

ENSAYO 1: MODELOS UNIVARIADOS - ARIMA

Roger Alejandro Banegas Rivero^σ

UNIVERSIDAD AUTÓNOMA GABRIEL RENÉ MORENO

Este diseño, 15 de abril de 2022

1. MODELOS UNIVARIADOS DE SERIES DE TIEMPO

Los modelos *ARIMA* (*AR = autoregresivo, I = integrado y MA = media móvil*) corresponden a una estimación econométrica de tipo univariada, lo cual significa la con una sola variable (y_t). En consecuencia, a partir de los valores pasados de una variable, se puede predecir el presente y el futuro de la misma. Este tipo de modelo corresponden a estimaciones de corto plazo; por lo cual, la serie de tiempo (y_t) debe ser “*estacionaria*”. Esto implica que una serie de tiempo para su modelación no debe tener tendencias, rumbos (al alza o a la baja): debe ser independiente del tiempo.

Para que una serie sea estacionaria se deben cumplir tres condiciones:

- i. El valor medio de una serie es contante (no depende del tiempo).

$$E(y_t) = \mu \quad \forall t = 1, 2, 3, \dots, \infty \quad (1)$$

- ii. La varianza de la serie de tiempo (y_t) es independiente del tiempo. Es constante.

$$\text{var}(y_t) = E(y_t - \mu)(y_t - \mu) = E(y_t - \mu)^2 = \sigma^2 \quad \forall t \quad (2)$$

- iii. La autocovarianza no depende del tiempo sino de la longitud o distancia entre los rezagos del tiempo.

$$E(y_{t_1} - \mu)(y_{t_2} - \mu) = \gamma_{t_2 - t_1} \quad \forall t_2, t_1 \quad (3)$$

Notación:

ARIMA (p, d, q)

AR(p) = Número de términos autoregresivos.

I(d) = Orden de integración en la estacioneridad de las variables.

MA(q) = Número de términos de media móvil

^σ Correo electrónico: rogerbanegas@uagrm.edu.bo

Especificación de un modelo $ARIMA(p, d, q)$

Vamos a partir de una serie de tiempo, la serie y_t :

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (4)$$

En la expresión (4), se deberá verificar que:

$$\phi_1 < 1 \wedge \theta_1 < 1$$

Existen tres órdenes de integración ($I = d = 0, 1, 2$):

- a) $I = 0$ -> La variable es estacionaria en niveles (y_t).
- b) $I = 1$ -> La variable es estacionaria después de una primera diferencia ($\Delta y_t = y_t - y_{t-1}$).
- c) $I = 2$ -> La variable o serie de tiempo es estacionaria después de una segunda diferencia ($\Delta \Delta y_t$).

Una segunda notación alternativa en modelos $ARIMA(p, d, q)$:

$$y_t = \mu + \sum_{i=0}^p \phi_i y_{t-i} + \sum_{j=0}^q \theta_j \varepsilon_{t-j} + \varepsilon_t \quad (5)$$

Una tercera notación de modelos $ARIMA(p, d, q)$, con operadores polinomiales $L^i y_t$ en rezago y_{t-i}

Ejemplos:

$$y_{t-1} = L y_t$$

$$y_{t-2} = L^2 y_t$$

$$y_{t-3} = L^3 y_t$$

$$y_{t-4} = L^4 y_t$$

De igual manera, existen operadores polinomiales de rezagos (L^i) para los términos MA o de media móvil: $L^i \varepsilon_t$ en rezago ε_{t-i} :

$$\varepsilon_{t-1} = L \varepsilon_t$$

$$\varepsilon_{t-2} = L^2 \varepsilon_t$$

$$\varepsilon_{t-3} = L^3 \varepsilon_t$$

$$\varepsilon_{t-4} = L^4 \varepsilon_t$$

A partir de una notación (5), introduciendo, términos de operadores polinomiales de rezago:

$$y_t = \mu + \phi_1 L y_t + \phi_2 L^2 y_t + \dots + \phi_p L^p y_t + \theta_1 L \varepsilon_t + \theta_2 L^2 \varepsilon_t + \dots + \theta_q L^q \varepsilon_t + \varepsilon_t$$

$$y_t - \phi_1 L y_t - \phi_2 L^2 y_t - \dots - \phi_p L^p y_t = \mu + \theta_1 L \varepsilon_t + \theta_2 L^2 \varepsilon_t + \dots + \theta_q L^q \varepsilon_t + \varepsilon_t$$

$$y_t(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) = \mu + (\theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q + 1) \varepsilon_t$$

(6)

a) Condición de estacionariedad e invertibilidad

Los componentes $AR(p)$ deben ser estacionarios y los componentes $MA(q)$ deben ser invertibles. En cada caso las raíces y soluciones deben estar fuera del círculo unitario o raíces inversas dentro del círculo unitario.

Ejemplo de estacionariedad de un modelo $AR(1)$:

$$y_t = \mu + \phi_1 y_{t-1} + \varepsilon_t, \text{ asumiendo que } \mu = 0, \phi_1 = 0.7 \quad (7)$$

Llevando a operadores polinomiales de rezago:

$$y_t = 0.7 L y_t + \varepsilon_t \quad (8)$$

$$y_t - 0.7 L y_t = \varepsilon_t$$

$$y_t(1 - 0.7 L) = \varepsilon_t$$

Encontrando la raíz o solución de L :

$$1 - 0.7 L = 0 \quad (9)$$

$$L = \frac{1}{0.7} = 1.43 \rightarrow |L| > 1 \quad \vee \quad |L|^{-1} < 1; \therefore \text{El modelo es estacionario}$$

Ejemplo de no invertibilidad de un modelo $MA(1)$:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t, \text{ asumiendo que } \mu = 0, \theta_1 = 1.2 \quad (10)$$

Llevando a operadores polinomiales de rezago:

$$y_t = 1.2L\varepsilon_t + \varepsilon_t \quad (11)$$

$$y_t = \varepsilon_t(1.20L + 1)$$

Encontrando la raíz o solución de L :

$$1.20L + 1 = 0 \quad (12)$$

$$L = -\frac{1}{1.2} = 0.83 \rightarrow |L| < 1 \quad \vee \quad |L|^{-1} > 1; \therefore \text{El modelo no es invertible}$$

b) Especificación de modelos *SARIMA*

Una extensión de los modelos *ARIMA* se basa en la incorporación de la estacionalidad (*seasonality*) ; en tanto, se incluye a una especificación de tipo *SARIMA*:

AR = Autoregresivos (p)

I = Integrados (d)

MA = Media móvil (q)

S = estacional (*Seasonality*)

SARIMA (p, d, q)(P, D, Q) $_M$

(p, d, q) < -procesos *arima*

(P, D, Q) $_M$ < -procesos *estacionales*

$M = 12$; *datos mensuales*

$M = 4$; *datos trimestrales*

$M = 2$; *datos semestrales*

$M = 3$; *datos cuatrimestrales*

$M = 365$; *datos diarios*

Lectura de un modelo *SARIMA*

Ejemplo: SARIMA (2,0,1)(1,0,0)₁₂

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \theta_1 \varepsilon_{t-1} + \phi_s y_{t-12} + \varepsilon_t \quad (13)$$

c) Descomposición de series de tiempo

La descomposición de factores de una serie de tiempo (y_t) encontrar cuatro componentes de forma respectiva mediante ARIMA CENSUS X-11 o TRAMO-SETS:

τ_t = Componente tendencial o componente de largo plazo.

ς_t = Componente cíclico.

s_t = Componente estacional.

ε_t = Componente irregular-estocástico

Existen tres tipos de descomposición de series de tiempo:

a) **Descomposición aditiva**

$$y_t = \tau_t + \varsigma_t + s_t + \varepsilon_t \quad (14)$$

La serie puede presentar negatividad o constituirse como no negativa.

Ej. LA inflación (>0; <0 deflación en caso contrario)

b) **Descomposición multiplicativa**

$$y_t = \tau_t * \varsigma_t * s_t * \varepsilon_t \quad (15)$$

La serie y_t es no negativa o en otras palabras es positiva $y_t > 0$

c) **Descomposición mixta**

Combina ambas descomposiciones: aditiva y multiplicativa

$$y_t = \tau_t * (1 + \varsigma_t)(1 + s_t) + \varepsilon_t \quad (16)$$

Se menciona que para obtener el componente cíclico se debe utilizar previamente filtros de series de tiempo (Hodrick-Prescot, Christiano & Fitzgerald, Baxter-King, Nadaraya-Watson, entre otros).

d) La metodología Box-Jenkins

Box y Jenkins (1976) establecieron una metodología para estimar los modelos ARIMA en cuatro pasos:

Cuadro 1. Metodología Box-Jenkins (1976)

Paso	Definición	Herramientas
1. Identificación	Consiste en identificar los posibles procesos $ARIMA(p, d, q)$.	Correlogramas: -Función de autocorrelación (FAC) para identificar procesos $MA(q)$; permite apreciar visualmente el orden de integración de una variable $I = \{0, 1, 2\}$. -Función de autocorrelación parcial (FACP) para identificar procesos $AR(p)$.
2. Estimación	Se refiere a encontrar las estimaciones de los parámetros o coeficientes	- Mínimos Cuadrados Ordinarios (MCO). - Máxima verosimilitud (ML).

3. Verificación o diagnóstico	Consiste en verificar que los coeficientes de autocorrelación se han de forma conjunta igual a cero.	- Prueba Ljung-Box (para muestras pequeñas.) - Prueba Box-Bierce (para muestras grandes).
4. Pronóstico	Si se cumplen los tres pasos previos, se puede utilizar el módulo para fines de pronóstico.	- Error medio (<i>ME</i>). - Error cuadrático medio o desviación de la raíz cuadrática media (<i>RMSE</i>). - Error medio absoluto (<i>MAE</i>). - Error medio porcentual (<i>MPE</i>). - Error medio absoluto porcentual (<i>MAPE</i>). - Error medio absoluto escalado (<i>MASE</i>).

Paqueterías y librerías utilizadas en RStudio, de forma inicial con
install.packages("...") (*Script 1*):

```
#### Script 1#####
#### Paqueterías y librerías en RStudio#####
library(psych) #Para análisis estadístico previo
library(colorspace) #Para gráficos y colores
library(ggplot2) #Para gráficos
library(tseries) #Para series de tiempo
library(forecast)#Para pronósticos y fanchart
library(fpp)#Para pronósticos y fanchart
library(seasonal) # Para ARIMA estacional
library(seasonalview)#Para ARIMA estacional
```

Simulación de ARIMA

- a) Se estiman cuatro series de modelos *ARIMA* (1,0,0) o *AR*(1) con diferentes coeficientes de términos autoregresivos (ϕ) (*Script 2*):

$$\hat{y}_t(1 - \hat{\phi}L) = \hat{\epsilon}_t \quad (17)$$

Donde $\epsilon_t \sim N_{iid}(0,1)$ para 100 realizaciones

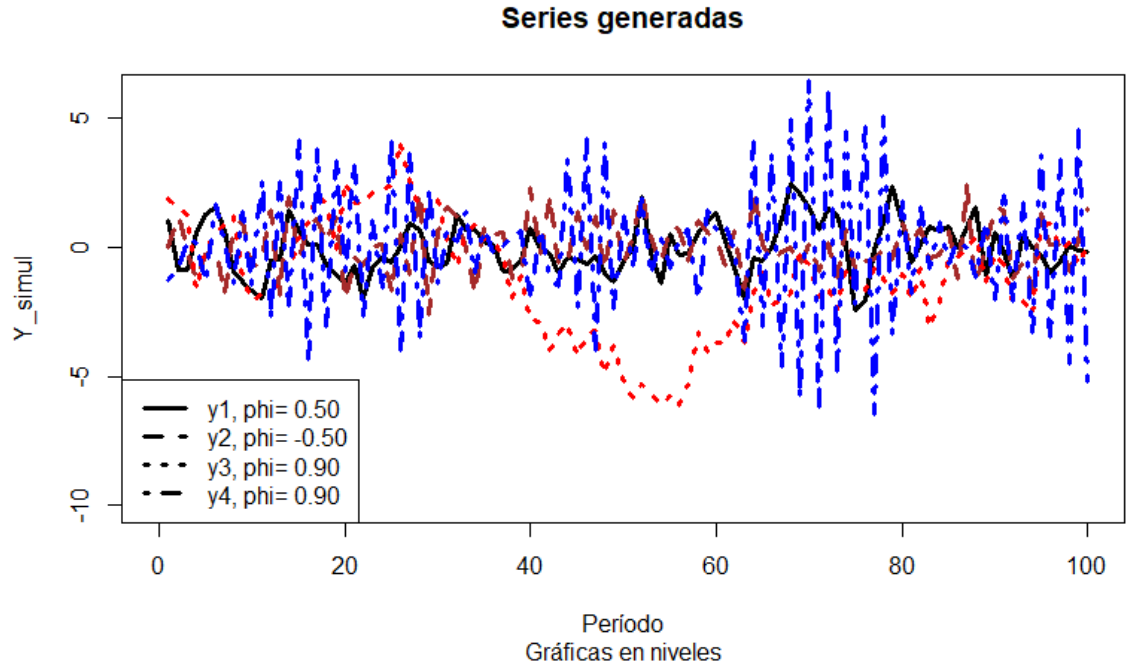
$$\hat{\phi} = 0.50; \hat{\phi} = -0.50; \hat{\phi} = 0.90; \hat{\phi} = -0.90$$

```
#### Script 2#####
#### Simulación de AR(1) en RStudio#####
set.seed(124)
y1_sim <- arima.sim(list(order = c(1,0,0), ar = 0.50), n = 100)
ts.plot(y1_sim)
##### Ejemplo, no ejecute#####

#### Script 3#####
#### Correlogramas en RStudio#####
#Gráfica en niveles
par(mfrow=c(2,1))
acf(y1_sim); pacf(y1_sim)

#Gráfica en primera diferencia
acf(diff(y1_sim)); pacf(diff(y1_sim))
par(mfrow=c(1,1))
##### Ejemplo, no ejecute#####
```

Figura 1. Cuatro series simuladas



y1 corresponde a un modelo ARIMA (1,0,0) $\phi_1 = 0.50$, línea color negra; y2 corresponde a un modelo ARIMA (0,0,1) $\phi_1 = -0.50$, línea color café; y3 corresponde a un modelo ARIMA (1,0,0) $\phi_1 = 0.90$, línea color roja; y4 corresponde a un modelo ARIMA (0,0,1) $\phi_1 = -0.90$, línea color azul.

De acuerdo a la figura 1, las series simuladas con los coeficientes más bajos, en términos absolutos, tienden a tener un comportamiento más estacionario. De las figuras 2 al 5, se identifican los correlogramas para procesos $AR(1)$ en función de cada coeficiente simulado.

Figura 2. Correlograma AR(1) $\hat{\phi} = 0.50$

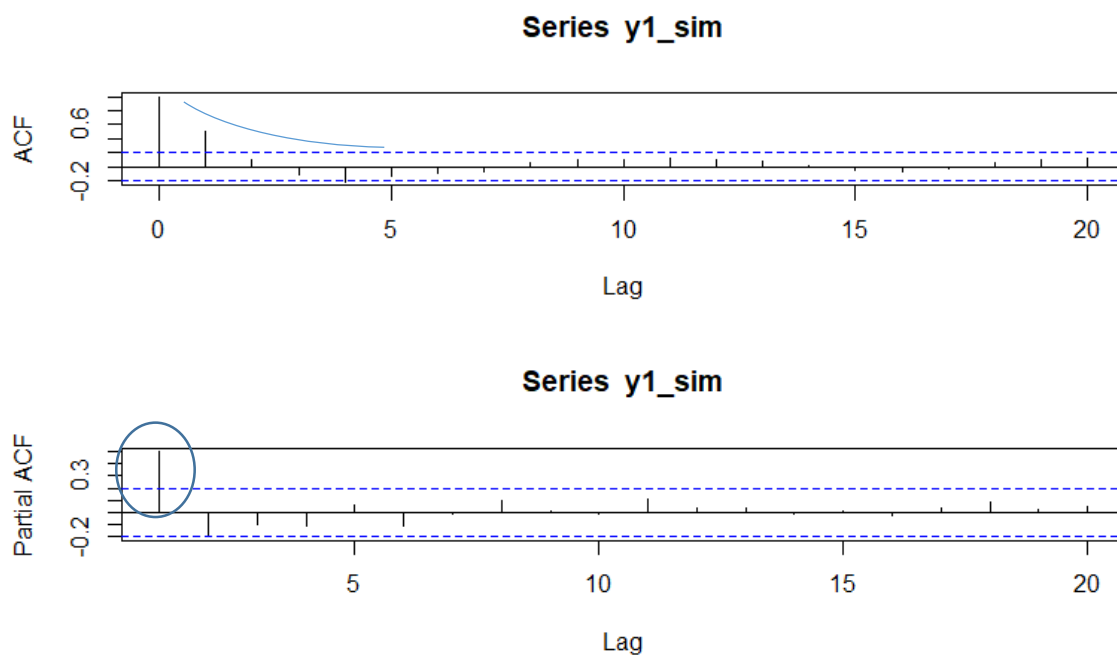


Figura 3. Correlograma AR(1) $\hat{\phi} = -0.50$

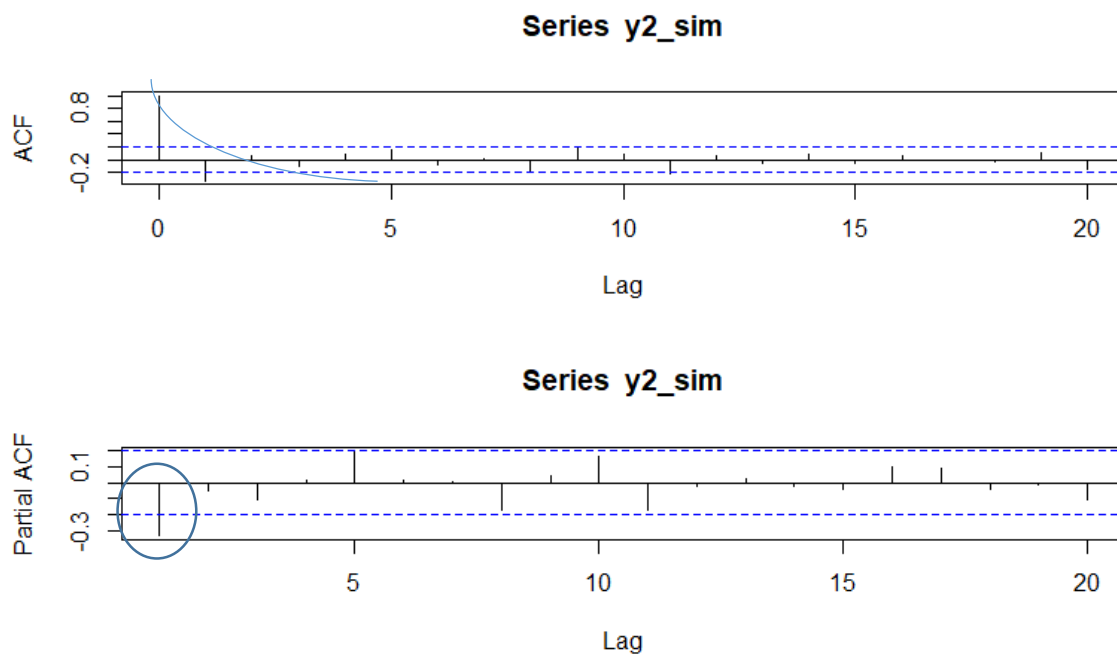


Figura 4. Correlograma AR(1) $\hat{\phi} = 0.90$

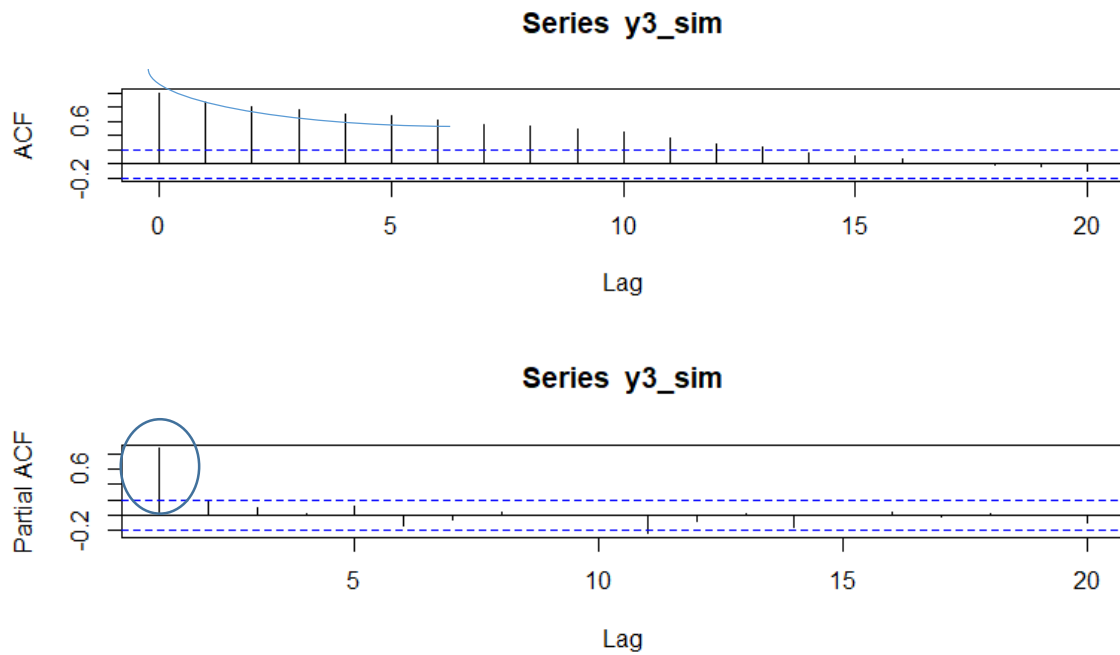
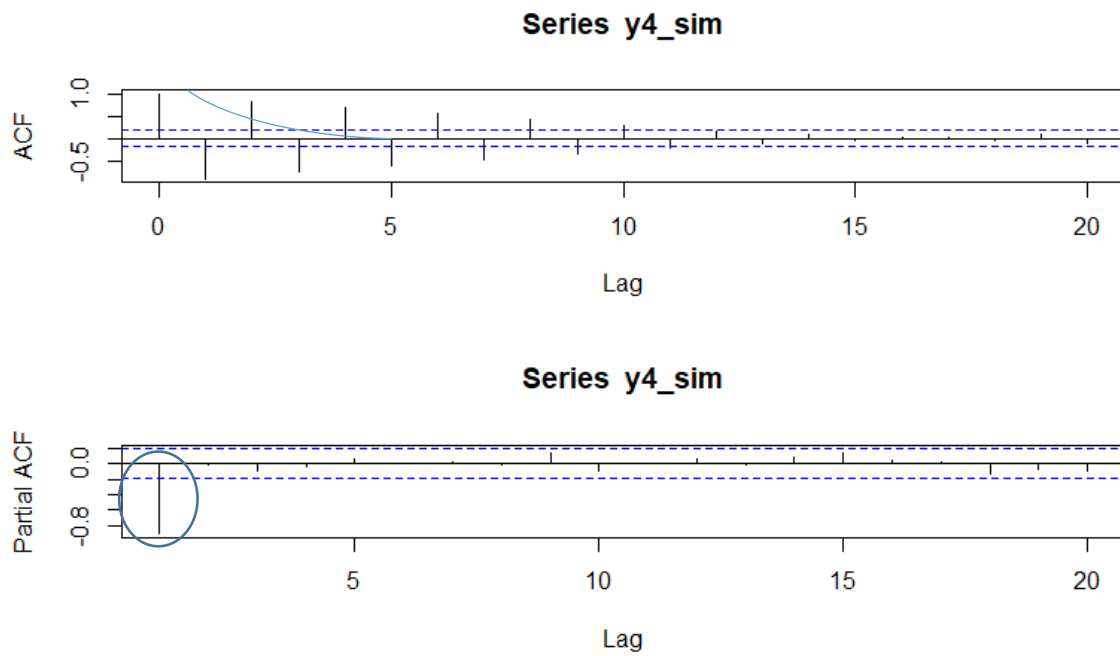


Figura 5. Correlograma AR(1) $\hat{\phi} = -0.90$



Para las figuras 4 y 5 respectivamente, la función de autocorrelación (FAC o ACF) va cayendo lentamente, lo cual indicaría que tendería a constituirse en una serie no estacionaria.

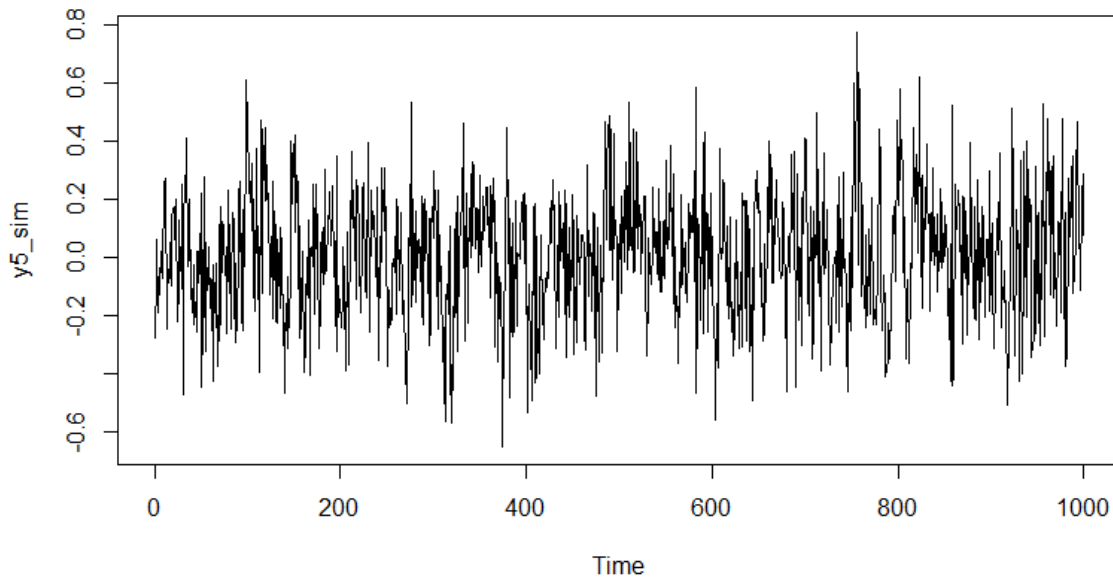
- b) Se estima un modelo ARIMA (3,0,0) o AR(3) con los siguientes coeficientes
(Script 3):

$$\hat{y}_t(1 - \hat{\phi}_1 L - \hat{\phi}_2 L^2 - \hat{\phi}_3 L^3) = \hat{\epsilon}_t \quad (18)$$

Donde $\epsilon_t \sim N_{iid}(0,0.2)$ para 1000 realizaciones

$$\hat{\phi}_1 = 0.10; \phi_2 = 0.20; \phi_3 = 0.10$$

Figura 6. ARIMA (3,0,0) o AR(3) simulado



De acuerdo con la figura 7, la FCA (AFC) cae rápidamente y se vuelve igual a cero con la identificación de tres saltos en la FACP (PAFC).

```
#### Script 3#####
#### Simulación ARIMA (3,0,0) en RStudio#####
# Construcción de una serie
# 1000 valores simulados de media 0 y sd. 0.2 asumiendo un modelo
AR (3)
```

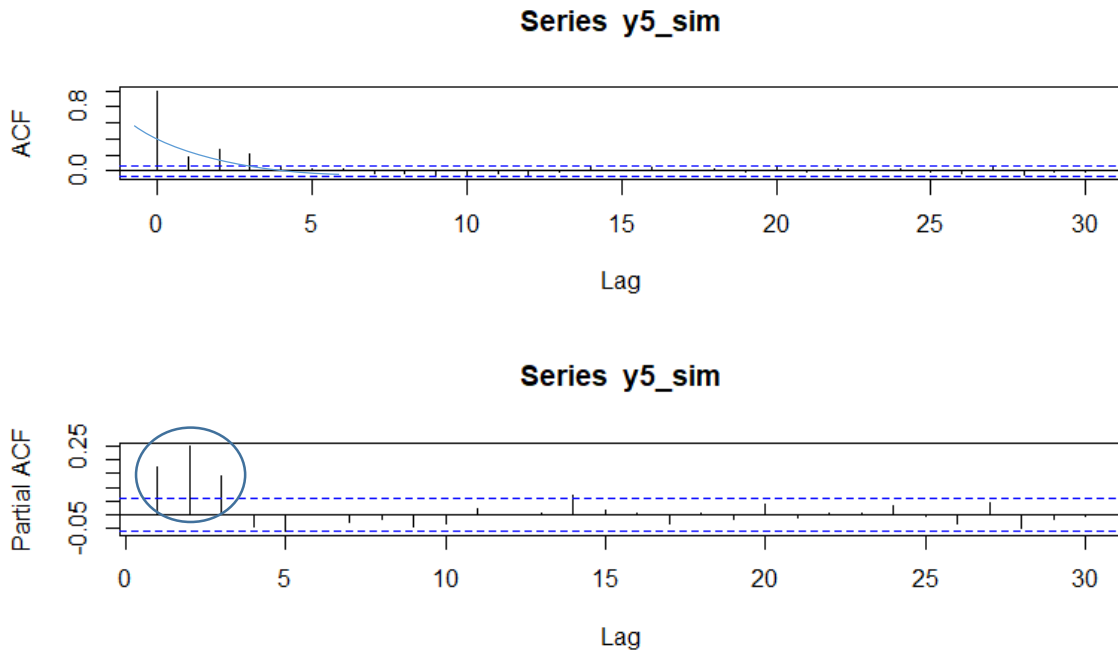


```

error.model=function(n){rnorm(n, mean=0, sd=.2)}
y5_sim=arima.sim(model=list(ar=c(0.1, 0.2, 0.1)), n=1000,
                 rand.gen=error.model )
##### Ejemplo, no ejecute#####

```

Figura 7. Correlograma de AR(3)



- c) Se estima un modelo ARIMA (0,0,2) o MA(2) con los siguientes coeficientes
(Script 4):

$$\hat{y}_t = \hat{\epsilon}_t(1 + \hat{\theta}_1 L + \hat{\theta}_2 L^2) \quad (19)$$

Donde $\epsilon_t \sim N_{iid}(0,0.5)$ para 1000 realizaciones

$$\hat{\theta}_1 = 0.80; \theta_2 = 0.70$$

Según la figura 9, la FACP cae rápidamente y se vuelve igual a cero con dos saltos significativos en la FAC por lo cual se identifican dos procesos MA.

```
#### Script 4#####
#### Simulación ARIMA (0,0, 2) en RStudio#####
# Se estima un modelo=list(MA=c(0.8, 0.7))
# con desv. Est. de .5
error.model=function(n){rnorm(n, mean=0, sd=.5)}
y6_sim=arima.sim(model=list(ma=c(0.8, 0.7)), n=1000,
                  rand.gen=error.model )
##### Ejemplo, no ejecute#####
```

Figura 8. ARIMA (0,0,2) o MA(2) simulado

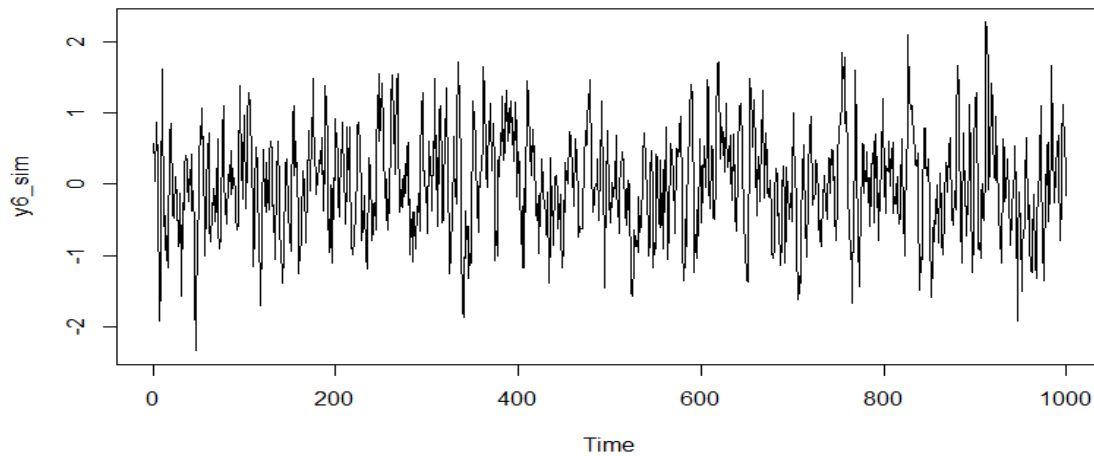
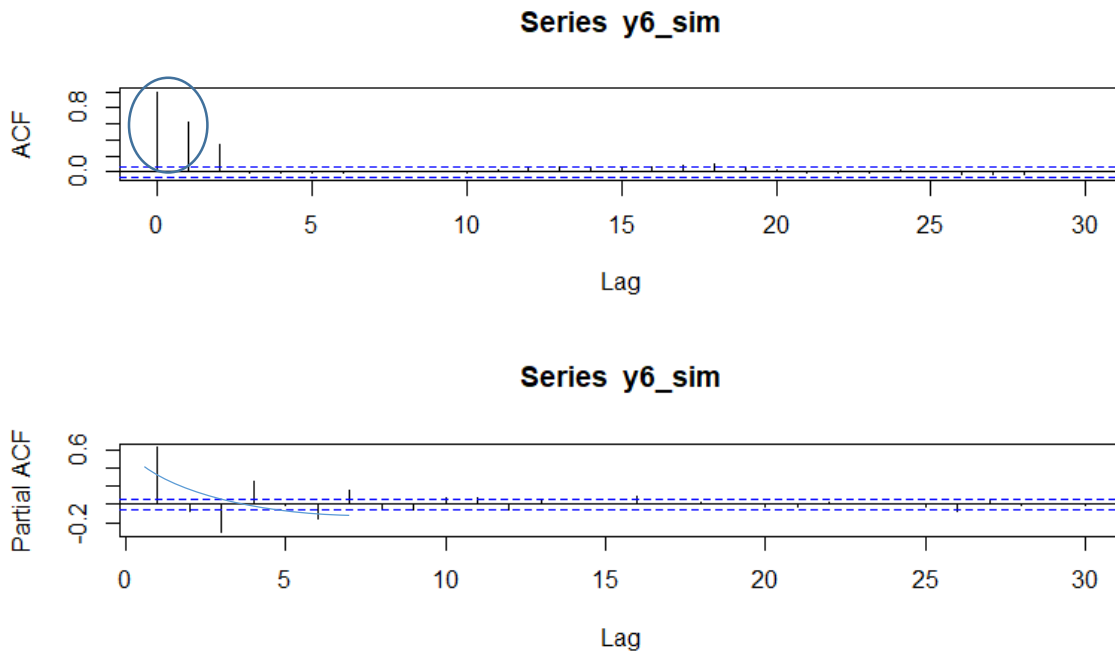


Figura 9. Correlograma de MA(2) o ARIMA(0,0,2)



- d) Se estima un modelo ARIMA (3,0,2) o ARMA(3,2) con los siguientes coeficientes (*Script 5*):

$$\hat{y}_t(1 - \widehat{\phi}_1 L - \widehat{\phi}_2 L^2 - \widehat{\phi}_3 L^3) = \hat{\epsilon}_t(1 + \widehat{\theta}_1 L + \widehat{\theta}_2 L^2) \quad (20)$$

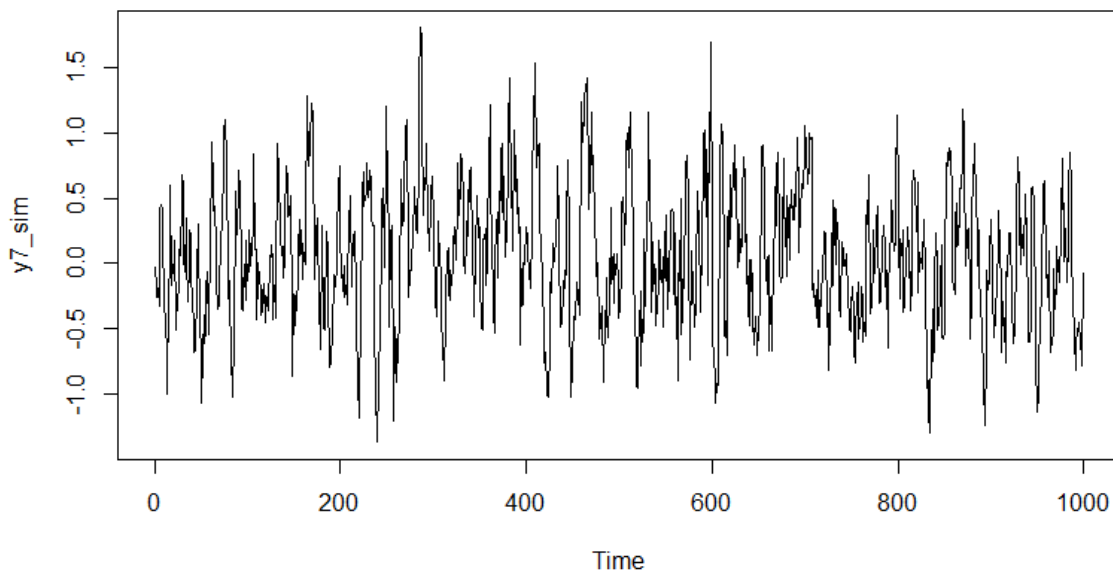
Donde $\epsilon_t \sim N_{iid}(0,0.3)$ para 1000 realizaciones

$$\widehat{\phi}_1 = 0.10; \phi_2 = 0.20; \phi_3 = 0.10;$$

$$\widehat{\theta}_1 = 0.80; \theta_2 = 0.70$$

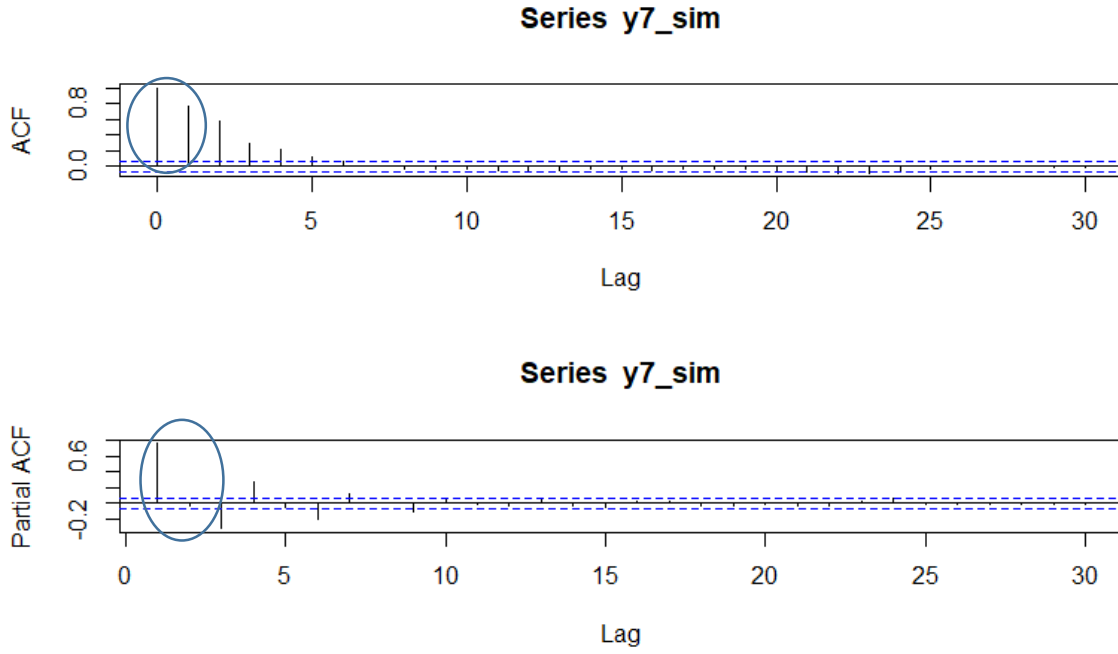
```
#### Script 5#####
#### Simulación ARIMA (3,0,2) en RStudio#####
# Se simula un ARIMA(3,0,2) con ruido blanco y DS de 0.3,
# con phi=0.1, 0.2, 0.1 and theta=0.8, 0.7
error.model=function(n){rnorm(n,mean=0, sd=.3)}
y7_sim=arima.sim(model=list(ar=c(0.1, 0.2, 0.1), ma=c(0.8, 0.7)),
n=1000, rand.gen=error.model)
##### Ejemplo, no ejecute#####
```

Figura 10. ARIMA (3,0,2) o ARMA(3,2) simulado



En la *figura 11*, se vuelve más compleja la identificación toda vez que la FAC (ACF) y FACP (PACF) caen rápidamente con saltos significativos ARIMA(3,0,2), el lector también podría identificar otros modelos alternativos rivales.

Figura 11. Correlograma de un ARIMA (3,0,2) o ARMA(3,2) simulado



e) Inclusión de variables de impulso o de intervención

Toda vez que los errores del modelo ARIMA deben seguir una distribución normal con media cero y varianza unitaria: $\epsilon_t \sim N_{iid}(0,1)$; en economía y finanzas es frecuente la presencia de valores atípicos que causan anormalidad en los residuos. Para ello, es pertinente la inclusión de variables de impulso, también llamadas de intervención que son de tipo booleanas o dicotómicas para corregir el problema de la no distribución normal en períodos anómalos de los residuos:

$$D_t = \begin{cases} 1 & \forall \text{ obs. Atípica} \\ 0 & \text{En otro caso} \end{cases} \quad (21)$$

f) Elección de modelos rivales ARIMA

En reiteradas ocasiones podemos enfrentarnos a modelos ARIMA rivales que compiten por realizar el mejor pronóstico univariado; en tal sentido, se presenta una priorización de especificaciones:

- Parámetros estimados de forma significativa.
- Cumplimiento del supuesto de estacionariedad e invertibilidad según corresponda.
- Cumplimiento de diagnóstico: Ljung-Box o Box-Pierce (*Script 6*).
- Cumplimiento del supuesto de distribución normal de los residuos (*Script 6*).
- Elegir el modelo cuyos estadísticos o criterios de información se minimicen: Aikake (AIC); Schwarz (SBIC) y Hanan-Quinn (HQ) (menor valor):

Minimización de criterios de información:

$$\text{Akaike: } AIC = \ln(\hat{\sigma}^2) + 2k/T \quad (22)$$

$$\text{Schwarz } SBIC = \ln(\hat{\sigma}^2) + k/T \ln T \quad (23)$$

$$\text{Hannan-Quinn } HQIC = SBIC = \ln(\hat{\sigma}^2) + 2k/T \ln(\ln T) \quad (24)$$

Donde: $k = p + q + 1$; $T =$

$n^\circ \text{ de Obs}$; $(\hat{\sigma}^2)$ corresponde a la varianza de los errores del modelo. Usualmente, AIC selecciona modelos ARIMA con mayor orden (más grandes).

SBIC es consistente pero ineficiente (selecciona modelos más parsimoniosos).

- Elegir el modelo cuyo estadístico o criterio de razón de verosimilitud (LL) se maximice (mayor valor).
- Elegir el modelo con la mayor precisión o menores estadísticos de errores en los pronósticos: error cuadrado medio (MSE), error medio absoluto (MAE o el error medio absoluto porcentual (MAPE):

$$MSE = \frac{1}{N} \sum_{t=1}^N (y_{t+s} - f_{t,s})^2 \quad (25)$$

$$MAE = \frac{1}{N} \sum_{t=1}^N |y_{t+s} - f_{t,s}| \quad (26)$$

$$MAPE = \frac{1}{N} \sum_{t=1}^N \frac{|y_{t+s} - f_{t,s}|}{y_{t+s}} \times 100 \quad (27)$$

$$\% \text{ correcto de signos de predicción} = \frac{1}{N} \sum_{t=1}^N z_{t+s} \quad (28)$$

Donde:

$$z_{t+s} = \begin{cases} 1 & \text{Si } (y_{t+s}, f_{t,s}) > 0 \\ 0 & \text{En otro caso} \end{cases} \quad (29)$$

y_{t+s} = variable observada en s períodos hacia adelante;

$f_{t,s}$ = variable pronosticada en s períodos hacia adelante.

```
#### Script 6#####
#### Prueba de normalidad de residuos en RStudio#####
#### Prueba de Ljung-Box y Box-Pierce#####
jarque.bera.test(Resid3)
shapiro.test(Resid3)
Box.test(Resid3, type="Ljung-Box")
Box.test(Resid3, type="Box-Pierce")
##### Ejemplo, no ejecute#####
```

g) Pronóstico fuera de muestra y muestra completa

En modelos *ARIMA* (p, d, q) para probar la capacidad de predicción del modelo es recomendable utilizar o dejar un período fuera de muestra para efectos de evaluación de los pronósticos. Para realizar pronósticos en *RStudio* fíjese en el *Script* 7.

Ejemplo:

- Muestra total para la inflación: Enero-2010 a Dic. – 2020 (120 meses).
- Período de muestra (*in-sample*): Ene. -2010 a Jun. -2020 (114 meses), para efectos de estimación del modelo ARIMA.
- Período fuera de muestra (*out-sample*): Jul. 2020 – Dic. 2020 (Seis meses), para efectos de comparación del pronóstico con la variable observada.

```

#### Script 7#####
#### Pronóstico de 10 períodos hacia adelante####
fit_ajus3 <- auto.arima(Ex3)
summary(fit_ajus3)

plot(forecast(fit_ajus3, h=10, level = c(10, 50, 75,90)), xlim=
c(980, 1010), ylim= c(-2,2), main= "Pronóstico de 10 períodos", col
= 1)
abline(h = 0., lty=2)
abline(h = 1, lty=2)
abline(h = -1, lty=2)
##### Ejemplo, no ejecute#####

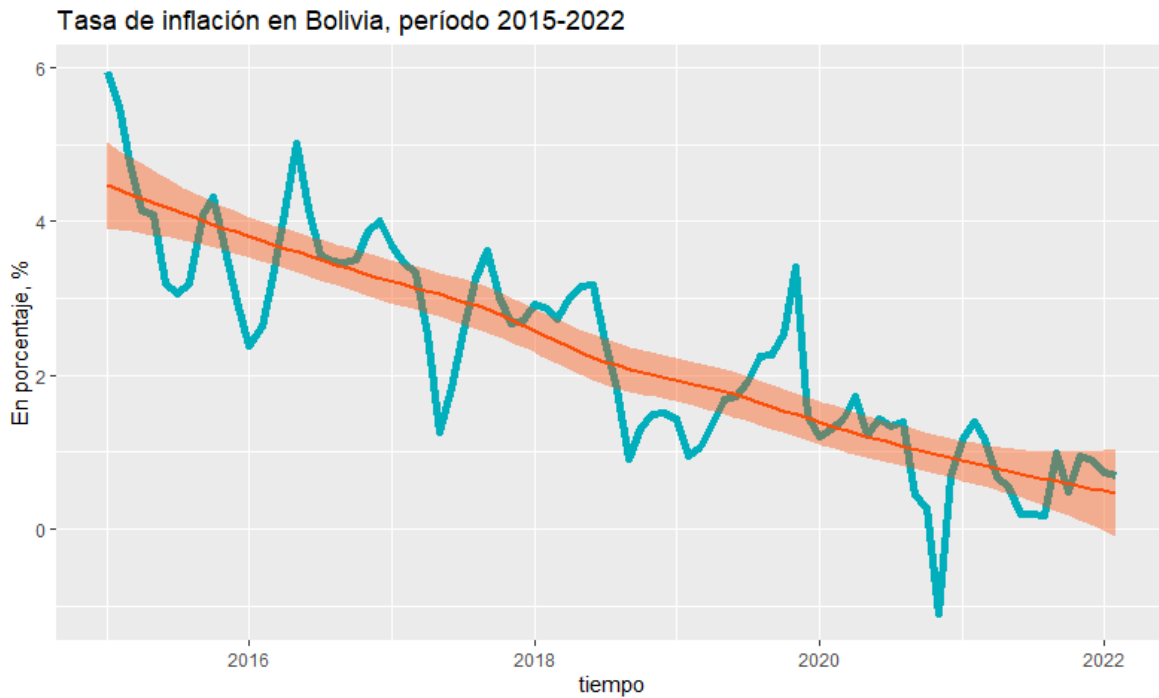
```

2. UNA APLICACIÓN A LA INFLACIÓN EN BOLIVIA CON RSTUDIO

A partir de datos del Instituto Nacional de Estadística (INE), se pretende realizar un modelo ARIMA para la tasa de inflación en Bolivia, expresada como variación a 12 meses, con datos mensuales durante el período: *ene.-2015 a Feb.-2022* (muestra total de 86 observaciones). Para ello, se pretende descomponer la muestra en:

- a) Período de estimación: *ene.-2015 a ago-2021*: 80 observaciones (*in sample*).
- b) Período fuera de muestra: *sep.-2021 a feb-2022*: 6 observaciones (*out-of-sample*).

Figura 12. Gráfica en niveles



Fuente: INE.

El propósito final en la estimación del ARIMA, es realizar un pronóstico de la inflación hacía el mes de *jun./2023*. En tal sentido y en función de la *figura 12*, se puede apreciar que la variable ha observado una tendencia decreciente en el tiempo;

por lo cual, se podría interpretar que es una variable no estacionaria: su media, varianza y autocovarianza no son constantes y varían en el tiempo.

En el cuadro 2, se reportan los estadísticos descriptivos de la tasa de inflación a 12 meses, con medidas de tendencia central entre 2.26 y 2.36%, una desviación estándar de 1.39% con relación a la media; con asimetría positiva o mayor cantidad de observaciones por encima de la media (sesgo positivo) y curtosis de tipo leptocúrtica (presencia de valores atípicos). Durante el período observado, más del 95% de las observaciones ha tenido una inflación inferior al nivel del 5%.

Cuadro 2. Estadísticos descriptivos
Tasa de inflación, variación a 12 meses

Estadísticos	Tasa de inflación (%)	Percentil	Tasa de inflación (%)	Umbrales	Porcentaje
N° Obs	86.00	0.10	0.67	Inflación > 5%	3.49
Media	2.29	0.20	1.06	Inflación < 5%	96.51
Mediana	2.32	0.25	1.24		
Media robusta	2.26	0.30	1.35		
Desv. Abs. Media	1.60	0.40	1.51		
Desv. Estándar	1.39	0.50	2.32		
Min	- 1.10	0.60	2.73		
Max	5.94	0.70	3.18		
Rango	7.03	0.75	3.32		
Asimetría	0.22	0.80	3.49		
Curtosis	- 0.52	0.90	4.08		
Error estándar	0.15				
Prob. J-B	0.48				

Fuente: Elaboración propia con base en datos del *INE*.

De igual manera, es posible descomponer la tasa de inflación en sus componentes: estacional (*seasonal*), tendencial (*trend*) e irregular (*figura 13*). Donde

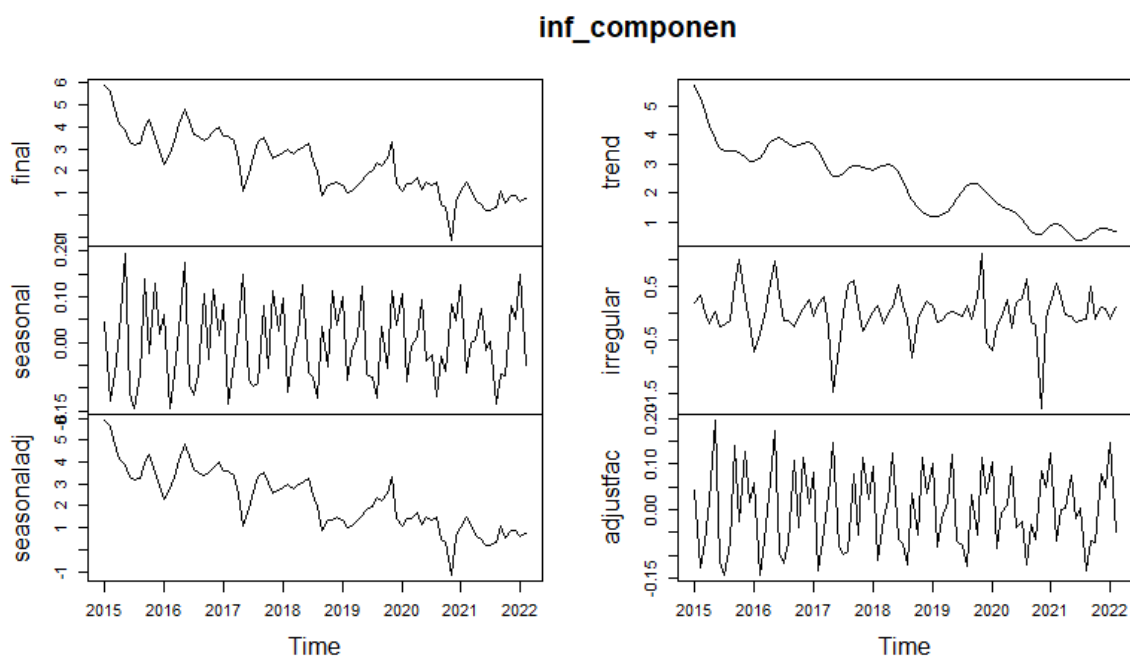
la serie ajustada estacionalmente (*seasonaladj*) corresponde a la tasa de inflación restando el efecto estacional (*Script 8*).

```
#### Script 8#####
#### Ajuste estacional y descomposición de series####
#### de tiempo usando X11#####
inf_sa <- seas(inf,x11 = "")
inf_sa
inf_sa$data
inf_componen<-inf_sa$data

plot(inf_componen)
inf_componen2<-as.data.frame(inf_componen)
inf_componen2
inf_ajus<-ts(inf_componen2$seasonaladj, start=c(2015, 1), freq=12)
plot(inf_ajus)

view(inf_sa)
plot(inf)
##### Ejemplo, no ejecute#####
```

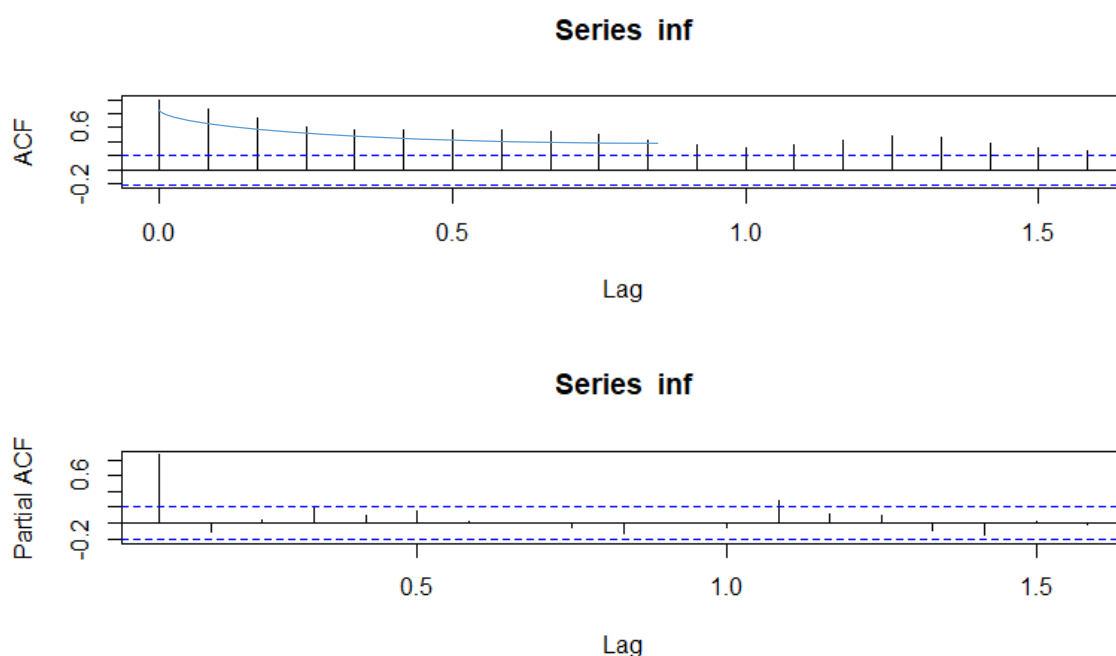
Figura 13. Descomposición de la inflación



Fuente: Estimaciones propias con base en el *INE*.

Según la metodología Box-Jenkins (1970), el primer paso es identificar los posibles procesos ARIMA mediante el correlograma. En la *figura 14*, se observa que la función de autocorrelación (ACF) va cayendo lentamente, lo cual señalaría que es una serie no estacionaria; este patrón cambia en primeras diferencias, por lo cual se podría interpretar un posible modelo *ARIMA* (1,1,0) o *ARIMA*(0,1,1) (*Script 9*).

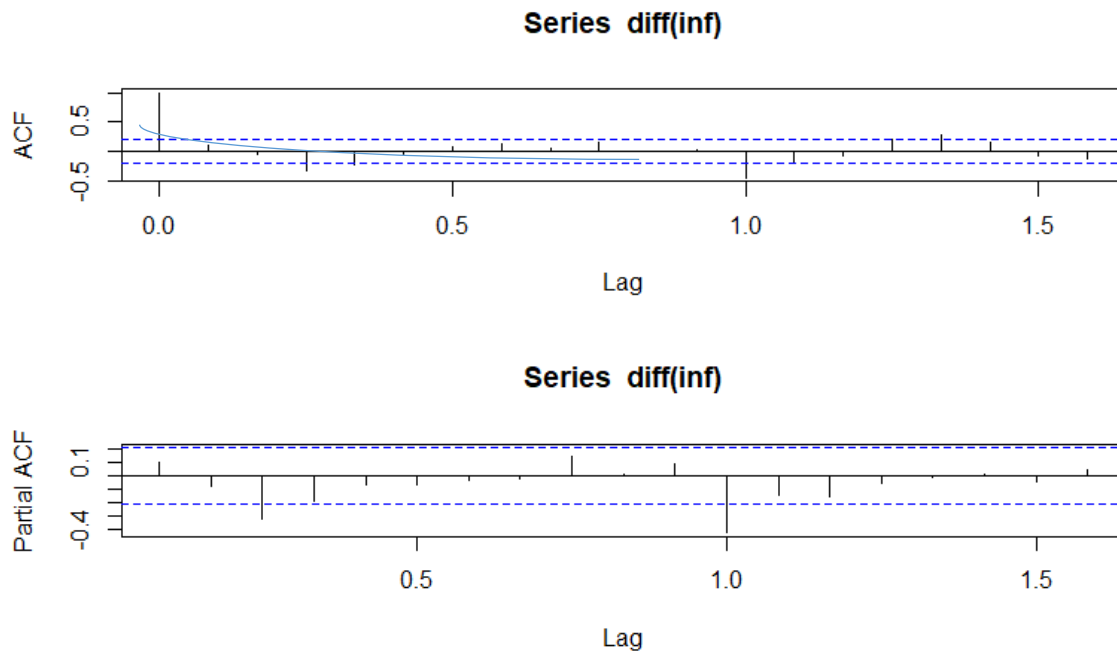
Figura 14. Correlograma de la inflación (en niveles)



```
#### Script 9#####
#### Correlogramas de la inflación#####
#En niveles
par(mfrow=c(2,1))
acf(inf); pacf(inf)

#En primera diferencia
acf(diff(inf)); pacf(diff(inf))
par(mfrow=c(1,1))
##### Ejemplo, no ejecute#####
```

Figura 15. Correlograma de la inflación (en primeras diferencias)



Fuente: Estimaciones propias con base en el *INE*.

Según el *cuadro 3*, se estima un modelo *ARIMA* (1,1,0), el cual no presenta residuos normales. La probabilidad de las pruebas Jarque-Bera (J-B) y Shapiro-Wilk (Sh-W) respectivamente son menores al nivel del 0.05. De igual manera, el estadístico t del coeficiente *AR*(1) es menor a 2; por lo cual no sería estadísticamente significativo. Para ello se procede a identificar las observaciones anómalas y la creación de variables de impulso para corregir el problema de la normalidad en la distribución de los errores (*Script 10 y 11*).

Cuadro 3. Estimación inicial, modelo univariado

Tasa de inflación, variación a 12 meses

Modelo ARIMA (1, 1, 0)

		Coeficiente			
AR(1)	ϕ_1	0.1393			
		0.1111			
		1.2538			
Prob. J-B		0.00	AIC		139.01
Prob. Sh-W		0.00	BIC		143.74
Prob. Ljung-Box		0.99	RV	-	67.50

```
#### Script 10#####
#### Estimación ARIMA(1,1,0) #####
inf_f<-window(inf, end=c(2021,08)) # Seis meses fuera de muestra

arima1<-Arima(inf_f, order=c(1,1,0), seasonal = c(0,0,0),
              xreg = NULL, include.mean = TRUE)

jarque.bera.test(arima1$residuals)
shapiro.test(arima1$residuals)
Box.test(arima1$residuals, type="Ljung-Box")
accuracy(arima1)
plot(arima1$residuals)
##### Ejemplo, no ejecute#####
```

Para identificar las observaciones atípicas, se sugiere según las *figuras 14, 15 y 16*:

- Graficar los residuos del modelo ARIMA, identificando las observaciones máximas y mínimas.
- Graficar un diagrama de cajas y observando los puntos fuera de los bigotes y percentiles 25% inferior y superior respectivamente.
- Crear variables dicotómicas para los valores anómalos identificados.

```

#### Script 11#####
#### Creando variables de impulso/intervención #####

# Identificando Outlier

myboxplot1 <- boxplot(arima1$residuals)
names(myboxplot1)
myboxplot1$out  # dos valores atípicos
3*sd(arima1$residuals)
max(arima1$residuals)
min(arima1$residuals)
plot(arima1$residuals)
arima1$residuals

# Creamos variables de impulso en el archivo excel
d19_12<- ts(d19_12, start=c(2015, 01), freq=12)
d20_12<- ts(d20_12, start=c(2015, 01), freq=12)

exogen<-ts.intersect(d19_12, d20_12)
plot(exogen)

exogen1<-window(exogen, start=c(2015,01), end=c(2021, 08))
exogen2<-window(exogen, start=c(2021,09), end=c(2022,02))
exogen2

##### Modelo 2, corrigiendo problemas de normalidad #####

arima2<-Arima(inf_f, order=c(1,1,0), seasonal = c(0,0,0), xreg =
exogen1, include.mean = TRUE)
arima2
summary(arima2)
jarque.bera.test(arima2$residuals)
shapiro.test(arima2$residuals)
Box.test(arima2$residuals, type="Ljung-Box")
plot(arima2$residuals)
arima2$residuals
##### Ejemplo, no ejecute#####

```

Figura 14. Gráfico de residuos, modelo ARIMA (1, 1, 0)

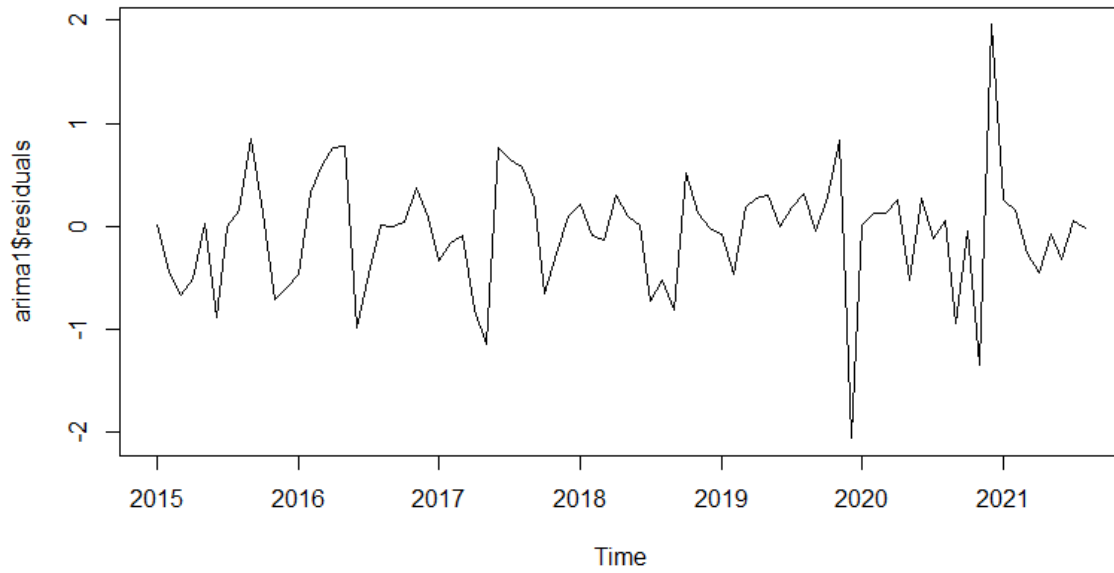


Figura 15. Diagrama de caja para identificación de valores atípicos,
modelo ARIMA (1, 1, 0)

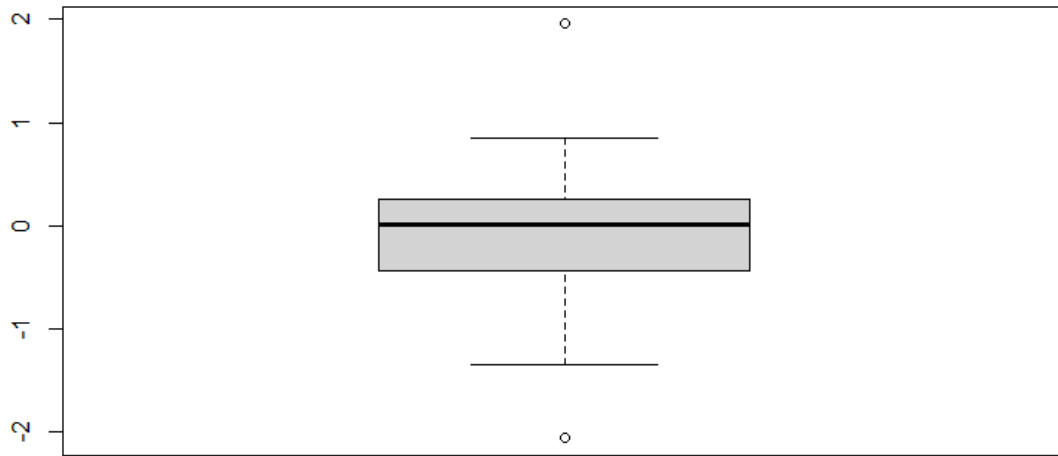
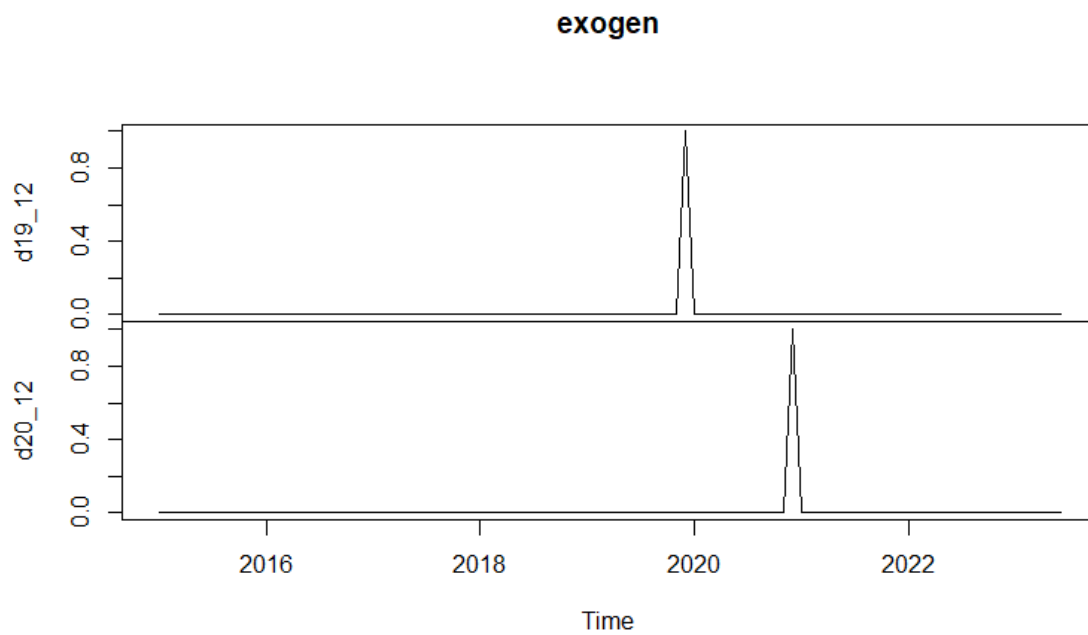


Figura 16. Variables booleanas de impulso/intervención



En resumen, para corregir el problema de normalidad en la distribución de los residuos, se crearon dos variables dicotómicas: para diciembre 2019 y diciembre 2020 respectivamente (valor de 1) y cero en otro caso. De forma seguida, estas variables fueron introducidas como variables explicativas en el modelo ARIMA en muestra completa y en muestra con período de evaluación (*out-of sample*).

En la *figura 17*, se observa el comportamiento de la tasa de inflación observada, misma que cae dentro del gráfico de abanico (*fanchart*) en el umbral superior con una tasa observada entorno al 1% en el período de evaluación fuera de muestra. Para la *figura 18*, se realiza un pronóstico a jun./2023 con una predicción central entre el 0 y el 1%, así como una expectativa de mayor inflación en los umbrales entre el 2 y el 4%. En un escenario de deflación, se observan posibles contextos de deflación hasta el -3% respectivamente (*Script 12*).

```
#### Script 12#####
#### Modelo de pronóstico para 16 meses hacia adelante ##### al 90%
de confianza#####

arima3<-Arima(inf_comp, order=c(1,1,0), seasonal = c(0,0,0), xreg =
exogen1f, include.mean = TRUE)

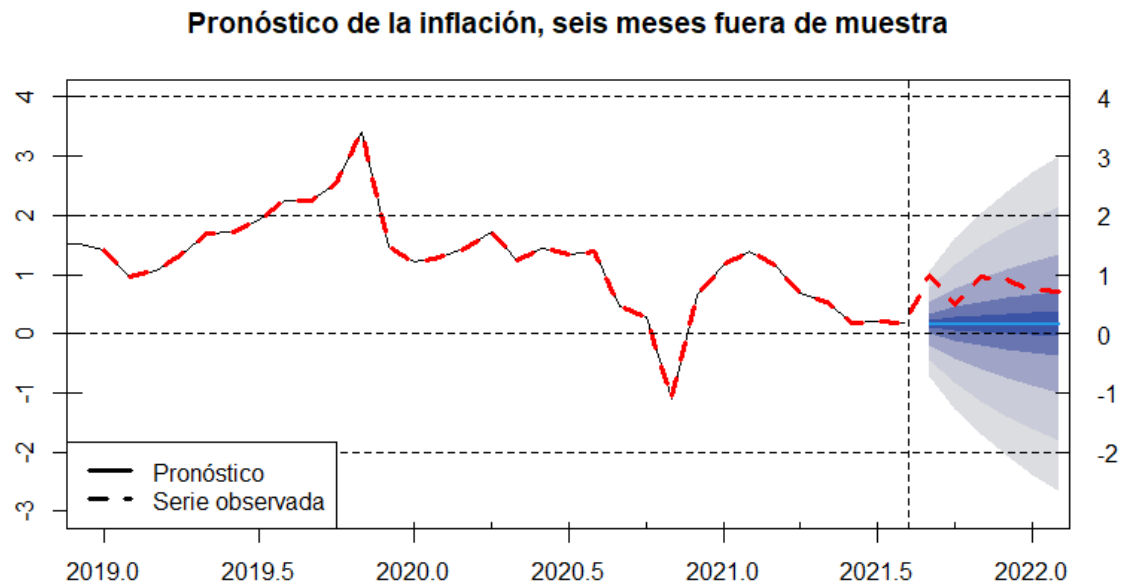
forecast(arima3, h=16, xreg=exogen2f, level = c(10,25,50,75,90))%>%
  autoplot(main="Pronóstico de la inflación en Bolivia
(Jun/2023)") +
  ylab("En porcentaje") + xlab("Año")+
  xlim(2020, 2023)+ ylim(-4, 6)
##### Ejemplo, no ejecute#####
```

Cuadro 4. Estimación final, modelo univariado

Tasa de inflación, variación a 12 meses

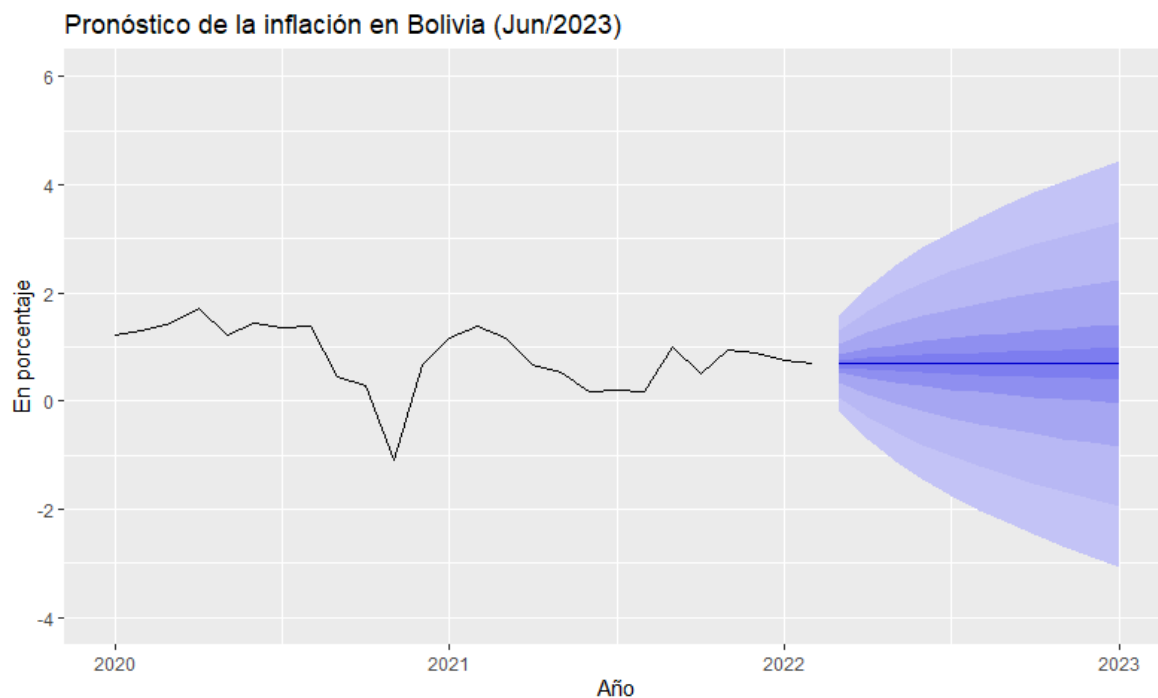
Con variables de impulso, seis observaciones fuera de muestra				Con variables de impulso, muestra completa			
Modelo ARIMA (1, 1, 0)				Modelo ARIMA (1, 1, 0)			
Coeficiente				Coeficiente			
AR(1)	ϕ_1		0.2799	AR(1)	ϕ_1		0.2397
			0.1088				0.1063
			2.5726				2.2549
d19_12	λ_1	-	0.9226	d19_12	λ_1	-	0.9144
			0.3209				0.3279
		-	2.8750			-	2.7887
d20_12	λ_2		0.7969	d20_12	λ_2		0.7801
			0.3231				0.3305
			2.4664				2.3604
Prob. J-B	0.85	AIC	131.27	Prob. J-B	0.85	AIC	140.31
Prob. Sh-W	0.48	BIC	140.75	Prob. Sh-W	0.48	BIC	150.08
Prob. Ljung-Box	0.56	RV	- 61.64	Prob. Ljung-Box	0.56	RV	- 66.15

Figura 17. Pronóstico fuera de muestra, modelo ARIMA (1, 1, 0)



Fuente: Estimaciones propias con base en el *INE*.

Figura 18. Pronóstico final, modelo ARIMA (1, 1, 0)



Fuente: Estimaciones propias con base en el *INE*.

Según el *cuadro 4*, el modelo estimado ARIMA (1, 1, 0) cumple con la normalidad en los residuos y con la prueba Ljung-Box (Prob. > 10%). Por lo cual, la $\Delta\pi_t$ depende en +0.24 y + 0.28% de lo que suceda con $\Delta\pi_{t-1}$. Con la introducción de las variables de impulso se observa una disminución en la variación de la inflación cercana al -1% para el mes de diciembre de 2019 y un aumento del +0.80% para el mes de diciembre de 2020.

Para finalizar, se brindan algunas reflexiones conclusivas:

- 1) A la medida que se aumenta el período del pronóstico hacia adelante, la precisión va disminuyendo.

- 2) Los Modelos ARIMA tienden a fallar en la detección de puntos de quiebre o cambios en la tendencia del pronóstico.
- 3) El pronóstico no es un sustituto de la intuición y juicio económico.
- 4) La ventaja de utilizar modelos de predicción en lugar del juicio de expertos, radica que en los últimos pueden estar sobre/sub-estimados, con sesgos, anclajes, inconsistencia o sujetos a un determinado grupo pensante.
- 5) Tal vez, el mejor enfoque de pronósticos consiste en combinar el juicio y proyección de expertos con los resultados de un modelo cuantitativo.

ENSAYO 2: MODELOS DE TENDENCIAS Y COINTEGRACIÓN UNIECUACIONAL

Roger Alejandro Banegas Rivero^σ

UNIVERSIDAD AUTÓNOMA GABRIEL RENÉ MORENO

Este diseño, 02 de mayo de 2022

INTRODUCCIÓN

En series de tiempo el análisis de corto plazo consiste en la modelización de variables estacionarias; sin embargo, en la mayoría de las series en economía y finanzas las variables son de tipo no estacionarias o presentan tendencias; así por ejemplo, el Producto Interno Bruto (*PIB*), el consumo privado, la cartera de créditos, los ahorros del sistema financiero, entre otras variables, presentan tendencias al alza; también existen otras variables con tendencias o rumbo a la baja (e.g. las tasas de interés); desde luego, también están las variables estacionarias, con media constante en el tiempo, varianza y autocovarianza como es la variación del tipo de cambio o los retornos financieros de las acciones entre otras.

De igual manera, en el largo plazo se puede realizar análisis econométrico con variables no estacionarias, si las mismas presenten co-movimientos o combinaciones lineales estacionarias, esto es la llamada cointegración. Una pregunta relevante consiste en responder: ¿Qué período de tiempo constituye corto y largo plazo respectivamente?

En economía, el corto plazo se define como aquel período de producción donde existe al menos un factor de producción fijo, como también variables; en cambio en el largo plazo todos los factores de producción son variables. Desde una perspectiva metodológica en series de tiempo, no existe una demostración en la amplitud

^σ Correo electrónico: rogerbanegas@uagrm.edu.bo

requerida temporal; sin embargo existen ciertos hechos estilizados. Por ejemplo, John Maynard Keynes señalaba: “en el largo plazo todos estaremos muertos”, por lo cual se puede interpretar que el mismo corresponde a décadas, asumiendo una esperanza de vida promedio desde luego. En tal sentido, como regla de dedo, un tanto arbitraria podemos hablar del largo plazo de 40, 50, 60, 70 o más años inclusive, de forma independiente a la cantidad de observaciones o frecuencia de tiempo (diaria, mensual, trimestral, etc). De igual manera, análisis temporales de 10, 20 o 30 es probable considerarlo como análisis de corto plazo.

1. ASPECTOS METODOLÓGICOS

Robert Engle & Clive Granger fueron laureados con el premio nobel en Economía (2003) por sus contribuciones a las series de tiempo en términos de relaciones de largo plazo (cointegración), causalidad y modelización en presencia de volatilidad para series financieras.

De esta forma [Granger \(1981\)](#) presentó por primera vez la relación entre cointegración y mecanismo corrector de error y la representación de teorema de [Granger \(1983\)](#) el cual establece que si existe una relación de largo plazo (cointegración) o combinación lineal estacionaria entre variables no estacionarios, puede existir un mecanismo corrector de error en relaciones de corto plazo (variables estacionarias).

En tal sentido, la contribución de [Engle & Granger \(1987\)](#) se fundamentó en una prueba de largo plazo de tipo uniecuacional (una sola ecuación). De forma posterior se han desarrollado otras pruebas de cointegración multivariadas (con dos o más ecuaciones) ([Johansen, 1988](#); [1991](#); [1995](#)). Por el momento se analizará la uniecuacional basada en cuatro pasos:

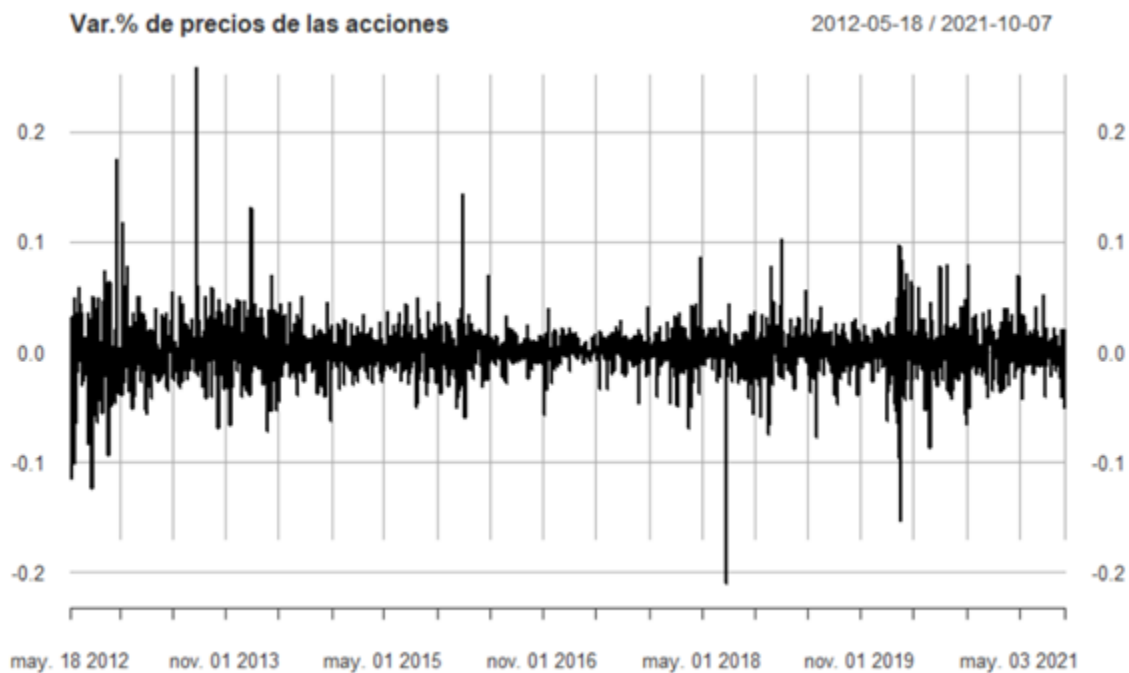
I. PASO 1: PRUEBAS DE RAÍZ UNITARIA

Al emplear series de tiempo, es importante determinar si las series de tiempo individuales son o no estacionarias, para ello, existen varias pruebas como Dickey-Fuller-Aumentada (DFA); Phillips-Perron (Ph-P) y KPSS).

Desde que la mayoría de las series en economía y finanzas presentan rumbos y tendencias se pueden identificar tres tipos:

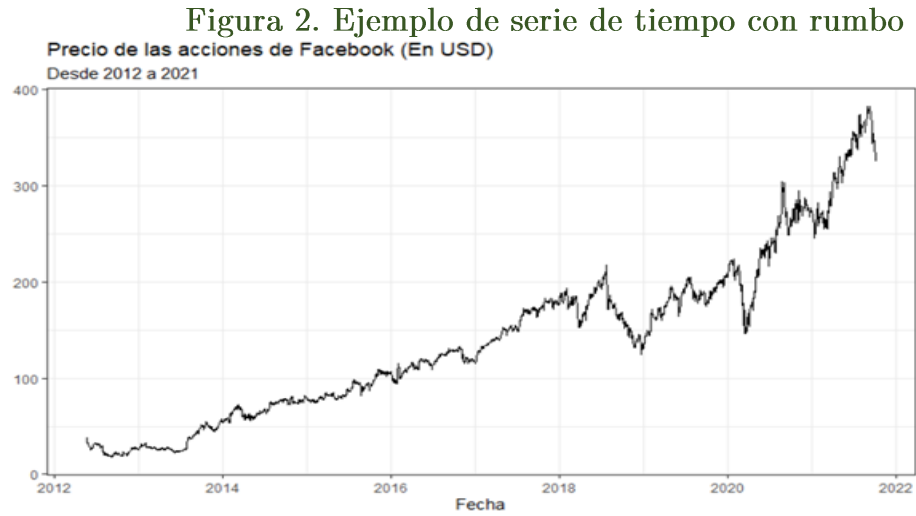
- a. Sin rumbo, ni deriva (*none*) – caminata aleatoria pura

Figura 1. Ejemplo de serie de tiempo sin rumbo



Fuente: *Yahoo Finance*

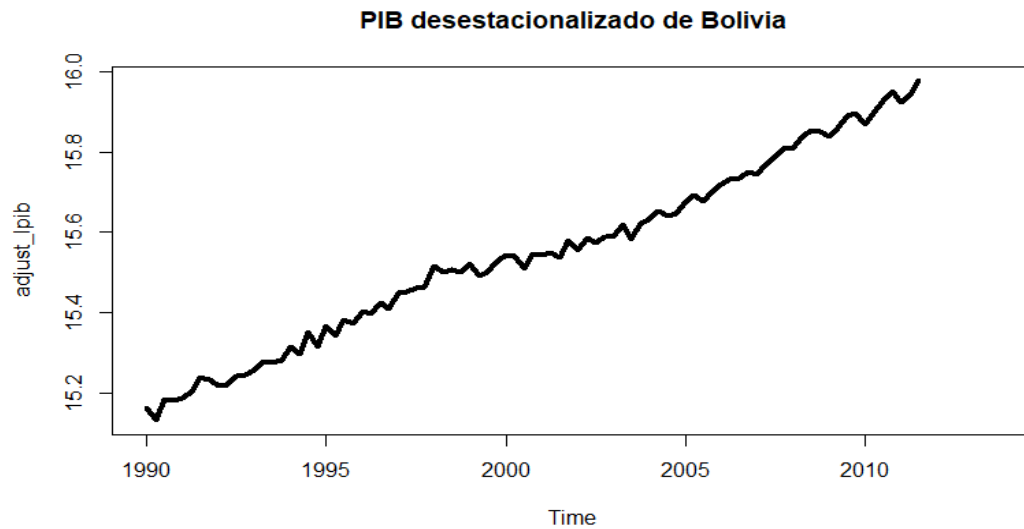
- b. Con rumbo o deriva (constant) – caminata aleatoria con deriva



Fuente: *Yahoo Finance*

- c. Con rumbo y tendencia determinista (constant +trend)

Figura 3. Ejemplo de serie de tiempo con rumbo y tendencia determinística



Fuente: *Instituto Nacional de Estadística (INE)*

En la *figura 1*, no se identifica un patrón de tendencia o rumbo (similar al proceso de ruido blanco; para la *figura 2*, existen fuertes choques y persistencias al alza o la baja. Finalmente para la *figura 3*, se vislumbra una tendencia más suavizada.

Asimismo, existen tres formas para realizar análisis de estacionariedad:

- a) Análisis gráfico (¿la serie de tiempo tiene una tendencia, rumbo al alza o un rumbo a la baja?
- b) Correlograma (Función de autocorrelación, FAC y la Función de autocorrelación parcial (FACP). FAC cae lentamente o nunca tiende cero, es la interpretación de una serie no estacionaria.
- c) Pruebas de raíces unitarias o pruebas de estacionariedad (Dickey Fuller Aumentada (DFA), Phillips-Perron (Ph-P) y KPSS.

En consecuencia, surge un relevante cuestionamiento: ¿Qué sucede en Economía y Finanzas analizamos series de tiempo que no cumple el supuesto de estacionariedad?

Nos lleva al problema de la regresión espuria, resultados falsos y conclusiones contrarias.

El ejemplo de Hendry:

$y_t \sim I(1)$ la serie no estacionaria en niveles o presenta una raíz unitaria. La variable es estacionaria después de realizar una primera diferencia.

y_t = “El número de alcoholólicos en Londres”

$x_t \sim I(1)$ la serie no es estacionaria en niveles o presenta una raíz unitaria. La variable es estacionaria después de realizar una primera diferencia.

x_t = “El número de sacerdotes ordenados en la Iglesia Anglicana, en Londres”

Se plantea un modelo econométrico:

$$y_t = \alpha + \beta x_t + \varepsilon_t \quad (1)$$

Esta es la demostración de la regresión espuria (sin sentido).

Lo correcto sería estimar una regresión con variables estacionarias:

$$\Delta y_t = \alpha + \beta \Delta x_t + \varepsilon_t \quad (2)$$

$$\beta = 0$$

La conclusión verdadera es que no existe relación alguna entre x_t y y_t , hay que neutralizar el efecto del tiempo. Debemos convertir las variables no estacionarias en variables estacionarias

En los supuestos clásicos del modelo de Regresión lineal:

- Los residuos (ε_t) tienen una distribución normal.
- Los residuos (ε_t) no tienen autocorrelación.
- Los residuos (ε_t) tienen una varianza constante (son homocedásticos).
- No existe multicolinealidad (las variables explicativas no están altamente correlacionadas).
- El modelo está bien especificado en su forma funcional y en la cantidad de variables incluidas.

Con econometría de series de tiempo añadimos dos supuestos adicionales:

- *Las variables son estacionarias.*
- Si las variables no son estacionarias, debe probarse la cointegración o relación de largo plazo.

Esquema 1. Dickey-Fuller como prueba de estacionariedad

Sinónimos

Serie no estacionaria = Serie con raíz unitaria

Veamos la demostración de la prueba, asumiendo un proceso AR(1):

$$y_t = \phi y_{t-1} + \varepsilon_t \quad (3)$$

$\phi < 1$ <- La serie y_t es estacionaria.

$\phi = 1$ <- La serie y_t no es estacionaria o tiene una raíz unitaria.

$\phi > 1$ <- La serie y_t es explosiva.

A la expresión le vamos restar y_{t-1} en ambos lados de la ecuación (3)

$$y_t - y_{t-1} = \phi y_{t-1} - y_{t-1} + \varepsilon_t \quad (4)$$

$$\Delta y_t = (\phi - 1) y_{t-1} + \varepsilon_t \quad (5)$$

Definimos un parámetro $\psi = (\phi - 1)$ en (5)

$$\Delta y_t = \psi y_{t-1} + \varepsilon_t \quad (6)$$

H_o = La serie y_t tiene una raíz unitaria o no es estacionaria

H_1 = La serie y_t es estacionaria

$$H_o = \psi = 0 \quad o \quad \phi = 1$$

$$H_1 = \psi < 0 \quad o \quad \phi < 1$$

La prueba de hipótesis Dickey-Fuller (DF) tenía autocorrelación, los residuos o errores estaban correlacionados entre sí.

Esquema 2. Dickey- Fuller Aumentada (DFA),

Se incorporan términos autoregresivos en la expresión 6:

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^p \rho_i \Delta y_{t-i} + \varepsilon_t \quad (7)$$

Tres tipos de series de tiempo:

a) Sin rumbo y sin tendencia (caminata aleatoria pura)

$$\Delta y_t = \psi y_{t-1} + \sum_{i=1}^p \rho_i \Delta y_{t-i} + \varepsilon_t \quad (8)$$

b) Caminata aleatoria con rumbo o deriva

$$\Delta y_t = \alpha + \psi y_{t-1} + \sum_{i=1}^p \rho_i \Delta y_{t-i} + \varepsilon_t \quad (9)$$

$\alpha > 0$; Rumbo al alza

$\alpha < 0$; Rumbo a la baja

c) Con rumbo y tendencia determinista o determinística

$$\Delta y_t = \alpha + \psi y_{t-1} + \beta t + \sum_{i=1}^p \rho_i \Delta y_{t-i} + \varepsilon_t \quad (10)$$

$\alpha > 0$; Rumbo al alza.

$\alpha < 0$; Rumbo a la baja.

$\beta > 0$; tendencia lineal (tiempo) a crecer.

$\beta < 0$; tendencia lineal (tiempo) a la disminución.

Esquema 3. La prueba de hipótesis de raíz unitaria, Phillips-Perron

H_0 = La serie y_t tiene una raíz unitaria o no es estacionaria

H_1 = La serie y_t es estacionaria

Esquema 4. La prueba de hipótesis de raíz unitaria, KPSS

H_0 = La serie y_t es estacionaria

H_1 = La serie y_t tiene una raíz unitaria o no es estacionaria

Existen los siguientes niveles de integración o estacionariedad $I(d)$,:

$I(0)$, la variable es estacionaria en niveles, y_t

$I(1)$, la variable es estacionaria después de una primera diferencia, Δy_t

$I(2)$, la variable es estacionaria después de una segunda diferencia, $\Delta\Delta y_t$

Una característica importante es el mismo orden de integración entre las variables $[I(d)]$ para evaluar la posible relación de largo plazo o cointegración.

II. PASO 2: PRUEBA DE COINTEGRACIÓN

Para definir la cointegración, supongamos un modelo econométrico con las siguientes variables:

$$y_t = \widehat{\alpha}_0 + \widehat{\alpha}_1 x_{1t} + \widehat{\alpha}_2 x_{2t} + u_t \quad (11)$$

En primer lugar, si $y_t \sim I(0)$, $x_{1t} \sim I(0)$, $x_{2t} \sim I(0)$, la combinación lineal de las variables tomará en el término de error con el mismo orden de integración: $u_t \sim I(0)$.

Por otra parte, la inexistencia o ausencia de cointegración, se presenta cuando en la expresión 11:

- Sea $y_t \sim I(0)$, $x_{1t} \sim I(1)$, $x_{2t} \sim I(1)$, la combinación lineal de las variables tomará en el término de error el máximo orden de integración: $u_t \sim I(1)$.
- Sea $y_t \sim I(1)$, $x_{1t} \sim I(1)$, $x_{2t} \sim I(2)$, la combinación lineal de las variables tomará en el término de error el máximo orden de integración: $u_t \sim I(2)$.
- Sea $y_t \sim I(1)$, $x_{1t} \sim I(1)$, $x_{2t} \sim I(1)$, generalmente la combinación lineal de las variables tomará en el término de error el máximo orden de integración: $u_t \sim I(1)$; es decir, el fenómeno de la regresión espuria.

Por otra parte, la existencia de cointegración o relación de largo plazo se presenta cuando en la expresión 11:

- Sea $y_t \sim I(1)$, $x_{1t} \sim I(1)$, $x_{2t} \sim I(1)$, la combinación lineal de las variables tomará en el término de error será estacionaria: $u_t \sim I(0)$.
- Sea $y_t \sim I(2)$, $x_{1t} \sim I(2)$, $x_{2t} \sim I(2)$, la combinación lineal de las variables tomará en el término de error será estacionaria: $u_t \sim I(0)$.

En consecuencia, si las series de tiempo son de tipo no estacionarias, pero su combinación lineal es estacionaria, se concluye que las series están cointegradas o mantienen movimientos en la misma dirección en el largo plazo. Para ello, las pruebas más utilizadas están basadas en [Engle & Granger \(1987\)](#) y [Phillips-Ouliaris \(1990\)](#), quienes proponen la evaluación de la hipótesis nula (H_0) de no cointegración versus la hipótesis alternativa (H_1) de cointegración o relación de largo plazo.

III. MECANISMO CORRECTOR DE ERROR

Si existe una relación de cointegración entre las variables, el teorema de representación de Granger establecer que puede existir un mecanismo de corrector de error para variables estacionarias en el corto plazo:

$$\Delta y_t = \widehat{\beta}_0 + \widehat{\beta}_1 \Delta x_{1t} + \widehat{\beta}_2 \Delta x_{2t} + \hat{\gamma} u_{t-1} + \epsilon_t \quad (12)$$

Donde: $u_{t-1} = y_{t-1} - \widehat{\alpha}_0 - \widehat{\alpha}_1 x_{1t-1} - \widehat{\alpha}_2 x_{2t-1}$; $\gamma < 0$ (equilibrio de largo plazo)

Asumiendo: $y_t \sim I(1)$, $x_{1t} \sim I(1)$, $x_{2t} \sim I(1)$, $u_t \sim I(0)$, $\epsilon_t \sim I(0)$

La expresión 12 es equivalente a:

$$\Delta y_t = \widehat{\beta}_0 + \widehat{\beta}_1 \Delta x_{1t} + \widehat{\beta}_2 \Delta x_{2t} + \hat{\gamma}(y_{t-1} - \widehat{\alpha}_0 - \widehat{\alpha}_1 x_{1t-1} - \widehat{\alpha}_2 x_{2t-1}) + \epsilon_t \quad (13)$$

En otras palabras, se guardan los residuos de la ecuación de largo plazo (u_t , expresión 11) y se estima una nueva ecuación con variables estacionaras

(expresión 12) incorporando los errores con un rezago (u_{t-1}) como variable explicativa; donde γ representa la velocidad de ajuste interpretado como la proporción de corrección en el error que es corregida para volver a su equilibrio, lo cual significa que el parámetro debe ser negativo ($\gamma < 0$).

IV. PRUEBAS DE DIAGNÓSTICO

La relación de cointegración y mecanismo corrector de error involucra el cumplimiento de los supuestos clásicos del modelo de regresión lineal (MCRL): residuos con distribución normal, homocedasticidad, no multicolinealidad, No autocorrelación o sin correlación serial, correcta especificación matemática y funcional, sin problemas de variables omitidas y redundantes; así como la incorporación de los supuestos de la econometría moderna (pruebas de estacionariedad y cointegración). Con este cumplimiento se lograría el teorema de Gauss-Markov para obtener los mejores estimadores lineales e insesgados (*MELI o BLUE*).

2. EXTENSIONES DE REGRESIÓN DE COINTEGRACIÓN UNIECUACIONAL

En esta sección se presentan tres extensiones en métodos econométricos, completamente eficientes, para evaluar relaciones de largo plazo (regresiones de cointegración): 1) Mínimos cuadrados ordinarios (MCO) completamente modificados (FMOLS); 2) Mínimos Cuadrados dinámicos (DOLS) y 3) estimador integrado y modificado de MCO (IM-OLS).

Desde que la mayoría de las series comprende una tendencia estocástica o determinista [$I(1)$], la estimación de análisis de dependencia en niveles [$I(1)$] podrá

llevar a relaciones estadísticas que serán estadísticamente significativas; sin embargo, de forma recurrente se pueden probar como relaciones espurias.

La propuesta de [Engle & Granger \(1987\)](#) señalaba que si dos o más series están integradas en orden $I(1)$, cuya combinación lineal o comportamiento de los residuos es estacionaria $[I(0)]$ se diría que las variables mantienen una relación de largo plazo o están cointegradas.

Como ejemplo, considere la presencia de un vector único de cointegración de variables $n + 1 \{y_t, X'_t\}$ mediante la expresión:

$$y_t = X'_t \beta + D'_{1t} \gamma_1 + u_{1t} \quad (13)$$

Donde $\{D'_{1t}, D'_{2t}\}$ corresponde a regresores de tendencia determinista y n regresores estocásticos X_t que son gobernadas por el siguiente sistema de ecuaciones:

$$X_t = \Gamma'_{21} D_{1t} + \Gamma'_{22} D_{2t} + \varepsilon_{2t} \quad (14)$$

$$\Delta \varepsilon_{2t} = u_{2t} \quad (15)$$

El vector p_1 de D_{1t} se incorpora en ambas ecuaciones, ecuación de cointegración (13) y la ecuación del regresor (14), mientras que el vector p_2 de D_{2t} son regresores de tendencia determinista incluidas en las ecuaciones de regresores (14) pero excluidas de la ecuación de cointegración (13) (si la constante está presente, se asume ser un elemento de D_{1t} , así que no está en D_{2t}).

En tal sentido, se asume que las innovaciones son estrictamente estacionarias $u_t = \{u_{1t}, u_{2t}'\}'$ y ergódicas¹ con media cero, matriz de covarianza contemporánea Σ , un lado de la matriz covarianza de largo plazo Λ y una matriz de covarianza Ω , las cuales se dividen conforme a u_t .

$$\begin{aligned}\Sigma &= \{u_t, u_t'\} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \Sigma_{22} \end{bmatrix} \\ \Lambda &= \sum_{j=0}^{\infty} \{u_t u_{t-j}'\} = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \Lambda_{22} \end{bmatrix} \quad (16) \\ \Omega &= \sum_{j=-\infty}^{\infty} \{u_t u_{t-j}'\} = \begin{bmatrix} \omega_{11} & \omega_{12} \\ \omega_{21} & \Omega_{22} \end{bmatrix} = \Lambda - \Lambda' - \Sigma\end{aligned}$$

En consideración, se asume un rango n de largo plazo que proviene de la matriz de covarianza Ω con matriz no singular Ω_{22} . De forma conjunta, se asume que y_t y X_t son $I(1)$ pero se excluye de la ecuación de cointegración los elementos de X_t así como la multicointegración.

Si las series presentan una relación de largo plazo, estimando β con MCO es consistente en convergencia a una tasa más rápida que su estándar. Una importante desventaja del método estático de MCO (SOLS) tienen distribución asintótica que generalmente es de tipo no-Gaussiano, exhibiendo sesgo asintótica, asimétricos y que

¹Proviene de la teoría ergódica que es estudio matemático del promedio de largo plazo de los sistemas dinámicos. De forma general se asume que el promedio en tiempo y el promedio en espacio son iguales, por tanto, la media es ergódica y la medida es invariante.

están en función de parámetros y ruidos no escalares. SOLS no es recomendable si se quieren realizar inferencias a partir de un vector de cointegración.

El problema de la distribución asintótica de SOLS se basa en la correlación de los términos de perturbación estocástica de largo plazo y las innovaciones de los regresores (ω_{12}) , además de la correlación entre la correlación cruzada entre los errores de la ecuación de largo plazo y los errores. En el sentido especial que X_t es estrictamente exógeno, entonces, $\omega_{12} = 0$ y $\lambda_{12} = 0$, el sesgo, la asimetría y la dependencia de parámetros ruidosos no escalares disminuye. Desde esta óptica SOLS serán eficientemente asintóticos, por tanto, se permitirá emplear la estándar prueba de Wald para distribuciones χ^2 .

De forma alternativa, SOLS tiene una mezcla de estimadores asintóticamente gaussianos, si el número de tendencias deterministas excluidas de la ecuación de cointegración (p_2) no es menor que el número de regresores estocásticos de n .

A manera de ejemplo, se menciona que $m_2 = \max(n - p_2, 0)$ representa el número de variables con relación de largo plazo menos el número de parámetros excluidos de la función de largo plazo. Luego si $m_2 = 0$, la tendencia determinista en los regresores dominará la tendencia estocástica en la ecuación de cointegración.

- Mínimos cuadrados ordinarios completamente modificados (FMOLS)

Este método por su acrónimo en inglés FMOLS (*Fully modified Ordinary Least Squares*) es una propuesta de [Phillips & Hansen \(1990\)](#) con el fin de eliminar la correlación de largo plazo entre los residuos de la ecuación de cointegración y las innovaciones de los regresores estocásticos. Las estimaciones de FMOLS son parámetros asintóticamente insesgados y presentan eficiencia normal asintótica para pruebas de hipótesis estándar como la prueba Wald que emplea la distribución χ^2 .

- Mínimos cuadrados dinámicos (DOLS)

Supongamos una relación de largo plazo entre dos variables:

$$y_t = \alpha_0 + \alpha_1 x_t + u_t \quad (17)$$

Donde $y_t \wedge x_t \sim I(1)$; $u_t \sim I(0)$

Stock & Watson (1993) proponen la siguiente especificación:

$$y_t = \alpha_0 + \alpha_1 x_t + \alpha_3 \Delta x_t + u_t \quad (18)$$

Una vez incluida la variable Δx_t ; se tendrá una correcta especificación e inferencia sobre α_1

Con esta especificación se incluyen adelantos y rezagos mismos que pueden determinarse de manera automática por los criterios de información AIC o SBIC.

Cuando existe problema de autocorrelación y heterocedasticidad se puede incluir una especificación de estimación robusta en la matriz de varianza-covarianza (*HAC*).

- *Estimador integrado y modificado de MCO*

A partir de [Vogelsang & Wagner \(2014\)](#), se obtiene un estimador integrado y modificado de MCO; sin embargo, cuando los regresores son endógenos existe el problema del sesgo. Para ello, se parte de las siguientes expresiones:

$$y_t = \mu + x'_t \beta + u_t \quad (19)$$

$$x_t = x_{t-1} + v_t \quad (20)$$

De igual manera, existe un vector de errores $\eta_t = [u_t, v_t']' \sim I(0)$, con una matriz de varianza-covarianza para los errores (Ω). La propuesta es la estimación de un estimador basado en la suma parcial transformada de una regresión aumentada por los regresores originales integrados, la transformación deriva en media cero de los errores sin la necesidad de utilizar ancho de bandas o adelantos y rezagos. Por tanto, se requiere ortogonalizar los residuos a partir de los regresores adicionales contruidos.

3. UNA APLICACIÓN DE COINTEGRACIÓN UNIECUACIONAL CON RSTUDIO

A partir de información de Bloomberg, se analizan tres series temporales: -el tipo de cambio de dólar americano por dólar canadiense, los precios de Estados Unidos y los precios de Canadá, con información trimestral desde 1971.T1 al 2021.T4 (≈ 50 años):

Variables en la base de datos USDCAD.xls

- USDCAD <- Tipo de cambio de dólares americano (USD)/ dólar canadiense (CAD).
- CPI_CAD <- IPC del Canadá.
- CPI_US <- IPC de EEUU.

Teoría del Paridad de Poder Compra (PPC),

La ley de un solo precio:

$$P_t = E_t P_t^* \quad (21)$$

Los precios domésticos (P_t) están conectados con los precios internacionales (P_t^*) a través del tipo de cambio (E_t).

¿Qué sucede si en (1) despejamos el tipo de cambio (E_t)?

$$E_t = P_t/P_t^* \quad (22)$$

La definición del tipo de cambio se refiere a precios relativos de la economía doméstica en función con los precios externos o internacionales.

Ejemplo de aplicación:

- $\text{Log}(ex) = \text{Log del tipo de cambio del USDCAD.}$
- $\text{Log}(pcad_t) = \text{Log del IPC doméstico (Canadá).}$
- $\text{Log}(pus_t) = \text{Log del IPC externo(EEUU.)}$

Dado (2), se plantean tres modelos:

I. Un modelo de corto plazo

Involucra que las variables son estacionarias [$I(d)$], se deben utilizar las pruebas de raíces unitarias: DFA, Ph-P y KPSS.

Si se encuentra que las variables son $I(1)$, son estacionarias después de una diferencia; por lo tanto el modelo de corto plazo será el siguiente:

$$\Delta \text{Log}(ex_t) = \widehat{\alpha}_0 + \widehat{\alpha}_1 \Delta \text{Log}(pcad_t) + \widehat{\alpha}_2 \Delta \text{Log}(pus_t) + \epsilon_t \quad (23)$$

Donde $\widehat{\alpha}_1$ y $\widehat{\alpha}_2$ corresponden a elasticidades de corto plazo.

$\widehat{\alpha}_1 < 0$; $\widehat{\alpha}_2 > 0$. En otras palabras: si los precios domésticos de Canadá aumentan, *ceteris paribus*, el dólar americano se fortalecerá; si los precios externos de EEUU aumentan, *ceteris paribus*, el dólar canadiense se fortalecerá.

II. Un modelo de largo plazo

Básicamente, si todas las variables son $I(1)$, puede existir una relación de largo plazo entre las variables: una tendencia común, cuya combinación lineal es estacionaria $I(0)$.

$$\text{Log}(ex_t) = \widehat{\delta}_0 + \widehat{\delta}_1 \text{Log}(pcad_t) + \widehat{\delta}_2 \text{Log}(pus_t) + u_t \quad (24)$$

Donde $\widehat{\delta}_1$ y $\widehat{\delta}_2$ corresponden a elasticidades de largo plazo.

$$\widehat{\delta}_1 < 0; \widehat{\delta}_2 > 0.$$

Para evidenciar de que existe una relación de largo (cointegración), se deberán aplicar las metodologías de Engle & Granger (1987) y de Phillips-Ouliaris.

Si no existe cointegración, la estimación (24) será un resultado espurio o regresión falsa.

De otra forma si las variables de precios tienen el mismo grado de cointegración $I(d)$, se puede evidenciar si tienen una relación de largo plazo:

$$\text{Log}(pcad_t) = \widehat{\tau}_0 + \widehat{\tau}_1 \text{Log}(pus_t) + u_t \quad (25)$$

Donde $\widehat{\tau}_1$ corresponde a elasticidad de largo plazo. $\widehat{\tau}_1 > 0$.

III. Un modelo con corrección de error

Si existe cointegración, podemos señalar que la dinámica de largo plazo influye o se incorpora en la dinámica de corto plazo, a través de un mecanismo del término de perturbación estocástica como retroalimentación o ajuste en un período:

$$\Delta \text{Log}(ex_t) = \widehat{\alpha}_0 + \widehat{\alpha}_1 \Delta \text{Log}(pcad_t) + \widehat{\alpha}_2 \Delta \text{Log}(pus_t) + \widehat{\gamma} u_{t-1} + \epsilon_t \quad (26)$$

Donde: $u_{t-1} = \text{Log}(ex_{t-1}) - \widehat{\delta}_0 - \widehat{\delta}_1 \text{Log}(pcad_{t-1}) - \widehat{\delta}_2 \text{Log}(pus_{t-1})$; $\gamma < 0$

En otras palabras, el mecanismo de corrección de error desarrollado por Engle y Granger sirve para conciliar el comportamiento de corto plazo de una variable económica con su comportamiento de largo plazo.

Por lo cual, si Y_t se incrementa por encima de sus determinantes de largo plazo (X_t), existirá un mecanismo de ajuste en el período siguiente ($\gamma < 0$).

4. RESULTADOS

Antes de explicar los resultados obtenidos, el análisis de cointegración uniecuacional en *RStudio* requiere instalar previamente las siguientes paqueterías y librerías de forma respectiva (*Script 1*):

```
#### Script 1#####  
#### Paqueterías y librerías en RStudio#####  
library(PerformanceAnalytics)  
library(urca)  
library(aTSA)  
library(cointReg)
```

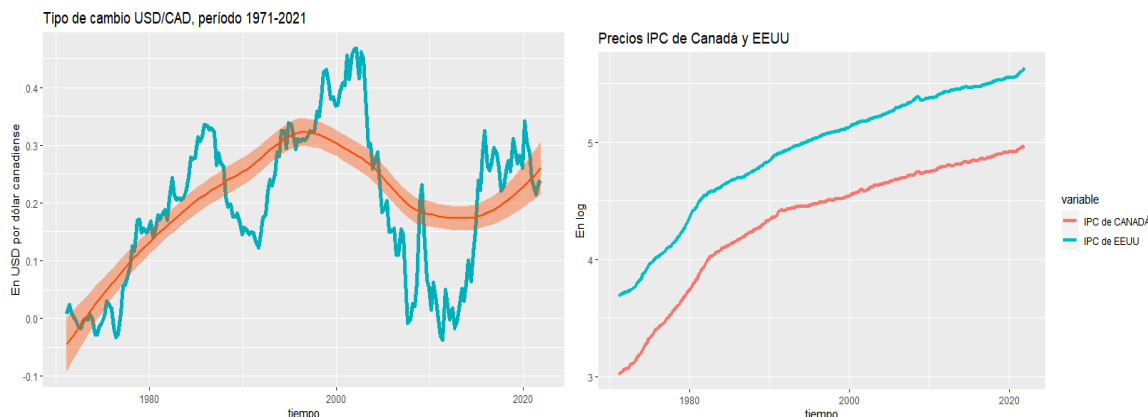
De forma seguida, el siguiente paso fue graficar las series de tiempo (*figura 1*). Para el tipo de cambio USD/CAD se asumió una tendencia con rumbo; mientras que para los IPC de Canadá y EEUU respectivamente se asumió una tendencia con rumbo y tendencia determinística. De forma a priori, no se vislumbraría movimientos entre las tres variables, solamente entre los niveles de precios de ambos países respectivamente (*figura 1.b*). De igual manera, se espera que cada una de las

series no sea estacionaria en niveles y en consecuencia, se tengan que aplicar diferencias.

Figura 1. Series de tiempo utilizadas

a. Tipo de cambio USD/CAD

b. IPC de Canadá y EEUU



Fuente: Elaboración propia con base en el *Bloomberg*.

El siguiente paso consistió en aplicar las pruebas de raíces unitarias Dickey-Fuller-Aumentada (DFA), Phillips-Perron (Ph-P) y KPSS. Tanto DFA como Ph-P evalúan la H_0 que la serie es no estacionaria versus la H_1 que la serie estacionaria, ya sea en niveles $I(0)$, primera $I(1)$, o segunda diferencia $I(2)$ respectivamente; en cambio, KPSS evalúa la H_0 que la serie es estacionaria, versus la H_1 que la serie no es estacionaria (*Scripts: 2, 3 y 4*):

```
#### Script 2#####
#### Ejemplo de prueba DFA en RStudio#####
help("ur.pp")
# Ho: La serie no es estacionaria/ presenta una raíz unitaria
# H1: La serie es estacionaria
#Tipos de tendencias "none", "constant", "trend"

log_ex.pp <- ur.pp(log_ex, type="Z-tau", model="constant",
lags="long")
summary(log_ex.pp)
##### Ejemplo, no ejecute#####
```



```
#### Script 3#####
#### Ejemplo de prueba Ph-P en RStudio#####
help("ur.df") # <- Nos ayuda a conocer los argumentos de la función
a utilizar

ur.df.log_ex<-ur.df(log_ex, type='drift', lags=12, selectlags =
'AIC')
summary(ur.df.log_ex)
##### Ejemplo, no ejecute#####

#### Script 4#####
#### Ejemplo de prueba KPSS en RStudio#####
help("ur.kpss")# <- Nos ayuda a conocer los argumentos de la
función a utilizar

log_ex.kpss <- ur.kpss(log_ex, type="tau", lags="long")
summary(log_ex.kpss)
##### Ejemplo, no ejecute#####
```

Cuadro 2. Pruebas de raíces unitarias (no estacionariedad)

Dickey-Fuller-Aumentada (DFA)									
Series	En niveles			En primera diferencia			En segunda diferencia		
	Especificación	Nro. De rezagos	Estadístico t	Especificación	Nro. De rezagos	Estadístico t	Especificación	Nro. De rezagos	Estadístico t
Log USD/CAD	Con rumbo	1	-2.4033	Ninguna	1	-10.0161***	Ninguna		
Log IPC CAD	Con rumbo y tendencia determinística	12	-3.4129*	Ninguna	11	-1.8861*	Ninguna	10	-6.2356***
Log IPC US	Con rumbo y tendencia determinística	3	-2.9854	Ninguna	4	-1.949	Ninguna	12	-6.2466 ***

Phillips-Perron					KPSS			
Series	En niveles	En primera dif.	En segunda dif.	Orden de integración [I(d)]	En niveles	En primera dif.	En segunda dif.	Orden de integración [I(d)]
Log USD/CAD	-2.1889	-12.7338***		1	0.1867***	0.0608		1
Log IPC CAD	-7.0294***			0	0.3299***	0.2192***	0.0475	2
Log IPC US	-5.282***			0	0.3276***	0.1444*	0.0602	2

Fuente: Estimaciones propias.

Las pruebas fueron consistentes para el *Log USD/CAD*, misma que es $I(1)$, mientras existieron discrepancias para el resto de las series de tiempo; sin embargo, dos de tres pruebas señalan que el *Log IPC CAD* es $I(2)$, al igual que el *LOG IPC US* es $I(2)$. En consecuencia, al tener distintos de grados de integración, no es posible aplicar la cointegración univariada para las tres series (ecuación cuatro), sólo relación

de largo plazo; pese a ello, si es posible evaluar la relación de largo plazo para las series que tienen el mismo grado de integración (ecuación 5).

En el *cuadro 3*, se estima un modelo con las variables en sentido estacionario $[I(d)]$, donde la variación de la inflación de Canadá y EEUU no inciden significativamente en la variación porcentual cambiaria del tipo de cambio USD/CAD.^ψ De igual manera, se introdujeron dos variables de impulso (= 1) para disminuir el problema de la no normalidad en la distribución de los residuos para 2003.T2 y 2008.T4 respectivamente.^φ

Las estimaciones de corto plazo y la presentación de los resultados econométricos se realizaron con los *Scripts 5 y 6*:

```
#### Script 5#####
#### Ejemplo de estimación de corto plazo #####

muestra3<-ts.intersect(dlog_ex,ddlog_pcad, ddlog_pus, d08q4, d03q2)
View(muestra3)

mod_cp2 <- lm(dlog_ex ~ ddlog_pcad+ ddlog_pus + d08q4 +d03q2,
muestra3)
summary(mod_cp2)

resid2 <- mod_cp2$residuals
resid2 <- ts(resid2, end=c(2021,4), freq=4)
plot(resid2)

# Pruebas de normalidad
jarque.bera.test(resid2)
shapiro.test(resid2)
```

^ψ Una diferencia logarítmica equivale a un cambio porcentual.

^φ Dada la cantidad de observaciones atípicas para este tipo de modelos sería pertinente modelizar en presencia de volatilidad ARCH-GARCH (se verá más adelante).

```
#### Script 6#####
#### Presentación de resultados econométricos, corto plazo
#####

library(stargazer)

stargazer(mod_cp2, type="text", title="Modelos de regresión de
corto plazo")
#### Ejemplo, no ejecute#####
```

Cuadro 3. Modelo de corto plazo

Modelos de regresión de corto plazo

Dependent variable:	
$\Delta \log \text{ USDCAD}$	
$\Delta \log \text{ pcad}$	0.099 0.255
$\Delta \log \text{ pcad}$	-0.289 0.326
d08q4	0.125*** 0.032
d03q2	0.089*** 0.03
Constant	0.001 0.002
Prob. Jarque-Bera	<0.10
Prob. Shapirlo-Wilk:	<0.05
Observations	202
R2	0.135
Adjusted R2	0.118
Residual Std. Error	0.029 (df = 197)
F Statistic	7.701*** (df = 4; 197)
Note:	*p<0.1; **p<0.05; ***p<0.01

Fuente: Estimaciones propias.

De igual manera, se procedió en evaluar una ecuación de largo plazo, con variables en sentido no estacionario (cuadro 4); donde un cambio de +1% en los precios de EEUU se relaciona con cerca de +1% en los precios de Canadá al nivel del 0.01 de significancia estadística; no obstante se evaluó la cointegración respectiva.

Cuadro 4. Modelo de largo plazo

Dependent variable:	
Log de los precios de Canadá	
log_pus	0.990*** 0.006
Constant	-0.540*** 0.03
Prob. Jarque-Bera	<0.05
Prob. Shapirlo-Wilks	<0.05
Prob. Cointegr. E&G	>=0.10
Prob. Cointegr. Ph.-O	>=0.10
Observations	204
R2	0.992
Adjusted R2	0.992
Residual Std. Error	0.046 (df = 202)
F Statistic	26,572.210*** (df = 1; 202)

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

Fuente: Estimaciones propias.

En consecuencia, los resultados del cuadro 4, corresponde a una regresión espuria (falsa), al evaluar la evaluación de cointegración uniecuacional de Engle & Granger y Phillips-Ouliaris; por lo cual, no se rechaza la H_0 . De no cointegración o no relación de largo plazo ($P.Value > 0.10$; *Valor estadístico < Valores críticos*) (Scripts: 7, 8 y 9):

```
#### Script 7#####
#### Ejemplo de estimación de largo plazo #####
mod_lp <- lm(log_cad ~ log_us)
summary(mod_lp)

res_lp <- residuals.lm(mod_lp)
res_lp <- ts(res_lp, end=c(2021,4), freq=4)
plot(res_lp)

#analizamos la normalidad de los residuos
jarque.bera.test(res_lp)
shapiro.test(res_lp)
##### Ejemplo, no ejecute#####

#### Script 8#####
#### Evaluación de cointegración#####
#### Engle & Granger #####

coint.df <- ur.df(y=res_lp, type='none',lags=8,
selectlags=c("AIC"))
summary(coint.df)

X<-ts.intersect(log_pus)

coint.test(log_pcad, X, d = 0)
##### Ejemplo, no ejecute#####

#### Script 9#####
#### Evaluación de cointegración#####
#### Phillips-Ouliaris #####

help("ca.po")
z<-ts.intersect(log_pcad, log_pus)
test.po<-ca.po(z, demean = 'constant', lag = 'long', type = 'Pu',
tol = NULL)
summary(test.po)
##### Ejemplo, no ejecute#####
```

Cuadro 5. Evaluación de cointegración uniecuacional
Engle & Granger (1987); Philiips-Ouliaris

Engle-Granger Cointegration Test

Ho: Las series no están cointegradas

H1: Las series están cointegradas o relación de largo

Type 1: no trend

lag	EG	p.value
4.00	-1.35	0.10

Type 2: linear trend

lag	EG	p.value
4.00	-1.22	0.10

Type 3: quadratic trend

lag	EG	p.value
4.00	2.21	0.10

Note: p.value = 0.01 means p.value $\leq$ 0.01

p.value = 0.10 means p.value $\geq$ 0.10

Phillips-Ouliaris Cointegration test

Ho: Las series no están cointegradas

H1: Las series están cointegradas o relación de largo

Valor estadístico

4.3171

	10pct	5pct	1pct
Valores críticos	27.8536	33.713	48.0021

Fuente: Estimaciones propias.

En caso de haber existido cointegración, es decir con rechazo de H_0 ($P.Value < 0.10$; $Valor\ estadístico > Valores\ críticos$), se podría haber estimado el mecanismo corrector de error (cuadro 6); donde $ECM < 0$. Para efectos de ejemplificación el mismo fue estimado *Script 10*).

Cuadro 6. Mecanismo corrector de error (ECM)

ECM model

Var. Dependiente $\Delta \log P CAD$

	Estimate	Std. Error	t value	Pr(> t)	
$\Delta \log P US$	0.89	0.04	23.04	0.00	***
ECM	0.02	0.01	- 2.98	0.00	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.006492 on 201 degrees of freedom

Multiple R-squared: 0.775 Adjusted R-squared: 0.7729

F-statistic: 346.5 on 2 and 201 DF, p-value: < 2.2e-16

```
#### Script 10#####
#### Mecanismo corrector de error (ECM)#####
help("ecm")

mod_ecm<-ecm(log_pcad,X)
mod_ecm

res_ecm <- mod_ecm$residuals

jarque.bera.test(res_ecm)
shapiro.test(res_ecm)##### Ejemplo, no
ejecute#####
```

El lector deberá verificar que el correcto orden de integración de *Log P CAD* y *Log P US* es $I(2)$; por ello, el lenguaje de programación *RStudio* realiza la estimación ECM en $I(1)$ de forma automática; por lo cual, deberá realizarse una estimación manual: guardando los residuos de la ecuación de largo plazo e introduciendo como una variable explicativa respectivamente.

Otras extensiones de regresión de largo plazo fueron calculadas para efectos de ejemplificación: MCO- completamente modificados (FM-OLS); MCO-dinámicos

(FM-OLS) y MCO- integrados y modificados (IM-OLS) (*cuadro 7*). Pesé a ello, se observas que sus gráficos residuos no son estacionarios, sino que presentan tendencias (*figuras: 2, 3 y 4* respectivamente) *Scripts: 11, 12 y 13*):

```
#### Script 11#####
#### Estimación de largo plazo, FM-OLS#####
help("cointRegFM")
xfm<-cbind(log_pus)
xfm

# Rumbo
const<-ts(1, start=c(1971,1), end=c(2021,4), freq=4)
length(const)

# Rumbo y tendencia determinística
const_trend <- cbind(level = 1, trend = 1:204)

# Especificación con rumbo
eq_fm1 <- cointRegFM(xfm, log_pcad, deter=const, kernel = "ba",
bandwidth = "and")
print(eq_fm1)
plot(eq_fm1)

# Especificación con rumbo y tendencia
eq_fm2<- cointRegFM(xfm, log_pcad, deter = const_trend, kernel =
"ba", bandwidth = "and")
print(eq_fm2)
plot(eq_fm2)
##### Ejemplo, no ejecute#####

#### Script 12#####
#### Estimación de largo plazo, D-OLS#####
help("cointRegD")

# Especificación con rumbo
eq_D <- cointRegD(xfm, log_pcad, deter=const, kernel = "ba",
bandwidth = "and")
print(eq_D)
plot(eq_D)

# Especificación con rumbo y tendencia
eq_D2<- cointRegD(xfm, log_pcad, deter = const_trend, kernel = "ba",
bandwidth = "and")
print(eq_D2)
plot(eq_D2)

##### Ejemplo, no ejecute#####
```



```

#### Script 13#####
#### Estimación de largo plazo, IM-OLS#####
help("cointRegIM")

# Especificación con rumbo
eq_IM <- cointRegIM(xfm, log_pcad, deter=const, kernel = "ba",
bandwidth = "and")
print(eq_IM)
plot(eq_IM)

#Especificación con rumbo y tendencia
eq_IM2<- cointRegIM(xfm, log_pcad, deter = const_trend, kernel =
"ba", bandwidth = "and")
print(eq_IM2)
plot(eq_IM2)
##### Ejemplo, no ejecute#####

```

Cuadro 7. Extensiones de regresiones de largo plazo
MCO-Completamente modificados (FM-OLS)
MCO-dinámicos (D-OLS)
MCO-integrados y modificados (IM-OLS)

a.1) MCO-Completamente modificados

Especificación: con rumbo

	Estimate	Std.Err	t value	Pr(t >0)
deter	-0.506913	0.138199	-3.668	0.0003126 ***
x.coint	0.981328	0.027902	35.17	2.20E-16 ***

a.2) MCO-Completamente modificados

Especificación: con rumbo y tendencia determinística

	Estimate	Std.Err	t value	Pr(t >0)
level	- 1.56	0.17	- 9.35	0.00 ***
trend	- 0.00	0.00	- 6.54	0.00 ***
x.coint	1.25	0.04	30.29	0.00 ***

b.1) MCO-dinámicos

Especificación: con rumbo

	Estimate	Std.Err	t value	Pr(t >0)
deter	- 0.54	0.15	- 3.67	0.00 ***
x.coint	0.99	0.03	33.27	0.00 ***

b.2) MCO-dinámicos

Especificación: con rumbo y tendencias determinística

	Estimate	Std.Err	t value	Pr(t >0)
level	- 1.45	0.18	- 8.02	0.00 ***
trend	- 0.00	0.00	- 6.19	0.00 ***
x.coint	1.23	0.04	28.17	0.00 ***

c.1) MCO- Integrado y modificado

Especificación: con rumbo

	Estimate	Std.Err	t value	Pr(t >0)
S.deter	- 0.26	0.19	- 1.41	0.16
S.x.coint	0.94	0.04	25.40	0.00 ***

c.2) MCO- Integrado y modificado

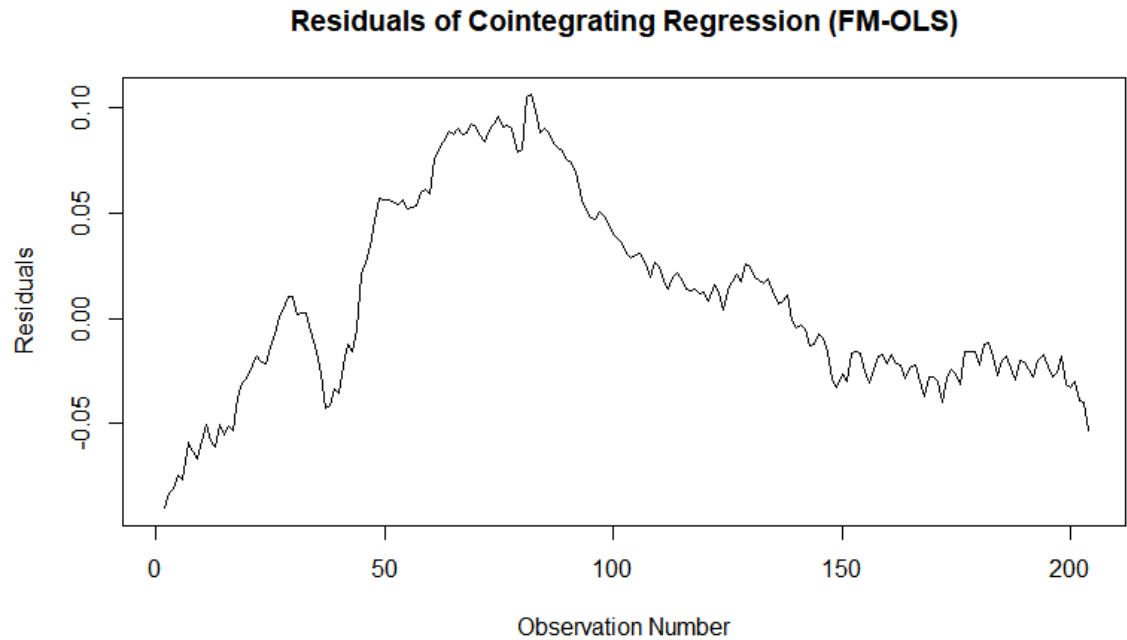
Especificación: con rumbo y tendencias determinística

	Estimate	Std.Err	t value	Pr(t >0)
S.level	- 1.72	0.22	- 7.95	0.00 ***
S.trend	- 0.00	0.00	- 6.35	0.00 ***
S.x.coint	1.29	0.05	24.43	0.00 ***

Signif. codes: 0 '*' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1**

Figura 2. Residuos de ecuación de cointegración (FM- OLS)

a. Especificación con rumbo



b. Especificación con rumbo y tendencia determinística

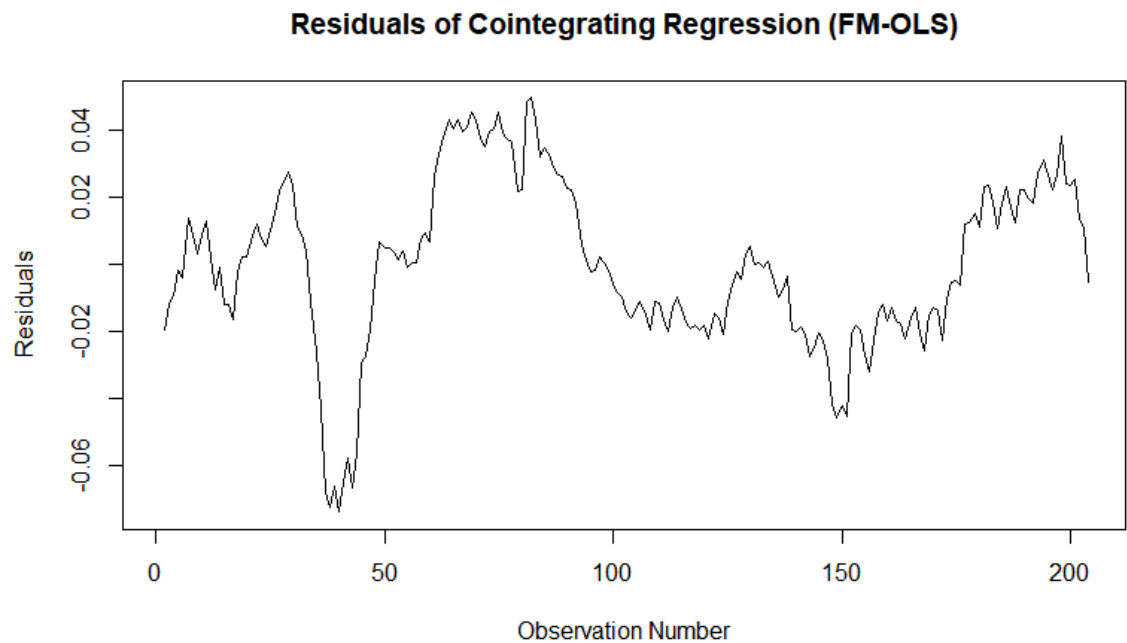
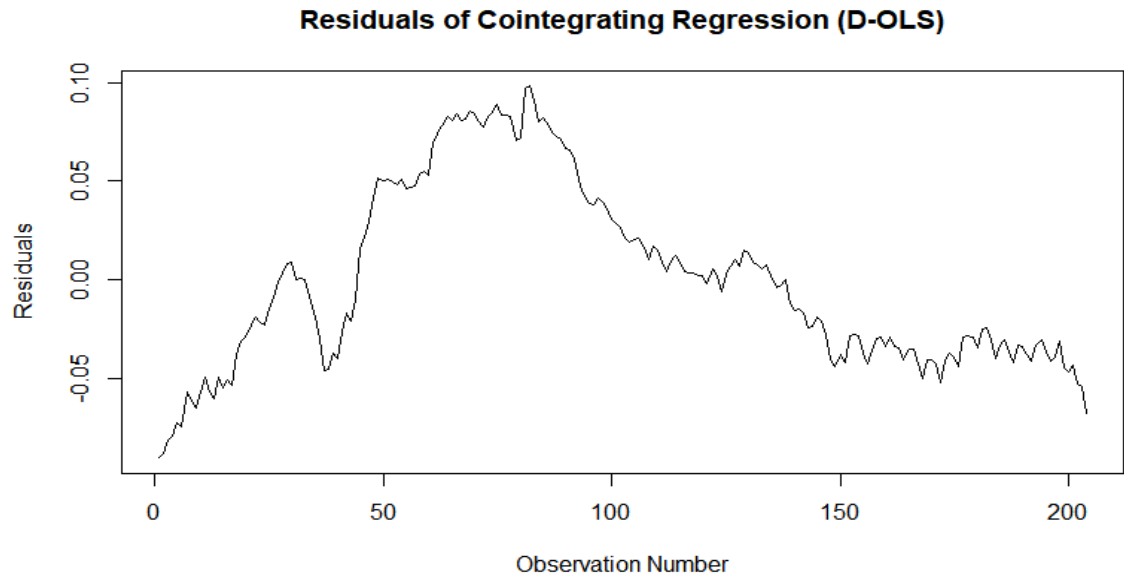


Figura 3. Residuos de ecuación de cointegración (D- OLS)

a. Especificación con rumbo



b. Especificación con rumbo y tendencia determinística

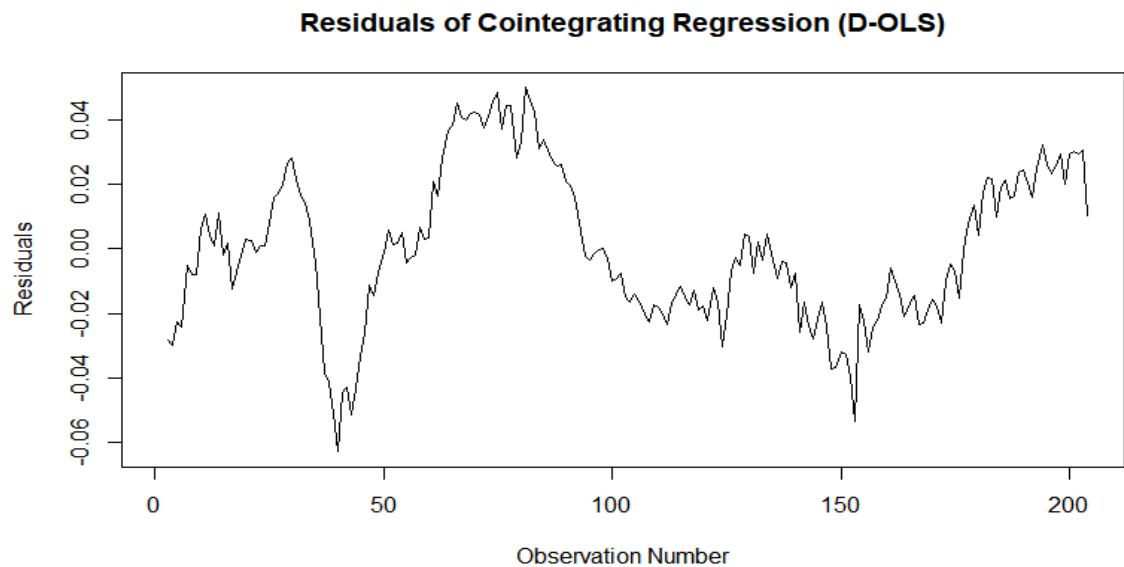
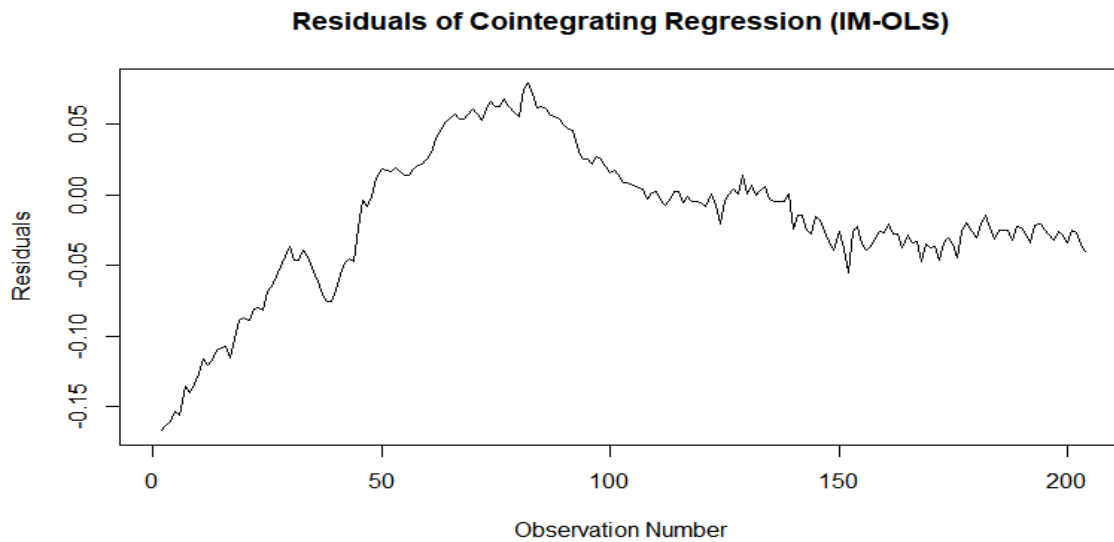
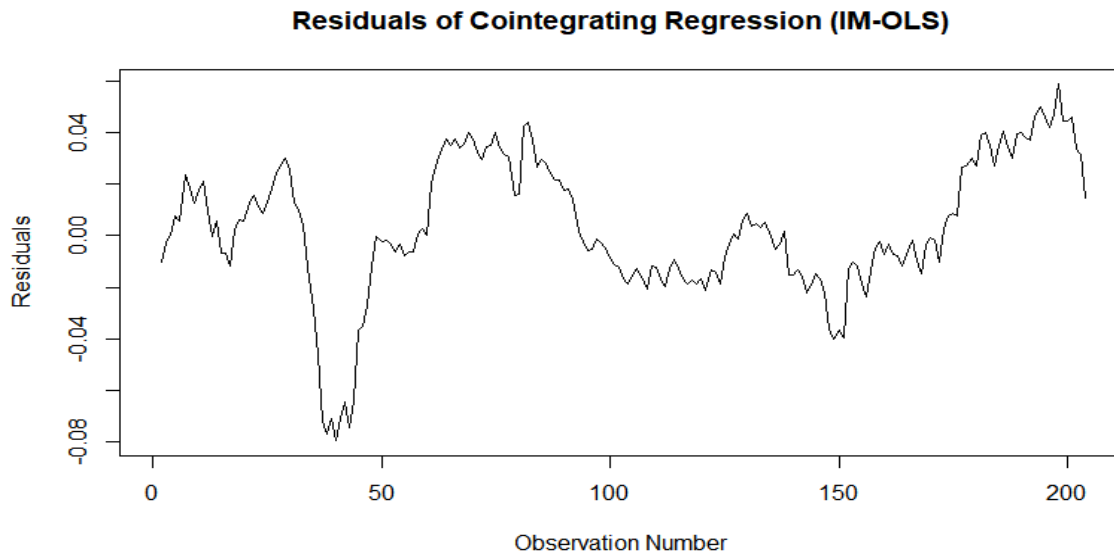


Figura 4. Residuos de ecuación de cointegración (D- OLS)

a. Especificación con rumbo



b. Especificación con rumbo y tendencia determinística



SÍNTESIS

La cointegración es un supuesto moderno de la econometría de series de tiempo que evalúa la hipótesis de co-movimientos de largo plazo entre las series económicas y financieras, puede ser de dos tipos: uniecuacional y multiecuacional. En esta sección, se abordó la primera.

La propuesta metodológica de [Engle & Granger \(1987\)](#) consistió en cuatro pasos para cointegración uniecuacional: la realización de pruebas de raíz unitaria o no estacionariedad, la realización de la estimación y prueba de cointegración, el mecanismo de corrección de error (ECM) y la validación de pruebas y supuestos econométricos.

El primer punto de partida es el análisis de estacionariedad de las series, usualmente mediante Dickey-Fuller-Aumentada (DFA), Phillips-Perron (Ph-P) y KPSS (existen otras pruebas aditivas). En suma, la cointegración se presenta cuando las series a pesar de ser no estacionarias, su combinación lineal lo es. Como punto de partida la cointegración uniecuacional generalmente funciona mejor entre dos variables. En consecuencia, de forma posterior, se abordará la cointegración multivariada, generalmente aplicada para más de dos variables.

Finalmente, se tienen otras extensiones de cointegración uniecuacional por ciertas limitaciones de mínimos cuadrados ordinarios estáticos (SOLS) como son los mínimos cuadrados completamente modificados (FMOLS), los mínimos cuadrados dinámicos (DOLS) y los mínimos cuadrados integrados y modificados (IM-OLS). En todos estos casos se deberán apreciar residuos estacionarios para evidenciar relaciones de largo plazo.

REFERENCIAS

- Engle, R. F., & Granger, C. W. (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251-276.
- Granger, C. W. (1981). Some properties of time series data and their use in econometric model specification. *Journal of econometrics*, 16(1), 121-130.
- Granger, C. W. (1983). Co-integrated variables and error correcting models. UCSD Discussion paper number 83-13a.
- Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of economic dynamics and control*, 12(2-3), 231-254.
- Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica: journal of the Econometric Society*, 1551-1580.
- Johansen, S. (1995). Identifying restrictions of linear equations with applications to simultaneous equations and cointegration. *Journal of econometrics*, 69(1), 111-132.
- Phillips, P. C., & Hansen, B. E. (1990). Statistical inference in instrumental variables regression with I (1) processes. *The Review of Economic Studies*, 57(1), 99-125.
- Phillips, P. C., & Ouliaris, S. (1990). Asymptotic properties of residual based tests for cointegration. *Econometrica: journal of the Econometric Society*, 165-193.
- Vogelsang, T. J., & Wagner, M. (2014). Integrated modified OLS estimation and fixed-b inference for cointegrating regressions. *Journal of Econometrics*, 178(2), 741-760.

ENSAYO 3: MODELOS DE REGRESIÓN NO LINEAL (MRNL)

Roger Alejandro Banegas Rivero
UNIVERSIDAD AUTÓNOMA GABRIEL RENÉ MORENO^σ

Este diseño, 04 de mayo de 2022

INTRODUCCIÓN

Uno de los términos principales en la econometría consiste en el término de linealidad, el cual se refiere que los parámetros del modelo son de grado uno: $\beta^1, \delta^1, \gamma^1$, etc. En consecuencia, las variables explicativas pueden ser lineales o no lineales y el modelo seguirá siendo de tipo lineal. Por consiguiente, los modelos de regresión no lineal (*MRNL*) son aquellos cuyos estimadores son diferentes de la unidad: $\beta^{1/2}, \delta^2, \gamma^{3/2}$, etc.

Por otra parte, una de las características más importantes en estos modelos se centra en sus propiedades asintóticas (aplicable a muestras grandes); es decir, en la medida que aumenta la muestra, los parámetros estimados convergerán hacia sus verdaderos valores poblacionales; en caso contrario o en presencia de muestras pequeñas, los resultados deberán interpretarse con precaución.

De igual manera, las aplicaciones en *MRNL* están relacionadas con estudios previos en casos tradicionales y relaciones matemáticas sugeridas por la teoría. Asimismo, sus estimaciones se caracterizan por contar un elevado coeficiente de determinación o bondad de ajuste (Ej. $R^2 > 0.70$) por el ajuste requerido de la estimación a los datos observados.

En lo relacionado con los supuestos clásicos del modelo de regresión lineal (*MCRL*), tales como cero autocorrelación, residuos homocedásticos, errores con

^σ Correo electrónico: rogerbanegas@uagrm.edu.bo

distribución normal y demás pruebas de especificación son aplicables a los MRNL; en caso de incumplimiento de los supuestos puede afectar a sus resultados.

En tal sentido, este documento está conformado por tres secciones. La primera aborda los temas esenciales metodológicos relacionados con definiciones y procedimientos alternativos de estimación; la segunda involucra los aspectos relacionados con las estimaciones en RStudio; La tercera sección aborda una aplicación No Lineal a la estimación de cantidad de infectados por COVID-19 según el modelo de Gompertz. Al final, se presenta una síntesis de los principales tópicos abordados.

1. ASPECTOS ESENCIALES BÁSICOS

Otra definición básica consiste en aquellas especificaciones que de forma inicial o aparente no son modelos lineales, sin embargo a través de algebra básica y operaciones matemática se transforma en un modelo lineal, lo cual se conoce como modelos intrínsecamente lineales. En caso, de no lograr la transformación respectiva se conoce como modelo intrínsecamente no lineal.

Veamos el ejemplo de un modelo macroeconómico de tipo Cobb-Douglas:

$$\hat{y}_t = \widehat{A}_t k_t^{\hat{\beta}} l_t^{\hat{\delta}_t} e^{\hat{\epsilon}_t} \quad (1)$$

En (1) $\hat{y}_t$ representa la producción agregada (*PIB*); k_t simboliza el acervo de capital físico; l_t representa el número de trabajadores; $\widehat{A}_t$ representa una constante tecnológica. Por lo cual, $\hat{\beta}_t$ es el aporte del capital al producto y $\hat{\delta}_t = 1 - \hat{\beta}_t$ es la contribución del trabajo respectivamente; $\hat{\epsilon}_t$ es un término de perturbación estocástico. La pregunta consiste en responder si la expresión (1), se constituye en un modelo de regresión no lineal (*MRNL*).

Para este caso, se asume rendimientos constantes a escala: el aumento de 1% en el acervo del capital físico y el aumento de 1% en el factor trabajo conlleva al aumento del +1% en la producción agregada de la economía ($\widehat{\beta}_t + \widehat{\delta}_t = 1$). De forma alternativa, se podría asumir rendimientos decrecientes ($\widehat{\beta}_t + \widehat{\delta}_t < 1$) a escala o rendimiento crecientes a escala ($\widehat{\beta}_t + \widehat{\delta}_t > 1$).

Aplicando logaritmo ambas partes en (1), se obtiene *un modelo intrínsecamente lineal*:

$$\ln \left(\frac{y}{l} \right)_t = \widehat{\alpha} + \widehat{\beta} \ln \left(\frac{k}{l} \right)_t + \widehat{\epsilon}_t \quad (2)$$

En la expresión (2) se llega la conclusión relevante en la economía que el producto por trabajador $\left(\frac{y}{l} \right)_t$ depende del capital por trabajador $\left(\frac{k}{l} \right)_t$, represando $\widehat{\beta}$ una elasticidad respectiva. En suma, la expresión (1) no se constituye en modelo de regresión no lineal. En otros casos, cuando no es posible realizar la transformación en modelos lineales, se asumiría que son *modelos de regresión intrínsecamente no lineales*.

Se menciona la elasticidad constante de sustitución (*CES*) con dos y tres factores productivos respectivamente de forma alternativa a la función de producción de tipo Cobb-Douglas:

$$\widehat{y}_t = e^{\widehat{\lambda}t} \left[\widehat{\beta} k_t^{-\widehat{\rho}} + (1 - \widehat{\beta}) l_t^{-\widehat{\rho}} \right]^{-\widehat{\nu}/\widehat{\rho}} \quad (3)$$

$$\widehat{y}_t = e^{\widehat{\lambda}t} \left[\widehat{\beta} \left(\widehat{\beta}_1 k_t^{-\widehat{\rho}_1} + (1 - \widehat{\beta}_1) l_t^{-\widehat{\rho}_1} \right)^{-\rho/\rho_1} + (1 - \widehat{\beta}) r r n n_t^{-\widehat{\rho}} \right]^{-\widehat{\nu}/\widehat{\rho}} \quad (4)$$

En (3) y (4) la elasticidad de sustitución (*CES*): $s = \frac{1}{1+\rho}$; donde es el tiempo y corresponde a la abundancia de recursos naturales; $\widehat{\lambda}$, $\widehat{\rho}$, $\widehat{\nu}$, $\widehat{\rho}_1$ y $\widehat{\beta}_1$ son parámetros adicionales estimados del modelo. Más allá del mecanismo de incorporar el término de perturbación estocástico en las expresiones señaladas, no existe ninguna

alternativa en realizar alguna transformación para linealizar los parámetros; por tanto, se constituyen en modelos de regresión intrínsecamente no lineales.

Para estimar los parámetros en *MRNL*, existen tres formas de estimar MRNL:

- a. Ensayo y error: es un proceso iterativo y se probando en nuevas estimaciones secuenciales hasta minimizar la suma de cuadrados residuales.
- b. Mínimos cuadrados no lineales (MCNL): se emplea un método de optimización y pasos hacia descendentes (iteración); es decir, ajuste a partir de valores iniciales (al igual que el método de ensayo y error).
- c. Expansión de series de Taylor: Se parte de una linealización de parámetros iniciales a través de MCO, ajustándose los valores iniciales estimados hasta que los parámetros no tengan variaciones significativas (convergencia). Se cuentan con herramientas computacionales potentes: rutinas integradas, como las de Gauss-Newton, Newton-Raphson y la de Marquard.

2. ESTIMACIÓN DE MODELOS LINEALES Y NO LINEALES EN RSTUDIO

Dependiendo del tipo de estimaciones, existen dos paqueterías y librerías aplicables en la estimación de modelos de regresión no lineal (*MRNL*) en el *Script 1*: ‘*easynls*’ que permite estimar 13 modelos lineales y no lineales; así como ‘*micEconCES*’ exclusivamente para la estimación de funciones de producción de tipo CES. De igual manera, se pueden incluir otras paqueterías y librerías, tales como: ‘*tidyverse*’ para análisis de ciencias de datos, así como ‘*ggplot2*’ y ‘*colorspace*’ para presentar gráficos elegantes respectivamente.

```
#### Script 1#####
#### Paqueterías y librerías en RStudio#####
install.packages("easynls")
library(easynls)

install.packages("micEconCES")
library("micEconCES" )

install.packages("tidyverse")
library(tidyverse)

install.packages("ggplot2")
library(ggplot2)

install.packages("colorspace")
library(colorspace)
```

En tal sentido y conforme a lo señalado previamente, se expresan los 13 tipos de especificaciones con ‘*easynls*’:

Cuadro 1. Comparación de modelos de regresión lineal (***MRL***) y modelos de regresión no lineal (***MRNL***)

Ejemplificación	Modelos
a) $y_t = \text{Salarios en USD/Mes}$ $x_t = \text{Edad del trabajador (en años)}$	1, 2, 4, 6, 8 y 9 Mayor ajuste: modelo 2, $R^2 = 0.7289$ Menor ajuste: modelo 6, $R^2 = 0.3322$
b) $y_t =$ Peso de raza bovina Angus (En kg) $x_t = \text{Edad del ganado(en meses)}$	3, 5, 7 y 10 Mayor ajuste: modelo 7 y 10, $R^2 = 0.993$ Menor ajuste: modelo 12, $R^2 = 0.986$
c) $y_t = \text{Cantidad de lactancia (leche)}$ $x_t = \text{Edad del bebé (en días)}$	11 Ajuste: $R^2 = 0.9396$
d) $y_t = \text{Degradación}; x_t = \text{Tiempo}$	12 Ajuste: $R^2 = 0.966$

e) $y_t = \text{Producción de gas}$; $x_t = \text{Tiempo}$

13
Ajuste: $R^2 = 0.99$

Creamos cuatro bases de datos en RStudio (*Script 2*):

```
#### Script 2#####  
#### Bases de datos#####  
# data1 representa edad (años) y salarios (USD/mes)  
salarios=c(280,400,800,1000,2000,2500,1800,1900,1500,1300)  
edad=c(21,25,30,36,38,40,45,48,50,55)  
data1=data.frame(edad, salarios)  
  
# data2 represent weights of an Angus cow at ages from 8 to 108  
months (Kaps and Lamberson, 2009)  
weight=c(280,340,430,480,550,580,590,600,590,600)  
age=c(8,12,24,36,48,60,72,84,96,108)  
data2=data.frame(age, weight)  
  
# data3  
milk=c(25,24,26,28,30,31,27,26,25,24,23,24,22,21,22,20,21,19,  
       18,17,18,18,16,17,15,16,14)  
days=c(15,15,15,75,75,75,135,135,135,195,195,195,255,255,255,  
        315,315,315,375,375,375,435,435,435,495,495,495)  
data3=data.frame(days,milk)  
  
#data4  
time=c(2,6,9,24,48,72,96)  
deg=c(20,33,46,55,66,72,76)  
data4=data.frame(time,deg)  
  
# data5 (gas production)  
time=c(0,12,24,36,48,60,72,84,96,108,120,144,168,192)  
gas=c(0.002,3.8,8,14.5,16,16.5,17,17.4,17.9,18.1,18.8,  
      19,19.2,19.3)  
data5=data.frame(time,gas)  
##### Ejemplo, no ejecute#####
```

Estimamos las 13 especificaciones señaladas en ‘*easynls*’:

- RStudio (*Script 3*):

Script 3#####
Estimaciones en RStudio#####

```
# linear
nlsfit(data1, model=1)

nlsplot(data1, model = 1, start = c(a = -320.93342981,
                                   b = 43.01374819),
        xlab = "Edad (años)" ,
        ylab = "Salario/ Mes en USD", position = 7)

# quadratic
nlsfit(data1, model=2)
nlsplot(data1, model = 2, start = c(a = -5.057174e+03,
                                   b = 3.151726e+02,
                                   c=-3.603665e+00),
        xlab = "Edad (años)" , ylab = "Salario/ Mes en USD",
        position = 7)

# linear plateau
nlsfit(data2, model=3)

nlsplot(data2, model = 3, start = c(a = 592.00000000, b = 6.38728325,
                                   c=53.15475108),
        xlab = "X" , ylab = "Y", position = 7)

# logistic
nlsfit(data2, model=7, start=c(600,4,0.05))
nlsplot(data2, model = 7, start = c(a = 600, b = 4, c=0.05),
        xlab = "X" , ylab = "Y", position = 7)

# quadratic plateau
nlsfit(data1, model=4)
nlsplot(data1, model = 4, start = c(a = -4287.5170228,
                                   b = 269.6214416,
                                   c=-3.0130196),
        xlab = "Edad (años)" , ylab = "Salario (USD/mes)", position =
7)

# two linear
nlsfit(data2, model=5, start=c(250,6,2,50))
nlsplot(data2, model = 5, start = c(a = 250, b = 6,
                                   c=2, d=50),
        xlab = "X" , ylab = "Y", position = 7)

# exponential
nlsfit(data1, model=6, start=c(450,0.05))
nlsplot(data1, model = 6, start = c(a = 450, b = 0.05),
        xlab = "Edad (años)", ylab = "Salario (USD/mes)", position = 7)

# logistic
nlsfit(data2, model=7, start=c(600,4,0.05))
nlsplot(data2, model = 7, start = c(a = 600, b = 4, c=0.05),
```

```

        xlab = "X" , ylab = "Y", position = 7)

# van bertalanffy
nlsfit(data1, model=8, start=c(1.785964e+03,8.157733e+01,2.057388e-01))
nlsplot(data1, model = 8, start = c(a = 1.785964e+03,
                                     b = 8.157733e+01,
                                     c=2.057388e-01),
        xlab = "Edad (años)" , ylab = "Salario (USD/mes)",
        position = 7)

# brody
nlsfit(data1, model=9, start=c(1.785964e+03,8.157733e+01,2.057388e-01))
nlsplot(data1, model = 9, start = c(a = 1.860069e+03, b = 7.995949e+00,
                                     c=1.008199e-01),
        xlab = "Edad (años)" , ylab = "Salario (USD/mes)",
        position = 7)

# gompertz
nlsfit(data2, model=10, start=c(600,4,0.05))
nlsplot(data2, model = 10, start = c(a = 600, b = 4, c=0.05),
        xlab = "X" , ylab = "Y", position = 7)

#lactancy
nlsfit(data3, model=11, start=c(16,0.25,0.004))

nlsplot(data3, model = 11, start = c(a = 16, b = 0.25, c=0.004),
        xlab = "días" , ylab = "lactancia", position = 7)
#ruminal degradation
nlsfit(data4, model=12)

#ruminal degradation
nlsplot(data4, model = 12, start = c(a = 17, b = 56, c=0.06),
        xlab = "X" , ylab = "Y", position = 7)

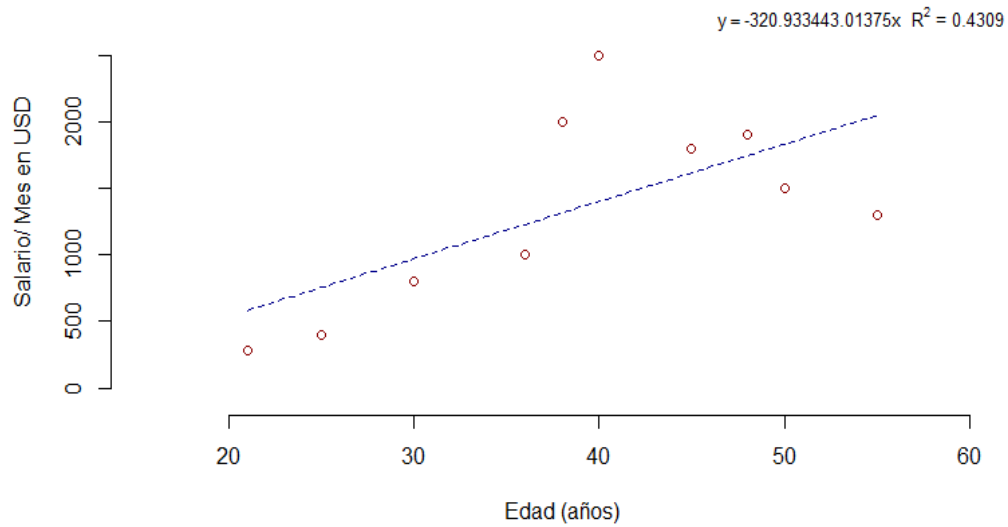
# logistic bi-compartmental (gas production)
nlsfit(data5, model=13, start=c(19,4,0.025,0.004,5))
nlsplot(data5, model = 13, start = c(a = 19, b = 4, c=0.025,
                                     d= 0.004, e=5),
        xlab = "X" , ylab = "Y", position = 7)

##### Ejemplo, no ejecute#####

```

1. $y = \alpha + \beta x + \epsilon$, modelo de regresión lineal (*MRL*).

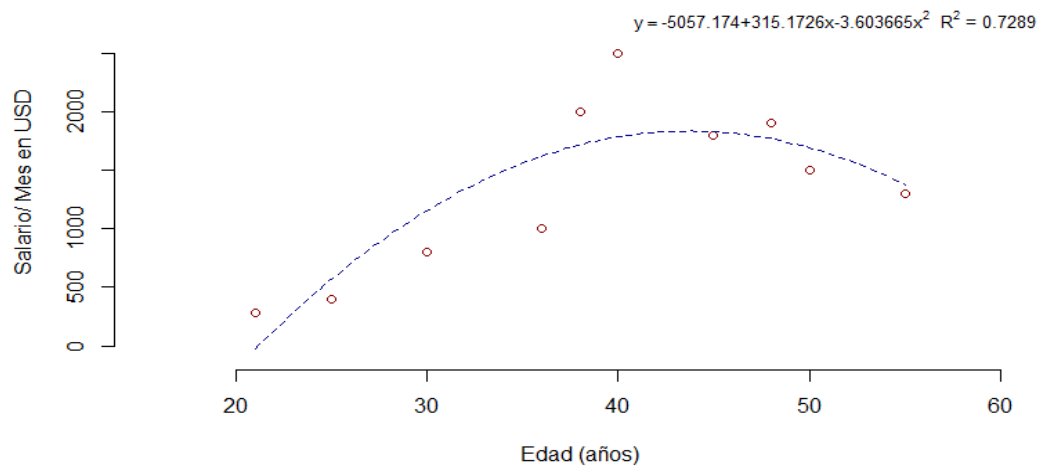
Figura 1. Relación lineal



Nota: MRL a un modelo de salarios.

2. $y = \alpha + \beta_1 x + \beta_2 x^2 + \epsilon$, modelo cuadrático o relaciones curvilíneas (*MRL*).

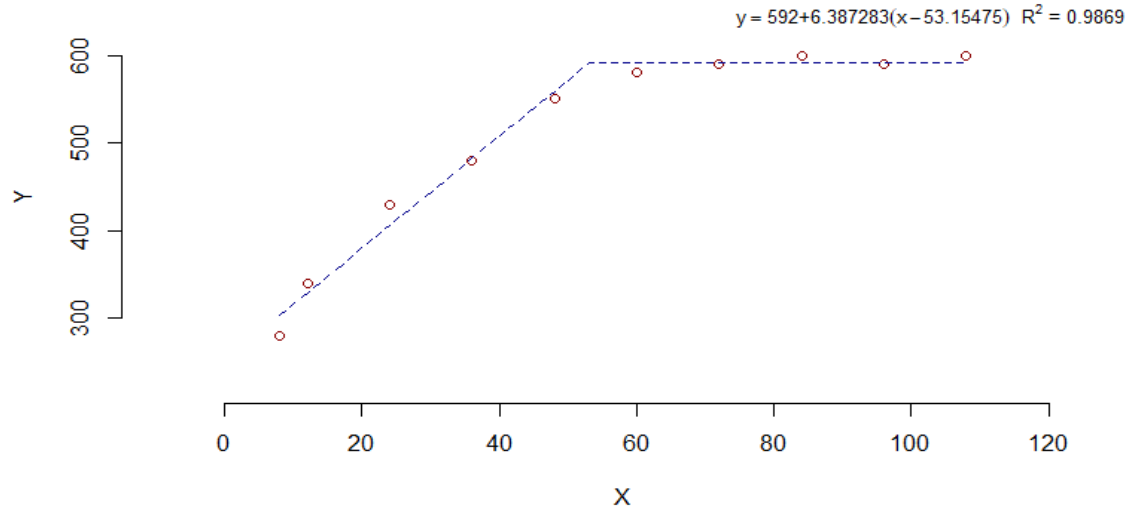
Figura 2. Relaciones curvilíneas



Nota: MRL a un modelo de salarios.

3. $y = \alpha + \beta(x - \delta) * (x \leq \delta) + \epsilon$, modelo de meseta (*MRL*).

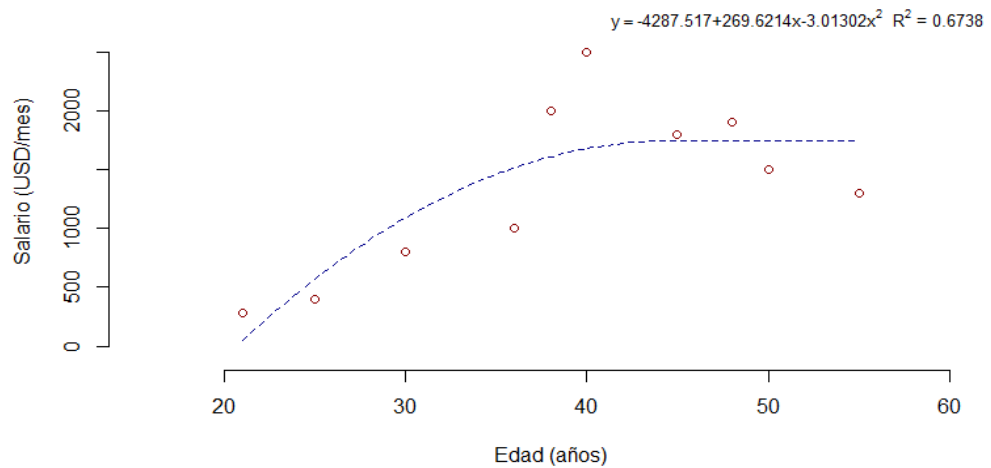
Figura 3. Modelo Meseta



Nota: MRL a un modelo para determinar el peso del ganado vacuno.

4. $y = \alpha + \beta x + \delta I(x^2) * \left[\left(x \leq -0.5 \frac{\beta}{\delta} \right) + (\alpha + I(-\beta^2)/(4\delta)) * \left(x > -0.5 \frac{\beta}{\delta} \right) \right] + \epsilon$,
modelo de meseta cuadrático (*MRLN*).

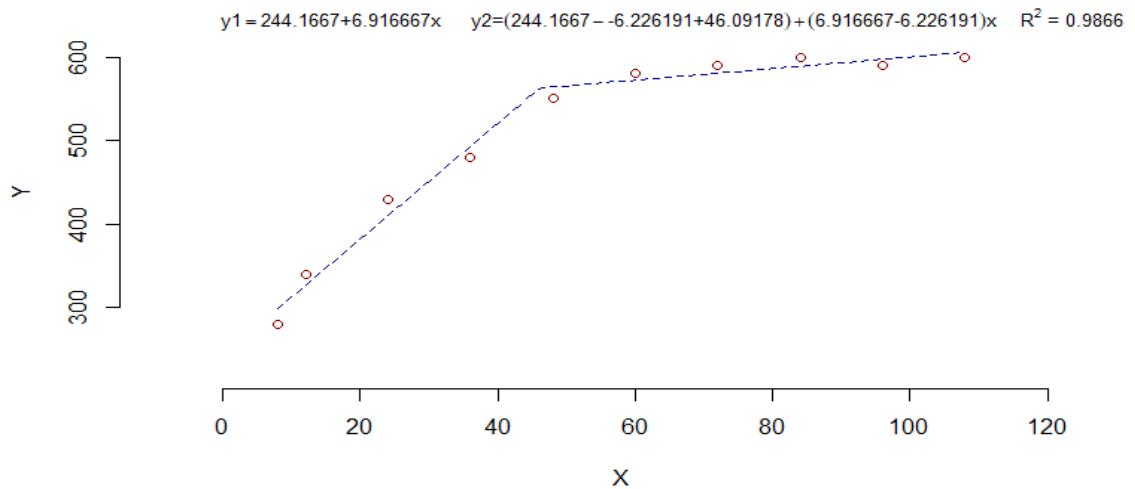
Figura 4. Modelo de Meseta cuadrático



Nota: MRLN a un modelo de salarios.

5. Si $(x \geq \gamma, (\alpha - \delta\gamma) + (\beta + \delta)x + \epsilon, \alpha + \beta x + \epsilon)$, modelo condicional lineal (MRL).

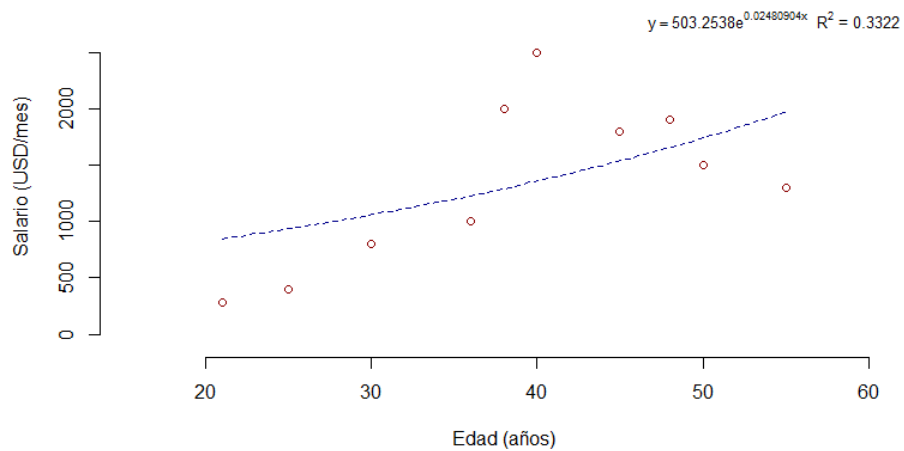
Figura 5. Modelo condicional lineal



Nota: MRL a un modelo para determinar el peso del ganado vacuno.

6. $y = ae^{(\beta x)} + \epsilon$, modelo exponencial (MRNL).

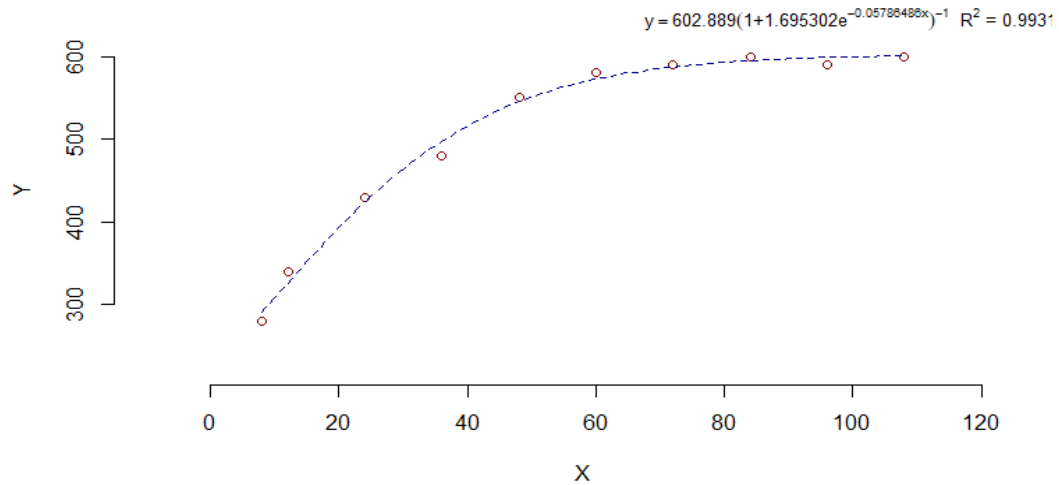
Figura 6. Modelo exponencial



Nota: MRLN a un modelo de salarios.

7. $y = \alpha[1 + \beta e^{(-\delta x)}]^{-1} + \epsilon$, modelo logístico (*MRNL*).

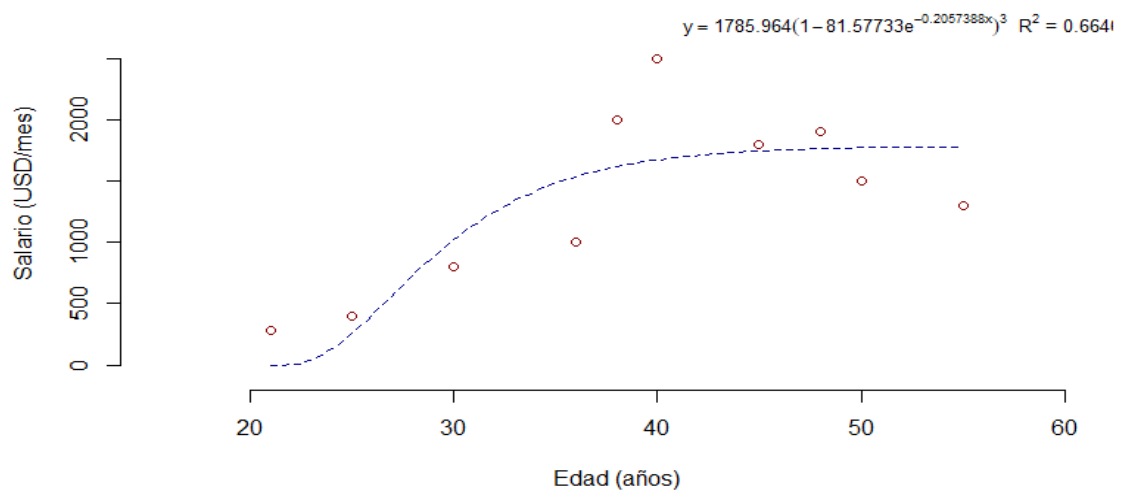
Figura 7. Modelo logístico



Nota: MRLN a un modelo para determinar el peso del ganado vacuno.

8. $y = \alpha * [1 - \beta e^{(-\delta x)}]^3 + \epsilon$, modelo van Bertalanffy (*MRNL*).

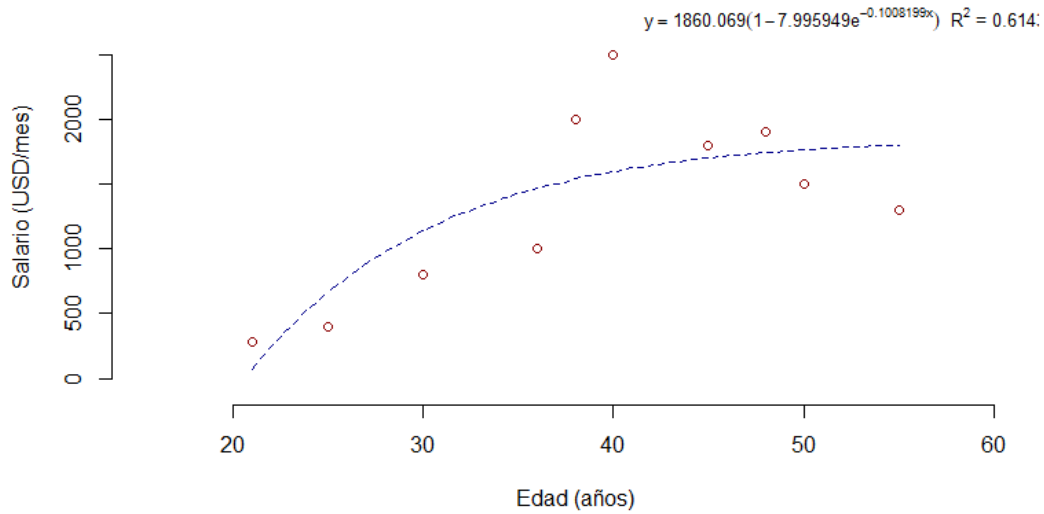
Figura 8. Modelo Bertalanffy



Nota: MRLN a un modelo de salarios.

9. $y = \alpha[1 - \beta e^{(-\delta x)}] + \epsilon$, modelo Brody (*MRNL*).

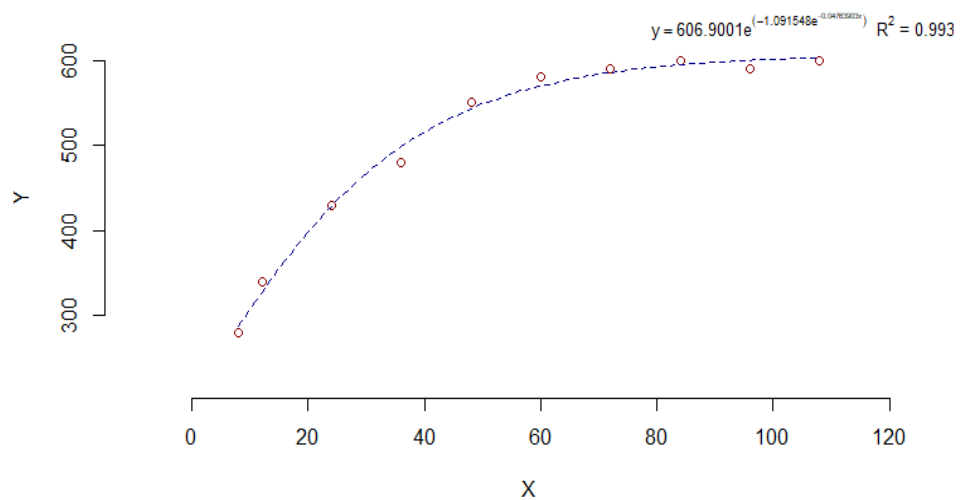
Figura 9. Modelo Brody



Nota: MRLN a un modelo de salarios.

10. $y = \alpha e^{(-\beta e^{(-\delta x)})} + \epsilon$, modelo de Gompertz (*MRNL*).

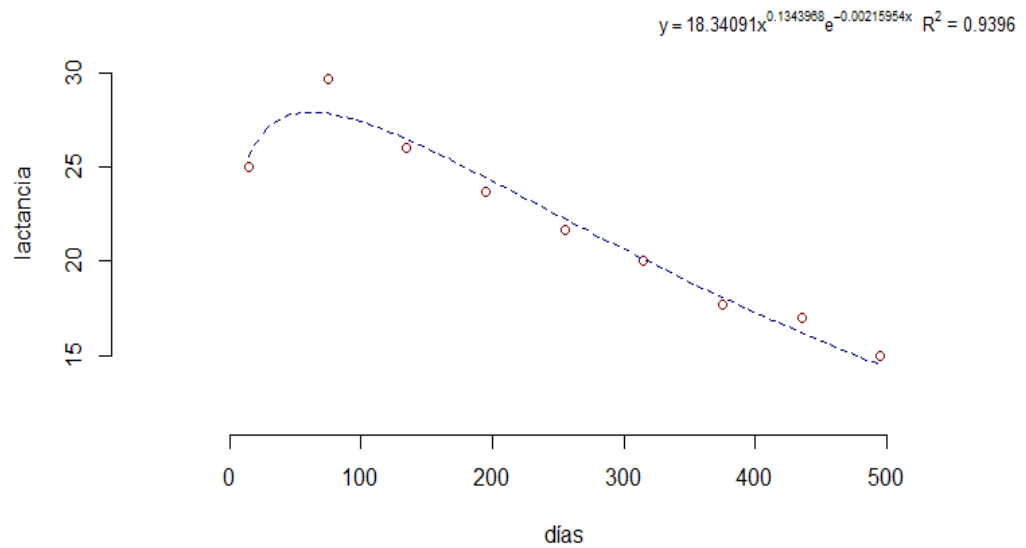
Figura 10 Modelo de Gompertz



Nota: MRLN a un modelo para determinar el peso del ganado vacuno.

11. $y = (\alpha x^\beta)e^{(-\delta x)} + \epsilon$, curva de lactancia (*MRNL*).

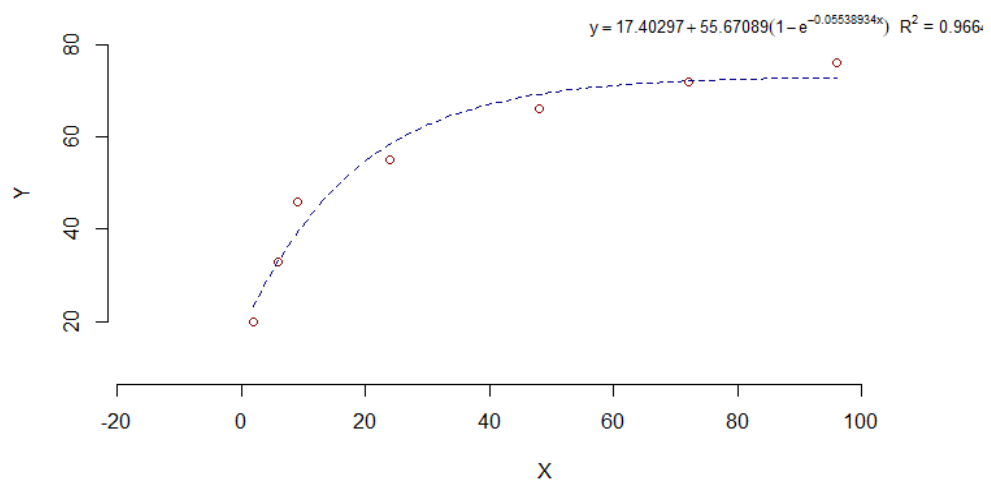
Figura 11. Modelo de lactancia



Nota: MRLN a un modelo de lactancia

12. $y = \alpha + \beta(1 - e^{(-\delta x)}) + \epsilon$, curva de degradación ruminal (*MRNL*).

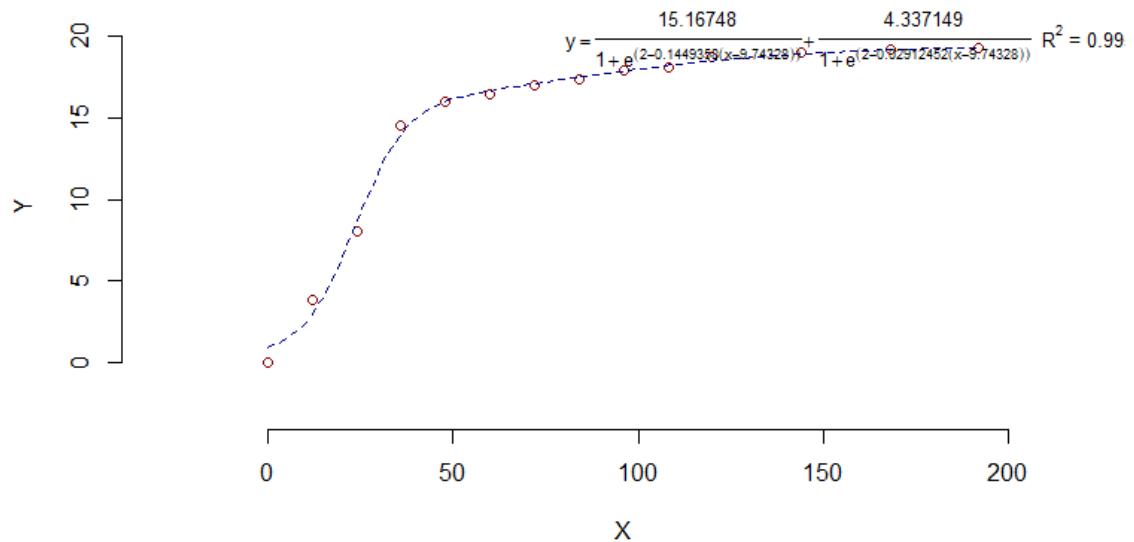
Figura 12. Modelo de degradación ruminal



Nota: MRLN a un modelo para determinar la degradación ruminal.

13. $y = \frac{\alpha}{1+e^{(2-4\delta)(x-e)}} + \frac{\beta}{1+e^{(2-4\gamma)(x-e)}} + \epsilon$, modelo logístico bi-compartimental (*MRNL*).

Figura 13. Modelo bi-logístico o logístico bi-compartimental



Nota: MRLN a un modelo para determinar la producción de gas.

En suma para el modelo de salarios, el mejor ajuste o especificación es de tipo curvilíneo (*MRL*): a mayor edad, existen mayores salarios; sin embargo, después de cierta edad los salarios comienzan a disminuir; para el modelo de peso de ganado vacuno es de tipo no lineal (*MRNL*) de tipo logístico o de Gompertz de forma alternativa con presencia de valores asintóticos y para el modelo de lactancia tiene un buen ajuste a sus datos con un comportamiento decreciente según la edad (*MRNL*); el modelo de degradación es no lineal, al igual que el modelo de producción de gas respectivamente (*MRNL*).

3. APLICACIÓN DEL MODELO DE GOMPERTZ, REGRESIÓN NO LINEAL AL COVID- 19

A partir de datos obtenidos de *Our World in data*, se obtuvo información diaria del número de casos de infectados por COVID-19 en Bolivia, del 11/03/2020 al 19/01/2022, donde se desea estimar el siguiente modelo de Gompertz (*model = 10*, paquetería *easynls*):

$$\hat{y}_t = ae^{(-be^{(-ct-t_0)})} + \hat{\epsilon}_t \quad (5)$$

En (5), y_t corresponde al número total de casos infectados por COVID-19; a es el valor máximo de predicción de infectados, corresponde a una asíntota; b es un coeficiente de ajuste exponencial; c es la tasa de declinación de la curva; t es el número de días desde que ocurrió el primer caso y t_0 es cuando ocurrió el primer caso. De forma adicional a la estimación de los parámetros, se desea realizar un pronóstico al 30/06/2022. La función *nlsfit* permite obtener la estimación respectiva con definición de los parámetros iniciales a , b y c (*script 4*).

```
#### Script 4#####
#### Modelo de Gompertz aplicado al COVID-19#####

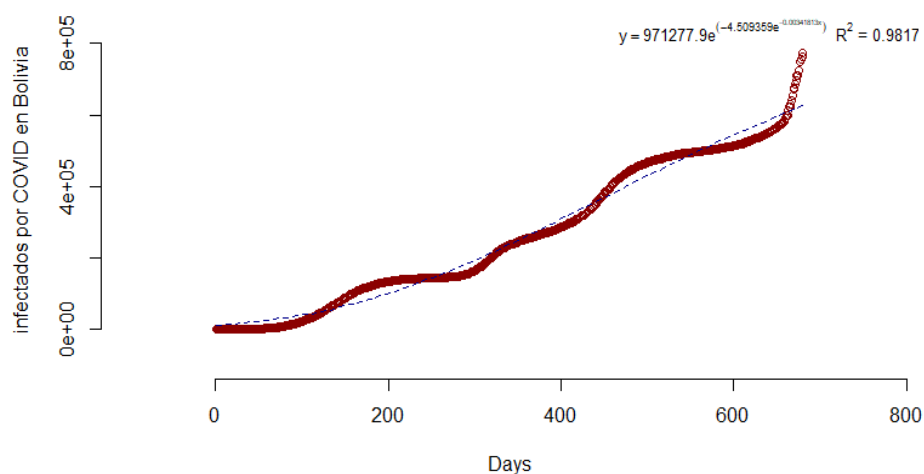
d<-cbind(day, casos)
d<-as.data.frame(d)
d
model = nlsfit(d, model = 10, start = c(a = 775000, b = 2, c = 0.05))
model
nlsplot(d, model = 10, start = c(a = 775000, b = 2, c = 0.05),
        xlab = "Days" , ylab = "infectados por COVID en Bolivia",
        position = 7)
##### Ejemplo, no ejecute#####
```

En tal sentido, se tendrían los siguientes resultados:

Coeficiente “ α ” 971 277.9***
 Coeficiente “ β ” 4.5093***
 Coeficiente “ δ ” 0.0034***
 Nivel de significancia estadística: *** al 1%
 $R^2 = 0.9817$

Según las estimaciones realizadas, el valor asintótico como cantidad máxima de infectados por COVID-19 llegaría a 971 mil personas aproximadamente, con un ajuste exponencial de 4.51 y una tasa de declinación en la curva de 0.34%; todos los parámetros estadísticamente significativos al nivel del 0.01.

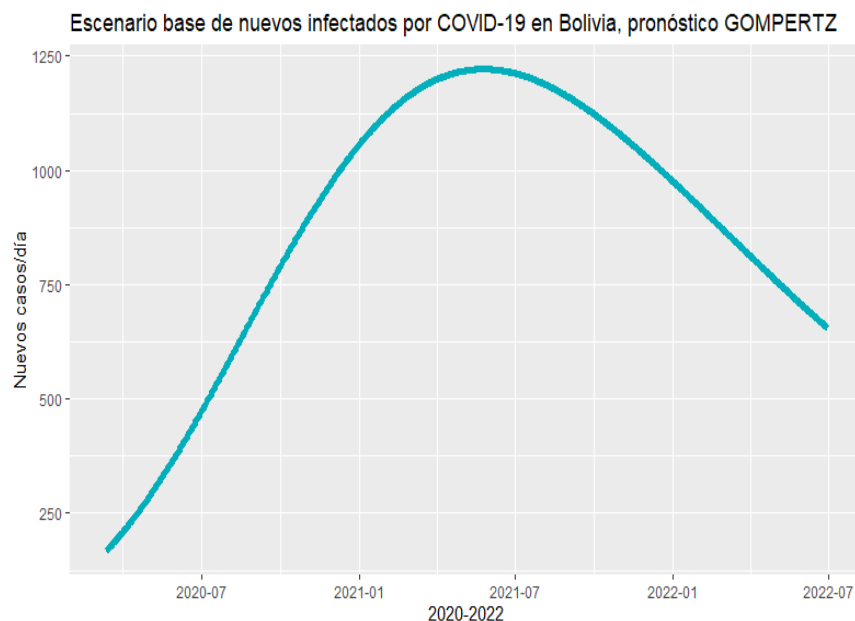
Figura 14. Modelo de Gompertz aplicado al COVID-19 en Bolivia



Fuente: estimación propia con datos hasta el 19.01.2022.

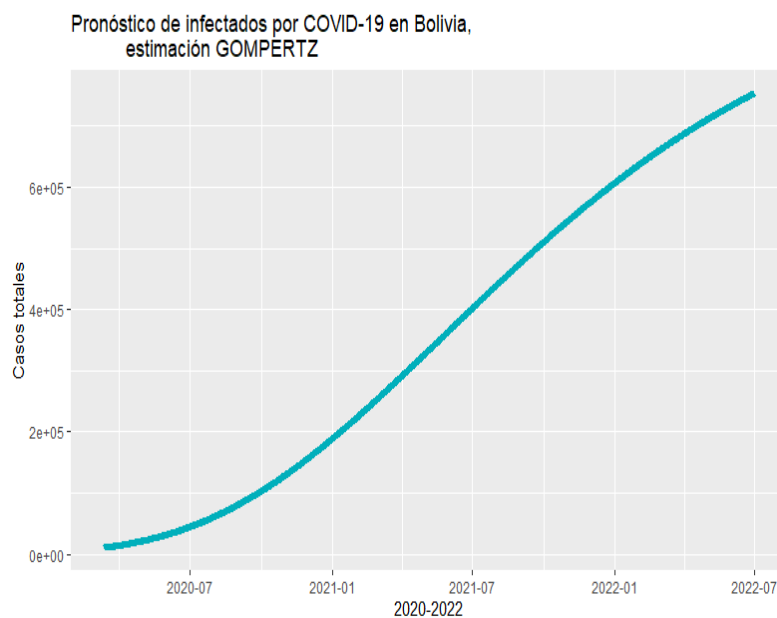
Sí bien el ajuste de la regresión no lineal a los datos es alta ($R^2 = 0.98$); los datos observados al final de la muestra mostraron un rebrote por encima de lo sugerido por el modelo (descontrol de la pandemia) (figura 14).

Figura 15. Pronóstico de nuevos casos COVID-19 al 30.06.2022



Fuente: estimación propia con datos hasta el 19.01.2022.

Figura 16. Pronóstico de casos acumulados por COVID-19 al 30.06.2022



Fuente: estimación propia con datos hasta el 19.01.2022.

Según un pronóstico de la pandemia al mes de junio/2022, se esperaría un comportamiento con tendencia a la disminución para los nuevos casos, mismos que al final del período de predicción no deberían superar los 750 casos diarios y no deberían superar los 800 mil casos acumulados (control de la pandemia) (*figuras 15 y 16* respectivamente); en caso contrario, se evidenciaría un descontrol epidemiológico. En suma el modelo de Gompertz tiene a bien capturar todas las fases de la epidemia: aparición, crecimiento exponencial, pico o cima, decrecimiento y aplanamiento o disminución de nuevos casos con descenso significativo.

ENSAYO 4: ‘DESCOMPOSICIÓN DE SERIES DE TIEMPO Y CICLOS ECONÓMICOS’

‘TIME SERIES DECOMPOSITION AND BUSINESS CYCLE’

Roger Alejandro Banegas Rivero^σ
UNIVERSIDAD AUTÓNOMA GABRIEL RENÉ MORENO

Este diseño, 01 de junio

INTRODUCCIÓN

En macroeconomía aplicada usualmente se considera al crecimiento potencial de la economía o también llamada crecimiento natural como la variación porcentual en el componente tendencial o de largo plazo (τ_t): variable latente no observada directamente en el Producto Interno Bruto real (PIB) de la economía.

Una definición más formal del producto potencial se vincula con aquel nivel de producción que puede alcanzar un país en plena utilización de sus factores productivos. De igual manera, se considera al ciclo económico, también llamado fluctuaciones cíclicas o componente cíclico (ς_t) como a la diferencia entre el producto efectivo y el producto potencial. Para ello, cobra especial interés la aplicación en la descomposición de series de tiempo, misma que se encuentra caracterizada por cuatro componentes respectivos:

τ_t = Componente tendencial o componente de largo plazo.

ς_t =Componente cíclico.

s_t =Componente estacional.

^σ Profesor-investigador a tiempo completo, Economía-UAGRM. Correo electrónico: rogerbanegas@uagrm.edu.bo

ε_t =Componente irregular-estocástico

Para poder extraer los componentes, existen tres tipos de descomposición de series de tiempo:

d) Descomposición aditiva

$$y_t = \tau_t + \varsigma_t + s_t + \varepsilon_t \quad (1)$$

La serie puede presentar negatividad o constituirse como no negativa.

Ej. La inflación (> 0 ; < 0 *deflación en caso contrario*).

e) Descomposición multiplicativa

$$y_t = \tau_t * \varsigma_t * s_t * \varepsilon_t \quad (2)$$

La serie y_t es no negativa o en otras palabras es positiva $y_t > 0$

f) Descomposición mixta

Combina ambas descomposiciones: aditiva y multiplicativa

$$y_t = \tau_t * (1 + \varsigma_t)(1 + s_t) + \varepsilon_t \quad (3)$$

1. MEDICIÓN DE LAS FLUCTUACIONES CÍCLICAS

Siguiendo a [Banegas, Núñez & Valdez del Río \(2020\)](#) en la medición del ciclo económico o fluctuaciones cíclicas, se emplean los filtros de series de tiempo, donde los más utilizados son: [Christiano & Fitzgerald, 2003](#); [Baxter & King, 1999](#); [Hodrick & Prescott, 1980](#) con el propósito de evaluar las consistencias en las fluctuaciones económicas y evitar ciclos espurios ([Kydland & Prescott, 1990](#)). En consecuencia, se procederá con la descomposición de cada serie original (y_t) – Producto real, producto real – en dos componentes: fluctuaciones cíclicas (ς_t) y tendencia de largo plazo (τ_t):

$$y_t = \varsigma_t + \tau_t \quad (4)$$

1.1. Filtro de Hodrick-Prescott (*HP*)

Uno de los filtros pioneros de las series de tiempo es el de [Hodrick & Prescott \(1980\)](#), presentado en la versión original [*HP*]:

$$\min_{\{\tau_t\}} \sum_{t=1}^T (y_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \quad (5)$$

$$\varsigma_{y_t} = brecha \ y_t = ly_t - ly_t^* \quad (6)$$

Donde ly_t corresponde al PIB real efectivo (observado) y ly_t^* representa al PIB real potencial (mediante filtros de series de tiempo; Donde λ es un parámetro de suavización entre la serie original (y_t) y el componente de largo plazo (τ_t). Mientras más grande sea λ , las fluctuaciones cíclicas serán menores. Otra crítica adicional de λ es que ha sido calibrado a partir de datos de la economía estadounidense:

Cuadro 1. Parámetro de penalización o suavización, filtro HP

	Mensual	Trimestral	Semestral	Anual
λ	14.400	1.600	400	100

Fuente: a partir de Hodrick-Prescott (1980)

El filtro *HP* es uno de los métodos más ampliamente usados en el ámbito académico, estudios de políticas, análisis del sector privado y sectores económicos; de forma posterior se verán sus limitaciones e inconvenientes ([Hamilton, 2018](#)).

1.2. Filtro de Christiano-Fitzgerald (CF)

El filtro de [Christiano & Fitzgerald \(2003\)](#) es un método de optimización lineal, el cual consiste en seleccionar ponderaciones de filtro para que τ_t se aproxime de mejor manera a la serie de interés, y_t , de tal forma que se minimice el error esperado cuadrático:

$$E[(y_t - \tau_t)^2 | x], x \equiv [x_1, x_2, \dots, x_T] \quad (7)$$

Donde τ_t es la tendencia lineal de y_t sobre cada elemento de la base de datos, x_t , donde el problema es el tiempo de proyección t , por lo cual, la solución se centra en las ponderaciones del filtro. El núcleo de este filtro consiste en que el ciclo económico (ς_t) contiene un componente de alta frecuencia. En la práctica, este filtro contiene dos versiones simétrica y asimétrica: en la primera se eliminan observaciones al inicio y al final de la muestra para evitar el problema del punto inicial y final (paso de banda)²; la versión asimétrica no implica la eliminación de observaciones.

1.3. Filtro de Bakker-King (BK)

El filtro de [Baxter & King \(1999\)](#) también es un filtro de paso de banda que permite capturar las fluctuaciones cíclicas (ς_t) de una serie de tiempo (y_t), en sentido estacionario, así como su componente de tendencia (τ_t). Los componentes cíclicos se construyen a través de medias móviles en una banda específica de frecuencia:

$$\varsigma_t = a(L)y_t \quad (8)$$

$$a(L)^s_{-r} = a_{-r}L^{-r} + \dots + a_0 + a_1L + a_sL^s \quad (9)$$

² Se utilizan períodos adelantados y tres períodos atrás para los filtros de paso de banda. De igual forma, se suele incorporar un escenario esperado para el crecimiento del PIB real en j períodos hacia adelante.

$$\tau_t = [1 - a(L)]y_t \quad (10)$$

Donde $a(L)$ es un operador polinomial en rezago L , el cual especifica el tamaño del rezago $[-r, s]$ del modelo, que de forma usual son simétricos y a dos bandas; en consecuencia el tamaño de los valores futuros y pasados determina el valor de la tendencia (τ_t).

1.4. Filtro de Nadaraya-Watson (NW)

La versión no paramétrica del filtro de [Nadaraya \(1964\)](#) y [Watson \(1964\)](#), también conocido por su abreviatura (NW). está en función de Kernel con una media local (m) y un núcleo (K) de ponderación para los componentes de ciclo y tendencia en una serie de tiempo desestacionalizada y sin el componente irregular de la siguiente forma:

$$m_h(x) = \frac{\sum_{i=1}^n K_h(x-X_i)Y_i}{\sum_{i=1}^n K_h(x-X_i)} \quad (11)$$

En la expresión (11), K representa el núcleo; h simboliza el ancho de bando del núcleo. Para ello, se tiene una función de distribución $f(x, y)$ y $f(x)$ del núcleo, tal que:

$$E\left(\frac{y}{x}\right) = \int y \frac{f(x, y)}{f(x)} dy \quad (12)$$

En (12), se encuentra el estimador propuesto por NW:

$$f(x) = n^{-1}h^{-1} \sum_{i=1}^n K\left(\frac{x-X_i}{h}\right) \quad (13)$$

1.5. El filtro de Hamilton (HA)

La propuesta de [Hamilton \(2018\)](#) se origina a través de una crítica a la práctica estándar del filtro *HP*, mismo que en el análisis realizado:

- i) Genera relaciones dinámicas espurias sin sustento en la generación intrínseca de datos.

- ii) Presenta el problema del punto final: los valores al final de la muestra son muy diferentes de los valores del medio;
- iii) El parámetro de suavización está muy alejado de la práctica común (e.g. 1600 para datos trimestrales).

En contraposición, una regresión de la variable en $t + h$ con base en los cuatro valores más recientes ofrece un enfoque robusto para destendenciar a diferencia del filtro *HP*. Para ello, el filtro de Hamilton, se fundamenta en las siguientes propiedades:

- a) Pronosticar cuando el verdadero proceso es desconocido.

$$y_{t+h} = y_t + w_t^{(h)}$$

Donde $w_t^{(h)}$ corresponde a los procesos estacionarios

- b) El pronóstico se puede realizar con base en los últimos cuatro valores recientes:

$$y_{t+h} = \beta_0 + \beta_1 y_t + \beta_2 y_{t-1} + \beta_3 y_{t-2} + \beta_4 y_{t-3} + v_{t+h}$$

Un punto de decisión importante es el horizonte de proyección de h y el tamaño del rezago p . Por ello, con una muestra T grande, se puede requerir un período h grande. Si estamos interesados en ciclos económicos, el horizonte estándar es de dos años. Por tanto para datos trimestrales es recomendable $p = 4$ y $h = 8$.

2. PAQUETERÍAS Y LIBRERÍAS EN *RStudio*

Antes de explicar la aplicación respectiva, para utilizar los filtros de series de tiempo en *RStudio*, se requiere instalar previamente las siguientes paqueterías y librerías de forma respectiva (*Script 1*):

```
#### Script 1#####  
#### Paqueterías y librerías en RStudio#####  
library(fpp)// Ej. Para descomposición aditiva de series de tiempo  
library(Ecdat)// Ej. Para descomposición multiplicativa  
library(seasonal)// Para Desestacionalización ARIMA Census-X11  
library(mFilter)// Para filtros de series de tiempo  
library(neverhpfiler)// Para el filtro de Hamilton
```

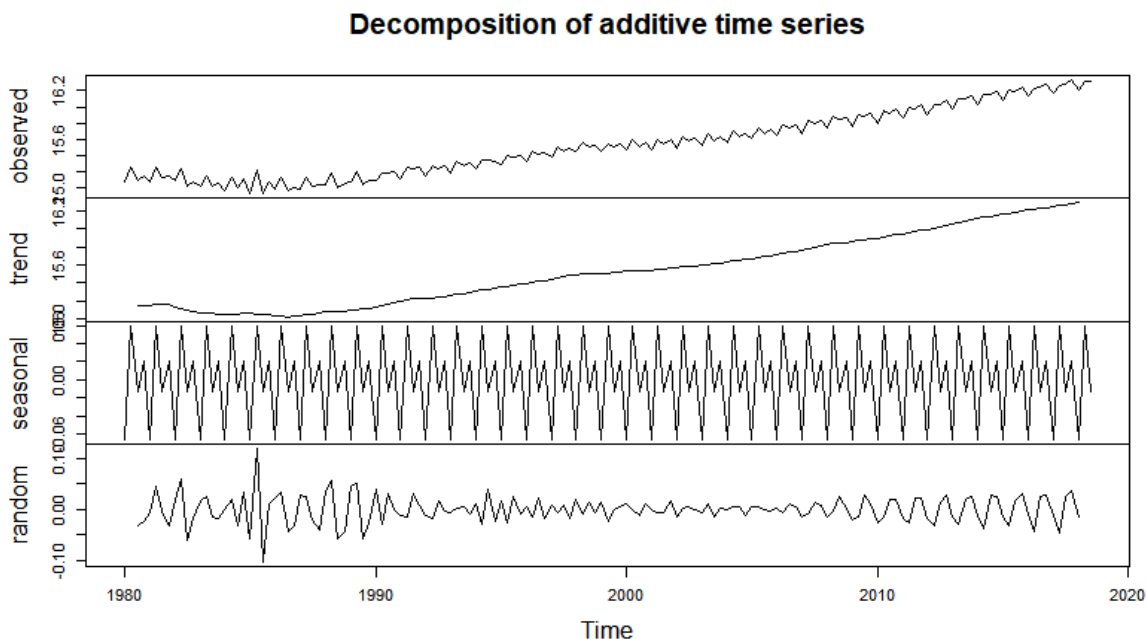
3. UNA APLICACIÓN DE DESCOMPOSICIÓN Y FILTROS DE SERIES DE TIEMPOS CON RSTUDIO

A partir de datos del Instituto Nacional de Estadística (INE) y UDAPE se construyó una serie encadenada para el *PIB real* de Bolivia, con información trimestral de 1980.T1 al 2018.T3, con el propósito de descomponer la serie en componente tendencial-cíclico, estacional e irregular mediante el método de descomposición aditiva, multiplicativa y *ARIMA Census X – 11*.

De igual manera el propósito central es obtener las fluctuaciones cíclicas (brecha del producto) con base en los filtros: Hodrick-Prescott, Christiano-Fitzgerald (versión simétrica y asimétrica), Baxter-King, Nadaraya-Watson y Hamilton. Finalmente, se desea obtener la tasa de crecimiento natural o crecimiento potencial de la economía boliviana con base el período reciente 2006.T1-2016T3 para la muestra común obtenida después de eliminar observaciones del punto final (ocho trimestres).

En las *figuras 1 y 2*, se presenta la descomposición de la serie *Log del PIB real de Bolivia*, en tres componentes: tendencia, estacional y aleatoria según el método aditivo y multiplicativo respectivamente. En la *figura 3*, se realiza el ajuste estacional o Desestacionalización de la serie mediante *ARIMA Census X – 11*, eliminándose el componente estacional errático.

Figura 1. Descomposición aditiva del PIB real de Bolivia



Fuente: Elaboración propia.

En tal sentido, a través de los *Scripts 2,3 y 4* respectivamente, se realiza la descomposición de la serie de tiempo *lpibr* en mecanismo aditivo, multiplicativo y por ARIMA Census X-11. Donde corresponde `adjust_lpibr`, `adjust_lpibr2` y `lpibr_sa` corresponden a la serie desestacionalizada en cada código correspondiente.

```

#### Script 2#####
#### Descomposición aditiva RStudio#####
library(fpp)
decompose_lpibr = decompose(lpibr, "additive")
adjust_lpibr = lpibr - decompose_lpibr$seasonal
plot(adjust_lpibr)

adjust_lpibr #serie desestacionalizada
plot(decompose_lpibr)

s<-decompose_lpibr$seasonal
trend_cycle<-decompose_lpibr$trend
e<-decompose_lpibr$random

plot(s, main="Componente estacional del PIB real")
plot(trend_cycle, main = "Componente cíclico y tendencial")
plot(e, main="componente irregular")
##### Ejemplo, no ejecute#####

#### Script 3#####
#### Descomposición multiplicativa RStudio#####
library(Ecdat)

decompose_lpibr2 = decompose(lpibr, "multiplicative")
adjust_lpibr2 <- lpibr / decompose_lpibr2$seasonal
plot(adjust_lpibr2) #serie desestacionalizada

s2<-decompose_lpibr2$seasonal
trend_cycle2<-decompose_lpibr2$trend
irreg2<-decompose_lpibr2$random
##### Ejemplo, no ejecute#####

#### Script 4#####
#### Desestacionalización con ARIMA CENSUS-X11 en RStudio###

library(seasonal)

# seasonal adjust time series using X11.
lpibr<-log(lpibr)

lpibr_sa <- seas(lpibr,x11 = "")
lpibr_sa

view(lpibr_sa)
plot(lpibr_sa)

# plot original and seasonally adjusted series.
plot(lpibr, main= "PIB real de Bolivia",
      ylab= "En escala logarítmica",

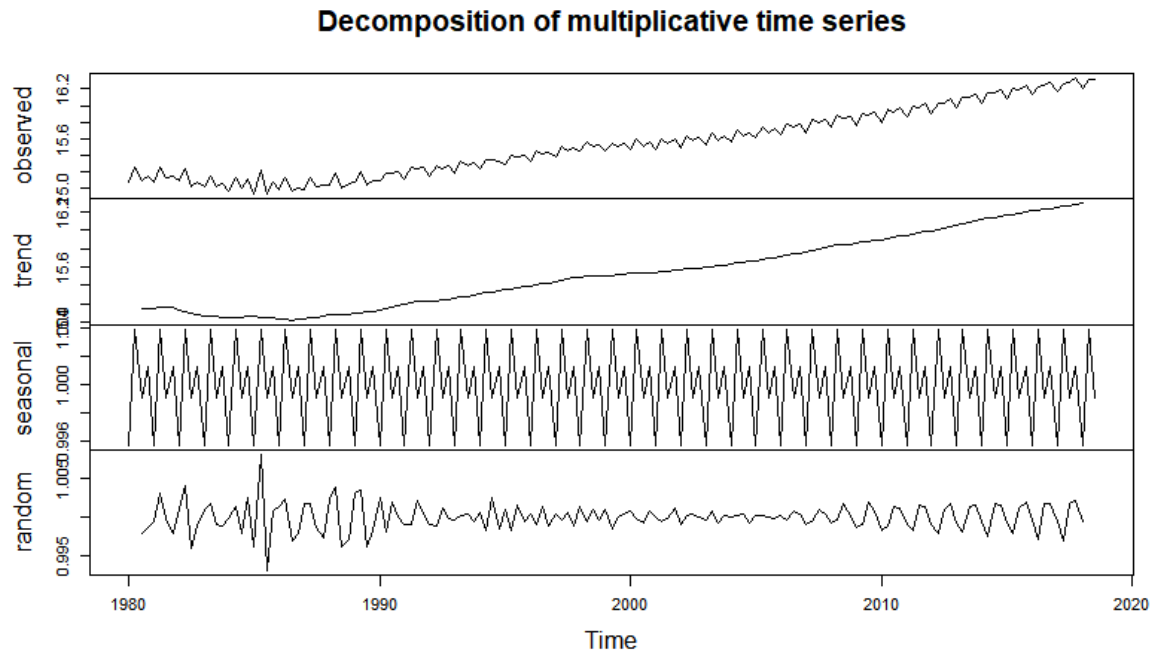
```

```

sub="Método ARIMA CENSUS x-11",
lwd=3)
lines(final(lpibr_sa),col="red", lwd=3, lty=1)
legend("topleft", c("Rojo= serie desestacionalizada",
                    "negro=serie original"), lwd=3)
##### Ejemplo, no ejecute#####

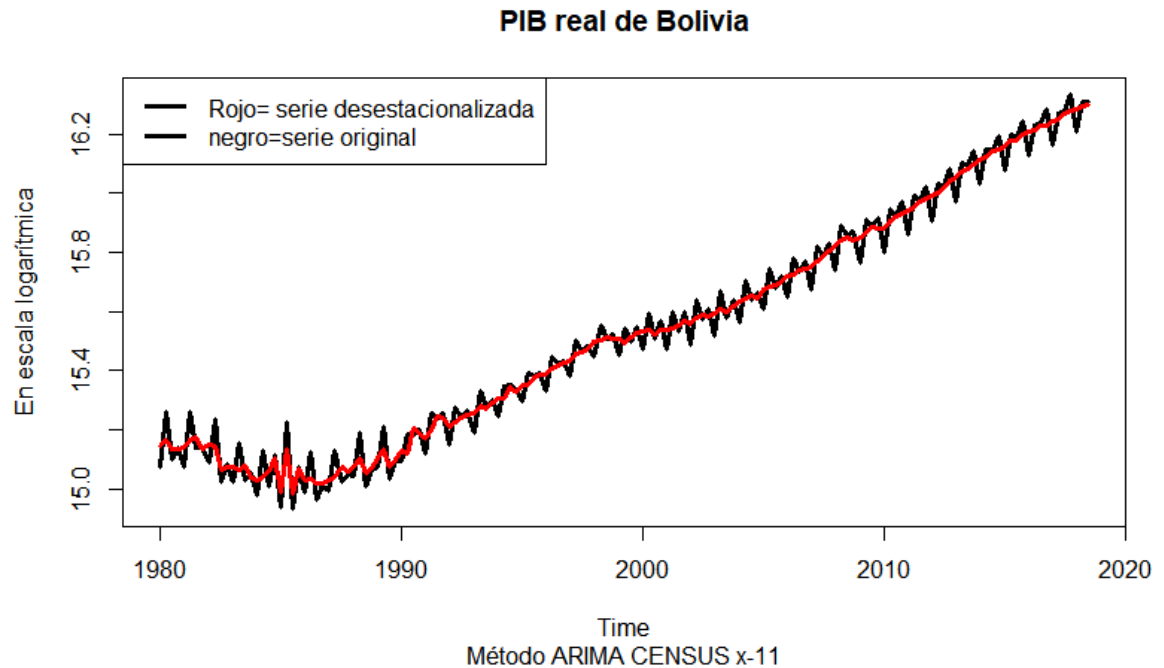
```

Figura 2. Descomposición multiplicativa del PIB real de Bolivia



Fuente: Elaboración propia.

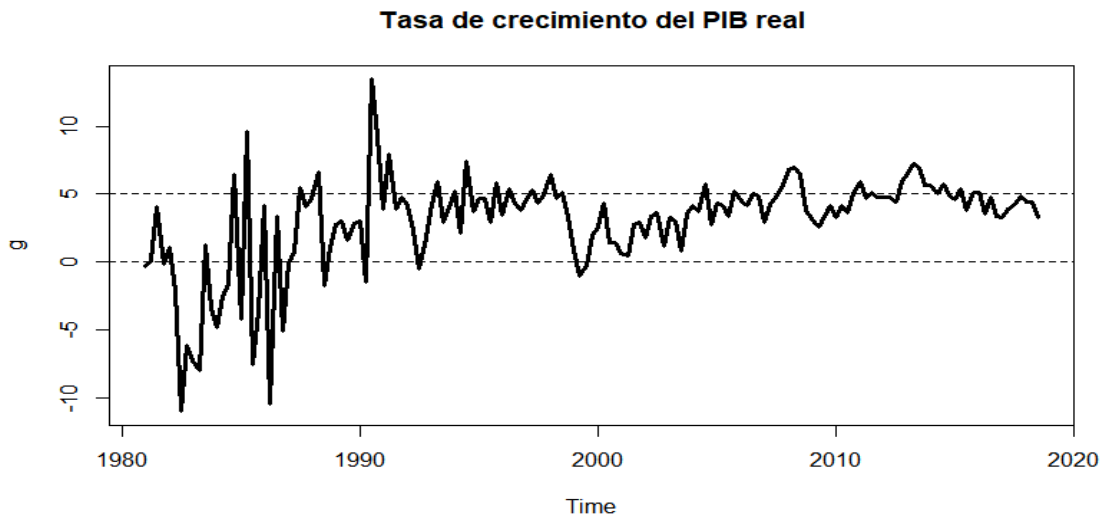
Figura 3. Desestacionalización del PIB real, método ARIMA Census-X11



Fuente: Elaboración propia.

Toda vez que se compararán fases del ciclo económico, es necesario abordar la historia económica en la trayectoria del crecimiento real de la economía (*figura 4*), a través del comportamiento de su tasa de crecimiento a 12 meses, el cual refleja distintos períodos: crisis de los 80's; estabilización económica (85- 2000) y mayor expansión del crecimiento (después del año 2000). De forma posterior, las fluctuaciones cíclicas deberán capturar estas fases señaladas previamente.

Figura 4. Crecimiento real de la economía boliviana



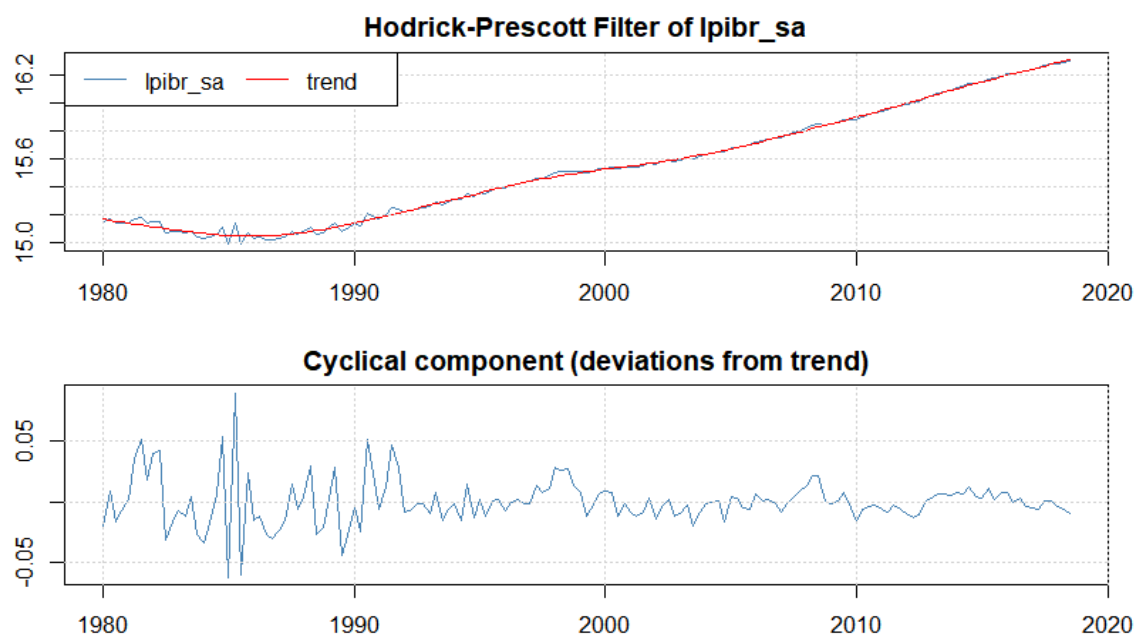
Fuente: Elaboración propia.

A manera de ejemplo el primer filtro en emplearse es del Hodrick-Prescott. El cual descompone el PIB real desestacionalizado (*lpibr_sa*) en componente tendencial o de largo plazo (*trend*) y el componente cíclico, medido como desviación porcentual de la tendencia, también llamadas brechas del producto. Se observa desviación negativa del -5% para mediados de los 80s, algunos periodos de volatilidad en las fluctuaciones a principios de los 90s y después en general la economía evidencia una tendencia de 0% de brecha del producto (*figura 5*), misma que es obtenida a través del *Script 5*:

```
#### Script 5#####
#### Filtro Hodrick-Prescott en RStudio#####
lpib_ra.hp <- hpfilter(lpibr_sa)
lpib_ra.hp
plot(lpib_ra.hp)
plot(lpib_ra.hp$cycle*100,
     main="Filtro HP, PIB Bolivia",col=1,ylab="Desv. %")

# Tasa de crecimiento potencial, filtro HP
GDP_pot_HP<-diff(lpib_ra.hp$trend,lag=4, differences = 1)*100
##### Ejemplo, no ejecute#####
```

Figura 5. Componente de largo plazo y cíclico de la economía Boliviana, 1980-2018. Filtro Hodrick-Prescott



Fuente: Elaboración propia.

En el mismo sentido, se realiza la descripción del *Script 6* para el filtro Baxter-King, estableciendo ocho rezagos y adelantes, estableciendo la serie tiene una raíz unitaria y la amplitud de las fluctuaciones cíclicas entre 6 y 32 trimestres respectivamente, así como el cálculo de la tasa de crecimiento potencial con variación a 12 meses:

```
#### Script 6#####
#### Filtro Baxter-King en RStudio#####

GDP_bk<-mFilter(lpibr_sa,filter = "BK", nfix=8) #recomendable 1
GDP_bk2<-mFilter(lpibr_sa,filter = "BK", nfix=8, root=TRUE)
#recomendable 2
GDP_bk3<-mFilter(lpibr_sa,filter = "BK", nfix=12, root=TRUE, pl=6,
pu=32)

plot(GDP_bk$cycle, col="blue",
     lwd=3,
     main="Filtro BK, PIBR Bolivia",
     ylab="Desviaciones % de la tendencia")
abline(h=0)
```

```
#Tasa de crecimiento potencial

GDP_bk_pot<-diff(GDP_bk$trend, lag=4, differences = 1)*100

library(psych)
describe(GDP_bk_pot)
##### Ejemplo, no ejecute#####
```

De la misma manera, se efectúa el filtro Christiano-Fitzgerald en su versión simétrica y asimétrica para la variable `lpibr_sa`, así como el cálculo de la tasa de crecimiento potencial con variación a 12 meses (*Script 7*):

```
#### Script 7#####
#### Filtro Christiano-Fitzgerald en RStudio#####

# Simétrico
GDP.cf.sym <- cffilter(lpibr_sa, type = "symmetric", nfix=8,
root=TRUE)

# Asimétrico
GDP.cf <- mFilter(lpibr_sa,filter="CF", root=TRUE)

# Tasa de crecimiento potencial, filtro CF, simétrico y asimétrico
GDP_pot_CFsym<-diff(GDP.cf.sym$trend, lag=4, differences = 1)*100
GDP_pot_CF<- diff(GDP.cf$trend, lag=4, differences = 1)*100
##### Ejemplo, no ejecute#####
```

Los *Script 8 y 9* corresponden a los filtros de *Nadaraya – Watson (NW)* y *Hamilton (HA)* respectivamente. En *NW*, se utiliza una función de distribución de Kernel para la variable `lpibr_sa`, En *HA*, se utiliza el criterio de pronóstico en ocho trimestres hacia adelante ($h = 8$) a partir de cuatro rezagos ($p = 4$) respectivamente y su descomposición de la serie de tiempo `lpibr_sa` en componente tendencial y cíclico:


```
#### Script 8#####
#### Filtro Nadaraya-Watson en RStudio#####
```

```
plot(lpibr_sa)
k <- kernel("daniell", 1) # for lags
plot(k)
k
w <- kernapply(lpibr_sa, k)
plot(w)
lines(lpibr_sa, col="red")
df.kernel(k)
bandwidth.kernel(k)
is.tskernel(k)
plot(lpibr_sa-w)

GDP_NW.cycle<-lpibr_sa-w
plot(GDP_NW.cycle)

#Tasa de crecimiento potencial NW
GDP_pot_NW<-diff(w,lag=4,differences = 1)*100

plot(GDP_pot_NW,
      main="Crec. Potencial Bolivia, NW",
      ylab="En %")
##### Ejemplo, no ejecute#####
```

```
#### Script 9#####
#### Filtro Hamilton en RStudio#####
```

```
library(psych)
library(neverhpfilter)
library(xts)

t<-seq(as.Date("1980/3/1"), as.Date("2018/9/1"), by = "quarter")
t
describe(t)

xts1<-xts(x=lpibr_sa, order.by=t)
xts1

GDP_hamil<-xts::to.quarterly(xts1["1980/"], OHLC = FALSE)
GDP_hamil
GDP_hamil4 <- xts(x = GDP_hamil)
GDP_hamil4

GDP_cycle_hamilton <- yth_filter(GDP_hamil4, h = 8, p = 4, output =
c("cycle", "random"),
                                family = gaussian)
main <- "Log del ciclo del GDP y la caminata aleatoria"
plot(GDP_cycle_hamilton, grid.col = "white", legend.loc =
"topright", main = main)

GDP_cycle_hamilton
```

```

GDP_trend_hamilton <- yth_filter(GDP_hamil4, h = 8, p = 4, output =
c("x", "trend"),
                                family = gaussian)

GDP_trend_hamilton
plot(GDP_trend_hamilton$x.trend, grid.col = "white", legend.loc =
"topleft",
     main = "Descomposición del lpib, Filtro de Hamilton, en
millones de USD")

GDP_hamilton_cycle_stochas<-GDP_trend_hamilton$x.cycle
# Tasa de crecimiento potencial
GDP_pot_HA<-diff(GDP_trend_hamilton$x.trend, lag=4, differences =
1)*100
GDP_pot_HA<-ts(GDP_pot_HA, end=c(2018,3), freq=4)

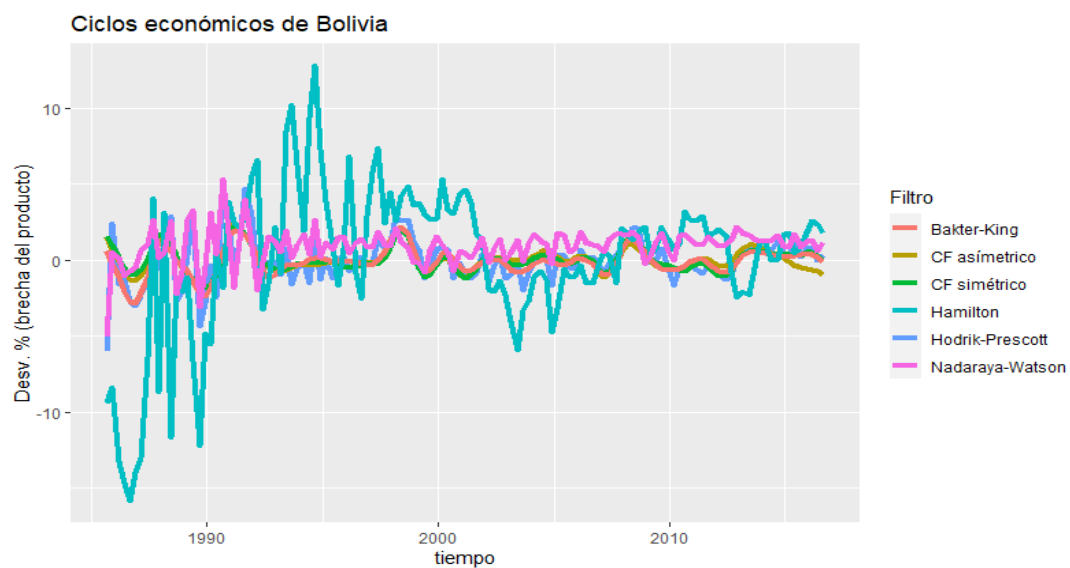
GDP_hamilton_cycle_stochas<-ts(GDP_hamilton_cycle_stochas,
start=c(1982, 4), freq=4)

plot(GDP_hamilton_cycle_stochas)

plot(GDP_pot_HA,
     main="Crec. Potencial PIB Bolivia, Filtro de Hamilton")
##### Ejemplo, no ejecute#####

```

Figura 6. Medición alternativa de las fluctuaciones cíclicas en Bolivia, 1980-2018. Diversos filtros



Fuente: Elaboración propia.

Una de las estrategias más utilizadas para comparar los filtros de series de tiempo, consiste en el análisis de correlaciones (ρ) para poder clasificar las magnitudes de las asociaciones fuertes ($[\rho] > 0.70$), moderadas ($[\rho] \geq 0.50$); o asociaciones débiles ($[\rho] < 0.50$) ([Mejía, 2013](#)).

De igual manera, en la *figura 6* se incorporan otros filtros de series de tiempo para evaluar la consistencia en el comportamiento cíclico: similitud y diferencias entre los mismos. El filtro de Hamilton exhibe mayores fluctuaciones cíclicas en comparación con los demás. Para ello, se empleó el *Script 10*, a partir de la base de datos, llamada ‘base_m’ con el propósito de evaluar las correlaciones respectivas; de la misma manera, se empleó la base de datos ‘base_potf’ para comparar y correlacionar las tasas de crecimiento potencial. En ambos caso, se empleó una muestra común con la misma cantidad de observaciones para todos los filtros.

```
#### Script 10#####
#### Comparamos los filtros en RStudio#####

#Componente cíclico
base<-ts.intersect(lpib_ra.hp$cycle,
                  GDP.cf$cycle,
                  GDP.cf.sym$cycle,
                  GDP.bk$cycle,
                  GDP_hamilton_cycle_stochas,
                  GDP_NW.cycle )

base<-window(base, start=c(1985,3), end=c(2016,3))

cor(base)
describe(base)
t0<-seq(as.Date("1985/9/1"), as.Date("2016/9/1"), by = "quarter")
length(t0)

base_m<-as.data.frame(base)
base_m
```

```

a<-ggplot(base_m, aes(t0)) +
  geom_line(aes(y = base_m$`lpib_ra.hp$cycle`*100, colour =
"Hodrik-Prescott"), size = 1.5) +
  geom_line(aes(y = base_m$`GDP.cf$cycle`*100, colour = "CF
asimétrico"), size = 1.5) +
  geom_line(aes(y = base_m$`GDP.cf.sym$cycle`*100, colour = "CF
simétrico"), size = 1.5) +
  geom_line(aes(y = base_m$`GDP.bk$cycle`*100, colour = "Bakter-
King"), size = 1.5) +
  geom_line(aes(y = base_m$GDP_hamilton_cycle_stochas*100, colour =
"Hamilton"), size = 1.5) +
  geom_line(aes(y = base_m$GDP_NW.cycle*100, colour = "Nadaraya-
Watson"), size = 1.5) +
  scale_colour_hue("Filtro")+
  ggtitle("Ciclos económicos de Bolivia")+
  xlab("tiempo") + ylab("Desv. % (brecha del producto)")
a

```

```

library(PerformanceAnalytics)
chart.Correlation(base_m[1:6])

```

```

#Componente tendencial o potencial

```

```

base_pot<-ts.intersect(GDP_pot_BK,
                      GDP_pot_CF,
                      GDP_pot_CFsym,
                      GDP_pot_HP,
                      GDP_pot_NW,
                      GDP_pot_HA)

```

```

base_potf<-window(base_pot, start=c(2006,1), end=c(2016,3))
base_potf<-as.data.frame(base_potf)

```

```

describe(base_potf)
quantile(base_potf$GDP_pot_HP, 1:9/10)
quantile(base_potf$GDP_pot_BK, 1:9/10)
quantile(base_potf$GDP_pot_CF, 1:9/10)
quantile(base_potf$GDP_pot_CFsym, 1:9/10)
quantile(base_potf$GDP_pot_NW, 1:9/10)
quantile(base_potf$GDP_pot_HA, 1:9/10)

```

```

chart.Correlation(base_potf[1:6])

```

```

t1<-seq(as.Date("2006/1/1"), as.Date("2016/9/1"), by = "quarter")
length(t1)

```

```

b<-ggplot(base_potf, aes(t1)) +
  geom_line(aes(y = base_potf$GDP_pot_HP, colour = "Hodrik-
Prescott"), size = 1.5) +
  geom_line(aes(y = base_potf$GDP_pot_CF, colour = "CF
asimétrico"), size = 1.5) +

```

```

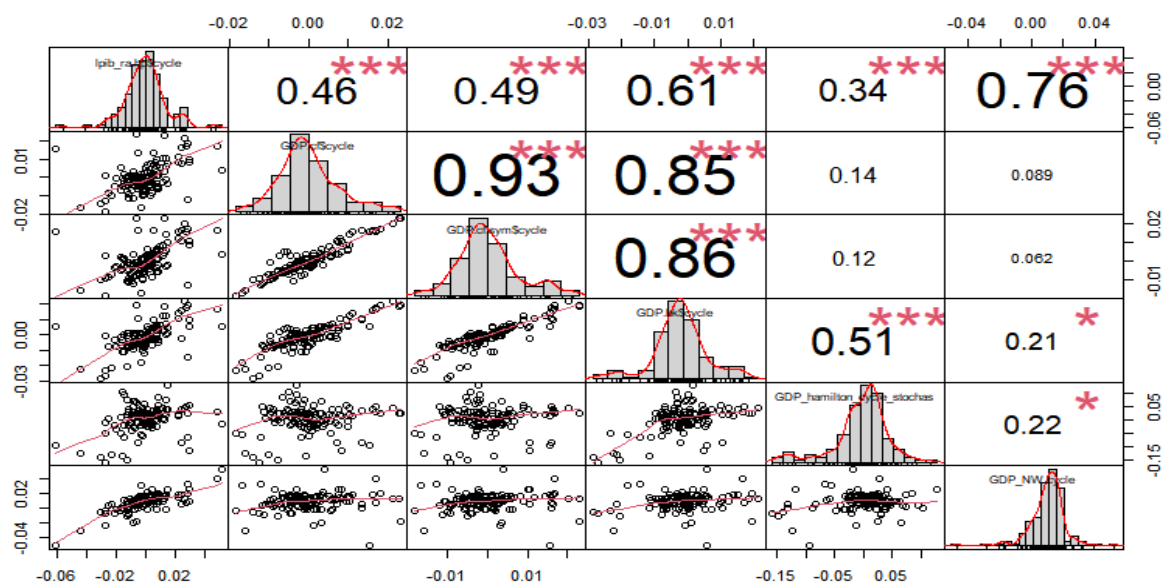
    geom_line(aes(y = base_potf$GDP_pot_CFsym, colour = "CF
simétrico"), size = 1.5) +
    geom_line(aes(y = base_potf$GDP_pot_BK, colour = "Bakter-King"),
size = 1.5) +
    geom_line(aes(y = base_potf$GDP_pot_HA, colour = "Hamilton"),
size = 1.5) +
    geom_line(aes(y = base_potf$GDP_pot_NW, colour = "Nadaraya-
Watson"), size = 1.5) +
    scale_colour_hue("Filtro")+
    ggtitle("Crecimiento potencial del PIB real en Bolivia")+
    xlab("tiempo") + ylab("En porcentaje")
b
##### Ejemplo, no ejecute#####

```

Acorde a la *figura 7*, se concluye que los filtros más sincronizados corresponden a los pasos de banda: *Christiano-Fitzgerald (versión asimétrica y simétrica)*, así como el filtro de *Baxter-King con grado de asociación fuerte* ($[\rho] > 0.70$). El filtro de *Hodrick-Prescott*, uno de los filtros más usados y criticados a la vez, tiene una sincronización estadísticamente significativa con todos los filtros con *grado entre moderado y fuerte*. Los filtros *Nadaraya-Watson* y *Hamilton* son los que tienen menor semejanza con los demás filtros señalados previamente con grado predominante entre *débil y moderado* (excepto la asociación entre *HP* y *NW*).

De igual manera, se procedió en estimar el crecimiento natural o potencial de la economía, según la *figura 8*, el filtro de *Hamilton* presenta la mayor variabilidad en la tasa de crecimiento sin asociación significativa con el resto de los filtros en componente tendencial, asimismo, la mayor asociación fuerte corresponde entre los filtros de *Baxter-King* y *Christiano-Fitzgerald* (simétrico), así como con el componente asimétrico. Los filtros de *Hodrick-Prescott* y *Nadaraya-Watson* presentan una asociación positiva y significada en grado débil-moderado (*figura 9*).

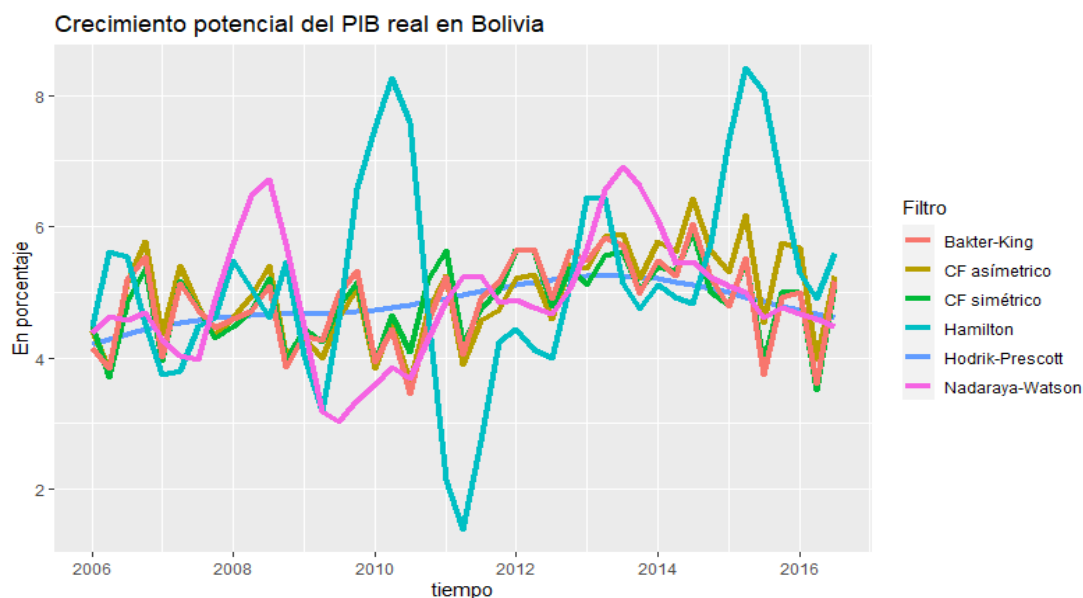
Figura 7 Correlaciones entre los componentes cíclicos de distintos
filtros de series de tiempo



Orden de los componentes cíclicos en la diagonal principal: HP.cycle =Hodrick-Prescott, CF.cycle=Christiano-Fitzgerlad, CFsym.cycle =Christiano-Fitzgerlad simétrico, BK.cycle = BakterKing, NW.cycle =Nadaraya-Watson, HA.cycle =Hamilton.

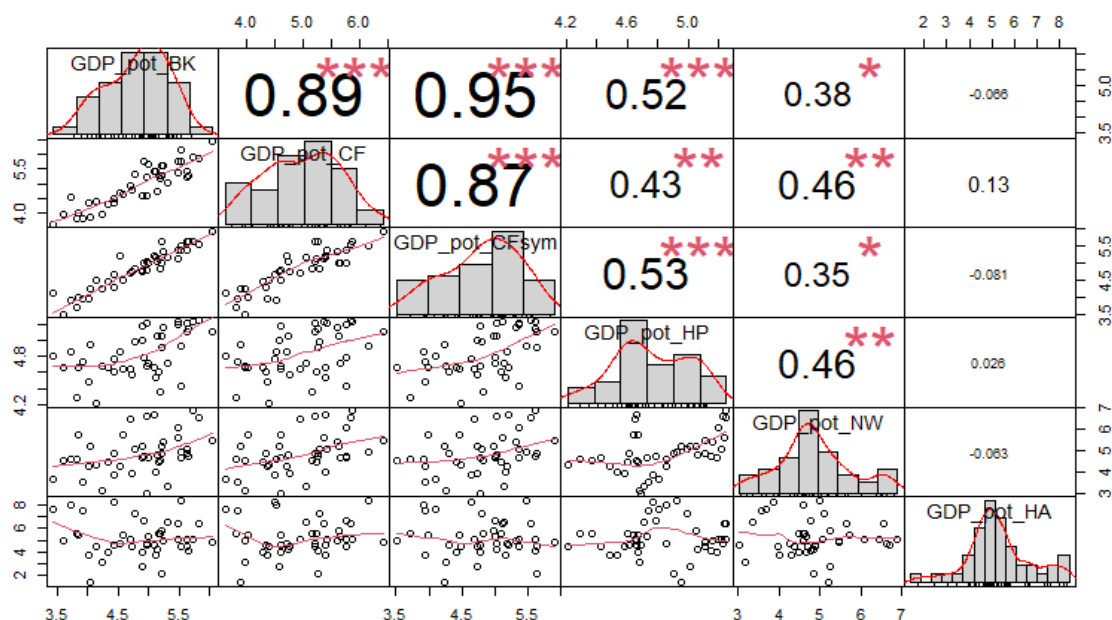
Fuente: Elaboración propia.

Figura 8. Crecimiento natural (potencial) de la economía boliviana



Fuente: Elaboración propia.

Figura 9. Correlaciones entre los componentes de largo plazo de distintos filtros de series de tiempo



Orden en las variaciones de componentes de largo plazo (potencial) en la diagonal principal: HP=Hodrick-Prescott, CF=Christiano-Fitzgerlad, CF=Christiano-Fitzgerlad simétrico, BK = BakterKing, NW=Nadaraya-Watson, HA =Hamilton.

Fuente: Elaboración propia.

Acorde al *cuadro 2*, los estadísticos de los compontes cíclicos, el filtro de Hamilton presenta la mayor variabilidad en torno al 5% del producto (D.S), mientras el resto de los filtros presenta desviaciones entre el 1 y 2% respectivamente. *HP*, *CF* presentan sesgo positivo, mientras *BK*, *NW* y *HA* presentan sesgo negativo. Todos los filtros presentan curtosis leptocúrtica o presencia de valores atípicos.

Cuadro 2. Estadísticas de componentes cíclicos

Filtro	n	Media	Mediana	Media robusta	D.S.	Mad	Min.	Máy	Rango	Asimetría	Curtosis	Error estim.
Hodrick-Prescott	125	0.00	0.00	0.00	0.02	0.01	-0.06	0.05	0.11	0.00	2.51	0.00
Christiano-Fitzgerald	125	0.00	0.00	0.00	0.01	0.01	-0.02	0.02	0.04	0.62	0.59	0.00
Christiano-Fitzgerald simétrico	125	0.00	0.00	0.00	0.01	0.01	-0.02	0.02	0.04	0.64	0.51	0.00
Baxter-King	125	0.00	0.00	0.00	0.01	0.01	-0.03	0.02	0.05	-0.26	1.44	0.00
Hamilton	125	0.00	0.01	0.01	0.05	0.03	-0.16	0.13	0.29	-0.98	1.81	0.00
Nadaraya-Watson	125	0.01	0.01	0.01	0.01	0.01	-0.05	0.05	0.10	-0.95	5.58	0.00

Fuente: Elaboración propia.

Cuadro 3. Crecimiento natural de la economía boliviana. Submuestra: 2006.T1 – 2016.T3

Filtro	n	Media	Mediana	Media robusta	D.S.	Mad	Min.	Máy	Rango	Asimetría	Curtosis	Error estim.
Hodrick-Prescott	43	4.8	4.8	4.8	0.3	0.3	4.2	5.3	1.1 -	0.1 -	0.9	0.0
Christiano-Fitzgerald	43	5.0	5.1	5.0	0.7	0.8	3.6	6.4	2.8 -	0.0 -	1.0	0.1
Christiano-Fitzgerald simétrico	43	4.8	5.0	4.9	0.6	0.6	3.5	5.9	2.4 -	0.4 -	0.8	0.1
Baxter-King	43	4.9	5.0	4.9	0.7	0.8	3.5	6.1	2.6 -	0.3 -	0.9	0.1
Hamilton	43	5.2	4.9	5.1	1.5	1.0	1.4	8.4	7.0	0.2	0.1	0.2
Nadaraya-Watson	43	4.9	4.8	4.9	0.9	0.7	3.0	6.9	3.9	0.3 -	0.3	0.1
		4.9	4.9	4.9								

Fuente: Elaboración propia.

Cuadro 4. Crecimiento natural de la economía boliviana.

En percentiles

Filtro	Percentiles								
	10%	20%	30%	40%	50%	60%	70%	80%	90%
Hodrick-Prescott	4.5	4.6	4.7	4.7	4.8	4.9	5.0	5.1	5.2
Christiano-Fitzgerald	4.0	4.4	4.6	4.7	5.1	5.2	5.4	5.6	5.8
Christiano-Fitzgerald simétrico	4.0	4.3	4.6	4.8	5.0	5.1	5.2	5.4	5.6
Baxter-King	3.9	4.2	4.5	4.8	5.0	5.1	5.2	5.5	5.6
Hamilton	3.7	4.3	4.5	4.7	4.8	4.9	5.2	5.6	6.4
Nadaraya-Watson	3.8	4.2	4.5	4.7	4.9	5.2	5.6	6.4	7.5
Promedio	4.0				4.9				6.0
Desv. Estándar	0.3				0.1				0.8

Fuente: Elaboración propia.

Finalmente, según los *cuadros 3 y 4*, la tasa de crecimiento natural de la economía boliviana, estaría con una tendencia central del 4.9% con una desviación estándar del 0.1%. Al percentil superior, la economía boliviana podría crecer hasta el 6.0% y en 4.0% respectivamente en el percentil inferior de manera consistente.

CONCLUSIONES Y SÍNTESIS

Las series de tiempo se pueden descomponer a través de varios métodos. Los más utilizados son la descomposición aditiva, multiplicativa, mixta y a través de *ARIMA Census X – 11* en componente tendencial, irregular y estacional. En *RStudio*, las paqueterías y librerías más utilizadas para estas estimaciones son ‘*fpp*’, ‘*Ecdat*’ y ‘*seasonal*’.

Para poder obtener el componente cíclico es necesario la utilización de filtros de series de tiempo: Hodrick-Prescott (*HP*), Christiano-Fitzgerald (*CF*, versión simétrica y asimétrica), Baxter-King (*BK*), Nadaraya-Watson (*NW*) y Hamilton

(*HA*). Usualmente estos filtros se aplican sobre series desestacionalizadas o libres de estacionalidad encontrando el componente tendencial o de largo plazo y el componente cíclico. En *RStudio*, las paqueterías y librerías más utilizadas para estas estimaciones son '*mFilter*' y '*neverhpfilter*'.

Asimismo, en macroeconomía aplicada el componente tendencial es interpretado como la producción potencial de la economía o crecimiento natural; mientras que el componente cíclico es interpretado como las fluctuaciones cíclicas, brecha del producto o ciclo económico.

La evaluación de los componentes cíclicos está ligada con la historia económica y con las fases de expansión y contracción de la economía. De igual manera, es pertinente evaluar la consistencia de los resultados para brindar confiabilidad sobre las estimaciones.

De la aplicación realizada, se concluyó que los filtros más sincronizados o con resultados similares conciernen a Christiano-Fitzgerald (*CF*) y Baxter-King (*BK*), mismos que corresponden a especificaciones de paso de banda. Los filtros Hodrick-Prescott (*HP*) y Nadaraya-Watson (*NW*) presentan una asociación débil y moderada con el resto de los filtros; mientras que el filtro de Hamilton (*HA*) presenta mayor volatilidad en los componentes cíclicos y de largo plazo, además de evidenciar la menor sincronización con el resto de los filtros (asociación débil).

REFERENCIAS

- Banegas, R.A., Núñez, M.A. & Váldez del Río, S. (2020). Business Cycle, Asymmetries and Non-Linearity: the Bolivian Case. *Economics & Sociology*, 13(2), 26-42.
- Baxter, M., & King, R. G. (1999). Measuring business cycles: approximate band-pass filters for economic time series. *Review of economics and statistics*, 81(4), 575-593.
- Christiano, L. J., & Fitzgerald, T. J. (2003). The band pass filter. *international economic review*, 44(2), 435-465.
- Hamilton, J. D. (2018). Why you should never use the Hodrick-Prescott filter. *Review of Economics and Statistics*, 100(5), 831-843.
- Hodrick, R. J., & Prescott, E. C. (1980). Postwar US business cycles: an empirical investigation. *Carnegie Mellon University*, discussion paper Nro. 50.
- Kydland, F. E., & Prescott, E. C. (1990). Business cycles: Real facts and a monetary myth. *Federal Reserve Bank of Minneapolis Quarterly Review*, 14(2), 3-18.
- Mejía, P. (2013). *Fluctuaciones cíclicas y crecimiento económico en México*. Madrid: Plaza y Valdes, S.A de CV.
- Nadaraya, A. (1964). On estimating regression. *Theory of Probability & Its Applications* 9 (1): 141-2. doi:10.1137/1109020.
- Watson, G. S. (1964). Smooth regression analysis. *Sankhyii Series A* 26:359-72.