

Pronóstico de ventas a través de series de tiempo orientada a micro-empresa de comida rápida hondureña

Mario Padilla

Escuela de Matemática y Ciencias de la Computación
Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras

ÍNDICE

I.	Introducción	2
II.	conceptos preliminares	2
II-A.	prueba de dicker-fuller	2
II-B.	modelo	2
II-C.	modelo autorregresivo de orden p	2
II-D.	Función de autocorrelacion simple	3
II-E.	Función de autocorrelación parcial	3
II-F.	Criterios de información	3
II-G.	pronostico de procesos AR	3
III.	planteamiento del problema	3
IV.	resultados preliminares	4
IV-A.	resultados preliminares actualizados	4
IV-B.	resultados preliminares actualizados 3	6
V.	resultado central	7
	Referencias	9

ÍNDICE DE FIGURAS

1.	salida de Función plot(lynx)	4
2.	salida de Función de correlación parcial (PACF)	4
3.	gráfica del modelo	5
4.	gráfica del la serie	5
5.	salida de función tso()	5
6.	salida de función tsclean()	5
7.	gráfica de la serie diferenciada de orden 3	5
8.	salida de Función de correlación parcial (PACF)	6
9.	gráfica del modelo	6
10.	gráfica del modelo	6
11.	serie de tiempo elecsales	7
12.	salida de función PACF()	7
13.	gráfica del modelo	7
14.	Pronostico con media distinta de cero	7
15.	pronostico con Drift	7
16.	gráfica del modelo	8
17.	ventaspiegigante diferenciada orden 1	8
18.	salida de función PACF()	8
19.	gráfica del modelo	9
20.	pronostico $l = 4$ pasos	9

ÍNDICE DE CUADROS

I.	comparación	4
II.	comparación	6
III.	comparación	8
IV.	puntos de predicción	9

Pronóstico de ventas a través de series de tiempo orientada a micro-empresa de comida rápida hondureña

Resumen—

Keywords—*Serie de tiempo, raíz unitaria, tendencia lineal, serie estacionaria, modelo autorregresivo*

I. INTRODUCCIÓN

II. CONCEPTOS PRELIMINARES

en este estudio siempre que nos refiramos a una serie de tiempo consideraremos una serie de tiempo univariada discreta denotada por $\{Y_t\} = \{Y_t : t \in T, T \subset Z\}$ en el cual se definirá como proceso estocástico, El conjunto T es llamado conjunto de índices de Y que representa el tiempo en el cual muestra datos estadísticos se recopilan, observan o registran en intervalos de tiempo t regulares (diario, semanal, semestral, anual, entre otros) con respecto a una variable de interés Y_t [8]. para este estudio de caso tendremos que tener claro las características fundamentales que clasifican las series según su comportamiento en el tiempo. en ese sentido el problema de pronostico de ventas en el sector alimenticio seguirá la metodología de Box-jenkins de series estacionarias para su estudio, modelacion y predicción. para ello se debe asegurar la estacionariedad de la serie de tiempo.

definición 1.A Sea Y_t una serie de tiempo con las siguientes propiedades se dice que es débilmente estacionaria o simplemente estacionaria si: $E(Y_t) = E(Y_{t+k}) = \mu$
 $Var(Y_t) = Var(Y_{t+k}) = \sigma^2$
 $Cov(Y_t, Y_{t+k}) = Y_k$

II-A. prueba de dicker-fuller

este es una prueba para determinar la existencia de raíces unitarias (determinar si la serie es o no estacionaria) en el cual representaremos la forma mas simple del test de dickey-fuller que es equivalente a estimar

$$Y_t = P_a Y_{t-1} + \epsilon_t \quad (1)$$

o

$$(1 - L)Y_t = (P_a - 1)Y_{t-1} + \epsilon_t. \quad (2)$$

Cualquiera de estas dos variantes de la prueba es aplicable, siendo el nulo $H_0 : P_a = 1$ contra la alternativa $H_1 : P_a < 1$. La ventaja de la segunda ecuación es que esto es equivalente a prueba $(P_a - 1) = P_a^* = 0$ contra $P_a^* < 0$; lo que es más importante, sin embargo, también simplifica las cosas para utilizar esta segunda forma de la prueba cuando un más complicado Se considera el proceso $AR(p)$. si la serie de tiempo no es estacionaria es posible convertirla en estacionaria

mediante una transformación que se denomina diferenciación **definición 2.A** Sea $\{Y_t\}$ un proceso estocástico. El operador diferencia de orden 1, denotado por ∇ , transforma el proceso original en un nuevo proceso Z_t resultado de la resta entre Y_t y Y_{t-1} , es decir

$$Z_t = \nabla Y_t = Y_t - Y_{t-1}. \quad (3)$$

El operador diferencia de orden d denotado por ∇^d funciona de la siguiente forma:

$$\nabla^d Y_t = (I - B)^d Y_t. \quad (4)$$

Existe una relación entre el operador de retraso B^k y el operador ∇^d el cual es

$$\nabla^d = \nabla(\nabla^{d-1}). \quad (5)$$

Entonces el nuevo proceso obtenido mediante la primera diferencia del proceso original es un proceso de ruido blanco y por lo tanto es estacionario. Es posible regresar al proceso $\{Y_t\}$ a partir de $\{Z_t\}$ ya que se obtiene sumando los procesos $\{Z_t\}$ es decir que

$$Y_t = Z_t + Z_{t-1} + Z_{t-2} + \dots + Z_{t-k}. \quad (6)$$

II-B. modelo

modelo es una representación simbólica de un proceso aleatorio, que se da en periodos de tiempo en el cual sus expresiones matemáticas consisten en una variable de interés que depende de sus observaciones pasadas. Específicamente, la variable de interés o de salida, depende linealmente de sus valores anteriores.

II-C. modelo autorregresivo de orden p

es un modelo en el cual el valor de la serie en el momento t depende de los valores de la serie en el momento $(t - 1, t - 2, \dots, t - p)$, comúnmente denominado $AR(p)$ cuyo proceso puede escribirse como:

$$Y_t = \sum_{i=1}^p \phi_i Y_{t-i} + \epsilon_t. \quad (7)$$

donde ϕ_1, \dots, ϕ_p son parámetros de la serie temporal que nos revelaran información de tendencia, estacionariedad y significancia, p es el numero de rezagos de los cuales la serie depende significativamente y ϵ_t es un término de error (ruido blanco). el

modelo $AR(p)$ se puede reescribir en forma equivalente más compacta mediante el operador de retraso B como [2]

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)Y_t = \epsilon_t \quad (8)$$

o

$$\phi(B)Y_t = \epsilon_t \quad (9)$$

un proceso AR sera estacionario o no estacionario dependiendo, de los valores que tomen las raíces de la ecuación característica $\phi(x) = 0$ la cual rige el comportamiento del proceso autorregresivo.[1] el cual como resultado de la ecuación en diferencia se sabe que $\phi(B) = 0$ se escribe como

$$(1 - g_1 B)(1 - g_2 B) \cdots (1 - g_2 B) \quad (10)$$

de tal manera que el proceso AR definido por $\phi(B)$ sera estacionario siempre y cuando

$$|g_i| < 1, i = 1, 2, \dots, p \quad (11)$$

II-D. Función de autocorrelacion simple

Definición 3.E El coeficiente de autocorrelacion simple en el retardo k , denotado por ρ_k , es el coeficiente de correlación simple entre las variables aleatorias Y_t y Y_{t-k} .

$$\rho_k = \frac{Cov(Y_t, Y_{t-k})}{\sqrt{Var(Y_t)Var(Y_{t-k})}} = \frac{E(Y_t - \mu)(Y_{t-k} - \mu)}{\sqrt{E(Y_t - \mu)^2(Y_{t-k} - \mu)^2}} \quad (12)$$

Definición 4. El coeficiente de autocorrelacion parcial en el retardo k , denotado por ϕ_{kk} , es el la correlación simple entre Y_t y Y_{t-k} después de extraer la influencia de los retardos intermedios

Definición 4.E El conjunto $\{\rho_k; k = 0, \pm 1, \dots\}$ se denomina función de autocorrelacion simple (ACF); y su gráfico, correlograma.

II-E. Función de autocorrelación parcial

La autocorrelación parcial indica la relación de la variable en el momento t con sus propios rezagos.

$$\phi_{kk} = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-1} & \cdots & 1 \end{vmatrix}} \quad (13)$$

Definición 5.F El conjunto $\{\phi_{kk}; k = 0, \pm 1, \dots\}$ se denomina función de autocorrelacion parcial (PACF)

II-F. Criterios de información

son medidas de la calidad relativa de un modelo estadístico, para un conjunto dado de datos. Como tal, el AIC y BIC proporcionan un medio para la selección del modelo.

Criterio de información de Akaike En el caso general, el AIC es:

$$AIC = 2k - 2\ln(L) \quad (14)$$

donde k es el número de parámetros en el modelo estadístico, y L es el máximo valor de la función de verosimilitud para el modelo estimado.

Criterio de información de Bayesiano

II-G. pronostico de procesos AR

propocision 1. La predicción de los procesos de AR (p) se realiza de forma iterativa. Formamos el pronóstico de 1 paso y, dado esto, formamos el pronóstico de 2 pasos y así sucesivamente. El pronóstico de 1 paso en el tiempo T de un proceso AR (p) se define como:

$$\begin{aligned} \hat{y}_{T+1} &= E[\phi_1 y_T + \cdots + \phi_p y_{T-p+1} + \epsilon_{T+1} | y_T, y_{T-1}, \dots] \\ &= \phi_1 y_T + \cdots + \phi_p y_{T-p+1} \\ \hat{y}_{T+2} &= E[\phi_1 y_{T+1} + \cdots + \phi_p y_{T-p+2} + \epsilon_{T+2} | \hat{y}_{T+1}, y_T, y_{T-1}, \dots] \\ &= \phi_1 \hat{y}_{T+1} + \cdots + \phi_p y_{T-p+2} \\ &\vdots \end{aligned} \quad (15)$$

continuamos así sucesivamente los l -pasos que deseamos pronosticar y varianza del l -paso del error de pronostico esta dada por:

$$Var(l) = (1 + \sum_{i=1}^{l-1} \phi_i^2) \sigma_\epsilon^2 \quad (16)$$

$$\begin{aligned} \sigma_\epsilon^2 &= E[(\hat{y}_{T+1} - y_{T+1})^2 | y_T, y_{T-1}, \dots] \\ &= E[\epsilon_{T+1}^2 | y_T, y_{T-1}, \dots]. \end{aligned} \quad (17)$$

Proposicion 2. los limites de probabilidad estan dados por:

$$\hat{Y}_T(l) \pm z_{\alpha/2} \sqrt{Var(l)} \quad (18)$$

con $z \sim N(0, 1)$

III. PLANTEAMIENTO DEL PROBLEMA

En la actualidad los países se han visto beneficiado y afectado por la globalización y las micro-empresas que han surgido deben de orientarse a la satisfacción de las necesidades de los consumidores y las expectativas que ellos demandan, debido al enorme competitividad que existe en el mercado, estas empresas deben de estar preparadas para cualquier eventualidad, la pandemia ha sido motivo por el cual una gran cantidad de micro-empresas que no se pudieron adaptar al cambio, tuvieron que cerrar sus puertas, pero esto no paso

con todas las micro-empresas, ya que alguna de estas lograron adaptarse y salir adelante ante esta fenómeno, y no solo esto, también lograron aumentar las ventas, como es el caso de la micro-empresas Pizzería Pie de Gigante que debido a su estilo de negocio, de vender comida a domicilio, tuvo un aumento en las ventas, y debido a esto, el gerente de la micro-empresa debe de aumentar el personal y los implementos necesarios para poder cumplir con la cuota de venta y poder mantenerse competitivo, es de suma importancia preguntarnos : ¿hasta qué punto la venta aumentará o disminuirá en un periodo de tiempo futuro?, aquí se presentan 3 modelos para explicar el comportamiento de las ventas a través de series de tiempo; modelo autorregresivo , modelo de medias móviles y modelo autorregresivo de media móvil con ellos pronosticaremos los valores de venta futuros para la micro-empresa de comida rápida (Pizzería Pie de Gigante) esto es de vital importancia para que la

IV. RESULTADOS PRELIMINARES

Los experimentos se han realizado utilizando usando la serie de tiempo procesada (Lynx) de la base de datos de R que nos muestra los Números anuales de trampas de lince para 1821-1934 en Canadá en segmentos de tiempo de 20 años

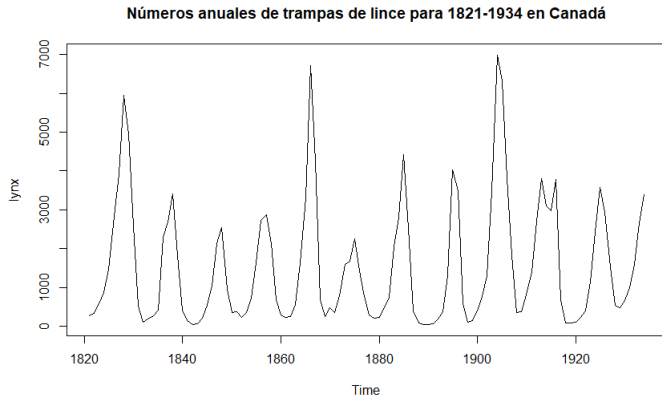


Figura 1. salida de Función plot(lynx)

con la gráfica anterior se observa la estacionariedad de la serie pero lo comprobaremos con el test de *dickey-fuller* calculado en [7] con un nivel de significancia $\alpha = 0,05$ que nos arroja un valor $p = 0,01 < 0,05$ que nos indica que la serie no tiene raíces unitarias (es estacionaria) con el cual nos permitiría continuar con el proceso AR de orden p

continuaremos con la identificación del orden del proceso AR(p) para ello utilizaremos una técnica experimental comparativa empírica de criterios de información en conjunto del método gráficos de la función de correlación parcial para la elección del orden p

la figura.2 nos muestra 4 rezagos significativos que se observan fuera de las bandas de confianza los que nos hace sospechar de un posible orden que nos proporciona un punto de partida para las comparaciones experimentales como se muestra en el cuadro II.[7]

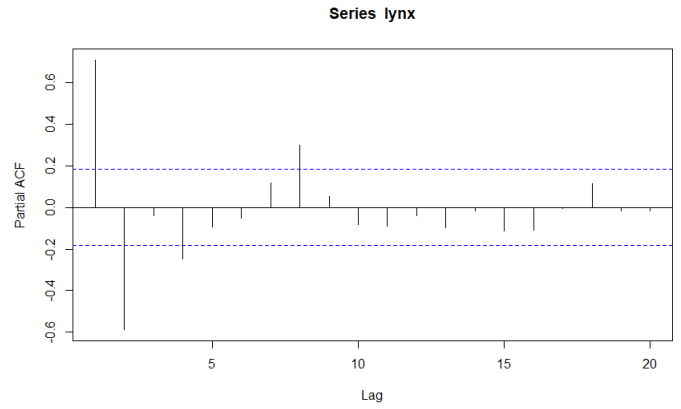


Figura 2. salida de Función de correlación parcial (PACF)

	criterio	
	AIC	BIC
orden 1	1926.991	1935.199
orden 2	1878.032	1888.977
orden 3	1879.957	1893.638
orden 4	1874.222	1890.639

Cuadro I. COMPARACIÓN

Para cada proceso ejecutado se ha calculado los criterios de informacion Akaike (AIC) y Bayesiano (BIC) y se toma el mínimo valor arrojado por los (IC) que corresponde al orden 4 que sería nuestro parámetro p para la simulación del proceso AR(p) con lo cual solo nos resta simular el modelo AR(4) que por las ecuaciones expresadas en (6) y los parámetros calculados en [7] tenemos que:

$$Y_t = 1,12Y_{t-1} - 0,72Y_{t-2} + 0,26Y_{t-3} - 0,25Y_{t-4} + 1547,386 \quad (19)$$

Cuya representación gráfica se observa en la figura. 3. todos las funciones utilizadas y paquetes utilizados para los cálculos en este documento están disponibles en [7].

IV-A. resultados preliminares actualizados

Los experimentos se han realizado usando la serie de tiempo ventaspiegigante.csv disponible en [7] que nos muestra los números mensuales de ventas de pizzas de la micro empresa hondureña pie de gigante desde mayo de 2017 hasta junio de 2021 en segmentos de tiempo de 12 meses.

iniciaremos con la intervención(limpieza) de la serie de tiempo en el cual aplicaremos una herramienta computacional llamada tsclean() proporcionada por R disponible en [7] para la detección y reemplazo de valores atípicos y valores que faltantes en una serie temporal y que es la combinación de tres

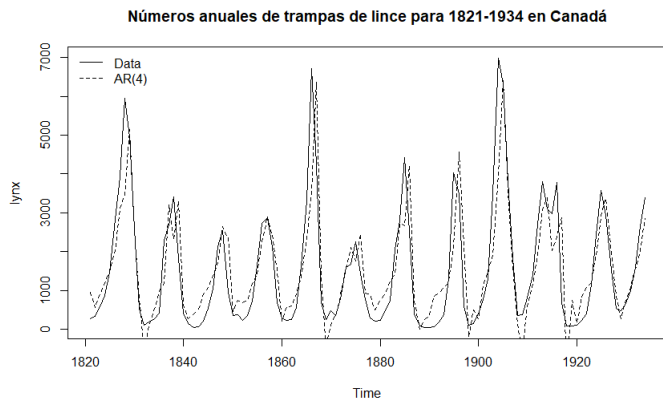


Figura 3. gráfica del modelo

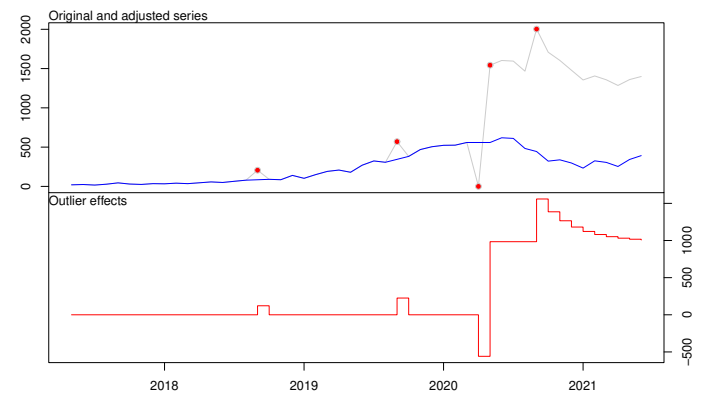


Figura 5. salida de función tso()



Figura 4. gráfica de la serie

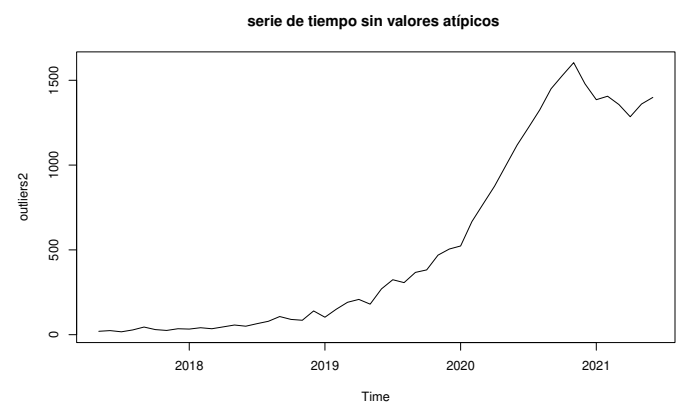


Figura 6. salida de función tsclean()

procesos: suavizado exponencial para series no estacionales [8], descomposición STL robusta para series estacionales y interpolación lineal en la serie (posiblemente ajustada estacionalmente) [10]. Para estimar los valores faltantes y los reemplazos de valores atípicos y que nos permite visualizar los valores atípicos con la figura 5

con la figura 6. se observa que la serie no es estacionaria pero lo comprobaremos con el test de *dickey-fuller* calculado en [7] con un nivel de significancia $\alpha = 0,05$ que nos arroja un valor $p = 0,3108 > 0,05$ que nos indica que la serie tiene raíces unitarias (es no estacionaria) con el cual debemos estacionarizar la serie para poder continuar con el proceso AR de orden p . para ello aplicaremos la técnica de diferencias finitas (3) calculado en [7]. fue necesario una diferenciación de orden 3 para poder alcanzar un valor $p = 0,01 < 0,05$ que nos proporciona una serie estacionaria que se muestra en la figura 7.

continuaremos con la identificación del orden del proceso $AR(p)$ para ello utilizaremos una técnica experimental comparativa empírica de criterios de información en conjunto

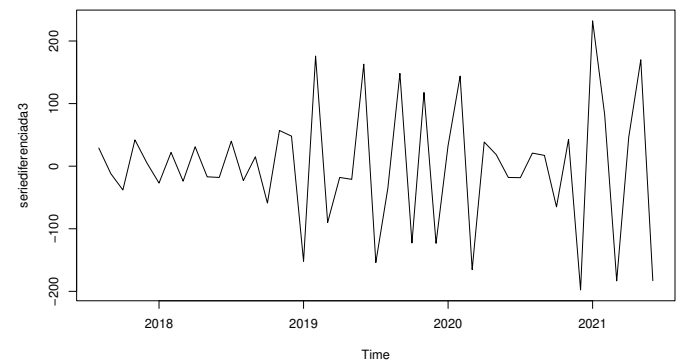


Figura 7. gráfica de la serie diferenciada de orden 3

del método gráficos de la función de correlación parcial para la elección del orden p

la figura.8 nos muestra 4 rezagos significativos que se observan fuera de las bandas de confianza los que nos hace

Ventas anuales de electricidad para Australia del Sur en GWh de 1989 a 2008.

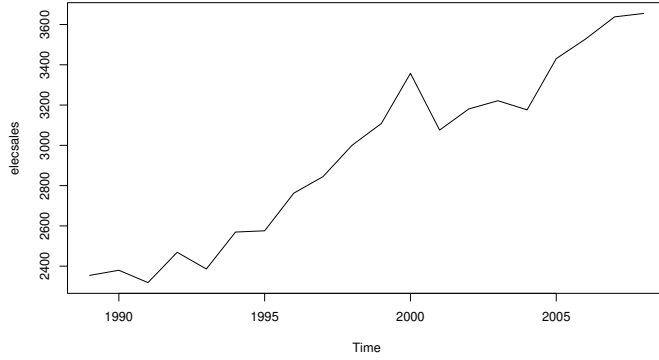


Figura 11. serie de tiempo elecsales

Ventas anuales de electricidad para Australia del Sur en GWh de 1989 a 2008.

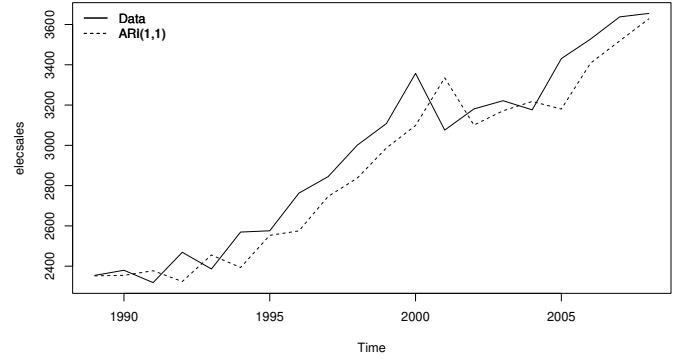


Figura 13. gráfica del modelo

Series eleciferenciada1

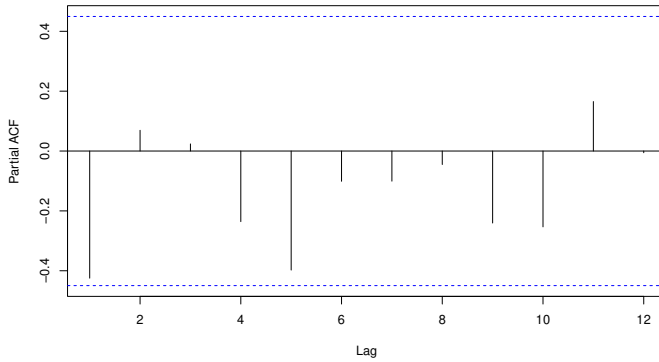


Figura 12. salida de función PACF()

que nos muestra que no hay rezagos significativos y por tanto nos induce directamente a un modelo AR de orden 1 con el cual solo nos restaría calcular el modelo como en sección (IV-A) que resultaría como en figura 13.

nos resta calcular el pronostico propuesto a $l = 12$ pasos (15) con media distinta de cero y su limites de probabilidad de pronostico (18) como lo muestra la figura 14.

con el fin de mejorar el pronostico se presenta el pronóstico de un modelo AR (1) con un término de deriva lineal (en efecto, se ajusta un modelo de regresión lineal) el modelo podría escribirse con una tendencia lineal como sigue:

$$\phi(B)Y_t = \gamma t + \epsilon_t$$

que es la ecuacion vista en (9) del proceso AR mas un parámetro γ que nos agrega una tendencia lineal (Drift) vea la figura 15.

Forecasts from ARIMA(1,1,0)

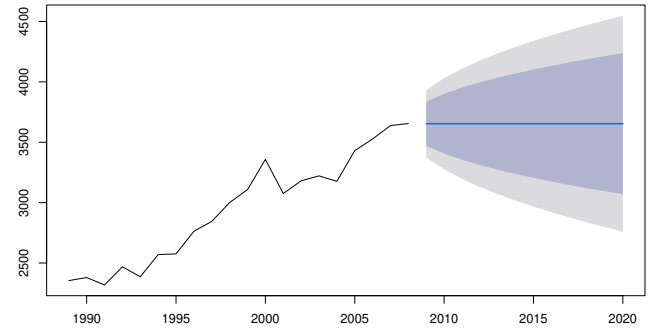


Figura 14. Pronostico con media distinta de cero

Forecasts from ARIMA(1,1,0) with drift

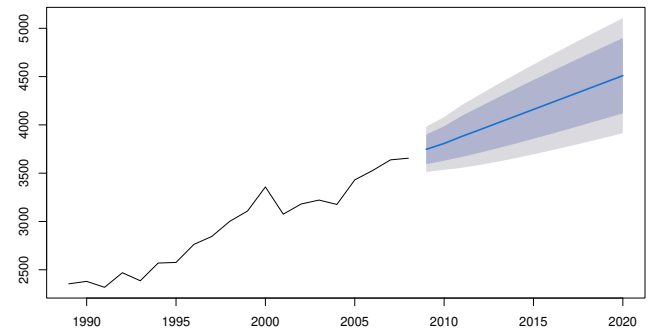


Figura 15. pronostico con Drift

V. RESULTADO CENTRAL

Los experimentos se han realizado usando la serie de tiempo ventaspiegigante.csv ya intervenida (limpia) vea la sección (IV-A) disponible en [7] que nos muestra los números mensuales

de ventas de pizzas de la micro empresa hondureña pie de gigante desde mayo de 2017 hasta junio de 2021 en segmentos de tiempo de 12 meses. figura 16.

anteriormente visto en sección (IV-A) esta serie no es estacio-

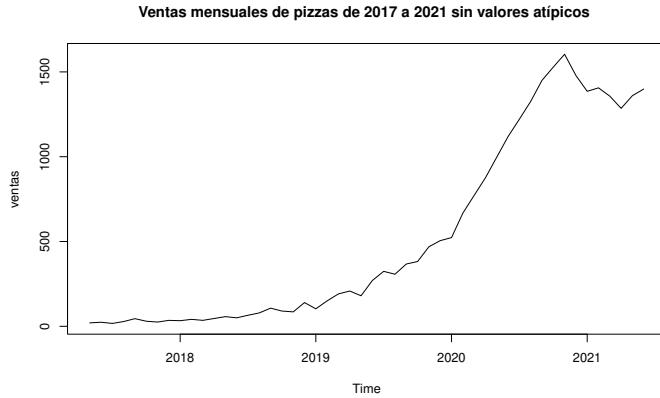


Figura 16. gráfica del modelo

naria pero lo comprobaremos con el test de *phillips – perron* basado en el test de *dickey – fuller* pero que aborda la cuestión de que el proceso de generación de datos para y_t podría tener un orden superior de autocorrelación que es admitido en la ecuación de prueba calculado en [7] con un nivel de significancia $\alpha = 0,05$ que nos arroja un valor $p = 0,861 > 0,05$ que nos indica que la serie tiene raíces unitarias (es no estacionaria) con el cual debemos estacionarizar la serie para poder continuar con el proceso AR de orden p . para ello aplicaremos la técnica de diferencias finitas (3) calculado en [7]. solo fue necesario una diferenciación de orden 1 para poder alcanzar un valor $p = 0,01 < 0,05$ que nos que nos proporciona una serie estacionaria. vea la figura 17

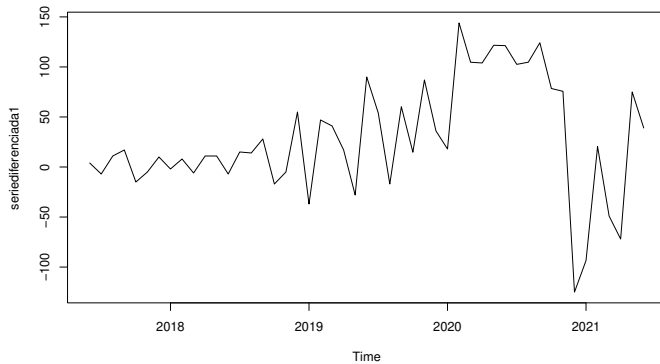


Figura 17. ventaspiegigante diferenciada orden 1

continuaremos con la identificación del orden del proceso $AR(p)$ ya estudiada en este documento vea sección (IV) con

una Función de correlación parcial figura 18.

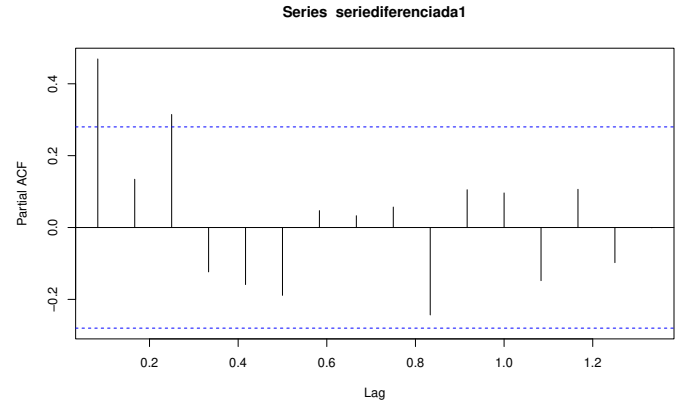


Figura 18. salida de función PACF()

la figura.18 nos muestra 2 rezagos significativos que se observan fuera de las bandas de confianza los que nos hace sospechar de un posible orden que nos proporciona un punto de partida para las comparaciones experimentales como se muestra en el cuadro III.[7]

proceso AR	criterio	
	AIC	BIC
orden 1	530.4544	534.2380
orden 2	530.6195	536.2949

Cuadro III. COMPARACIÓN

Para cada proceso ejecutado se ha calculado los criterios de información Akaike (AIC) y Bayesiano (BIC) y se toma el mínimo valor arrojado por los (IC) que corresponde al orden 1 que sería nuestro parámetro p para la simulación del proceso $AR(p)$ con el cual simularíamos el modelo $AR(1)$ que por las ecuaciones expresadas en (7) y los parámetros calculados en [7] son regresados a la serie original por medio de (6).

$$Y_t = 0,5648Y_{t-1} \quad (21)$$

Cuya representación gráfica se observa en la figura. 19.

solo nos restaría calcular el pronostico que con el fin de obtener la mejor precisión se presenta el pronóstico de un modelo $AR(1)$ con el término de deriva lineal visto en sección (IV-B) y propuesto a $l = 4$ pasos tomando en cuenta que nuestros puntos de entrenamiento son de 50 observaciones [7] que equivaldría al 8% de puntos de predicción que da lugar para minimizar los posibles errores de predicción vea la figura 20.

y podemos observar en el resumen de sus puntos como de sus intervalos de confianza del 95% una ligera tendencia de crecimiento en las ventas mensuales de pizzas en los meses de

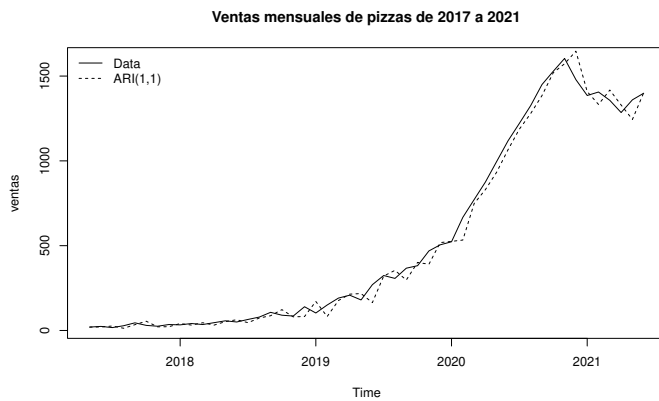


Figura 19. gráfica del modelo

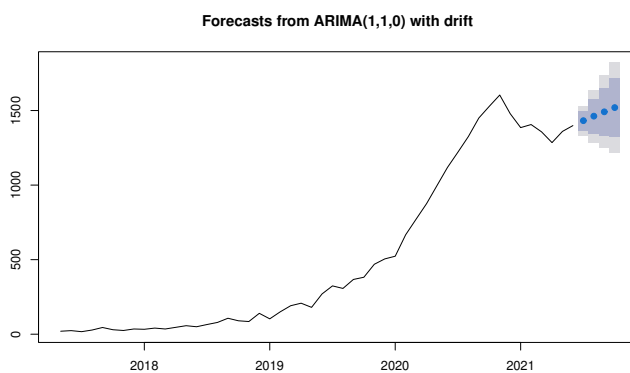


Figura 20. pronóstico $l = 4$ pasos

julio de 2021 a Octubre de 2021 como lo muestra el cuadro IV.

Fecha	Predicción de ventas	Min 95	Max 95
Jul 2021	1432	1331.8	1532.3
Aug 2021	1462	1284.7	1639.9
Sep 2021	1491	1246.9	1735.8
Oct 2021	1520	1217.5	1822.0

Cuadro IV. PUNTOS DE PREDICCIÓN

todos las funciones utilizadas y paquetes utilizados para los cálculos en este documento están disponibles en [7].

REFERENCIAS

- [1] Guerrero Guzmán, M. *análisis estadístico de series de tiempo económicas* [2 ed.] International Thomson Editores (2003)
- [2] George E. P. Box, Gwilym M. Jenkins, Gregory C. Reinsel, Greta M. Ljung *Time Series Analysis Forecasting and Control* [5ed] ed J. Wiley and Sons Inc. (2015)
- [3] Richard Harris, Robert Solis *Applied time series modelling and forecasting* - ed J. Wiley and Sons Inc. (2003)
- [4] J. Eduardo Alatorre B., Orlando R. Martínez *Modelos de series de tiempo y decisiones de política pública* septiembre 2011 Available: ebrary, https://www.cepal.org/sites/default/files/courses/files/03_modelos_de_series_de_tiempo_y_desiciones_de_politica_publ.pdf
- [5] Margarita R. Flores *análisis de series de tiempo univariadas y multivariadas mediante los modelos de ARIMA y MVAR* diciembre de 2016 Available: ebrary, http://mat.izt.uam.mx/mcmai/documentos/tesis/Gen.13-O/Margarita_Reyes_Flores.pdf
- [6] Jeffrey S. Racine *Reproducible Econometrics Using R* Oxford University Press (2019)
- [7] <https://github.com/Mariopadilla47/Seminario-de-Investigacion>
- [8] Friedman, J. H. *SMART User's Guide. Laboratory for Computational Statistics*, Stanford University Technical Report No.1.(1984)
- [9] Friedman, J. H. *A variable span scatterplot smoother. Laboratory for Computational*, Stanford University Technical Report No.5.(1984)
- [10] Bergmeir, C., Hyndman, R. J and Benítez, J. M. *Bagging exponential smoothing methods using STL decomposition and Box-Cox transformation*. International Journal of Forecasting, (2016).