales.

En resumen, el índice SIPSA se calcula mediante la ecuación

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

en que $CM_t = \text{Mediana}(CS_{1t}, \dots, CS_{kt})$ y k es el número de semanas que componen el mes t . A su vez, CS_{ij} es el consolidado de la semana i del mes j que se construye con los precios reportados por SIPSA y las respectivas ponderaciones por grupo de alimentos de la siguiente manera:

$$\begin{aligned}CS_{ij} = & 0,06 \times \text{mediana}(\text{VerduraYHortaliza}_i) + 0,04 \times \text{mediana}(\text{fruta}_i) \\ & + 0,03 \times \text{mediana}(\text{tubérculos}_i) + 0,12 \times \text{mediana}(\text{granoYCereal}_i) \\ & + 0,15 \times \text{mediana}(\text{huevosYLácteos}_i) + 0,17 \times \text{mediana}(\text{carnes}_i) \\ & + 0,02 \times \text{mediana}(\text{pescado}_i) + 0,11 \times \text{mediana}(\text{procesados}_i).\end{aligned}$$

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.

CAPÍTULO 5

Resultados

En el presente capítulo 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 4.2.

Para evaluar los modelos que serán presentados en esta sección, se utilizará el esquema de prueba por fuera de la muestra. Es decir, los datos serán divididos en dos partes: La primera parte de los mismos, que serán llamados de **in-sample**, contendrán los datos en el periodo de enero de 2013 hasta agosto de 2017. La segunda parte de los mismos, que serán llamados de **out-sample**, corresponden al periodo de septiembre de 2017 hasta abril 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. la sección 3.1.3 contiene las medidas utilizadas para evaluar tales pronósticos.

5.1. 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 5.1 se presentan las figuras de la variación en los precios de los alimentos y el índice SIPSA. Visualmente, se puede observar que, 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. 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 5.2 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 ($d = 1, D = 1$).

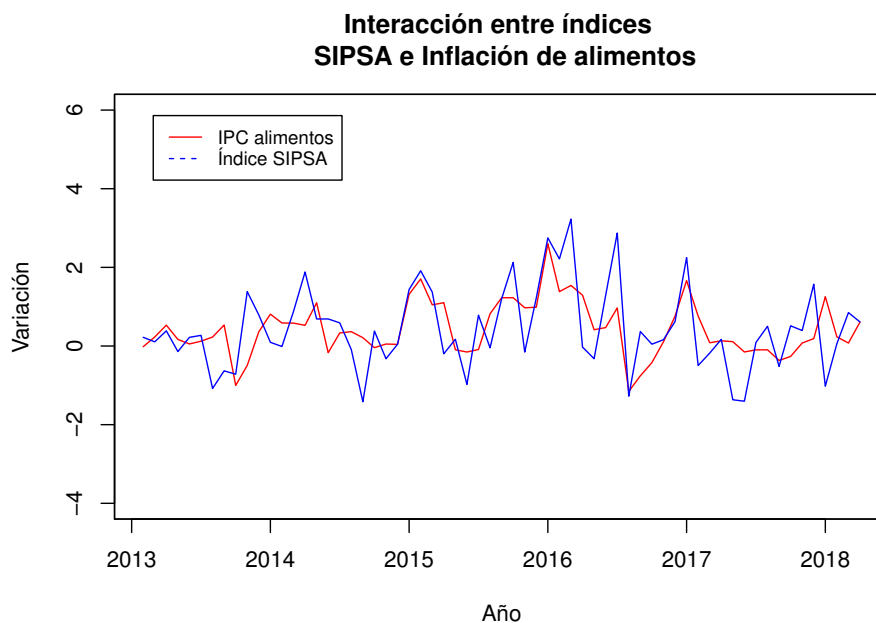


Figura 5.1: Interacción en el tiempo entre los índices considerados

Después de un proceso exhaustivo de búsqueda, en donde fueron tenidos en cuenta los siguientes criterios (se citan en orden aleatorio): 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 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, 1) × (0, 1, 1)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 5.1. La varianza estimada del modelo es 0.3523 y el BIC calculado es 80.75.

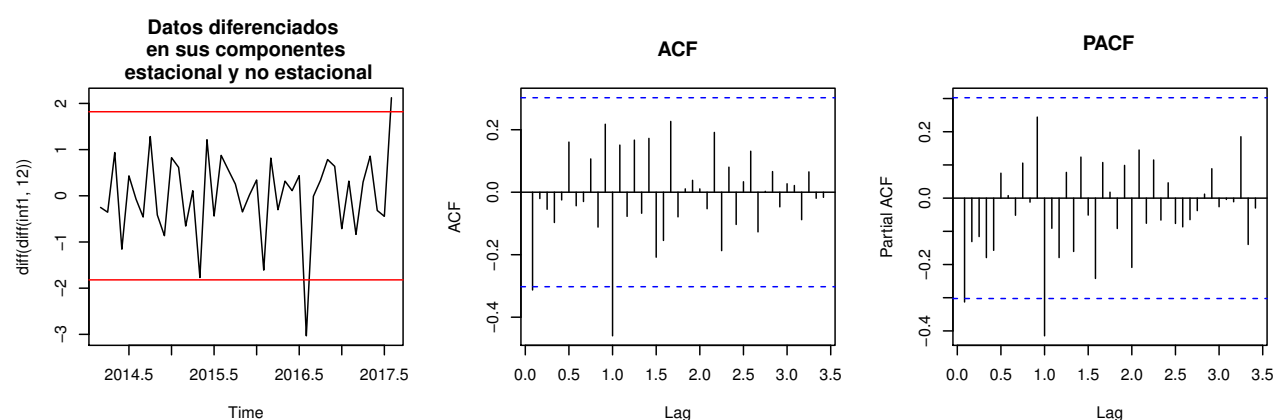


Figura 5.2: ACP y ACF para la serie IPC con diferencia en sus componentes estacional y no estacional

Parámetro	ϕ_{10}	θ_1	Θ_1	SIPSA _t	SIPSA _{t-1}
Estimación	-0.4	-0.665	-0.621	0.279	0.175
Error Estándar	0.148	0.202	0.263	0.059	0.058
t-valor	-2.696	-3.282	-2.361	4.684	3.019

Tabla 5.1: Estimativas para el modelo SARIMAX(10, 1, 1) × (0, 1, 1)₁₂

En la figura 5.3 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 es de $1,21 \times 10^{-9}$, sugiriendo, con una significancia del 5% que los residuos no se distribuyen normalmente.

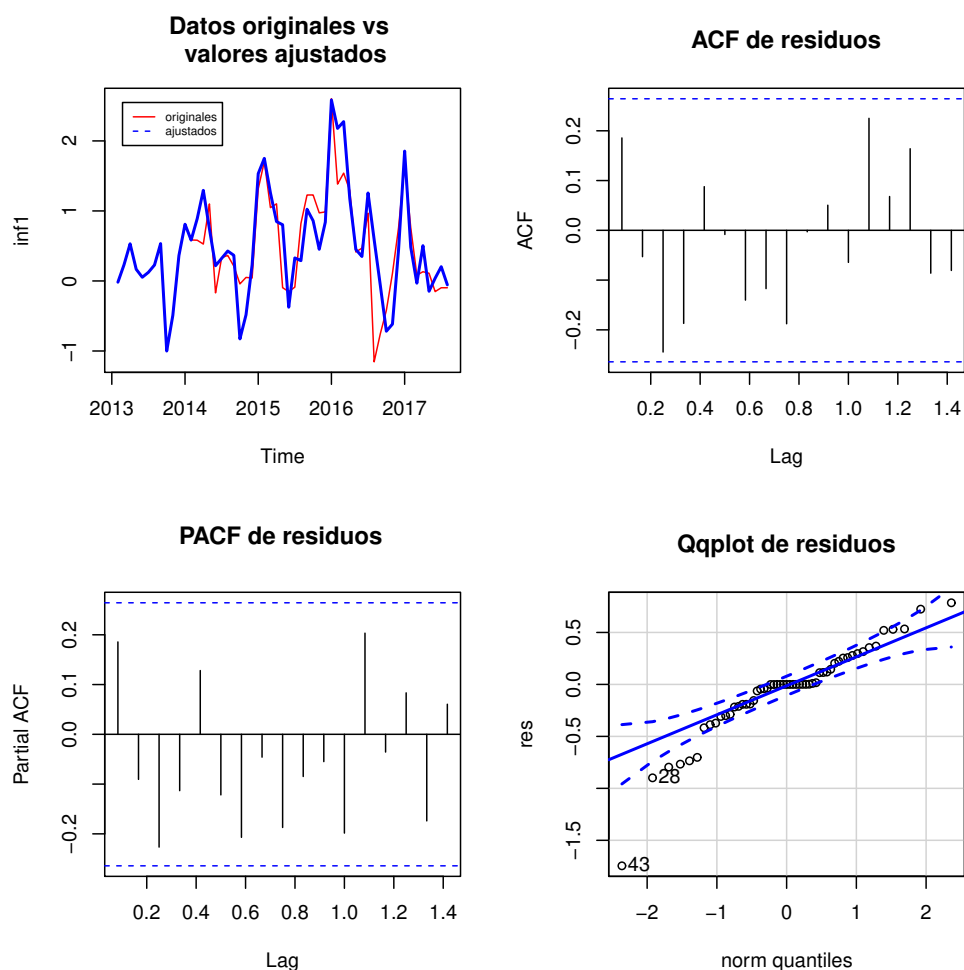


Figura 5.3: Validación del modelo $SARIMAX(10, 1, 1) \times (0, 1, 1)_{12}$ que incluye el valor contemporáneo y el primer rezago del índice SIPSA

Finalmente, para evaluar el modelo, de acuerdo a cómo fue dicho en el inicio de este capítulo, se evaluarán los pronósticos del modelo para ver si son razonablemente cercanos a los datos contenidos en la muestra out-sample. En la figura 5.4 se presentan, en rojo visualmente los pronósticos generados con el modelo dado en la tabla 5.1. Los datos originales observados se encuentran en rojo. Se observa que los valores estimados se encuentran razonablemente cerca a los valores originales de los datos. Adicionalmente, la tabla 5.2 presenta las medidas numéricas para evaluar los pronósticos. Nótese que este es el esquema de validación de modelos in-sample y out-sample comentado al inicio de este capítulo.

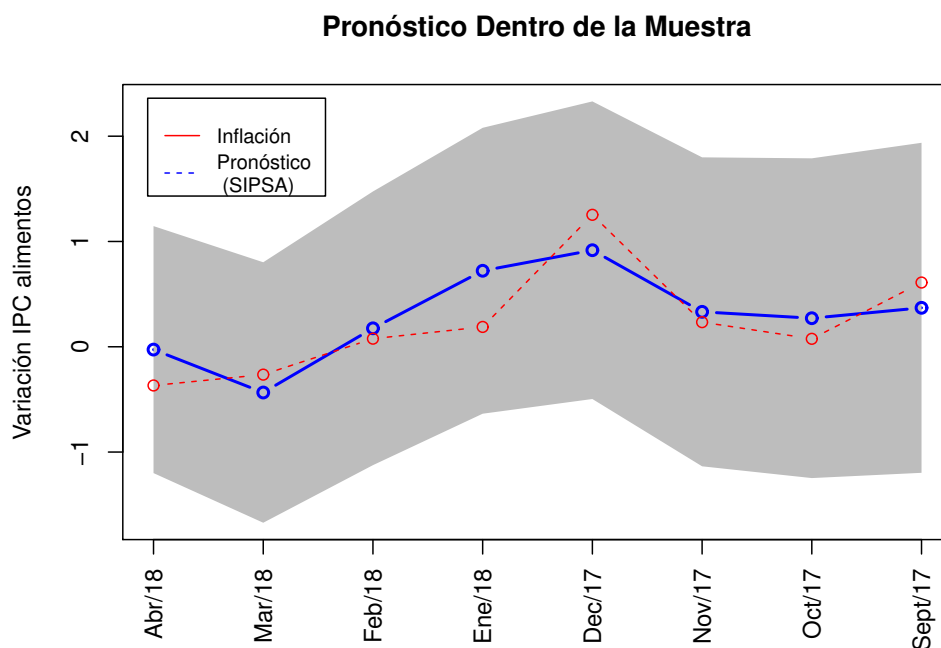


Figura 5.4: Pronóstico in-sample del modelo SARIMAX(10, 1, 1) × (0, 1, 1)₁₂ que incluye el valor contemporáneo y el primer rezago del índice SIPSA. Fancharts con un 95 % de confianza

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's
-0.065	0.286	0.251	-77.268	116.986	-0.470	0.847

Tabla 5.2: Medidas de error del modelo SARIMAX con out-sample

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 4.2, del valor contemporáneo (mismo mes) y el primer rezago (mes inmediatamente anterior).

5.2. 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. En la figura 5.5 se presenta el autocorrelograma cruzado 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 5.1 incluía la misma dinámica.

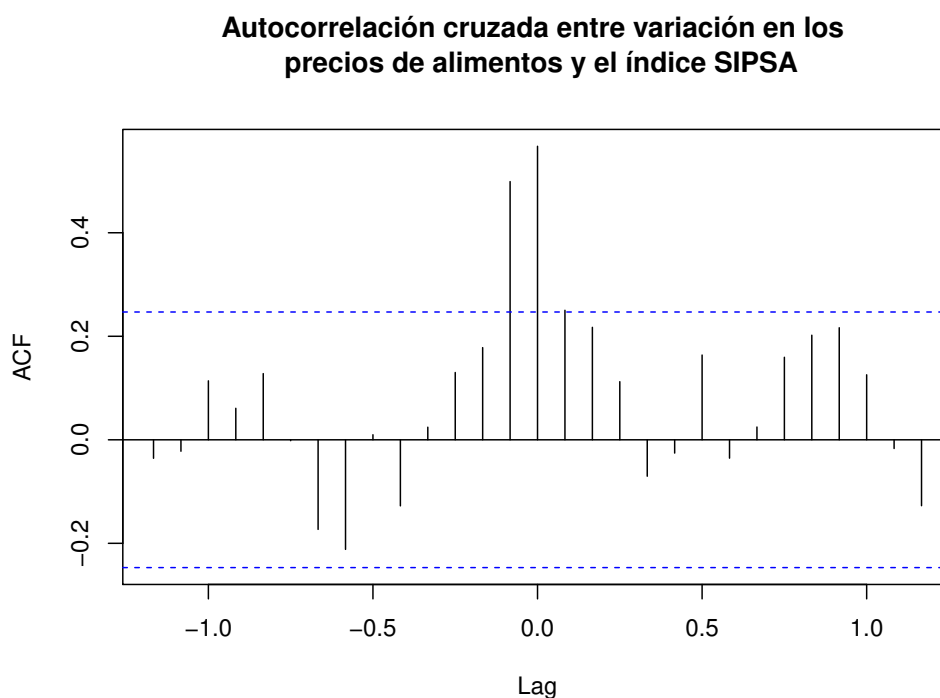


Figura 5.5: Autocorrelograma cruzado entre la variación en los precios de los alimentos y el índice SIPSA

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:

$$\begin{cases} \text{Inf}_t &= 0,234 + 0,0002t + 0,378\text{Inf}_{t-1} + 0,136\text{SISPA}_{t-1} + \sum_{i=1}^{11} \text{SDinf}_i \mathbf{1}_i, \\ \text{SIPSA}_t &= 0,220 + 0,002t + 0,234\text{Inf}_{t-1} + 0,195\text{SISPA}_{t-1} + \sum_{i=1}^{11} \text{SDSIPSA}_i \mathbf{1}_i \end{cases} \quad (5.1)$$

donde $\mathbf{1}_i$ es una variable dummy que toma el valor 1 cuando i representa un mes en particular y 0 en otro caso. Las constantes dadas en la tabla 5.3 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

$$\hat{\Sigma} = \begin{bmatrix} 0,270 & 0,275 \\ 0,275 & 1,050 \end{bmatrix}$$

Una vez establecido este modelo, se procede a validarlo. En las figuras 5.6 y 5.7 se presentan los ACF y PACF (lineal y al cuadrado) y el ACF cruzado para los rezagos de las variables de forma individual y de forma conjunta, respectivamente. 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, 8.559e-04, 0.3, con lo cual, con un 5% de significancia, hay evidencia estadísticamente significativa para concluir que los residuos no están correlacionados serialmente, son homoscedásticos pero no son normales. Con lo cual se concluye que los residuos son ruido blanco no gaussiano. En la figura 5.8 se presentan los valores observados contra los valores ajustados por el modelo.

i	1	2	3	4	5	6	7	8	9	10	11
SDinf $_i$	-1.021	-1.044	-0.996	-1.294	-1.394	-1.053	-1.49	-0.937	-1.459	-1.183	-0.861
SDSIPSA $_i$	-1.165	-0.644	-1.302	-1.821	-1.453	-0.450	-2.014	-1.383	-0.959	-1.182	-0.795

Tabla 5.3: valores para las variables dummies que explican la estacionalidad de los datos

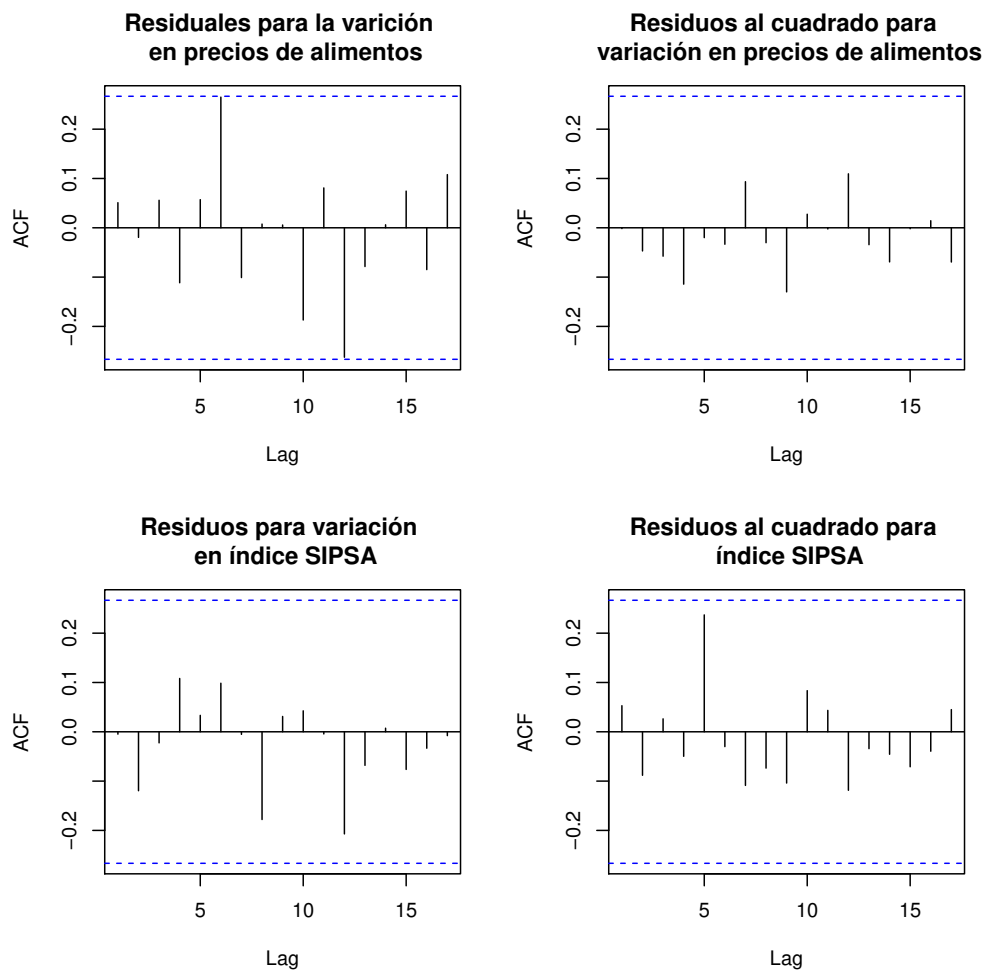


Figura 5.6: Correlogramas cruzados para el modelo ajustado

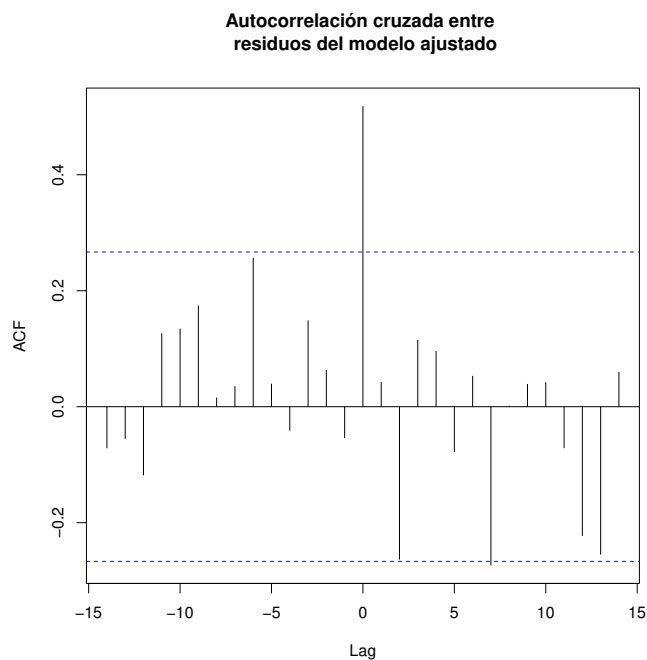


Figura 5.7: Autocorrelograma cruzado entre los residuos del modelo ajustado

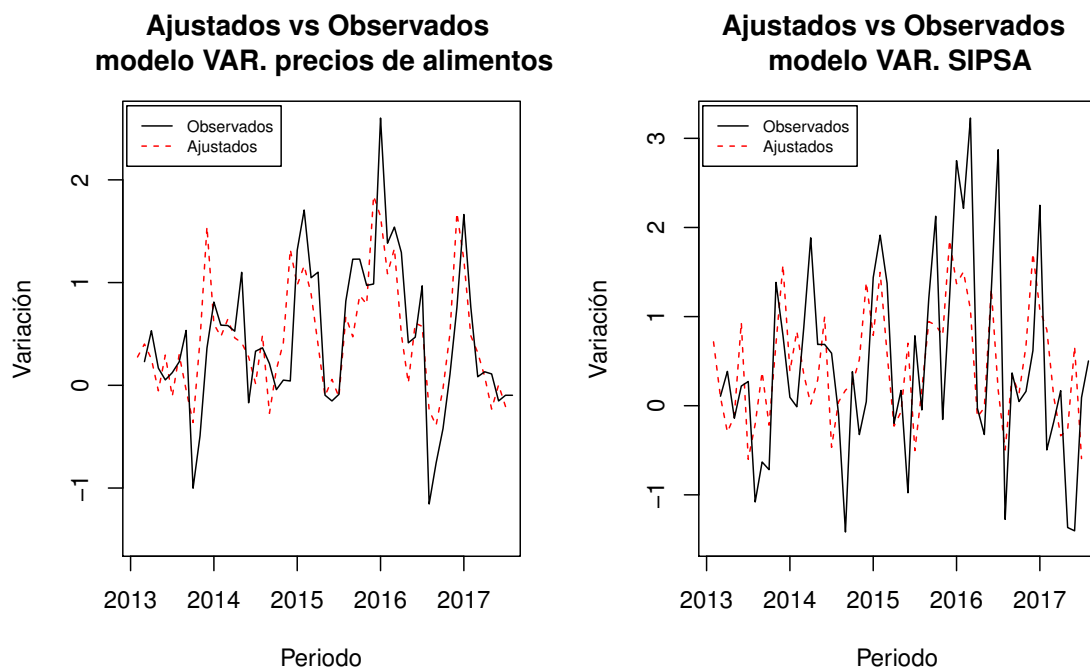


Figura 5.8: Estabilidad estructural del modelo propuesto

Una vez establecido el modelo VAR, se procede a hallar su forma estructural. Ésta será determinada con las matrices de restricciones A y B dadas por

$$A = \begin{bmatrix} \text{NA} & \text{NA} \\ 0 & \text{NA} \end{bmatrix} \quad B = \begin{bmatrix} \text{NA} & 0 \\ 0 & \text{NA} \end{bmatrix}$$

La idea general es integrar la relación en el corto plazo entre las variables que están siendo consideradas. El modelo SVAR está determinado por la ecuación (5.1), la forma estructural dada en la ecuación (3.4) 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 5.9 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 entonces que el modelo es estable y por tanto, válido para establecer las relaciones que se presentan a continuación.

En la figura 5.10 se presentan los impulsos asociados al modelo SVAR propuesto. En la figura (a) es 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 índice SIPSA. Adicionalmente, tales impactos son positivos lo que significa que ambos índices se mueven en la misma dirección en el tiempo.

En las figuras 5.11 y 5.12 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 se mueve libremente puesto que la contribución de la variación en los precios es apenas marginal. En conclusión, el índice SIPSA propuesto es útil para explicar la variación en los pronósticos en precios de alimentos.

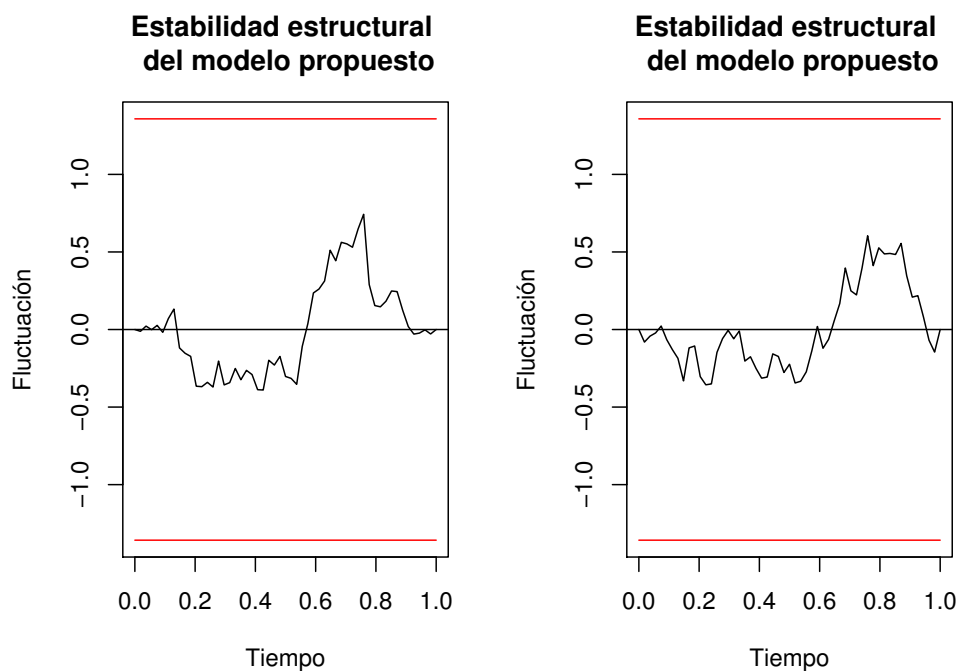
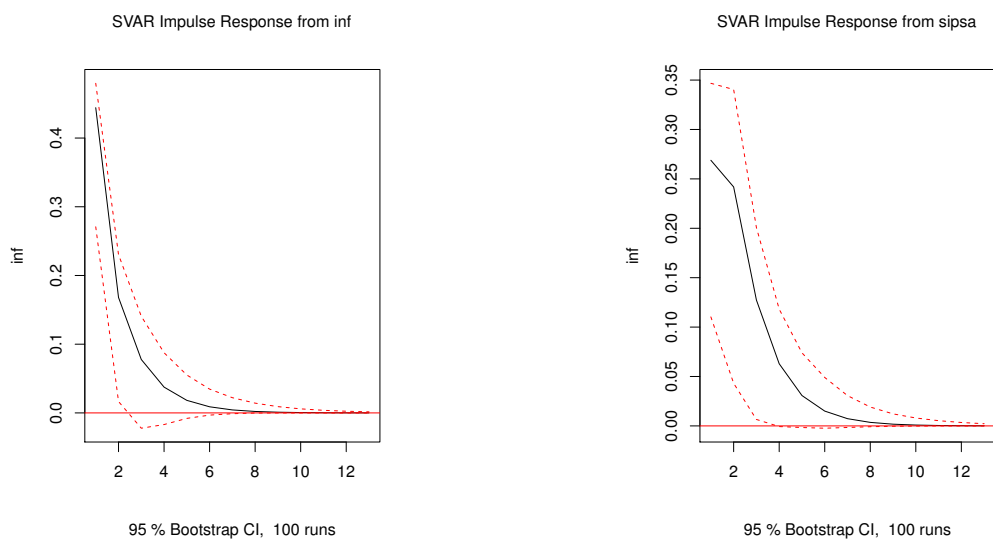


Figura 5.9: Estabilidad estructural del modelo propuesto



(a) Impulso de la variación en precios de alimentos

(b) Impulso de índice SIPSA

Figura 5.10: Impulsos asociados al modelo

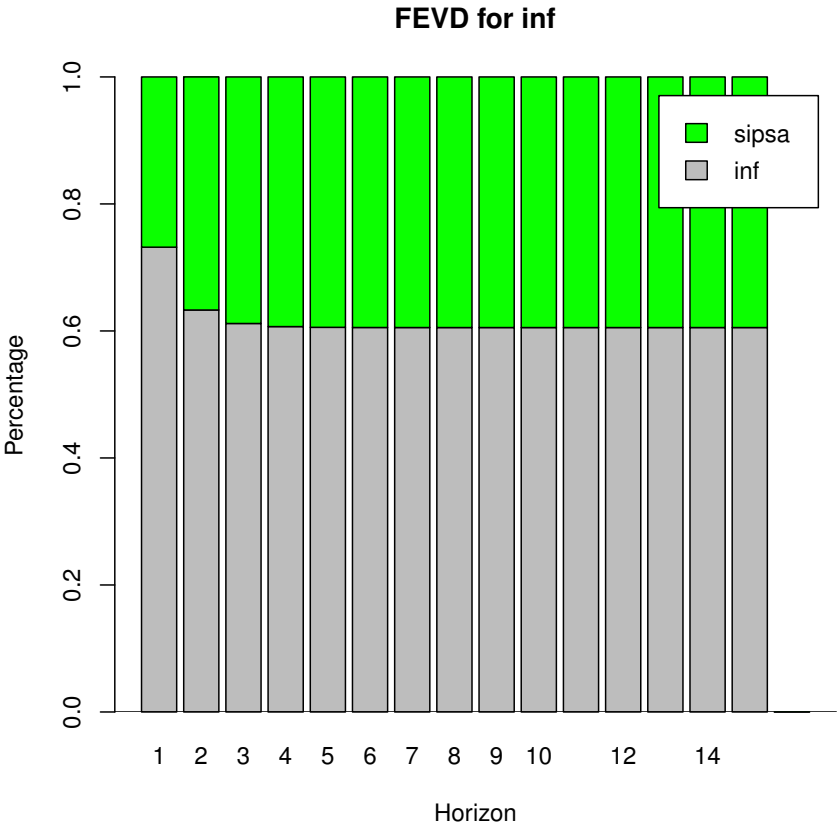


Figura 5.11: Descomposición en la varianza del error pronóstico para la variación en los precios de los alimentos

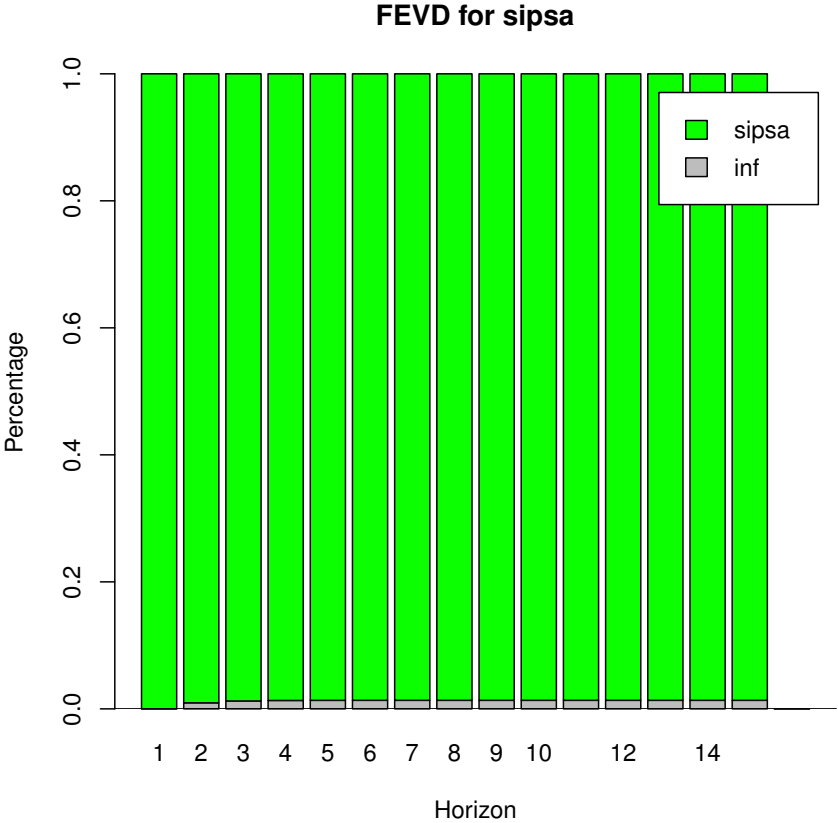


Figura 5.12: Descomposición en la varianza del error pronóstico para el SIPSA

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 validará la información pronostica con aquella que fue observada y se encuentra en el out-sample. Por limitaciones computacionales, no es posible la generación de pronósticos conjuntos a partir del modelo SVAR propuesto. Por lo tanto, serán generados tales pronósticos con el modelo VAR propuesto al inicio de esta sección. La figura 5.13 presenta el pronóstico para el índice SIPSA. Salvo por el punto correspondiente a Diciembre de 2017, los pronósticos se encuentran por dentro del fanchart del 95% de confianza. Análogamente, en la figura 5.14 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, las tablas 5.4 y 5.5 presentan las medidas numéricas para evaluar los pronósticos.

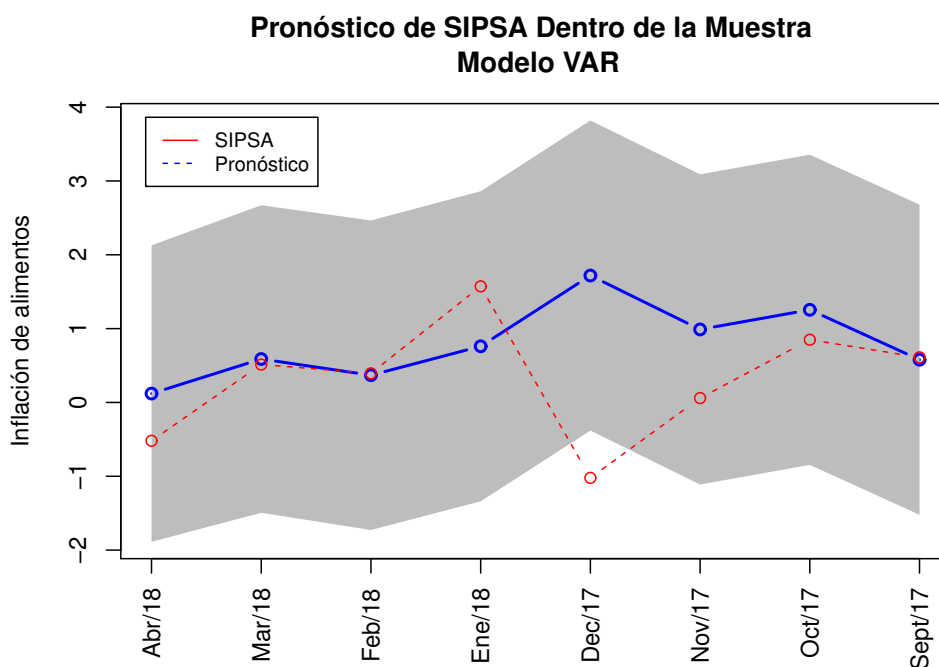


Figura 5.13: Pronóstico conjunto para SIPSA usando el modelo VAR

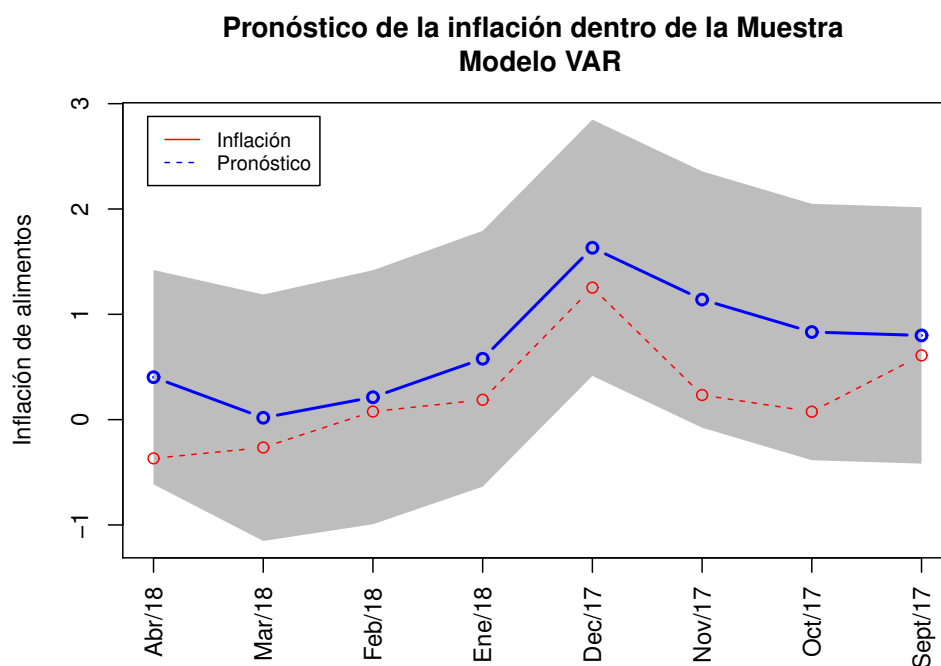


Figura 5.14: Pronóstico conjunto para la variación de los precios de los alimentos usando el modelo VAR

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

Tabla 5.4: Medidas de error del modelo VAR con out-sample para variación en los precios de los alimentos

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's
-0.289	1.03	0.736	-77.5	223	-0.155	0.45

Tabla 5.5: Medidas de error del modelo VAR con out-sample para SIPSA

5.3. 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 TRACE y MAXEIG que genera el test de cointegración de Johansen. En las dos pruebas deben aparecer vectores de cointegración para sustentar el uso del modelo.

Los resultados computaciones muestran el rechazo de la hipótesis nulo 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 5.6. 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. Para las pruebas citadas, se obtuvieron p -valores de 0.6, 0.0001 y 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 pero no son normales. Con lo cual se concluye que los residuos son ruido blanco no gaussiano.

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 uno. De acuerdo a (Engle and Granger, 1987) este vector presente equilibrio en el largo plazo por lo tanto se puede representar de la siguiente manera:

$$y_t = \text{Variación en precios de alimentos}_t - 1,2218\text{SIPSA}_t - 0,0007t \sim I(0) \quad (5.2)$$

La ecuación (5.2) 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. En el tiempo, se puede decir, que siempre estarán mostrando una interacción dinámica similar.

	Modelo Inf	Modelo SIPSA
ect1	0.09 (0.11)	0,85*** (0.21)
constant	0.02 (0.08)	0.11 (0.14)
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 5.6: Modelo VEC ajustado

La figura 5.15 presenta el pronóstico para el índice SIPSA. Salvo por el punto correspondiente a Diciembre de 2017, los pronósticos se encuentran por dentro del fanchart del 95 % de confianza. Análogamente, en la figura 5.16 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. 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 este capítulo es correcto pues se llegó a conclusiones similares de pronóstico con dos modelos diferentes. Adicionalmente, las tablas 5.7 y 5.8 presentan las medidas numéricas para evaluar los pronósticos.

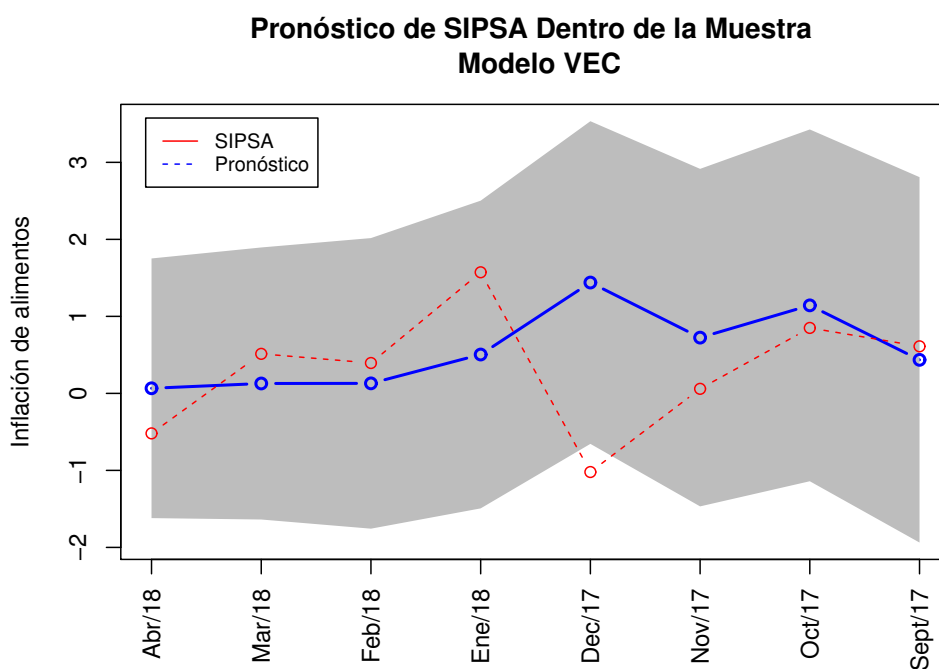


Figura 5.15: Pronóstico conjunto para SIPSA usando el modelo VEC

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

Tabla 5.7: Medidas de error del modelo VEC con out-sample para variación en los precios de los alimentos

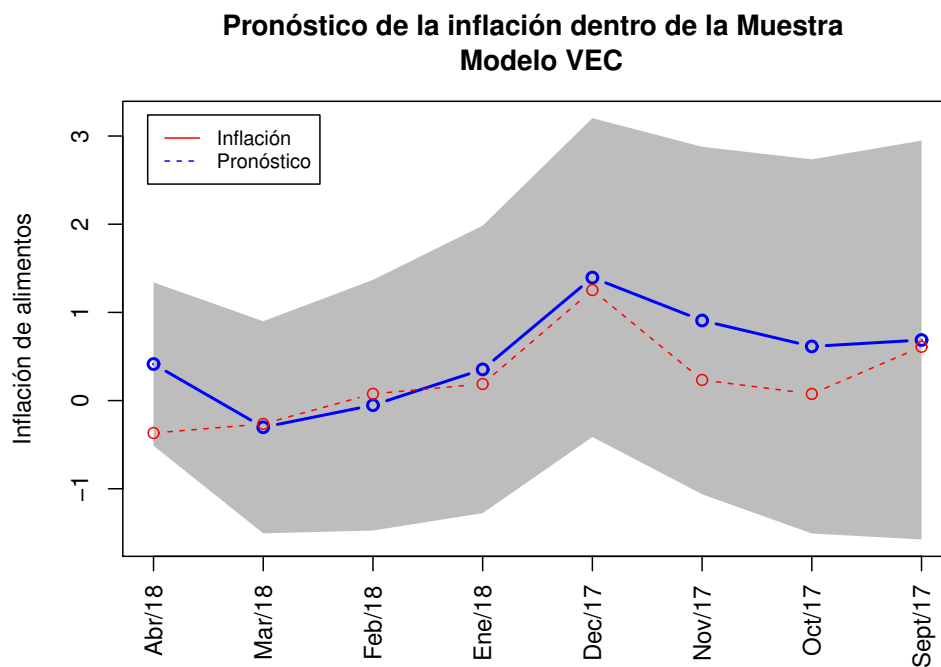


Figura 5.16: Pronóstico conjunto para la variación de los precios de los alimentos usando el modelo VEC

ME	RMSE	MAE	MPE	MAPE	ACF1	Theil's
-0.264	1.02	0.737	-70.6	219	-0.156	0.427

Tabla 5.8: Medidas de error del modelo VEC con out-sample para SIPSA

CAPÍTULO 6

Conclusión y recomendaciones

En este capítulo presentamos, de forma resumida, los principales resultados en el presente trabajo. Vale recordar que el objetivo principal es la obtención de información auxiliar que ayude a generar buenos pronósticos para la variación en los precios de los alimentos.

- 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 notablemente competitivos.

- 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 muestra competitivo 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.

CAPÍTULO 7

Trabajo Futuro

En general, el trabajo 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.
- Recomendar la creación de un índice institucional (por parte del DANE) de variación de 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 consume contribuye un 30 % en la inflación de alimentos).

Referencias bibliográficas

- Box, G.E.P.; Jenkins, G. (1970). *Time series analysis: forecasting and control*. Holden-Day series in time series analysis. Holden-Day.
- Cavaliere, G. (2003). Asymptotics for unit root tests under markov regime-switching. *The Econometrics Journal*, 6.
- Dickey, David A.; Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74.
- Engle, R. F. and Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55.
- Guerrero, V. (2003). *Análisis Estadístico de Series de tiempo económicas*. International Thomson Editores, 2 edition.
- Ihaka, R. (1996). R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics*, 5.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica*, 59.
- Knuth, D. E. (1986). *Computers Typesetting, Volume A: The TeXbook*. Addison-Wesley Professional.

- Lütkepohl, H. (2005). *New Introduction To Multiple Time Series Analysis*. Springer.
- Metcalf, A.V.; Cowpertwait, S. (2009). *Introductory Time Series with R. Use R*. Springer-Verlag New York, 1 edition.
- Pfaff, B. (2008). *Analysis of Integrated and Cointegrated Time Series with R (Use R)*. Springer, 2nd edition.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6.
- Tsay, R. S. (2012). *An Introduction to Analysis of Financial Data with R*. Wiley, 1 edition.
- Tsay, R. S. (2013). *Multivariate Time Series Analysis: With R and Financial Applications*. Wiley, 1 edition.