

# Metodología para la predicción con modelos SARIMA Bayesianos

Daniel Dala  
Departamento de Estadística  
Universidad Nacional Autónoma de Honduras  
e-mail: daniel.dala@unah.hn

## ÍNDICE

<b>I.</b>	<b>Introducción</b>	2
<b>II.</b>	<b>Definición del Problema</b>	2
<b>III.</b>	<b>Preliminares y Notación</b>	3
III-A.	Modelos $SARIMA(p, d, q) \times (P, D, Q)_m$	3
III-B.	Metodología Box-Jenkins	3
III-C.	Inferencia Bayesiana	4
III-D.	Modelos SARIMA Bayesianos	4
<b>IV.</b>	<b>Adaptaciones a modelos Bayesianos</b>	5
IV-A.	Estimación del Modelo	5
IV-B.	Evaluación de la inferencia	5
IV-C.	Diagnóstico del modelo	5
IV-D.	Comparación de los modelos	5
<b>V.</b>	<b>Ilustraciones</b>	5
V-A.	Temperatura promedio en Honduras	6
V-B.	Precio de cierre en la acciones de Pfizer	8
	<b>Referencias</b>	10

## ÍNDICE DE FIGURAS

1.	Metodología de Box-Jenkins (1970). El diagrama de flujo presenta el procedimiento a utilizar para un análisis de datos adecuado en un enfoque frequentista, para una mayor descripción de dicha metodología revisar [1].	2
2.	Metodología para la predicción con modelos SARIMA Bayesianos	4
3.	Temperatura promedio en Honduras de 1980 a 2013.	6
4.	Los gráficos superiores muestran los datos diferenciados estacionalmente con periodo 12 y luego diferenciados no estacionalmente. Los gráficos inferiores muestran las funciones ACF y PACF de los datos doblemente diferenciados	6
5.	Gráficos de las densidades y las cadenas simuladas para cada uno de los parámetros estimados	7
6.	La serie de los residuos (parte superior). El histograma y gráfico de cuantiles (parte media). Gráficos de autocorrelación y autocorrelación parcial de los residuos del modelo (parte inferior).	7
7.	Predicciones de los modelos intentando simular los datos del conjunto de prueba.	7
8.	Predicción final generada por el primer modelo.	7
9.	Precio de cierre promedio de enero 2010 hasta septiembre 2021.	8
10.	Los gráficos superiores muestran los datos diferenciados y doblemente diferenciados no estacionalmente. Los gráficos inferiores muestran las gráficas de los ACF y PACF de los datos doblemente diferenciados	8
11.	Gráficos de las densidades y las cadenas simuladas para cada uno de los parámetros estimados.	8
12.	La serie de los residuos (parte superior). El histograma y gráfico de cuantiles (parte media). Gráficos de autocorrelación y autocorrelación parcial de los residuos del modelo (parte inferior).	9
13.	Comparación de las predicciones de los modelos intentando simular el conjunto de prueba.	9
14.	Predicción final para los meses de octubre del 2021 hasta febrero del 2022.	9

# ÍNDICE DE CUADROS

I.	Resumen de los parámetros obtenidos por el modelo. . . . .	6
II.	Resumen comparativo de la precisión en cada modelo, en donde el modelo 1 muestra los mejores resultados y el modelo 3 muestra los peores. . . . .	7
III.	Resumen de los parámetros simulados. . . . .	8
IV.	Resumen comparativo de la precisión en cada modelo, en donde el modelo 1 muestra los mejores resultados y el modelo 3 muestra los peores. . . . .	9

# Metodología para la predicción con modelos SARIMA Bayesianos

## Resumen—

## I. INTRODUCCIÓN

## II. DEFINICIÓN DEL PROBLEMA

Una de las aplicaciones más importantes en el análisis de serie temporales es la predicción, esto es, estimar valores futuros que generalmente son desconocidos, para esto existen diferentes metodologías como los modelos de espacio y estado [2], Prophet [3], redes neuronales [4], splines [5], procesos Gaussianos, [6] entre otros. Una clase de modelos muy populares por su fácil interpretación y alta capacidad predictiva son los modelos SARIMA [7], [8], pero su implementación con datos reales es compleja debido a que seleccionar el orden del modelo es una tarea complicada. Box y Jenkins (1970) [1] propusieron una metodología para el uso adecuado de dichos modelos, la cual se basa en seis etapas iterativas: *visualización de los datos*, *selección del modelo*, *estimación de parámetros*, *diagnóstico*, *comparación de modelos* y *predicción*. Dicha metodología se ilustra en la Figura 1. Existen muchos esquemas para el proceso de inferencia, y en los últimos años la inferencia Bayesiana se ha vuelto una alternativa muy utilizada para el análisis de datos con muchas aplicaciones en economía, física, química, psicología, entre otras. Su creciente popularidad se debe a su capacidad de incorporar información externa al modelo mediante una distribución a priori, y actualizar las creencias mediante el Teorema de Bayes. Este enfoque de inferencia en la práctica es muy complicado, por lo cual en los últimos años se han aproximado los resultados mediante los métodos de Markov Chain Monte Carlo [9]. Estos métodos consisten en generar una cadena de Markov cuya distribución estacionaria es la distribución a posteriori del modelo, existen muchos procedimientos para implementar estos métodos, uno de los más comunes es el Monte-Carlo Hamiltoniano que por su flexible implementación en el lenguaje Stan ha sido de utilidad en múltiples aplicaciones [10].

El mayor obstáculo al momento de realizar un análisis de datos adecuado en un enfoque Bayesiano, es que los procedimientos de estimación, diagnóstico, y selección utilizados en Box y Jenkins (1970) no son válidos en este nuevo enfoque. Gelman, Vehtari et. al. (2020) [11] proponen una extensa y robusta metodología denominada "*Bayesian workflow*", que presenta diferentes herramientas para un análisis de datos adecuado. Esta metodología se basa en la propuesta por Box y Jenkins (1970), y se generaliza para cualquier tipo de modelamiento que involucre un enfoque de inferencia probabilístico.

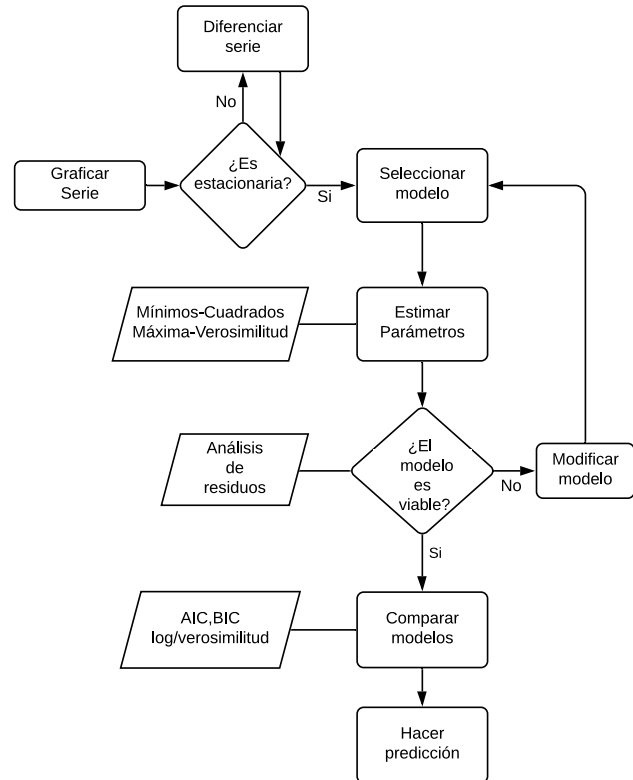


Figura 1. Metodología de Box-Jenkins (1970). El diagrama de flujo presenta el procedimiento a utilizar para un análisis de datos adecuado en un enfoque frequentista, para una mayor descripción de dicha metodología revisar [1].

Los dos principales problemas del método de Gelman, Vehtari et. al. (2020) al ser aplicados en el análisis de series temporales son su compleja estructura, y que algunas herramientas no son adecuadas para datos con supuestos de dependencia, por lo tanto, en este estudio presentamos una simplificación del *Bayesian Workflow* con ligeras variaciones en algunas de las herramientas para su adecuado uso en series temporales.

Finalmente, ilustrar la utilidad de la metodología propuesta con tres aplicaciones diferentes, el Índice de Precios al Consumidor en Honduras (IPC) de 1980 al 2018, la tasa de cambio de divisas entre Alemania y Reino Unido de 1984 a 1991 y la afluencia de turistas en Australia de 1995 al 2015, cada uno de estos conjuntos se encuentran en el paquete *bayesforecast* [12], que se especializa en el análisis Bayesiano de series temporales. Por último con los resultados de dichas pruebas se demostrará la funcionalidad del nuevo método propuesto.

### III. PRELIMINARES Y NOTACIÓN

Para los objetivos de este estudio un proceso estocástico es una colección arbitraria de variables aleatorias  $\{Y_1, Y_2, \dots\}$ , y una serie de tiempo o simplemente serie, es una realización o muestra finita  $\{y_1, y_2, \dots, y_n\}$  del proceso. Una propiedad importante a considerar es la estacionaridad, diremos que un proceso  $\{y_i\}_{i \in \mathbb{Z}}$  es *estacionario fuerte* si para cualquier colección finita del proceso su distribución conjunta se mantiene constante en el tiempo.

$$F_X(y_{t_1}, y_{t_2}, \dots, y_{t_n}) = F_X(y_{t_1+\tau}, y_{t_2+\tau}, \dots, y_{t_n+\tau}),$$

para  $t \in \mathbb{Z}_+$  con  $n \in \mathbb{N}$  y cualquier  $\tau \in \mathbb{Z}_+$ . Una propiedad menos restrictiva es la estacionaridad débil, diremos que el proceso  $\{y_i\}_{i \in \mathbb{Z}}$  es *estacionario débil* si el proceso tiene una media y varianza constante a través del tiempo, y la autocorrelación es una función lineal de la diferencia de dos tiempos.

$$\mu(t) = \mu, \quad \sigma^2(t) = \sigma^2, \quad \text{corr}(t, k) = \tau|t - k|.$$

Para  $t, k \in \mathbb{Z}$  y  $\tau > 0$ . Una serie  $\{y_t\}$  presenta tendencia sobre la media del proceso, si la media puede representarse como una función en el tiempo  $y_t = f(t) + \varepsilon_t$ , donde  $\{\varepsilon_t\}$  es un proceso con media cero. Para transformar un proceso con tendencia en uno estacionario, aplicamos el operador diferencia

$$\nabla y_t = y_t - y_{t-1}. \quad (1)$$

Para un proceso  $y$  con tendencia lineal, el proceso  $\nabla y$  obtenido en la ecuación 1 es estacionario. La ciclicidad en una serie, implica múltiples oscilaciones periódicas en la media del proceso, un caso particular es la estacionalidad, esta sucede cuando la serie presenta una oscilación periódica constante de periodo  $m$  en la media. Para transformar un proceso estacional en uno estacionario, aplicamos el operador diferencia estacional

$$\nabla_m y_t = y_t - y_{t-m}. \quad (2)$$

Donde  $m$  es un entero positivo que representa el periodo de la serie, y para una serie con estacionalidad  $y$ , el proceso  $\nabla_m y$  obtenido en la ecuación 2 es estacionario. Es importante recalcar que las series con tendencia o estacionalidad son no estacionarias, por los efectos de modelado, es necesario trabajar con procesos estacionarios. Un ejemplo de procesos estacionarios son los ruidos blancos, una colección de variables independientes con distribución normal, media cero, y varianza constante positiva.

#### III-A. Modelos SARIMA( $p, d, q$ ) $\times$ ( $P, D, Q$ ) $_m$

Sea  $\{Y_i\}_{i=i}^n$  una serie de tiempo, decimos que la serie sigue un modelo *Auto-regresivo Integrado de Medias móviles* ARIMA( $p, d, q$ ) si para cualquier tiempo  $Y_t$ , se puede escribir de la forma:

$$\nabla^d y_t = \mu_0 + \sum_{i=1}^p \phi_i \nabla^d y_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t \quad (3)$$

donde,  $\mu_0$  es la media inicial del proceso,  $p \in \mathbb{Z}_+$  y  $\{\phi_i\}_{i=1}^p$  son el orden y parámetros de la componente auto-regresiva respectivamente,  $q \in \mathbb{Z}_+$  y  $\{\theta_i\}_{i=1}^q$  son el orden y parámetros

de la componente de medias móviles respectivamente,  $d \in \mathbb{Z}_+$  representa el número de diferencias no estacionales, y  $\varepsilon_t \sim N(0, \sigma_0)$  es ruido blanco Gaussiano centrado en cero y con varianza constante positiva.

El modelo propuesto en la ecuación 3 se puede adaptar para analizar series de tiempo con estacionalidad, esto se puede logra agregando componentes auto-regresivas y de medias móviles para modelar la estacionalidad de forma aditiva, y adaptando la diferencia estacional de forma multiplicativa. Sea  $\{Y_i\}_{i=i}^n$  una serie de tiempo con estacionalidad y periodo  $m \in \mathbb{Z}_+$ , decimos que la serie sigue un modelo *ARIMA estacional Multiplicativo* SARIMA( $p, d, q$ )  $\times$  ( $P, D, Q$ ) $_m$  si para cualquier tiempo  $Y_t$ ,

$$Z_t = \mu_0 + \sum_{i=1}^p \phi_i Z_{t-i} + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^P \Phi_i Z_{t-im} + \sum_{i=1}^Q \Theta_i \varepsilon_{t-im} + \varepsilon_t, \quad (4)$$

$$Z_t = \nabla_m^D \nabla^d y_t,$$

donde  $\varepsilon_t \sim N(0, \sigma_0)$  es un ruido blanco Gaussiano con varianza constante positiva, los parámetros  $\mu_0, p, \{\phi_i\}_{i=1}^p, q, \{\theta_i\}_{i=1}^q, d$  son los mismos definidos en la ecuación 3,  $P \in \mathbb{Z}_+$  y  $\{\Phi_i\}_{i=1}^P$  son el orden y parámetros de la componente auto-regresiva estacional respectivamente,  $Q \in \mathbb{Z}_+$  y  $\{\Theta_i\}_{i=1}^Q$  son el orden y parámetros de la componente de medias móviles estacionales respectivamente, y  $D \in \mathbb{Z}_+$  representa el número de diferencias no estacionales. Note que  $Z_t$  es la transformación obtenida al aplicar diferencias y transformaciones estacionales de forma multiplicativa.

#### III-B. Metodología Box-Jenkins

El procedimiento para predicción de valores futuros en series de tiempo requiere de dos etapas fundamentales: análisis de los datos y selección del modelo de predicción que mejor se ajuste a los datos. Así que en las próximas secciones se expondrá a mayor profundidad cada uno de los pasos. Para un estudio más profundo del método leer [13].

1. *Visualización de los datos*: la visualización mediante gráficos, permite detectar patrones como tendencia, estacionalidad, ciclos u observaciones atípicas, que deben ser filtrados diferenciando la serie o aplicando diferencias estacionales. Los gráficos ACF (función de autocorrelación) [14] y PACF (función de autocorrelación parcial) [14] son de gran importancia para seleccionar los órdenes del modelo [2].
2. *Selección*: para definir un modelo inicial es necesario establecer los valores  $p, d, q, P, D, Q$  y  $m$ . Los valores  $d, D$  son el número de diferencias necesarias para que la serie sea estacionaria o un ruido blanco, esto se logra graficando la serie original y la serie diferenciada. Los valores  $(p, P)$  y  $(q, Q)$  se identifican con los gráficos PACF y ACF respectivamente, como el número de retardos (lags) diferentes de cero en la serie diferenciada, para más detalles ver [2].
3. *Estimación*: una vez se ha definido un modelo inicial, es necesario estimar los  $p + P + q + Q + 2$  parámetros, los métodos más usados son: *mínimos cuadrados* [15],

máxima verosimilitud [15], y la ecuación de Yule-Walker [16], [17].

4. *Diagnóstico*: Los modelos *SARIMA* siguen el supuesto que los errores siguen un ruido blanco Gaussiano, es decir, los errores son estacionarios con distribución normal. Para diagnóstico de estacionariedad, se utilizan las pruebas de Portmanteau [18] y Ljung-Box [19], o pruebas de raíz unitaria como la prueba Augmented Dickey-Fuller [20], Phillips-Perron [20], y KPSS [21]. Para medir normalidad pruebas como Epps [22], Lobato-Velasco [23] y proyecciones aleatorias [24] que miden normalidad en procesos estacionarios son las más adecuadas. Para más detalles ver [25].
5. *Comparación*: El criterio de selección de modelos más utilizado es el *Akaike's Information Criteria (AIC)* propuesto por Akaike en 1974 [26]. Sea  $n_p = p + q + P + Q + 2$  el número de parámetros estimados en el modelo, luego elegiremos los valores de  $p, q, P, Q$  que minimicen el AIC:

$$AIC = -2\log L + 2m, \quad (5)$$

donde  $L$  denota la verosimilitud. Existen varias modificaciones del AIC que también son usadas como el BIC (Bayesian Information Criteria) y la log-verosimilitud.

6. *Predicción*: Una vez elegido el modelo que mejor se ajuste a cada una de las pruebas anteriores se procede a hacer la predicción de las observaciones futuras.

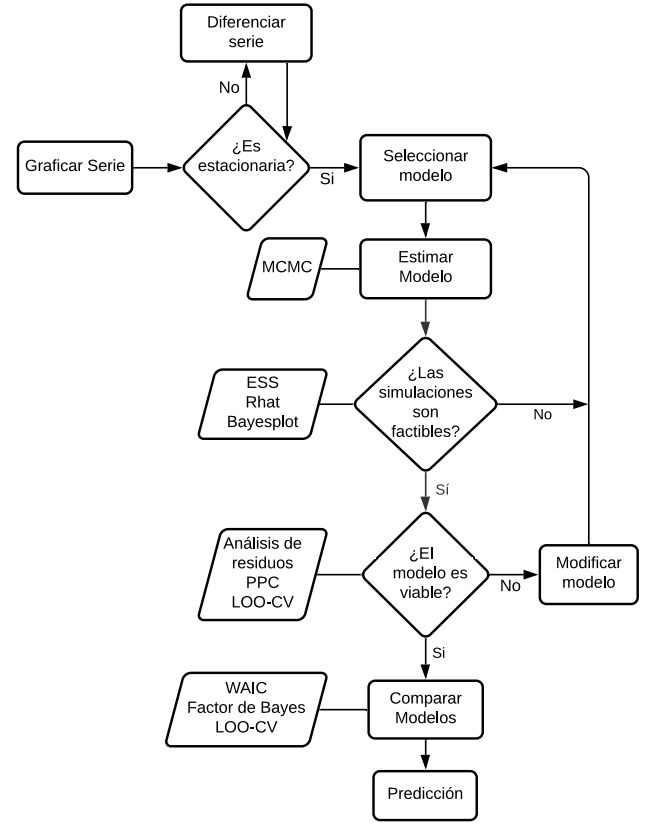


Figura 2. Metodología para la predicción con modelos *SARIMA* Bayesianos

### III-C. Inferencia Bayesiana

Para analizar la probabilidad de un parámetro  $\theta$  dado  $y$  debemos iniciar por establecer un modelo que proviene de una distribución de probabilidad conjunta para  $\theta$  y  $y$ . La función de probabilidad conjunta puede ser escrita como el producto de dos densidades:

$$p(\theta, y) = p(\theta)p(y|\theta)$$

donde a  $p(\theta)$  se le llama la distribución a *priori* y a  $p(y|\theta)$  la distribución muestral. Al condicionar el valor conocido de los datos  $y$  y usando el Teorema de Bayes obtenemos la distribución a *posteriori*:

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)}$$

Una forma equivalente de la ecuación anterior omite el factor  $p(y)$  el cual no depende de  $\theta$  y puede ser considerada una constante, lo que resulta en una distribución posteriori no normalizada, osea:

$$p(\theta|y) \propto p(\theta)p(y|\theta) \quad (6)$$

Al considerar  $p(y|\theta)$  como una función de  $\theta$  para  $y$  fija se le conoce como *función de verosimilitud*. Estas ecuaciones bastan para realizar inferencia Bayesiana en donde primero se establece un modelo  $p(\theta, y)$  y luego se desarrollan los cálculos computacionales para obtener  $p(\theta|y)$  el cual funciona como la información actualizada de los datos. Por otro lado, para hacer inferencia en observaciones desconocidas osea hacer inferencia predictiva, seguimos un procedimiento similar.

Antes que los datos  $y$  sean observados, podemos predecir una observación desconocida  $\tilde{y}$ , la distribución de  $\tilde{y}$  es llamada la *distribución predictiva posteriori* y se obtiene con la siguiente ecuación:

$$p(\tilde{y}|y) = \int p(\tilde{y}, \theta|y) d\theta = \int p(\tilde{y}|\theta)p(\theta|y) d\theta$$

### III-D. Modelos *SARIMA* Bayesianos

En base a la definición previa de un modelo *ARIMA* estacional podemos definir un Modelo *SARIMA* Bayesiano como:

$$\text{Modelo} \sim \text{SARIMA}(p, d, q) \times (P, D, Q)_m$$

$$\phi_i \sim p(\phi_i), \quad i = 1, \dots, p$$

$$\theta_j \sim p(\theta_j), \quad j = 1, \dots, q$$

$$\Phi_k \sim p(\Phi_k), \quad k = 1, \dots, P$$

$$\Theta_w \sim p(\Theta_w), \quad w = 1, \dots, Q$$

$$\mu_0 \sim p(\mu_0)$$

$$\sigma_0 \sim p(\sigma_0)$$

## RESULTADOS PRELIMINARES

### IV. ADAPTACIONES A MODELOS BAYESIANOS

Al proponer un modelo SARIMA Bayesiano para hacer la predicción debemos hacer modificaciones y adaptaciones al método Box-Jenkins, esto dado que las estimaciones y diagnósticos de parámetros se basan en inferencia Bayesiana. Inicialmente los pasos de *visualización de datos* y *selección del modelo* se mantienen igual, por lo que en las siguientes secciones se expondrán los pasos de *selección*, *diagnóstico* y *comparación de los modelos* que se verán modificados siguiendo la línea del *Bayesian Workflow* adaptado al análisis de series temporales.

#### IV-A. Estimación del Modelo

Luego de establecer un modelo inicial se estimarán los parámetros mediante inferencia Bayesiana. En este estudio la estimación será utilizando el método de simulación Monte Carlo Hamiltoniano el cual obtiene una distribución estacionaria que simula la distribución a posteriori para cada uno de los parámetros del modelo propuesto. Para una mejor comprensión de este proceso de inferencia se recomienda leer [27].

#### IV-B. Evaluación de la inferencia

Luego de hacer la inferencia es necesario evaluar la convergencia de las simulaciones midiendo la estacionariedad y la combinación de las cadenas simuladas para cada uno de los parámetros. De esta manera, para decidir la viabilidad de los resultados se utilizarán dos estadísticos, el Tamaño de muestra efectiva (*Effective Sample Size ESS*) y la Reducción de escala potencial (*potencial scale reduction  $\hat{R}$* ), el primero indica el tamaño suficiente de las simulaciones para aproximar correctamente los parámetros y el segundo es un indicador de convergencia de las cadenas que para este estudio se tomará como estimador factible si  $\hat{R} < 1,1$ . Las propiedades generales de ambos estimadores se encuentran en [27].

Es recomendable además, el análisis gráfico del ajuste a cada parámetro, esto es, observar los histogramas y gráficos de las cadenas generadas en busca de indicios de multimodalidad en la distribución de los parámetros y que las cadenas muestren convergencia. Este análisis gráfico se puede realizar con el paquete *bayesplot* [28] y se pueden observar ejemplos en el artículo [29].

#### IV-C. Diagnóstico del modelo

Al igual que en la Metodología Box-Jenkins es recomendable hacer un diagnóstico en los residuos, esto comprobando que exista un comportamiento de ruido blanco Gaussiano. Por otro lado en la inferencia Bayesiana existen métodos para constatar que el modelo representa efectivamente los datos observados, el principal método es la *Verificación predictiva a posteriori (PPC)* (Box, 1980, Rubin, 1984, Gelman, Meng, y Stern, 1996). Si el modelo se ajusta bien, este debería generar datos con el mismo comportamiento de las observaciones. No obstante si se comprueba que los residuos son iid y para evadir los problemas de no estacionariedad se recomienda hacer PPC en los residuos los cuales muestran condiciones óptimas

para la creación de histogramas y obtención de resultados concluyentes.

Generalmente la PPC es suficiente para encontrar errores en el ajuste del modelo, sin embargo, dado que usamos las observaciones para ajustar el modelo y hacer las evaluaciones es posible que en algunos casos se dejen pasar comportamientos anormales en los datos. Un camino alternativo es hacer el diagnóstico con validación cruzada *Leave-one-out cross-validation (LOO-CV)* en donde una parte de los datos es utilizada para ajustar el modelo y el resto se utiliza para medir la precisión de predicción. En [11] se aconsejan tres maneras de abordar la evaluación usando validación cruzada: 1. Verificaciones de calibración utilizando la distribución predictiva de validación cruzada 2. Identificar qué observaciones o grupos de observaciones son más difíciles de predecir 3. Identificar qué tan influyentes son las observaciones particulares, esto es, cuánta información proporcionan además de otras observaciones. Para una mayor comprensión de LOO se recomienda leer [30].

#### IV-D. Comparación de los modelos

En la comparación de modelos Bayesianos frecuentemente se utilizan LOO-CV y el Criterio de Información Watanabe-Akaike, WAIC (Watanabe, 2010). Ambos métodos estiman la precisión puntual en la predicción en un modelo Bayesiano usando la log-verosimilitud en las simulaciones extraídas de las posteriores de los parámetros. No obstante, el método de estimación de precisión para modelos en series de tiempo que mejores estimaciones propone es la validación cruzada de series de tiempo [31], sin embargo en este estudio se trabajará con LOO-CV y WAIC. La implementación de estos métodos se puede realizar con el paquete *loo* [32]. Para una mayor comprensión de los métodos se recomienda leer [30], [33].

Finalmente, luego de elegir el modelo que presente mejores resultados en la pruebas anteriores se procede a hacer la predicciones, sin embargo, es necesario tener en cuenta el contexto del estudio, esto es, series temporales. Las herramientas Bayesianas ofrecen la precisión suficiente para una buena estimación en los parámetros, no obstante hace falta un análisis cuidadoso a la hora de hacer una predicción de valores futuros en las series, de esta manera en las siguientes secciones se mostrarán ejemplos de como llevar a cabo el método Box-Jenkins adaptado a un enfoque Bayesiano junto con los análisis suficientes a tomar en cuenta en cada uno de los estadísticos para la toma de decisiones.

## V. ILUSTRACIONES

Aplicaremos la nueva metodología estudiando dos conjuntos de datos y realizando sus respectivas predicciones. Para la inferencia y análisis de los modelos utilizaremos el paquete *bayesforecast* generando 4 cadenas con un tamaño de muestra de 2,000 y un warmup de 1,000 iteraciones. Como diagnóstico de convergencia usaremos el estadístico  $\hat{R}$  [27] y la comparación de los modelos será con el paquete *loo* mediante CV-LOO [30].

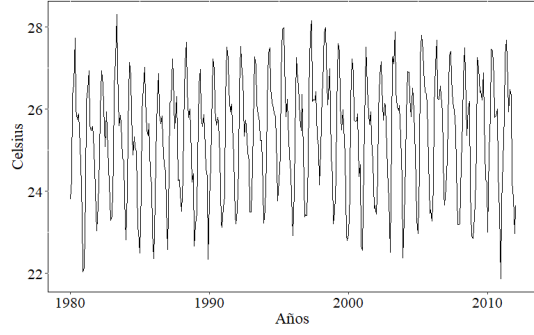


Figura 3. Temperatura promedio en Honduras de 1980 a 2013.

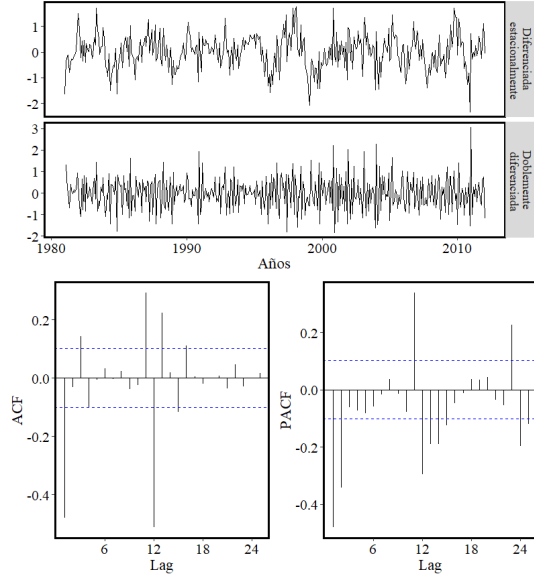


Figura 4. Los gráficos superiores muestran los datos diferenciados estacionalmente con periodo 12 y luego diferenciados no estacionalmente. Los gráficos inferiores muestran las funciones ACF y PACF de los datos doblemente diferenciados

#### V-A. Temperatura promedio en Honduras

El primer conjunto de datos muestra la temperatura promedio mensual en Honduras desde el año 1980 hasta el 2013, con un total de 405 observaciones en donde las primeras 385 servirán como conjunto de entrenamiento del modelo y el resto será el conjunto de prueba. Estos datos fueron obtenidos de *Berkeley Earth* [34]. Inicialmente, podemos observar en la Figura 3 que los datos no siguen un comportamiento estacionario, además dado que son observaciones climatológicas es claro que existen patrones estacionales anualmente, en base a estos hechos se hará una diferenciación estacional con periodo 12 y de ser necesario una diferenciación no estacional. Como podemos observar en la Figura 4 al aplicar una diferenciación estacional la serie de la parte superior aún no luce estacionaria dado que la media no parece ser constante sin embargo la estacionalidad ya no se observa lo que indica que basta hacer una segunda diferenciación en este caso no estacional, seguidamente se observa en la gráfica inferior la serie parece tener una media constante en cero y una varianza estable

	mean	se	5 %	95 %	ESS	$\hat{R}$
$\mu_0$	0.01	0.00	-0.01	0.03	3904.71	0.9998
$\sigma_0$	0.51	0.00	0.48	0.54	3489.52	0.9999
ar.1	-0.00	0.00	-0.13	0.13	4122.60	0.9998
ar.2	-0.02	0.00	-0.10	0.06	4087.97	1.0001
ma.1	-0.65	0.00	-0.81	-0.48	3967.56	0.9998
ma.2	0.03	0.00	-0.09	0.16	3888.24	0.9999
sar.1	-0.68	0.00	-0.77	-0.60	4093.16	0.9999
sar.2	-0.36	0.00	-0.44	-0.27	3934.67	0.9998
loglik	-277.24	0.03	-281.00	-274.51	4084.98	0.9998

Cuadro I  
RESUMEN DE LOS PARÁMETROS OBTENIDOS POR EL MODELO.

que es lo que buscamos. Por otra parte, una vía más sencilla y objetiva de determinar un comportamiento estacionario es graficar las funciones de autocorrelación y autocorrelación parcial. Al observar las funciones de auto correlación (ACF) y auto correlación parcial (PACF) de la Figura 4 estas muestran una leve correlación de a lo más dos retardos en los datos, el orden del modelo ARIMA a considerar entonces es  $p = 2$ ,  $d = 1$ ,  $q = 2$ . Por otra parte ambos gráficos muestran ciertos patrones periódicos, por lo tanto, el orden de la componente estacional es  $P = 2$ ,  $D = 1$ ,  $Q = 0$ . Consideramos distribuciones a priori poco informativas, usando las prioris por defecto para los parámetros de locación, escala, componentes autorregresiva y media móvil. De esta manera el modelo completo es:

$$\text{Modelo 1} \sim \text{SARIMA}(2, 1, 2) \times (2, 1, 0)_{12}$$

$$\mu_0 \sim t(0, 2, 5, 6)$$

$$\sigma_0 \sim t(7)$$

$$ar_i, ma_i \sim N(0, 0.5) \quad i = 1, 2$$

$$sar_i \sim N(0, 0.5) \quad i = 1, 2$$

En la Cuadro I se muestra un resumen de las características más importantes de la inferencia en cada uno de los parámetros. Al observar el estadístico  $\hat{R}$  de cada uno de ellos se puede concluir que estos tienen un comportamiento convergente ya que todos son menores a 1.1 y además los ESS son valores grandes indicando una tamaño factible para la representación efectiva de los parámetros lo cual comprueba una tendencia a converger. Por otra parte, podemos graficar las simulaciones generadas y observar si se muestra un comportamiento análogo al presentado en la Cuadro I. En la Figura 5 se observa que las cadenas presentan convergencia, además no se observa multimodalidad en la distribución de los parámetros por lo tanto podemos aceptar las estimaciones del modelo y continuar con el diagnóstico del ajuste en los datos. Luego, en la Figura 6 podemos observar en el histograma y gráfico de cuantiles (parte de en medio) que los residuos no tiene colas pesadas esto por la baja volatilidad de la serie lo que indica normalidad. Las gráficos ACF y PACF muestran una baja correlación manteniéndose de los intervalos lo que indica estacionaridad. Finalmente, luego que las pruebas diagnósticas fueran exitosas podemos creer que tenemos un modelo factible, sin embargo podemos hacernos la pregunta que ocurre si proponemos distribuciones a priori distintas para los modelos autorregresivos y de medias móviles además de modificar los parámetros del

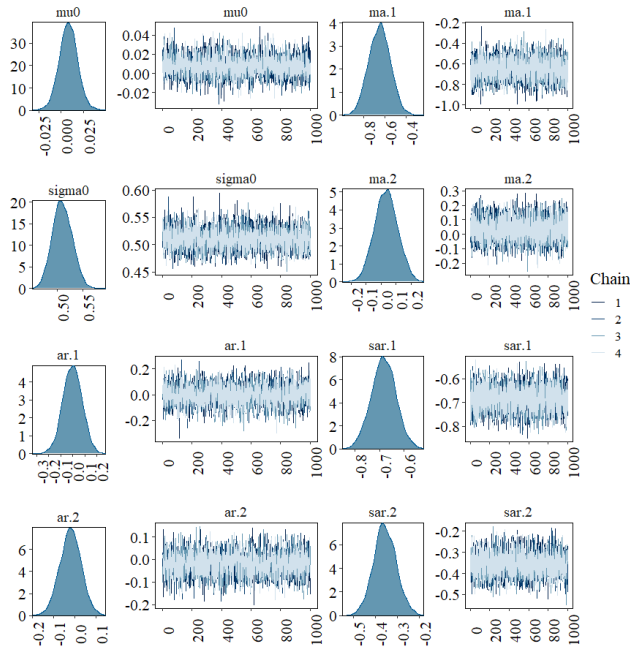


Figura 5. Gráficos de las densidades y las cadenas simuladas para cada uno de los parámetros estimados

	elpd_diff	se_diff
model1	0.00	0.00
model2	-2.99	8.89
model3	-29.68	12.28

Cuadro II

RESUMEN COMPARATIVO DE LA PRECISIÓN EN CADA MODELO, EN DONDE EL MODELO 1 MUESTRA LOS MEJORES RESULTADOS Y EL MODELO 3 MUESTRA LOS PEORES.

modelo SARIMA, así que se hará una comparación con los siguientes dos modelos:

$$\text{Modelo 2} \sim \text{SARIMA}(1, 1, 0) \times (1, 1, 1)_{12}$$

$$\text{Modelo 3} \sim \text{SARIMA}(1, 0, 0) \times (1, 1, 0)_{12}$$

Además, del orden del modelo SARIMA consideremos que los parámetros autorregresivos y de medias móviles en el modelo 2 tiene una distribución a priori  $\text{ar}, \text{sar}, \text{sam} \sim \text{Beta}(2, 2)$ , el resto de los parámetros se mantendrán con las mismas distribuciones que el modelo 1. Ahora, en la Figura 7 se observa que cada uno de los modelos generan una predicción muy confiable del conjunto de prueba, sin embargo no necesariamente se cumplirá que todos tengan la misma precisión. En el Cuadro II se muestra el resultado comparativo de los tres modelos, este indica que el modelo uno presenta una mayor precisión que los otros dos lo cual a medida que se íbamos haciendo los diagnósticos nos podía dar una leve intuición que era el modelo que mejor se ajustaba a los datos. Finalmente procedemos a realizar la predicción final, cabe mencionar que el conjunto total de los datos nos muestra observaciones hasta el mes de septiembre del 2013 así que vamos a predecir los siguientes 15 meses ósea hasta diciembre del 2015 como se muestra en la Figura 8.

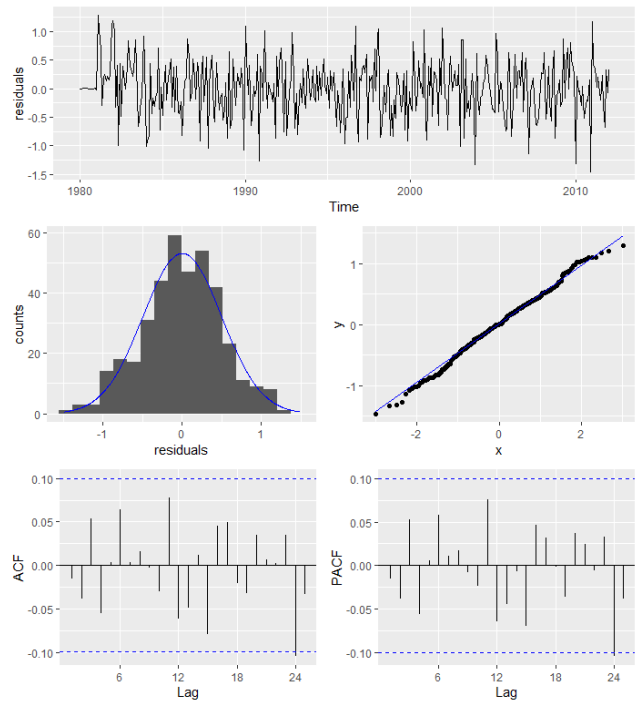


Figura 6. La serie de los residuos (parte superior). El histograma y gráfico de cuantiles (parte media). Gráficos de autocorrelación y autocorrelación parcial de los residuos del modelo (parte inferior).

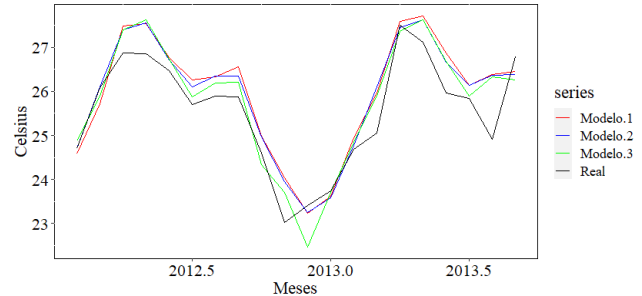


Figura 7. Predicciones de los modelos intentando simular los datos del conjunto de prueba.

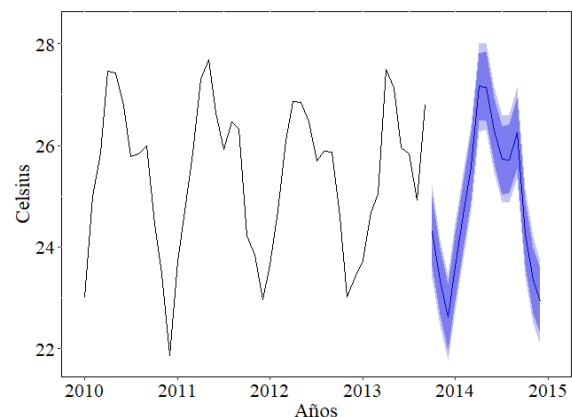


Figura 8. Predicción final generada por el primer modelo.



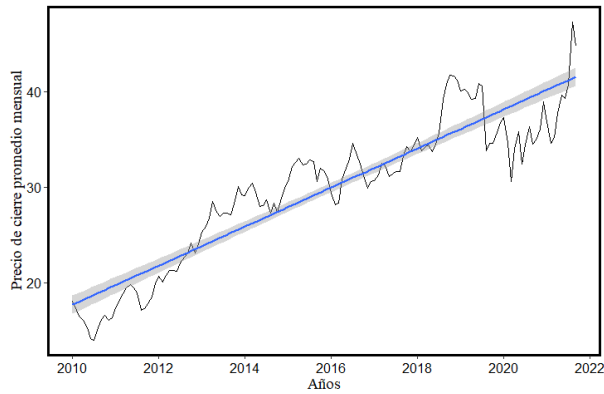


Figura 9. Precio de cierre promedio de enero 2010 hasta septiembre 2021.

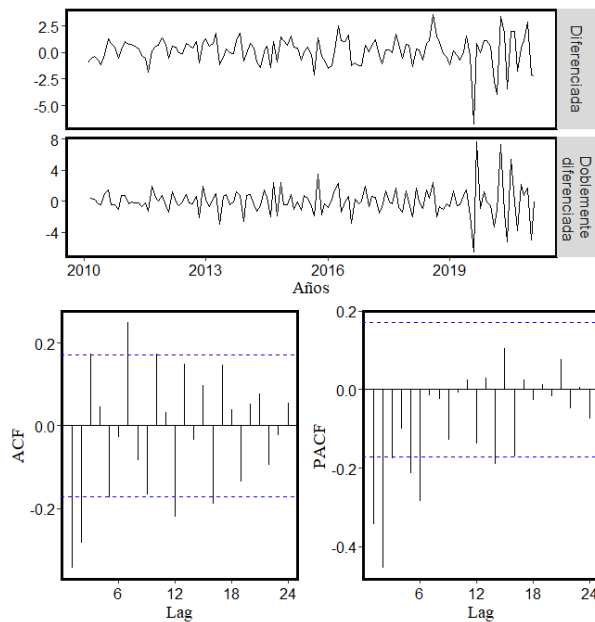


Figura 10. Los gráficos superiores muestran los datos diferenciados y doblemente diferenciados no estacionalmente. Los gráficos inferiores muestran las gráficas de los ACF y PACF de los datos doblemente diferenciados

#### V-B. Precio de cierre en la acciones de Pfizer

Para el segundo ejemplo estudiaremos el precio de cierre de las acciones en la empresa farmacéutica Pfizer, en este caso el conjunto de datos obtenidos de [35], muestra los atributos principales de las acciones desde junio de 1972 hasta septiembre de 2021 con 4 mediciones mensuales. En este caso, por conveniencia analizaremos el promedio mensual del precio de cierre desde enero del 2010. La serie final con que trabajaremos consta de 141 observaciones, en donde el 95 % (134 observaciones) será el conjunto de entrenamiento y el resto el conjunto de prueba. En primer lugar observamos en la Figura 9 que los datos muestran una clara tendencia creciente lo cual modelaremos mediante una regresión con errores ARIMA [36]. Por otra parte, en la Figura 10 observamos que los datos se estabilizan desde la primer diferenciación, además es importante notar la alta volatilidad que se presenta en el periodo entre 2019 y 2021. En base a estos hechos y tomando

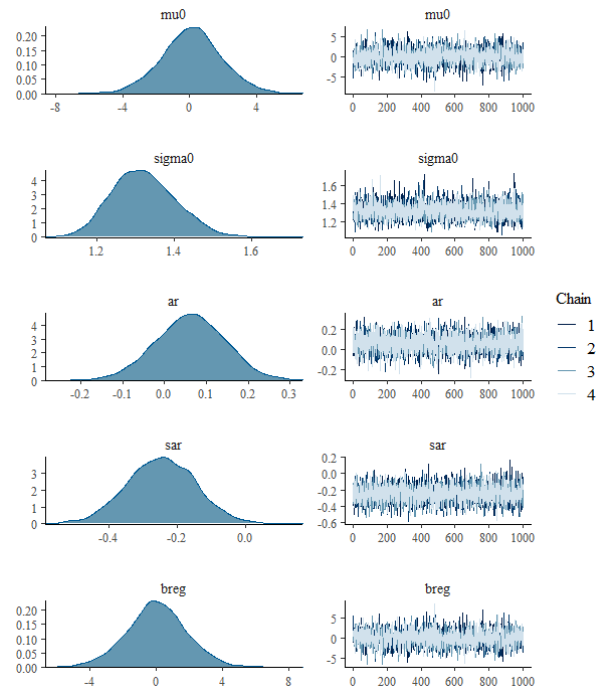


Figura 11. Gráficos de las densidades y las cadenas simuladas para cada uno de los parámetros estimados.

como referencia los gráficos ACF y PACF que indican un posible modelo autorregresivo definiremos el modelo a utilizar:

$$\begin{aligned} \text{Modelo 1: } y'_t &= \beta_1 x'_t + \eta'_t \\ \eta_t &\sim \text{SARIMA}(1, 1, 0) \times (1, 0, 0)_{12} \\ \mu_0 &\sim t(0, 2, 5, 6) \\ \sigma_0 &\sim t(7) \\ ar, sar &\sim N(0, 0, 5) \quad i = 1, 2 \\ \beta_1 &\sim t(0, 2, 5, 6) \end{aligned}$$

Notemos que el modelo  $y'_t$  es una modelo de regresión con sus error  $\eta'_t$ , de esta forma inferimos en cada uno de los parámetros del modelo SARIMA y el parámetro de regresión. En el Cuadro III se muestra un resumen de los simulaciones obtenidas y se observa que cada uno de los parámetros muestran convergencia lo cual se observa visualmente en la Figura 11, los gráficos de las distribuciones de las simulaciones no muestran multimodalidad y las cadenas se muestran convergentes, las cuales en este contexto son condiciones suficientes para suponer que se obtuvo una inferencia factible. Luego, al

	mean	se	5 %	95 %	ESS	$\bar{R}$
$\mu_0$	0.12	0.03	-2.92	3.17	3548.40	1.0012
$\sigma_0$	1.32	0.00	1.20	1.47	3884.63	1.0025
ar	0.06	0.00	-0.08	0.20	3738.73	1.0004
sar	-0.25	0.00	-0.41	-0.08	4392.88	1.0003
$\beta_1$	0.02	0.03	-3.03	3.07	3567.41	1.0012
loglik	-225.95	0.02	-228.70	-224.33	3227.89	1.0016

Cuadro III  
RESUMEN DE LOS PARÁMETROS SIMULADOS.

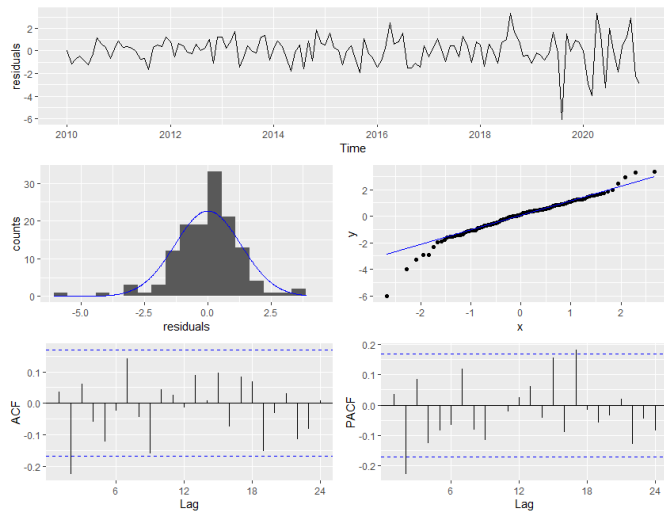


Figura 12. La serie de los residuos (parte superior). El histograma y gráfico de cuantiles (parte media). Gráficos de autocorrelación y autocorrelación parcial de los residuos del modelo (parte inferior).

observar la Figura 12 se muestra en la serie de los residuos que el modelo no explica correctamente el periodo entre 2019 y 2021 esto debido a la alta volatilidad en esos años, lo cual también se representa en la gráfica de densidad y cuantiles que muestran presencia de colas pesadas, sin embargo dada la baja autocorrelación mostradas en los gráficos ACF, PACF y el resto de la serie de los residuos este es un modelo factible para el ajuste de los datos. Por otra parte, los gráficos de autocorrelación muestran que un aumento en el orden del modelo autorregresivo podría mejorar los resultados. Para este ejemplo, se compararon otros dos modelos los cuales son:

$$\text{Modelo 2: } y'_t = \beta_1 x'_t + \eta'_t$$

$$\eta_t \sim ARIMA(1,1,0)$$

$$\text{Modelo 3: } y'_t = \beta_1 x'_t + \eta'_t$$

$$\eta_t \sim ARIMA(1,0,0)$$

Para el modelo 2, se prueba analizar los datos sin una componente estacional y para el modelo 3 se propone un modelo  $AR(1,0,0)$  sin hacer una diferenciación lo que indica que los errores se basan únicamente en el ajuste autorregresivo, por lo que la comparación se basa específicamente en el orden de los errores ya que las distribuciones a priori de los tres modelos son las mismas. Seguidamente, se observa en la Figura 13 que el modelo 3 presenta los peores resultados, sin embargo, los modelos 1 y 2 muestran predicciones similares por tanto a simple vista no se puede determinar cual presenta una mejor estimación. Por otro lado, al aplicar CV-LOO y calculando las diferencias de precisión en los modelos se obtiene que el modelo 1 presenta una mayor precisión que el modelo 2, por lo tanto se procede a realizar la predicción final mediante el primer modelo. Al observar la Figura 14 se muestra una predicción con comportamiento lineal, sin embargo en el contexto de las observaciones y la tendencia creciente es posible que los valores reales del futuro sean

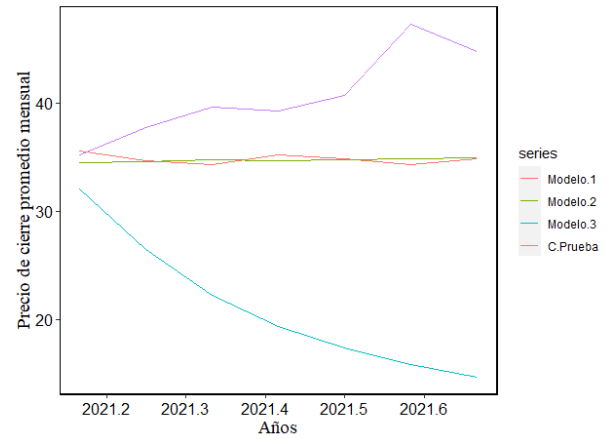


Figura 13. Comparación de las predicciones de los modelos intentando simular el conjunto de prueba.

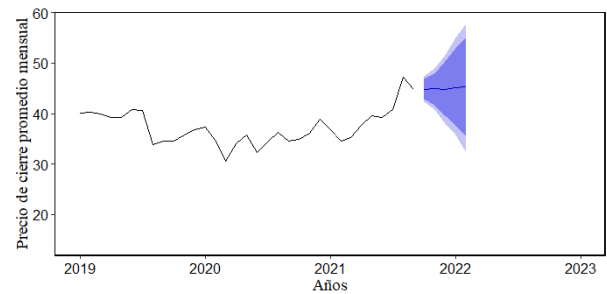


Figura 14. Predicción final para los meses de octubre del 2021 hasta febrero del 2022.

mayores a la predicción obtenida, no obstante, es suficiente para tomar una referencia del modelo y mejorarlo haciendo cambios en el orden de los errores o en las inferencias. Como

	elpd_diff	se_diff
modelo 1	0.00	0.00
modelo 2	-1.74	3.21
modelo 3	-47.19	35.44

Cuadro IV

RESUMEN COMPARATIVO DE LA PRECISIÓN EN CADA MODELO, EN DONDE EL MODELO 1 MUESTRA LOS MEJORES RESULTADOS Y EL MODELO 3 MUESTRA LOS PEORES.

podimos observar en los dos ejemplos antes analizados, cada una de las adaptaciones Bayesianas al Método Box-Jenkins proponen un mayor contexto en el análisis de los modelos, esto no necesariamente implica una mejor predicción, sin embargo, presentan una mejor visión del problema por lo que podemos asumir que en ambos ejemplos se obtuvo resultados satisfactorios y suficientes para crear conclusiones en nuestro estudio.

## DISPONIBILIDAD DE DATOS

El conjunto de herramientas y procedimientos explícitos realizados en las Ilustraciones se muestra en [37] además de los conjuntos de datos utilizados.

## REFERENCIAS

- [1] R. J. Hyndman, *Box-Jenkins modelling*. Hans Daellenbach and Robert Flood, 2002, ch. Informed Student Guide to Management Science. [Online]. Available: <https://robjhyndman.com/papers/BoxJenkins.pdf>
- [2] J. Durbin and S. Koopman, *Time Series analysis by state space methods*. Oxford University press, 2012, no. Second Edition.
- [3] S. Taylor and B. Letham, "Forecasting at scale," *PeerJ Preprints* 5:e3190v2, 2017. [Online]. Available: <https://doi.org/10.7287/peerj.preprints.3190v2>
- [4] O. Bogdan and C. Stefan, Cristian, "Time series forecasting using neural networks," *CoRR*, vol. abs/1401.1333, 2014. [Online]. Available: <http://arxiv.org/abs/1401.1333>
- [5] U. Lotrič and A. Dobnikar, "Using smoothing splines in time series prediction with neural networks," in *Artificial Neural Nets and Genetic Algorithms*. Vienna: Springer Vienna, 1999, pp. 121–126.
- [6] S. Roberts, M. Osborne, M. Ebdon, S. Reece, N. Gibson, and S. Aigrain, "Gaussian processes for time-series modelling," *Phil. Trans. R. Soc. A* 3712011055020110550, 2013. [Online]. Available: <http://doi.org/10.1098/rsta.2011.0550>
- [7] R. J. Hyndman, *ARIMA processes*. Hans Daellenbach and Robert Flood, 2002, ch. Informed Student Guide to Management Science. [Online]. Available: <https://robjhyndman.com/papers/ARIMA.pdf>
- [8] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 2nd ed. OTexts Melbourne, Australia, 2018, ch. 8.9. [Online]. Available: <https://otexts.com/fpp2/seasonal-arima.html>
- [9] J. S. Speagle, "A conceptual introduction to markov chain monte carlo methods," 2020. [Online]. Available: <https://arxiv.org/abs/1909.12313>
- [10] T. S. D. Team. Stan. [Online]. Available: <https://mc-stan.org/>
- [11] A. Gelman, A. Vehtari, D. Simpson, C. C. Margossian, B. Carpenter, Y. Yao, L. Kennedy, J. Gabry, P.-C. Bürkner, and M. Modrák, "Bayesian workflow," 2020. [Online]. Available: <https://arxiv.org/abs/2011.01808>
- [12] I. A. Alonzo and C. Cruz, "bayesforecast: Bayesian time series modeling with Stan," *ARXIV Preprint*, 2021. [Online]. Available: <https://CRAN.R-project.org/package=bayesforecast>
- [13] R. J. H. Spyros G. Makridakis, Steven C. Wheelwright, *Forecasting Methods and Applications*, 1998, ch. 7.
- [14] R. H. Shumway and D. S. Stoffer, *Time Series Analysis and Its Applications*, 4th ed. Springer International Publishing, 2017, ch. 3.3.
- [15] G. Box and G. M. Jenkins, *Time Series Analysis: Forecasting and Control*. Holden-Day, 1970, ch. 7.
- [16] P. Stoica, B. Friedlander, and T. Söderström, "A high-order yule-walker method for estimation of the ar parameters of an arma model," *Systems & Control Letters*, vol. 11, no. 2, pp. 99–105, 1988. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/0167691188900825>
- [17] C. Dimitriou-Fakalou, "Yule-walker estimation for the moving-average model," *International Journal of Stochastic Analysis*, vol. 2011, p. 151823, Aug 2011. [Online]. Available: <https://doi.org/10.1155/2011/151823>
- [18] E. Mahdi, "Portmanteau test statistics for seasonal serial correlation in time series models," *SpringerPlus*, vol. 5, 2016. [Online]. Available: <https://doi.org/10.1186/s40064-016-3167-4>
- [19] G. M. LJUNG and G. E. P. BOX, "On a measure of lack of fit in time series models," *Biometrika*, vol. 65, no. 2, pp. 297–303, 08 1978. [Online]. Available: <https://doi.org/10.1093/biomet/65.2.297>
- [20] R. R. Chaired, "Unit root tests," University of Washington. [Online]. Available: <https://faculty.washington.edu/ezivot/econ584/notes/unitroot.pdf>
- [21] D. Kwiatkowski, P. C. Phillips, P. Schmidt, and Y. Shin, "Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?" *Journal of Econometrics*, vol. 54, no. 1, pp. 159–178, 1992. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/030440769290104Y>
- [22] T. W. Epps and L. B. Pulley, "A test for normality based on the empirical characteristic function," *Biometrika*, vol. 70, no. 3, pp. 723–726, 1983. [Online]. Available: <http://www.jstor.org/stable/2336512>
- [23] I. N. Lobato and C. Velasco, "A simple test of normality for time series," *Econometric Theory*, vol. 20, no. 4, pp. 671–689, 2004. [Online]. Available: <http://www.jstor.org/stable/3533541>
- [24] M. E. Lopes, L. J. Jacob, and M. J. Wainwright, "A more powerful two-sample test in high dimensions using random projection," 2015.
- [25] A. A. Matamoros, A. Nieto-Reyes, R. Hyndman, M. O'Hara-Wild, and T. A., "nortstest: Assessing normality of stationary process," 2021, r package version 1.0.3. [Online]. Available: <https://CRAN.R-project.org/package=nortstest>
- [26] H. Akaike, "A new look at the statistical model identification," *IEEE Transactions on Automatic Control*, vol. 19, no. 6, pp. 716–723, 1974.
- [27] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 2021, ch. 11.
- [28] J. Gabry and T. Mahr, "bayesplot: Plotting for bayesian models," 2021, r package version 1.8.1. [Online]. Available: <https://mc-stan.org/bayesplot/>
- [29] J. Gabry, D. Simpson, A. Vehtari, M. Betancourt, and A. Gelman, "Visualization in bayesian workflow," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, vol. 182, no. 2, pp. 389–402, 2019. [Online]. Available: <https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/rssa.12378>
- [30] A. Vehtari, A. Gelman, and J. Gabry, "Practical bayesian model evaluation using leave-one-out cross-validation and waic," *Statistics and Computing*, vol. 27, no. 5, p. 1413–1432, Aug 2016. [Online]. Available: <https://arxiv.org/pdf/1507.04544.pdf>
- [31] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 3rd ed., 2021, ch. 5.10. [Online]. Available: <https://otexts.com/fpp3/tscv.html>
- [32] A. Vehtari, J. Gabry, M. Magnusson, Y. Yao, P.-C. Bürkner, T. Paananen, and A. Gelman, "loo: Efficient leave-one-out cross-validation and waic for bayesian models," 2020, r package version 2.4.1. [Online]. Available: <https://mc-stan.org/loo/>
- [33] M. Magnusson, M. R. Andersen, J. Jonasson, and A. Vehtari, "Leave-one-out cross-validation for bayesian model comparison in large data," 2020. [Online]. Available: <https://arxiv.org/abs/2001.00980>
- [34] R. Muller and E. Muller. Berkeley earth. [Online]. Available: <http://berkeleyearth.org>
- [35] Kaggle. Pfizer stock price (all time). [Online]. Available: <https://www.kaggle.com/kannan1314/pfizer-stock-price-all-time>
- [36] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*, 3rd ed., 2021, ch. 10. [Online]. Available: <https://otexts.com/fpp3/dynamic.html>
- [37] A. Dala. Seminario de investigación. [Online]. Available: [https://github.com/Andres-Dala/Seminario\\_de\\_Investigacion\\_MM700](https://github.com/Andres-Dala/Seminario_de_Investigacion_MM700)