



UNIVERSIDAD DE ANTIOQUIA

1 8 0 3

Análisis de series de tiempo para las RGUs de tres BD del Dashboard Pedido y Productos B2B

Leidy Johana Castro Vargas

Universidad de Antioquia
Facultad de Ciencias Exactas y Naturales
Instituto de Matemáticas
Medellín, Colombia
2021

Análisis de series de tiempo para las RGUs de tres BD del Dashboard Pedido y Productos B2B

Leidy Johana Castro Vargas

Trabajo de grado presentado como requisito parcial para optar al título de:
Estadístico

Sergio Luis Mercado Londoño

Orientador Interno, Instituto de Matemáticas

Andrés Felipe Valderrama González

Orientador externo, Tigo Colombia

Universidad de Antioquia
Facultad de Ciencias Exactas y Naturales
Instituto de Matemáticas
Medellín, Colombia
2021

Resumen

El desarrollo tecnológico cada día nos brinda más herramientas para la accesibilidad de la información. Compañías como Tigo, y en este caso el área B2B, se apoya de la visualización que nos brinda *Tableau Software*. Así, es importante para un equipo no solo visualizar sus datos, sino crear estrategias o conocer como es el comportamiento que mide; por ejemplo, los productos que adquieren o retiran sus clientes. En este trabajo y con la información del *Dashboard pedidos y productos B2B*, tablero contenido en la herramienta de visualización mencionada, se obtienen tres bases de datos que permiten crear modelos estadísticos por medio del análisis de series de tiempo con el objetivo de pronosticar unidades que generan un producto a partir de su ingreso (RGUs).

Palabras clave: (RGUs, análisis de series de tiempo, productos).

Abstract

Technological development gives us more and more tools for the accessibility of information. Companies like Tigo, and in this case the B2B area, rely on the visualization provided by *Tableau Software*. Thus, it is important for a team not only to visualize their data, but to create strategies or know how the behavior is that measures; for example, the products that their customers buy or remove. In this work and with the information of the *B2B orders and products Dashboard*, board contained in the above-mentioned display tool, three databases are obtained that allow statistical models to be created by means of time series analysis with the aim of forecasting units that generate a product from its income (RGUs).

Keywords: RGUs, time series analysis, products.)

Leidy Johana Castro Vargas *

Contenido

Resumen	3
1. Introducción	5
2. Marco Teórico	5
2.1. Análisis de series de tiempo	5
2.1.1. Componente determinística.	6
2.1.2. Componente estocástica.	7
3. Metodología	9
3.1. Base de datos	9
3.2. Implementación de modelos	9
4. Resultados	11
4.1. RGUs Pedidos	11
4.2. RGUs Retiros	14
4.3. RGUs Pedidos y Retiros	16
5. Conclusiones y Recomendaciones	18

*E-mail: leidyj.castro@udea.edu.co, [Instituto de Matemáticas](#), Universidad de Antioquia, Medellín, Colombia.

1. Introducción

El desarrollo tecnológico cada día nos brinda más opciones a la hora de trabajar con diferentes tipos de información. Nos encontramos con beneficios para cualquier sector, puesto que existe gran variedad de herramientas de programación y de visualización que aportan un método u opción ante seguimientos en pro de estrategias y conocimiento de cualquier compañía.

Por ende, compañías como Tigo buscan herramientas que ayudan a sus colaboradores en la accesibilidad de la información generalmente adquirida de sus clientes. Entonces, siguiendo esta idea Tableau Software hace parte de la visualización de los datos generando tableros que recolectan esta información, permitiendo encontrar desarrollos como el que ocuparemos en este proyecto, el cual fue adquirido a partir de productos que ofrecen o han dispuesto para los clientes desde el área B2B (business-to-business).

El **Dashboard pedidos y productos B2B**, tablero anteriormente mencionado, se obtiene de tres bases de datos: **Ocupaciones, Pedidos y Retiros**. Estas se utilizan con el fin de obtener en el tiempo dos comportamientos, los **Revenue**, es decir, los ingresos que generan cierto producto y las **RGUs** que nos indican unidades que generan un producto a partir de su ingreso. La obtención de estos valores se da por los productos brindados desde B2B y que están divididos en 9 líneas donde podemos categorizar por segmento. Es decir, si cierto producto está asociado a un cliente denominado Micro Small, Empresa, Multinacional, Gobierno, entre otros, se busca conocer a que regional pertenece, o en qué tecnología fue adquirido.

De acuerdo con lo anterior, es fundamental para el equipo de trabajo aprovechar la información adquirida en el Dashboard pedidos y productos B2B, así como poder obtener predicciones, tanto de lo que ingresa es decir pedidos, como de lo que sale es decir retiros, a partir del comportamiento en el tiempo, como ocurre con las RGUs.

El esquema de este documento es el siguiente: En la Sección 2 se describen los fundamentos teóricos de los modelos utilizados, es decir series de tiempo; en la Sección 3 se presenta la metodología abordada el proyecto; en la Sección 4 se describen los resultados obtenidos y, finalmente, en la Sección 5 se presentan las conclusiones y recomendaciones.

2. Marco Teórico

Series de tiempo

La serie de tiempo es una sucesión de valores observados para una variable en el tiempo que está ordenada cronológicamente y presente generalmente en una frecuencia regular.

2.1. Análisis de series de tiempo

El análisis de series de tiempo es el desarrollo de modelos estadísticos que nos permiten explicar el comportamiento de una serie de tiempo, su uso habitual es el de obtener predicciones y pronósticos que sean útiles para la planeación y toma de decisiones del estudio al que hemos asociado la serie de tiempo.

Así, las series de tiempo presentan 4 componentes en la modelación:

1. **Tendencia T_t** : Se puede definir como el patrón o cambio a largo plazo que se caracteriza por la persistencia al crecimiento o al decrecimiento de los valores de una serie, de esta forma nos evidencian el crecimiento o declinación de esta. La tendencia la podemos considerar determinista o estocástica, para la primera es una ocurrencia predecible, es decir, que podemos identificar los efectos permanentes sobre la serie temporal, por otro lado cuando no es determinista esta resulta impulsada por choques estocásticos y no presenta un nivel particular hacia el cual tienda a regresar, por ende se considera una caminata aleatoria.
Caminata aleatoria: Es como definimos en matemática a la trayectoria asociada que resulta de sucesivos pasos aleatorios; entre los ejemplos más usuales están los pasos que da un borracho en una acera.
2. **Estacionalidad S_t** : Es el patrón de cambio regular que se asocia a la variación definida por períodos donde observamos una frecuencia anual, mensual, diaria, entre otras. Igualmente está asociada a una estacionalidad determinista o estocástica, determinista es sólo aproximadamente constante, puede ser pronosticada con exactitud si es totalmente determinista y su aspecto permanece invariante en el tiempo, si es una componente estocástica está correlacionada con componentes no estacionales tal que presenta cambios abruptos en el tiempo.
3. **Ciclos**: Cambios o movimientos hacia arriba y hacia abajo con una larga duración, aunque no siempre iguales en toda la serie y debido a la influencia de fluctuaciones de esta, cambian la tendencia de los datos lo cual no es producido por la estacionalidad. Los ciclos son los componentes más difíciles de pronosticar pues no existe una función del tiempo para su modelación.
4. **Error E_t** : Fluctuaciones erráticas sin un patrón definido alrededor de una media constante que resultan de extraer la tendencia, la estacionalidad y los ciclos. Estas fluctuaciones son debidas a eventos externos que solo ocurren en un tiempo y de forma impredecible. Se dice que una serie que solo exhibe tal comportamiento es estacionaria en su media. También es conocido como el ruido de la serie.

Los modelos más básicos en series de tiempo consideran solo la tendencia o en tal caso tendencia y estacionalidad donde se ignoran los ciclos dejándolos al error del modelo, por otro lado las 2 primeras componentes anteriormente mencionadas pueden ser aditivas pues no presentan cambios en la fluctuaciones durante el tiempo de la serie o por el contrario presentan cambios y se denominan de componentes multiplicativos, igualmente tenemos modelos aditivos, parcialmente multiplicativos o multiplicativos.

El análisis de series de tiempo puede considerarse desde la componente determinista donde encontramos los modelos globales y los modelos locales; para la componente estocástica tenemos los modelos ARMA, ARIMA y SARIMA. Justamente como se mencionó anteriormente el proceso de este trabajo se hizo desde la componente determinista y concluimos con los procesos ARMA de la componente estocástica.

2.1.1. Componente determinística.

En la componente determinística están presentes los ajustes globales y los ajustes locales, para el primero podemos encontrar modelos globales aditivos o multiplicativos y para el segundo caso se puede considerar solo el comportamiento de la tendencia y antes de modelar se desestacionaliza la serie o por el contrario dejamos los efectos estacionales en el modelo y aplicamos suavizamiento exponencial, así

tenemos:

Ajuste global: Consideramos los modelos de regresión donde se asume y estima un modelo con parámetros constantes para toda la serie, bajo el supuesto de que las componentes estructurales (tendencia y estacionalidad), no sufren cambios en el tiempo.

Modelo global aditivo

Modelo polinomial de grado p con indicadoras

$$Y_t = \sum_{i=0}^p \beta_i t^i + \sum_{j=1}^{s-1} \delta_j I_{j,t} + E_t, \quad E_t \sim RB(0, \sigma^2).$$

Sean p el grado del polinomio con el que modelamos la tendencia, β_i los parámetros asociados al polinomio y s el periodo estacional o número mínimo de períodos que tarda el patrón estacional en repetirse.

$I_{j,t}$ es una función periódica discreta que permiten definir de forma completa el patrón estacional. δ_j es el coeficiente de regresión que asigna a las variables indicadoras.

Ajuste local: Esta estrategia admite cambios en los parámetros que definen las componentes estructurales (tendencia y estacionalidad), de manera que los modelos locales se adaptan a tales cambios.

Modelo local aditivo

Modelo LOESS cuadrático de descomposición aditiva:

En la vecindad de un tiempo t_k donde se quiere ajustar:

$$Y_t = \beta_{0,k} + \beta_{1,k}t + \beta_{2,k}t^2 + \sum_{i=1}^s \delta_i I_{i,t} + E_t, \quad E_t \sim N(0, \sigma^2)$$

para todo t vecino a t_k , con $\sum_{i=1}^s \delta_i = 0$, $\beta_{0,k}, \beta_{1,k}, \beta_{2,k}$ los parámetros de la recta local en la vecindad de t_k .

Modelo Holt-Winters de suavizamiento exponencial.

$$Y_{t+h} = \beta_{0,t} + \beta_{1,t}h + \sum_{i=1}^s \delta_i I_{i,t+h} + E_{t+h}, \quad E_t \sim N(0, \sigma^2)$$

con $\sum_{i=1}^s \delta_i = 0$; los parámetros $\beta_{0,t}, \beta_{1,t}$ y $\delta_{i,t}$ representan el nivel en t , la pendiente en t y los efectos estacionales en t , respectivamente, cambiando lentamente en el tiempo.

2.1.2. Componente estocástica.

Para la componente estocástica tenemos que una o las dos componentes estructurales no son impulsadas por resultados intrínsecos durante el proceso. Así, los patrones siguen una estructura que no permite retorno, por ende, no se espera que se presente en el futuro. Para seguir esta estructura debemos distinguir claramente los conceptos de estacionariedad, proceso estocástico, proceso estacionario, y ruido blanco.

1. **Estacionariedad:** para que una serie se considere estacionaria esta debe ser estable a lo largo del tiempo, es decir, la media y varianza son constantes en el tiempo. Esto se puede verificar gráficamente. Se busca que la serie oscile alrededor de una media constante y la variabilidad con respecto a esa media también permanece constante en el tiempo; cuando la media o la variabilidad no es constante decimos que no hay estacionariedad.
2. **Proceso estocástico:** se describe como una sucesión de variables aleatorias indexadas en el tiempo. Así, el proceso estocástico asociado a una serie de tiempo es una sucesión de variables aleatorias ordenadas en función del tiempo.
3. **Proceso estacionario:** un proceso estocástico es estacionario si su media y su varianza son constantes en el tiempo y si el valor de la covarianza entre dos periodos depende solamente de la distancia o rezago entre estos dos periodos de tiempo y no del tiempo en el cual se ha calculado la covarianza.
4. **Ruido blanco (RB):** se puede considerar como un caso de los procesos estocásticos estacionarios, y se define como una secuencia donde los valores son independientes e idénticamente distribuidos a lo largo del tiempo con media cero e igual varianza.

Ya con la información anterior podemos definir los procesos que consideramos.

Procesos Autorregresivos (AR(p)): Un proceso $Z_t, t \geq 1$ es un proceso autorregresivo de orden p tal que:

Ecuación:

$$Z_t = \sum_{j=1}^p \phi_j Z_{t-j} + a_t, a_t \sim RB(0, \sigma_a^2).$$

Polinomio AR:

$$\phi_p(B) = 1 - \sum_{j=1}^p \phi_j B^j.$$

Procesos de Medias Móviles (MA(q)): Un proceso $Z_t, t \geq 1$ es un proceso de medias móviles de orden q tal que:

Ecuación:

$$Z_t = \sum_{j=1}^q \theta_j a_{t-j} + a_t, a_t \sim RB(0, \sigma_a^2).$$

Polinomio MA:

$$\theta_q(B) = 1 - \sum_{j=1}^q \theta_j B^j.$$

A partir de los dos procesos estacionarios anteriores, se trabaja la componente estocástica para la modelación de la tendencia, y al unir los dos procesos se puede generar los **procesos autorregresivos y de medias móviles ARMA**. La ecuación y los polinomios se pueden expresar como

$$Z_t = \sum_{j=1}^p \phi_j Z_{t-j} + a_t + \sum_{i=1}^q \theta_i a_{t-i}, a_t \sim RB(0, \sigma_a^2).$$

$$\phi_p(B) = 1 - \sum_{j=1}^p \phi_j B^j, \quad \theta_q(B) = 1 + \sum_{i=1}^q \theta_i B^i.$$

donde $\phi_p(B)$ y $\theta_q(B)$ son los polinomios AR y MA.

Al considerar estos procesos podemos diferenciar la serie y obtener los Procesos Autorregresivos, Integrados y de Medias Móviles **ARIMA** (p, d, q) donde d representa las diferenciaciones realizadas a la serie hasta que esta sea estacionaria. Así mismo, si queremos considerar estacionalidad estocástica pasamos a los Procesos **ARIMA** estacionales o **SARIMA**. Los procesos anteriormente mencionados solo son para conocimiento de posibles modelos para aplicar a la serie desde la componente estocástica más no se aplicaron en el trabajo, es decir, fue suficiente la aplicación de modelos ARMA.

La información anterior fue tomada del libro [1] y las notas del curso de series de tiempo [4].

3. Metodología

La metodología que se desarrolló en este trabajo fue dividida en varias etapas. Se inició con la primera descarga de información bajo un tablero de prueba, se implementaron algunos modelos principalmente determinísticos y al salir el tablero oficial para uso del equipo y conociendo un poco los datos se hizo la siguiente descarga, tres veces, dado que estuvo en constante actualización. Este desarrollo se repitió hasta obtener los mejores modelos desde la componente determinística pasando a la componente estocástica.

3.1. Base de datos

Para el pronóstico de las RGUs indicado por la unidades que generan un producto a partir de su ingreso, y como se mencionó antes, se realizaron tres descargas de información, a medida que el tablero se iba actualizando. Así, considerando a nivel general tres situaciones:

1. **Pedidos:** RGUs que ingresan a la compañía totalmente nuevas.
2. **Retiros:** RGUs que salen de la compañía sea voluntariamente es decir por una migración o involuntario por no pagar el servicio, entre cada una pueden haber más casos.
3. **Pedido y retiros:** RGUs netas ante la entrada y salida de productos acá la entrada no es solamente lo nuevo, puede que sea una migración que ingresa por un cambio de tecnología.

La información de estas bases de datos se tomó en tres periodos. El que presentamos corresponde a la actualización comprendida entre el primero (1) de enero del 2020 y el catorce (14) de mayo del 2021. Para cada una de las tres situaciones anteriores se desarrolla un análisis en series de tiempo con frecuencia diaria.

3.2. Implementación de modelos

Para la implementación de los modelos tanto determinísticos como estocásticos se realizó el siguiente proceso a las bases de datos:

- Se dividió la base de datos en dos subconjuntos, uno de entrenamiento y otro de testeo.
- Se interpolaron valores perdidos con la función `na.approx` de la librería `zoo`. Esta aplica para pedidos pues es la base con 6 datos faltantes.

- Se limpiaron los valores atípicos y se tomaron los sugeridos por la función `tsoutliers` de la librería `forecast`.
- Se utilizó la transformación Box-Cox, (consideramos [2]) pero con una leve modificación ya que tenemos datos negativos.
En este caso, la función `sgn` hace referencia a la función signo, y x pertenece a los reales.

$$y_t = \begin{cases} \log(x_t) & \text{si } \lambda = 0; \\ \text{sgn}(x_t)(|x_t|^\lambda - 1)/\lambda & \text{en otro caso.} \end{cases} \quad \text{sgn}(x) = \begin{cases} 1 & \text{si } x > 0; \\ 0 & \text{si } x = 0; \\ -1 & \text{si } x < 0. \end{cases}$$

Para valores positivos de x_t , es lo mismo que la transformación de Box-Cox original. Tomado de [3]

- Se aplicó el respectivo modelo (un modelo global, tres modelos locales y tres modelos estocástico).
- Se realizó la predicción sobre el conjunto de testeo.
- Se calcularon las respectivas métricas para observar el desempeño de los modelos y compararlos respectivamente.

Las métricas usadas para medir y comparar los pronósticos de los modelos fueron utilizando la función `Accuracy`.

- Accuracy:** es una de las métricas más usadas a la hora de evaluar un modelo, acá podemos encontrar

RMSE: raíz del error cuadrático medio (error estándar).

$$\sqrt{\frac{1}{n} \sum_{t=1}^n [x_t - \hat{x}_t]^2}$$

MAE: error promedio absoluto de pronóstico.

$$\frac{1}{n} \sum_{t=1}^n |x_t - \hat{x}_t|$$

MAPE: porcentaje medio absoluto de error.

$$100\% \frac{1}{n} \sum_{t=1}^n \left| \frac{x_t - \hat{x}_t}{x_t} \right|$$

Donde x_t es el valor observado y \hat{x}_t es el valor estimado de la serie.

- Y para comparar los modelos se utilizaron los criterios de información

Akaike (AIC)

Bayesiana o de Schwarz (BIC)

Además de considerar los supuestos de cada modelo donde nos guiamos tanto de las pruebas gráficas como de las pruebas de hipótesis.

Para el desarrollo de los modelos se usó R Studio versión 4.0 librerías `TSA`, `fANCOVA`, `forecast`, `strucchange`, `car`, `FitAR`, `lmtest`, `sandwich`, `readxl` y `zoo`.

4. Resultados

El análisis de series de tiempo se desarrolla a partir de las dos componentes. Indicamos los modelos utilizados y damos los resultados del mejor o mejores modelos según corresponda. Los demás modelos están presentes en el código adjunto.

4.1. RGUs Pedidos

Los modelos para pronosticar las RGUs de los pedidos desde la componente determinística se dividen en un modelo global y tres modelos locales, para el primer caso consideramos un modelo polinomial de grado 3 con indicadoras; a partir de la prueba de estabilidad, que nos permite considerar si la serie requiere un análisis local, en este caso fue afirmativo, entonces consideramos dos modelos Loess uno lineal y otro cuadrático bajo el criterio GCV(criterio de validación cruzada generalizada) y el suavizamiento exponencial de Holt-Winters aditivo 2.1.1. Para la componente estocástica consideramos tres modelos un AR(3), AR(1), ARMA(2,1) 2.1.2.

Así, los mejores resultados fueron para el modelo local Holt-Winters aditivo y el modelo estocástico ARMA(2,1).

Holt-Winters aditivo - HW

Ecuación:

$$Y_{t+h} = \beta_{0,t} + \beta_{1,t}h + \sum_{i=1}^7 \delta_{i,t}I_{i,t+h} + E_{t+h}, \quad E_t \sim N(0, \sigma^2),$$

Bajo el modelo HW, obtuvimos

alpha(α)	beta(β)	gamma(γ)
0.1503	0.0396	0.0806

Tabla 1: Parámetros de suavizamiento

a	b	X	J	V	S	D	L	M
9.9862	0.0185	-1.0562	-5.2480	0.4229	0.4150	1.0829	1.2936	0.3979

Tabla 2: Coeficientes del modelo HW

Sean X, J, V, S, D, L y M, el coeficiente asociado al día de la semana, donde X hace referencia a el día miércoles y así van hasta el día martes(M).

Autorregresivo y de medias móviles ARMA(2,1)

Ecuación:

$$E_t = \phi_1 E_{t-1} + \phi_2 E_{t-2} + a_t + \theta_1 a_{t-1}, a_t \sim RB(0, \sigma_a^2).$$

Polinomios AR y MA

$$\phi_1(B) = 1 - \phi_1 B - \phi_2 B^2 \quad \theta_1(B) = 1 + \theta_1 B$$

Los resultados obtenidos son:

ar(1)	ar(2)	ma(1)	intercept	t	t^2	
	X	J	V	S	D	L
1.0247e+0	-1.1310e-0	-8.3211e-0	9.1630e+0	-5.5471e-0	1.6638e-0	
	2.8665e-0	2.6913e-0	-2.0041e-0	-1.8162e+0	-5.6095e+0	-9.6699e-0

Tabla 3: Coeficientes del modelo ARMA

Ahora, bajo los **Criterios de Información CI** presentados en la tabla 4, se observa que el mejor modelo a partir de los estimadores que están en las tablas 2 y 3 fue el ARMA(2,1). Este presenta significancia en 7 de los 12 coeficientes.

	HW	ARMA(2,1)
AIC	3.2724	2.8241
BIC	3.5085	3.1317

Tabla 4: CI

Siguiendo con los **pronósticos** tenemos el gráfico *RGUs-Ped con transformación*, obtenido del desarrollo de la serie ya que utilizamos la transformación Box-Cox, el segundo *RGUs-Ped sin transformaciones* es el resultado de devolver a la escala original la serie, los pronósticos para el subgrupo que dejamos de testeo se puede ver en color azul

Holt-Winters aditivo - HW

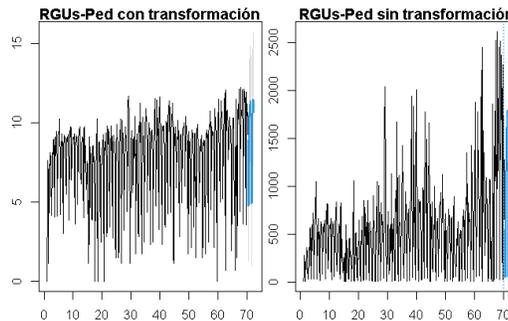


Figura 1: Gráficas con los pronósticos para las RGUs Pedidos HW

Autorregresivo y de medias móviles ARMA(2,1)

Consideramos primero el comportamiento de los residuales presentes en la figura 2 vemos un comportamiento errático que no nos sugiere considerar falta de ajuste o ciclos. En tal caso cumplen el supuesto de ruido blanco, igual la serie al tener tantos datos atípicos si hace que la normalidad no se cumpla en el modelo.

Por otro lado, los pronósticos para este modelo se pueden observar en la figura 3. Su construcción es la misma a la descrita en el modelo HW.

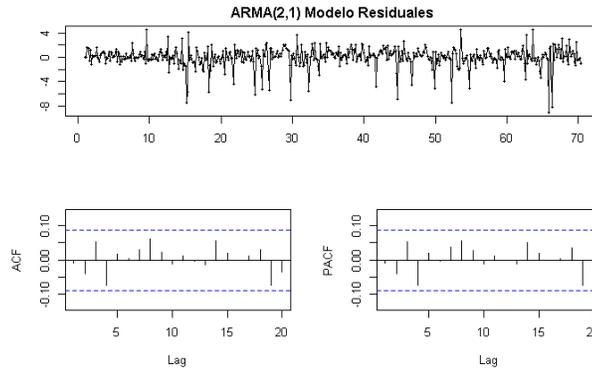


Figura 2: Gráficas de residuales RGUs Pedidos ARMA

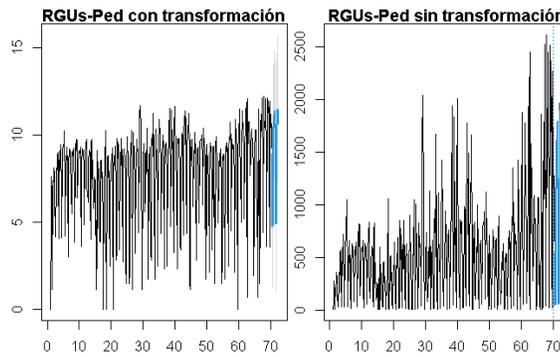


Figura 3: Gráficas con los pronósticos para las RGUs Pedidos ARMA

Por último tenemos los errores de pronóstico

	RMSE	MAE	MAPE
HW	1.05381	0.70694	7.99451
ARMA21	1.2266	1.0434	11.0223

Tabla 5: Error

Para los mejores dos modelos y según los pronósticos observados en las figuras 1 y 3 tenemos en la tabla 5 donde el RMSE (el error estándar mínimo) para modelo HW de 1.05 seguido del modelo ARMA con 1.2266 en RGUs para pedidos, ahora el MAPE (porcentaje medio de error) que comete cada modelo al realizar un pronóstico es de 7.99 en HW y 11.02 en ARMA, en general los resultados para esta base fueron los mejores.

Así, considerando los mejores resultados que hemos venido expresando tenemos un modelo local en este caso HW con los mejores criterios de información, pronósticos y residuos, pero no dejamos el modelo el ARMA(2,1) que sigue al igual en con excelentes resultados por ende, los modelos a escogencia de la empresa puede ser algunos los dos anteriormente mencionados que serán excelentes para pronosticar las RGUs en pedidos que tenemos del tablero de productos B2B.

4.2. RGUs Retiros

Seguimos con el mismo desarrollo expresado para la anterior base de datos, considerando un modelo global y tres modelos locales para el primer caso consideramos un polinomial de grado 3 con indicadoras; en los modelos locales tenemos dos modelos Loess uno lineal y otro cuadrático bajo el criterio GCV y el suavizamiento exponencial de Holt-Winters aditivo. Para la componente estocástica consideramos tres modelos un AR(27), AR(1), ARMA(1,7).

Con este desarrollo no obtuvimos tan buenos resultados como en pedidos. La serie fue complicada de predecir por tanto, solo presentamos un modelo, que según el análisis, presentó los mejores resultados y es el proceso estocástico ARMA(1,7).

Autorregresivo y de medias móviles ARMA(1,7)

Ecuación:

$$E_t = \phi_1 E_{t-1} + a_t + \sum_{i=1}^7 \theta_i a_{t-i}, a_t \sim RB(0, \sigma_a^2).$$

Polinomios AR y MA

$$\phi_1 = 1 - \phi_1 B, \quad \theta_7 = 1 + \sum_{i=1}^7 \theta_i B^i.$$

Los resultados para el modelo fueron

ar(1)	ma(1)	ma(2)	ma(3)	ma(4)	ma(5)	ma(6)	ma(7)
intercept	t	t ²					
X	J	V	S	D	L		
8.4481e-0	-5.1413e-0	-1.8382e-0	-4.8900e-0	1.1857e-0	-2.0668e-0	-4.8780e-0	1.2261e-0
1.8500e+0	-4.1932e-0	1.4194e-0					
1.7359e+0	1.1166e+0	1.1916e+0	-7.1097e+0	-1.2289e+0	6.5771e+0		

Tabla 6: Coeficientes del modelo ARMA

Ahora bajo los **criterios de información CI** presentes en la la tabla 7 y considerando en general fue el segundo mejor modelo en criterios de información, donde hubo una significancia en 10 de los 17 coeficientes presentes de la tabla 6, este así sea el segundo en CI es el más parsimonioso pues cuenta con menos coeficientes.

	HW
AIC	51647.33
BIC	57271.42

Tabla 7: CI

Siguiendo con los **pronósticos** tenemos el gráfico *RGUs-Ret con transformación*, obtenido del desarrollo de la serie. Dado que que utilizamos la transformación Box-Cox, el segundo *RGUs- Ret sin transformaciones* es el resultado obtenido al regresar la serie a la escala original. Los pronósticos se puede ver en color azul. En la figura 4

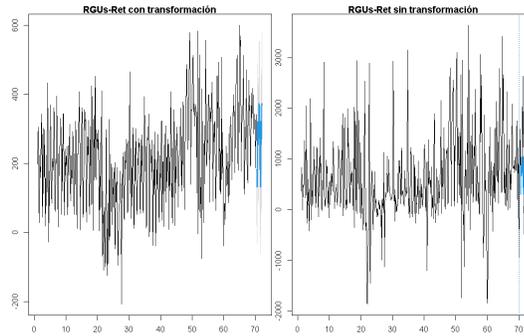


Figura 4: Gráficas con los pronósticos para las RGUs Retiros ARMA

Observando los errores del modelo en la figura 5, tenemos comportamientos erráticos, el supuesto de ruido blanco se cumple. No obstante, así se hayan cambiado algunos datos atípicos, hay datos que no permiten validar la normalidad en estos

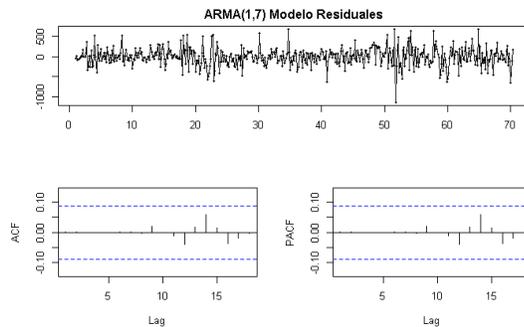


Figura 5: Gráficas de residuales RGUs Retiros ARMA

Por último tenemos los errores de pronóstico

	RMSE	MAE	MAPE
ARMA17	245.3602	160.1671	54.05899

Tabla 8: Error

Para todos los modelos implementados con la data de retiros no fue posible encontrar directamente ajustes, debido a la heteroscedasticidad en los datos. Aunque la transformación y la limpieza de algunos valores atípicos sirvió para que estos modelos se ajustaran mejor, nos da un indicio de que este análisis utilizando series de tiempo para buscar pronósticos no sea el más adecuado, uno porque considerando un modelo que tenga buenos errores donde se cumplan los supuestos el porcentaje de error para el pronóstico es de 54.06 con un error estándar para modelo que puede alejar el valor del pronóstico en la serie de 245.36 RGUs para retiros, ver tabla 8.

Así, considerando los mejores resultados tenemos con un buen criterio de información, pronósticos y residuos el ARMA(1,7), este modelo es relativamente aceptable para pronosticar las RGUs en retiros donde encontramos dificultad de modelar los datos obtenidos del tablero de productos B2B.

4.3. RGUs Pedidos y Retiros

Las RGUs netas al considerar tanto lo que ingresa a la compañía como lo que sale para el área B2B tuvo el mismo desarrollo, excepto que dado que el modelo no nos sugiere análisis local, consideramos el modelo global a partir de un polinomio de grado 2 con indicadores para la componente determinística. Para la componente estocástica consideramos tres modelos un AR(38), ARMA(1,1), ARMA(1,4).

A partir de este desarrollo obtuvimos los mejores ajustes para los procesos ARMA(1,1) y ARMA(1,4).

Autorregresivo y de medias móviles ARMA(1,1)

Ecuación:

$$E_t = \phi_1 E_{t-1} + a_t + \theta_1 a_{t-1}, a_t \sim RB(0, \sigma_a^2).$$

Polinomios AR y MA

$$\phi_1(B) = 1 - \phi_1 B \quad \theta_1(B) = 1 + \theta_1 B$$

Bajo el modelo ARMA(1,1), obtuvimos

ar(1)	ma(1)	intercept	t	t ²	
X	J	V	S	D	L
8.0678e-0	-6.4119e-0	1.7414e+0	5.3297e-0	-5.2836e-0	
1.9323e+0	6.3787e+0	7.4454e+0	-1.2345e+0	-2.1166e+0	-1.0357e+0

Tabla 9: Coeficientes del modelo ARMA(1,1)

Autorregresivo y de medias móviles ARMA(1,4)

Ecuación:

$$E_t = \phi_1 E_{t-1} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \theta_3 a_{t-3} + \theta_4 a_{t-4}.$$

Polinomios AR y MA:

$$\phi_2(B) = 1 - \phi_1 B \quad \theta_4(B) = 1 + \theta_1 B + \theta_2 B^2 + \theta_3 B^3 + \theta_4 B^4$$

Los resultados fueron

ar(1)	ma(1)	ma(2)	ma(3)	ma(4)	intercept	t	t ²
X	J	V	S	D	L		
-1.5893e-0	3.5013e-0	1.7830e-0	1.1070e-0	1.1440e-0	1.7738e+0	5.0297e-0	-4.6979e-0
2.0105e+0	6.3188e+0	7.4199e+0	-1.2384e+0	-2.1205e+0	-1.0402e+0		

Tabla 10: Coeficientes del modelo ARMA(1,4)

Ahora, bajo los **Criterios de Información CI** presentados en la tabla 11, tenemos que las mejores estimaciones se dieron el modelo ARMA(1,1) presentes en la tabla 9 donde podemos ver los resultados de esta estimación para la cual 7 de los 11 coeficientes fueron significativos en el modelo, siguiendo con los resultados el modelo ARMA(1,4) presentes en la tabla 10.

	ARMA(1,1)	ARMA(1,4)
AIC	31646.43	32054.00
BIC	34791.57	36162.13

Tabla 11: CI

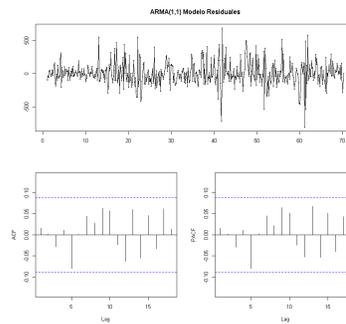


Figura 6: Gráficos de residuales RGUs Pedidos ARMA(1,1)

Consideramos ahora el comportamiento de los residuales presentes en la figura 6 y 7, podemos ver un comportamiento errático en cada modelo no nos sugiere ni falta de ajuste o ciclos, en tal caso cumplen el supuesto de ruido blanco.

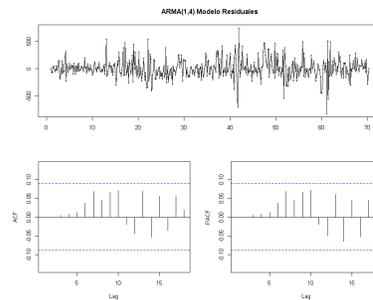


Figura 7: Gráficos de residuales RGUs Pedidos ARMA(1,4)

Ahora para los **pronósticos** tenemos el gráfico *RGUs-PedyRet con transformación*, obtenido del desarrollo de la serie ya que utilizamos la transformación Box-Cox, el segundo *RGUs-PedyRet sin transformaciones* es el resultado de devolver la serie, en color azul se presentan los pronósticos

Observando las figuras 8 y 9 donde están los pronósticos, no se detectan muchas diferencias. Aunque las observaciones a partir de los gráficos son subjetivas se pueden inferir que para los dos modelos los resultados presentes iban a la par, es decir fueron cercanos al pronósticar y esto se vió desde el ajuste.

Por último tenemos los errores de pronóstico

Para todos los modelos implementados con la data de Pedidos y Retiros siguen resultados muy parejos y en especial los capturados en la tabla 12 donde el porcentaje de error para el pronóstico es de 34.00 para un ARMA(1,1) y un 33,25 para un ARMA(1,4) con un error estándar para modelo que puede alejar a los valor del pronóstico en la serie de 200 y 198.86 respectivamente esto en niveles de

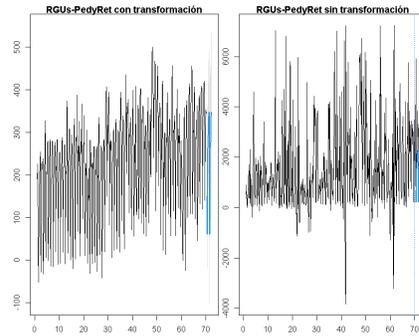


Figura 8: Gráficas con los pronósticos de las RGUs Pedidos y Retiros ARMA(1,1)

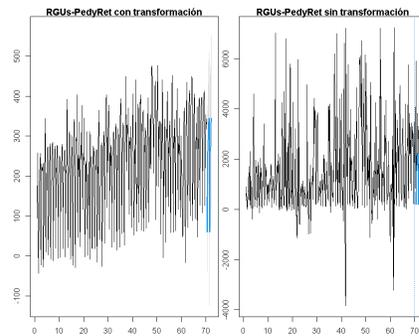


Figura 9: Gráficas con los pronósticos de las RGUs Pedidos y Retiros ARMA(1,4)

	RMSE	MAE	MAPE
ARMA11	200.0084	147.1972	34.00283
ARMA14	198.8615	143.4575	33.25956

Tabla 12: Error

RGUs para pedidos y retiros.

Así, considerando los mejores resultados tenemos dos modelos que complementan resultados y en general son buenos, uno el modelo ARMA(1,1) con mejor criterio de información, dos con buenos errores pero en el pronóstico por un punto está el ARMA(1,4), los modelos quedan a escogencia de la empresa y puede ser algunos los dos que serán buenos para pronosticar las RGUs en pedidos y retiros para el tablero de productos B2B.

5. Conclusiones y Recomendaciones

Los modelos propuestos anteriormente permiten tener un conocimiento previo a nivel general de tres bases de datos importantes para el área de producto, y en sí, son los primeros pasos. Además, se puede

evidenciar que los modelos logran captar el comportamiento en cada serie. En la base de datos para pedidos, por ejemplo, obtuvimos los mejores resultados. Sin embargo, el desbalanceo en la información afecta el desempeño de los modelos. Igual, se usaron estrategias para contrarrestarlo, como transformar los datos, identificar valores atípicos pero se requiere más conocimientos de las bases de datos para mejorar los modelos.

Para trabajos futuros, se recomienda categorizar la información que tiene la base. Así, tenemos la oportunidad de ver más modelos puesto que se nota la importancia y la diferencia por ejemplo de los clientes asociados a los distintos segmentos, y aunque no esté en este trabajo si busqué un modelo solo considerando los Pedidos para los clientes del segmento empresas, principalmente al ver buena información. Con esta categorización, el conocimiento del equipo y el desarrollo ya efectuado, los resultados serán mucho mejores.

Los modelos propuestos anteriormente permiten conocer como es la implementación desde el análisis por series de tiempo. Si se observa en el código se sigue un patrón que puede replicarse verificando los supuestos para los modelos. Por otro lado, se sugiere analizar esta información desde la inteligencia artificial por medio de las redes neuronales y observando que es un tablero que se actualiza permanentemente puede ser más rápida la implementación.

Referencias

- [1] Chan. Kung. D, Cryer. Jonathan. Sik. *Time Series Analysis With Applications in R*. Springer, 2 edition, 2008.
- [2] Luetkepohl. Tommaso, Proietti. Helmut. Does the box-cox transformation help in forecasting macroeconomic time series? *MPRA*, 2011.
- [3] Peter J Bickel and Kjell A Doksum. An analysis of transformations revisited. *American Statistical Association*, 76(374):296–311, 1981.
- [4] Impartido por el profesor Jonatan Andrey González Monsalve. Notas del curso series de tiempo, Semestre 20202.