



Universidad de Granada

**Modelización de Series Temporales modelos
clásicos y SARIMA**

Carlos Miranda Chinlli

Universidad de Granada

Departamento de Estadística e Investigación Operativa

Granada, España 2021

Modelización de Series Temporales modelos clásicos y SARIMA

Carlos Miranda Chinlli

Trabajo de Fin de Máster presentado como requisito parcial para optar al título de:

Máster en Estadística Aplicada

Director:

Doctor, Francisco Javier Alonso Morales

Línea de Investigación:

Análisis de series temporales.

Universidad de Granada

Departamento de Estadística e Investigación Operativa

Granada, España 2021

Agradecimiento

Particularmente deseo agradecer al Doctor Francisco Alonso, docente del Máster y tutor del presente TFM, por los lineamientos que permitieron la satisfactoria elaboración de este trabajo.

Resumen

Actualmente la evidencia científica, particularmente la meteorológica, el aumento de la temperatura, compromete a la integridad de numerosas especies vegetales y animales, constituyen una serie de amenazas para su conservación, además la consecuencia directa será la dificultad de cultivo y reducirá su productividad de alimentos y el aumento del hambre en el mundo. Por tanto, es primordial la predicción de la información meteorológica, para la planificación y toma de decisiones. Específicamente los objetivos fueron: (i) Determinar el modelo SARIMA adecuado para la situación (ii) Evaluar la significancia de dicho modelo (iii) Pronosticar la temperatura de la ciudad de New York.

Los datos utilizados en este estudio es la variable temperatura ambiente en °C de la ciudad de New York, recopilados de la estación meteorológica por el proyecto CRUTEM, es un conjunto de datos derivados de las temperaturas del aire cerca de la superficie terrestre registradas en las estaciones meteorológicas de todos los continentes de la tierra, ha sido desarrollado y mantenido por la Climatic Research Unit desde principios de los años 80, con financiación proporcionada principalmente por el Departamento de Energía de los Estados Unidos. La última versión de CRUTEM se llama CRUTEM4 y está disponible en formato de texto y netCDF en la Climatic Research Unit en el Met Office Hadley Centre. Pero también es posible acceder a CRUTEM4 a través de Google Earth para ello solamente se debe descargar un archivo KML y abrirlo en Google Earth, para realizar una consulta o descarga en cualquier parte del mundo, tan solo basta con hacer clic en el lugar deseado. Los datos son mensuales y se extiende desde 2000 hasta 2019, que incluyen 240 observaciones. Se utilizó la metodología Box - Jenkins, el proceso que se efectuó es el siguiente:

1. Identificación: Este primer paso busca establecer los valores apropiados para p, d, q, P, D, Q .
2. Estimación: Luego se deben estimar los parámetros incluidos en el modelo.
3. Verificación de diagnóstico: Al seleccionar el modelo ARIMA particular se debe comprobar si el modelo se ajusta a los datos (puede existir otro modelo ARIMA que presente mejor ajuste).

4. Predicción: El último plazo consiste en la elaboración de los pronósticos de la serie temporal en particular.

Por lo tanto, se concluye la metodología de Box & Jenkins nos otorga una manera sencilla y eficaz de ajustar series de tiempo de manera rápida y relativamente sencilla para realizar predicciones fiables para el futuro.

El modelo SARIMA (1,0,1) (0,1,1) [12] que se obtuvo, a pesar de ser un modelo apropiado para series de tiempo, es útil en el corto plazo, existe problemas debido a su falla en la predicción, especialmente si en el pasado la secuencia de series de tiempo tiene cambios anormales. En la predicción el error promedio (ME) es de 0.03 y el error absoluto medio (MAE) es de 1.33 esto significa que es un buen modelo de pronóstico.

Índice

Resumen.....	v
Introducción.....	1
1. Fundamentación sobre las series temporales.....	2
1.1 Objetivo general.....	2
1.2 Objetivo específico.....	2
1.3 Definiciones fundamentales.....	2
1.4 Procesos lineales generales.....	6
1.5 Procesos autorregresivos.....	6
1.6 Procesos de media móvil.....	11
1.7 Propiedades y funciones de la autocorrelación.....	14
1.8 Modelos ARMA, ARIMA Y SARIMA.....	22
2. Metodología Box – Jenkins.....	30
2.1 Identificación.....	31
2.2 Estimación.....	34
2.2.1 Estimación de máxima verosimilitud.....	34
2.2.2 Método de los mínimos cuadrados condicionales.....	36
2.2.3 Mínimos cuadrados incondicionales.....	36
2.2.4 Métodos de optimización no lineal.....	37
2.3 Diagnóstico del modelo.....	37
2.3.1 Diagnóstico de los coeficientes estimados.....	37
2.3.2 Diagnóstico de los residuos.....	38
2.4 Pronóstico.....	39
3. Aplicación usando el software R.....	41
3.1 Datos.....	41
3.2 Identificación del modelo.....	43
3.3 Estimación.....	47
3.5 Pronóstico.....	55
Conclusión.....	56
Bibliografía.....	57
ANEXO 1: Código R.....	58
ANEXO 2: Datos de temperatura ambiente de New York.....	59

Introducción

El análisis de Box Jenkins se refiere a un método sistemático de identificación, ajuste, verificación y uso integrado de modelos de series de tiempo de media móvil autorregresivo (ARIMA). El método es apropiado para series de tiempo de longitud media a larga (al menos 50 observaciones). Muchas series de tiempo encontradas en la práctica exhiben un comportamiento no estacionario. Por lo general, la no estacionariedad se debe a una tendencia, un cambio en la media local o variación estacional. Dado que la metodología, es solo para la aplicación de modelo estacionario tenemos que hacer algunos ajustes antes de que podamos modelar estas series no estacionarias. El modelo SARIMA se basa en la aplicación de los modelos ARMA a una serie temporal transformada donde se ha eliminado el comportamiento estacional y no estacionario, para una serie estacional con s periodos por año se utilizan los modelos SARIMA $(p, d, q) \times (P, D, Q)_s$.

El objetivo principal del siguiente trabajo es: elaborar un modelo SARIMA para pronóstico aplicando la metodología Box-Jenkins, de la variable temperatura de la ciudad de New York, utilizando datos históricos de series de tiempo. Si bien es evidencia científica, particularmente evidencia meteorológica, de que nos enfrentamos a un calentamiento global en el planeta que requiere una acción decisiva por parte del ser humano, creemos que debemos extremar las precauciones a la hora de investigar las causas de un cambio climático global que es posible que esté presenciando que algunos califican como únicos e irreversibles en la historia de nuestro planeta. Por tanto, el análisis de los datos de temperatura es primordial para la predicción de la información meteorológica, también para la planificación y toma de decisiones, ya que no solo afecta el crecimiento, reproducción de plantas y animales, además también influye en casi todas las otras variables meteorológicas, como la tasa de evaporación, la humedad relativa, la velocidad del viento, dirección del viento y patrones de precipitación.

Finalmente se realiza una aplicación práctica del modelo estudiado utilizando el software R Project, dicha aplicación se realiza con los datos utilizados en este estudio es de la variable temperatura ambiente en °C de la ciudad de New York, recopilados de la estación meteorológica por el proyecto CRUTEM.

1. Fundamentación sobre las series temporales

1.1 Objetivo general

Presentar la metodología de Box-Jenkins para modelos SARIMA, que ayude a pronosticar la temperatura de la ciudad de New York, utilizando datos históricos de series de tiempo, recopilados de la estación meteorológica por el proyecto CRUTEM. Para lograr este objetivo realizamos los siguientes pasos:

1.2 Objetivo específico

- Determinar el modelo SARIMA adecuado para la situación.
- Evaluar la significancia de dicho modelo.
- Pronosticar la temperatura de la ciudad de New York.

1.3 Definiciones fundamentales

Series de Tiempo

Por serie de tiempo nos referimos a datos estadísticos que se recopilan, observan o registran en intervalos de tiempo regulares (diario, semanal, semestral, anual). Dado que los acontecimientos pasados pueden tener influencia sobre acontecimientos futuros el tiempo es un parámetro importante en los conjuntos de series temporales. La serie temporal surge en muchos campos como: Meteorología, Economía y Finanzas, Medicina, Ingeniería, Control de Calidad, Modelización Ambiental, Hidrología y Demografía.:

La teoría clásica considera una serie de tiempo formado por cuatro componentes teóricas: tendencia, variaciones estacionales, variaciones cíclicas y variaciones residuales.

Tendencia

Son movimientos a largo plazo en la serie temporal sin efectos relacionados con el calendario y sin regularidad, y es un reflejo del nivel subyacente. Es el resultado de influencias como el crecimiento de la población, la inflación de precios y los cambios económicos generales. Es la componente de largo plazo que constituye la base del crecimiento o declinación de una serie histórica.

Efectos estacionales

Las fluctuaciones estacionales se encuentran en los datos clasificados por periodos de tiempo. La variación estacional se refiere a un patrón de cambio, regularmente recurrente a través del tiempo.

La componente estacional consiste en efectos que son razonablemente estables con respecto al tiempo, dirección y magnitud. Surge de influencias sistémicas, relacionadas con el calendario, tales como: condiciones naturales, trámites comerciales y administrativos y comportamiento social y cultural. Definen un patrón que se repite de forma aproximada pasadas “s” observaciones.

Ciclos

Conjunto de fluctuaciones en forma de onda o ciclos, de más de un “año “de duración, producidos por cambios en las condiciones económicas o naturales.

Variaciones irregulares (aleatorias)

El componente irregular (a veces también conocido como los residuales) es lo que queda después que los componentes estacionales y de tendencia se hayan estimado y eliminado series. Es el resultado de fluctuaciones a corto plazo. En la serie no son ni sistemáticos ni predecibles. Es una serie irregular. Estas fluctuaciones pueden dominar los movimientos, que enmascararán la tendencia y estacionalidad.

Proceso estocástico

Un proceso estocástico $\{X_t\}$ es una colección de variables aleatorias donde el parámetro t pertenece a un espacio de parámetros (o índice) T .

$$\{X_t\} = (\dots X_1, X_2, \dots, X_n, \dots).$$

El conjunto de valores posibles que puede asumir cada X_t se denomina comúnmente espacio de estados S . El espacio de estados puede ser discreto o continuo y también multidimensional.

Procesos estacionarios

Estacionariedad

Hacer inferencia estadística sobre la estructura de un proceso estocástico sobre la base de un registro observado de ese proceso, generalmente debemos hacer algunas

suposiciones simplificadoras (y presumiblemente razonables) acerca de esa estructura. La más importante es la estacionariedad. La idea básica de estacionariedad es que las distribuciones de probabilidad, las leyes que gobiernan el comportamiento del proceso, no cambian con el tiempo. En cierto sentido, el proceso está en equilibrio estadístico. Específicamente, se dice que un proceso $\{Y_t\}$ es estrictamente estacionario si la distribución conjunta de $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ es la misma que la distribución conjunta de $Y_{t_1-k}, Y_{t_2-k}, \dots, Y_{t_n-k}$ para todas las opciones de los puntos de tiempo t_1, t_2, \dots, t_n y todas las opciones de retardo de tiempo k .

La estacionariedad en sentido fuerte se debe contrastar por medio de distribuciones conjuntas para cualquier selección de variables del proceso, por ello se procede a definir la estacionariedad débil, la cual es más fácil comprobación.

Estacionariedad débil. Un proceso estocástico $\{X_t: t \in T\}$ de segundo orden se dice débilmente estacionario o estacionario en sentido amplio si, $\forall t \in T$ y, $\forall h \in \mathbb{R}$ tal que $t + h \in T$ se verifica:

$$E[X_t] = \mu \text{ (es constante),}$$

$$Cov(X_t, X_{t+h}) = \gamma_{t,t+h} = \gamma_h \text{ (depende de sólo de } h\text{).}$$

Un proceso estrictamente estacionario y con momentos de segundo orden finitos es débilmente estacionario. El recíproco no es cierto, debido a que la estacionariedad débil no impone restricción alguna sobre la distribución de las variables ni los momentos de orden superior a dos, sin embargo, en el caso de los procesos Gaussianos los dos tipos de estacionariedad equivalen.

Un concepto importante que se puede encontrar dentro del ámbito de los procesos estocásticos, es el de los procesos estacionarios. De una manera informal, una serie es estacionaria en sentido amplio o débil cuando se encuentra en equilibrio estadístico. Esto equivale a decir que la media y la varianza se mantienen constantes a lo largo del tiempo y, por lo tanto, la serie no presenta tendencias, que son los cambios a largo plazo de la media.

Ruido blanco

El ruido blanco o puramente aleatorio representa un buen ejemplo de proceso estacionario, ya que es una secuencia de variables aleatorias mutuamente ortogonales con media cero, varianza constante y covarianzas nulas entre valores correspondientes a distintos valores de t . El ruido blanco se utiliza en el análisis de series temporales para representar las innovaciones en una serie, ya que modela a la perfección el carácter impredecible de las variaciones aleatorias de una serie. Como ejemplo se puede ver en la figura 1.

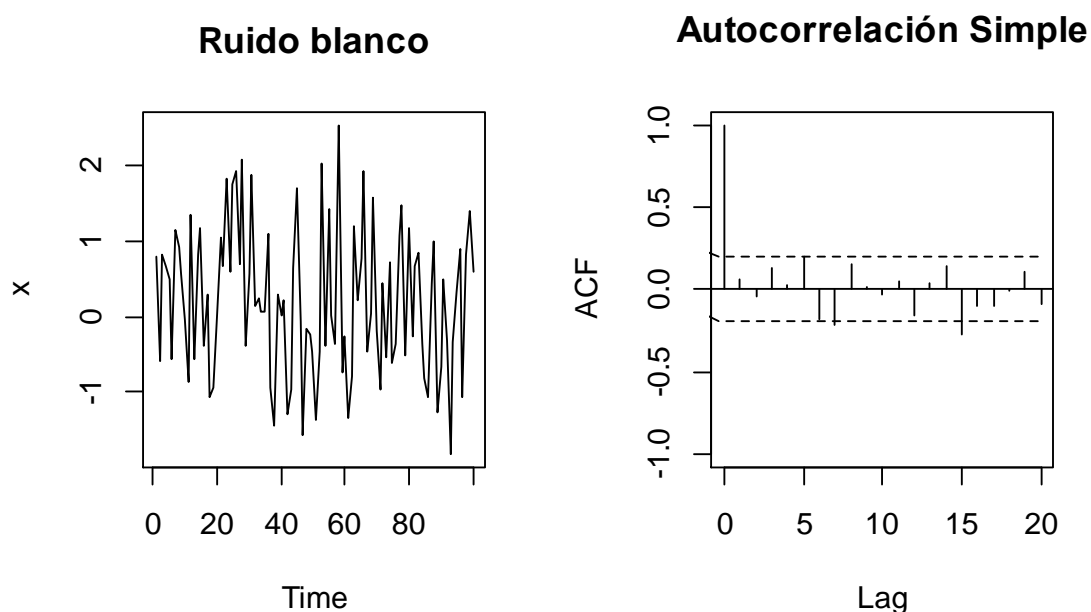


Fig. 1 Simulación de ruido blanco

Camino aleatorio

Se define camino aleatorio como un proceso estocástico X_t , donde $X_t = X_{t-1} + e_t$ y al obtener su primera diferencia se tiene $X_t - X_{t-1} = e_t$ es decir $\nabla X_t = e_t$, siendo este resultado un ruido blanco.

Sea $\{X_t; t \in T\}$ un proceso estocástico de segundo orden, se define la función de autocovarianza de X como:

$$\gamma_{t,t+h} = \text{cov}(X_t, X_{t+h}) = E[(X_t - \mu_t)(X_{t+h} - \mu_{t+h})].$$

Se define la función de autocorrelación de X como:

$$\rho_{t,t+h} = \frac{\gamma_{t,t+h}}{\sqrt{\gamma_{t,t}\gamma_{t+h,t+h}}} = \frac{\gamma_{t,t+h}}{\sigma_t\sigma_{t+h}}.$$

La función de autocorrelación es una medida adimensional de la dependencia lineal entre variables aleatorias de un proceso estocástico. En el caso estacionario las dos funciones dependen de h y la función de autocorrelación presenta las siguientes propiedades:

1. $\rho_0 = 1$
2. $-1 \leq \rho_h \leq 1 \quad \forall h$
3. $\rho_h = \rho_{-h}$ es una función par.

1.4 Procesos lineales generales

Sea $\{Y_t\}$ que denota la serie de tiempo observada. De aquí en adelante también notaremos $\{e_t\}$ representan una serie de ruido blanco no observada, es decir, una secuencia de variables aleatorias independientes distribuidas de forma idéntica, de media cero y varianza constante. Para gran parte de nuestro trabajo, la suposición de independencia podría ser reemplazado por el supuesto más débil de que el $\{e_t\}$ son variables aleatorias no correlacionadas, pero no buscaremos esa ligera generalidad.

Un proceso lineal general, $\{Y_t\}$, se puede representar como una combinación lineal ponderada de términos de ruido blanco presentes y pasados como:

$$Y_t = e_t + \psi_1 e_{t-1} + \psi_2 e_{t-2} + \dots \quad (1.1)$$

Si el lado derecho de esta expresión es verdaderamente una serie infinita, entonces ciertas condiciones debe colocarse en los pesos ψ para que el lado derecho sea significativo matemáticamente. Para nuestros propósitos, basta con suponer que:

$$\sum_{i=1}^{\infty} \psi_i^2 < \infty. \quad (1.2)$$

También debemos tener en cuenta que desde $\{e_t\}$ es inobservable, no hay pérdida en la generalidad de anterior si asumimos que el coeficiente de e_t es 1; efectivamente, $\psi_0 = 1$.

1.5 Procesos autorregresivos

Los procesos autorregresivos son, como sugiere su nombre, regresiones sobre sí mismos. Específicamente, un proceso autorregresivo de orden p $\{Y_t\}$ satisface la ecuación:

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t. \quad (1.3)$$

El valor actual de la serie Y_t es una combinación lineal de los p valores pasados más recientes de sí mismo más un término de innovación e_t que incorpora todo lo nuevo de la serie en tiempo t que no se explica por los pasados. Por lo tanto, para cada t asumimos que e_t es independiente de $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$. Yule (1926) llevó a cabo el trabajo original en procesos autorregresivos.

El proceso autorregresivo de primer orden

Nuevamente, es instructivo considerar en detalle el modelo de primer orden, abreviado AR (1). Suponga que la serie es estacionaria y satisface

$$Y_t = \phi Y_{t-1} + e_t. \quad (1.4)$$

Donde hemos eliminado el subíndice 1 del coeficiente ϕ por simplicidad, asumimos que la media del proceso se ha restado, de modo que la media de la serie es cero. Las condiciones de estacionariedad se consideran más adelante, primero tomamos las varianzas de ambos lados de la ecuación (1.4) y obtenemos

$$Y_0 = \phi^2 Y_0 + \sigma_e^2.$$

Resolviendo los rendimientos para Y_0

$$Y_0 = \frac{\sigma_e^2}{1-\phi^2}. \quad (1.5)$$

Obsérvese que como $Y_0 > 0$, $\phi^2 < 1$ es decir $-1 < \phi < 1$, para que sea estacionaria. Ahora toma la ecuación (1.5) multiplique ambos lados por Y_{t-k} ($k = 1, 2, \dots$) y tome los valores esperados, usando la independencia,

$$E(e_t Y_{t-k}) = E(e_t)E(Y_{t-k}) = 0.$$

Entonces

$$\gamma_k = \phi \gamma_{k-1} \quad \text{para } k = 1, 2, 3, \dots \quad (1.6)$$

Para $k = 1$ obtenemos $\gamma_1 = \phi\gamma_0 = \phi\sigma_e^2 / (1 - \phi^2)$. Con $k = 2$, obtenemos $\gamma_2 = \phi\gamma_1 = \phi^2\sigma_e^2 / (1 - \phi^2)$. Ahora es fácil ver que en general

$$\gamma_k = \phi^k \frac{\sigma_e^2}{1 - \phi^2} . \quad (1.7)$$

Por lo tanto

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi^k \quad \text{para } k = 1, 2, 3. \quad (1.8)$$

La magnitud de la función de autocorrelación disminuye exponencialmente a medida que aumenta el número de rezagos, k . Si $\phi > 0$, todas las correlaciones son positivas; si $\phi < 0$, la autocorrelación de retraso 1 es negativa ($\rho_1 = \phi$) y los signos de sucesivos las autocorrelaciones; alternan de positivas a negativas, con magnitudes decrecientes exponencialmente. Se muestra en la Fig. 2, el gráfico de autocorrelación de una serie AR(1) simulada con $\phi > 0$.

Autocorrelacion Simple

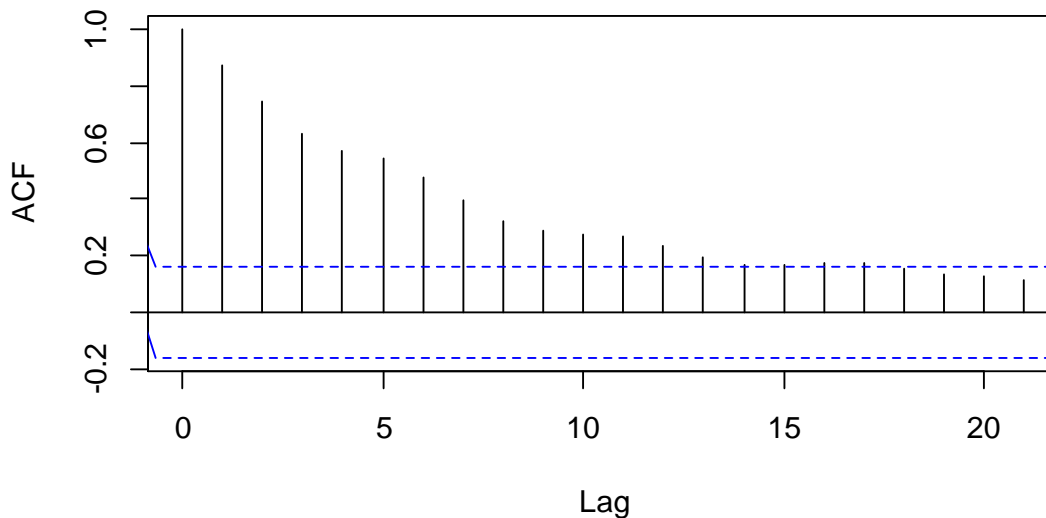


Fig. 2 Función de autocorrelación, AR(1) “simulado”.

El proceso autorregresivo de segundo orden

Ahora considere la serie que satisface

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + e_t, \quad (1.9)$$

donde, como siempre, asumimos que e_t es independiente de $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$. Para discutir estacionariedad, introducimos el polinomio característico AR

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2,$$

y la correspondiente ecuación característica AR

$$1 - \phi_1 x - \phi_2 x^2 = 0.$$

Recordamos que una ecuación cuadrática siempre tiene dos raíces (posiblemente complejas).

Estacionariedad del Proceso AR(2)

Puede demostrarse que, sujeto a la condición de que e_t es independiente de $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$ existe una solución estacionaria, si y solo si las raíces de AR ecuación característica excede 1 en valor absoluto (caso complejo). A veces decimos que las raíces deben estar fuera del círculo unitario en el plano complejo. Esta declaración se generalizará al caso de orden p de forma inmediata.

Por estacionariedad requerimos que estas raíces superen 1 en valor absoluto, ahora será cierto si y solo si se cumplen tres condiciones:

$$\phi_1 + \phi_2 < 1, \quad \phi_2 - \phi_1 < 1, \quad \text{y} \quad |\phi_2| < 1.$$

El proceso autorregresivo general

Considere ahora el modelo autorregresivo de orden p

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t. \quad (1.10)$$

Con AR polinomio característico:

$$\phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p, \quad (1.11)$$

y la correspondiente ecuación característica de AR

$$1 - \phi_1 x - \phi_2 x^2 - \dots - \phi_p x^p. \quad (1.12)$$

Como se señaló anteriormente, asumiendo que e_t es independiente de $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots$ existe una solución estacionaria de la ecuación de modelo si y solo si,

las p raíces de la ecuación característica AR, exceden de 1 en valor absoluto (módulo, en el caso complejo). Se pueden usar otras relaciones entre raíces polinomiales y coeficientes para mostrar que las siguientes dos desigualdades son necesarias para la estacionariedad. Es decir, para que las raíces sean mayores que 1 en módulo, es necesario, pero no suficiente, que ambos

$$\phi_1 + \phi_2 + \dots + \phi_p < 1 \quad \text{y} \quad |\phi_p| < 1. \quad (1.13)$$

Suponiendo estacionariedad y medias cero, podemos multiplicar la ecuación (1.10) por Y_{t-k} tomar esperanzas, dividir por γ_0 y obtener la importante relación recursiva

$$\rho_k = \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} + \phi_3 \rho_{k-3} + \dots + \phi_p \rho_{k-p} \quad \text{para } k \geq 1 \quad (1.14)$$

Poniendo $k = 1, 2, \dots, p$ en la ecuación (1.14) y usando $\rho_0 = 1$ y $\rho_{-k} = \rho_k$ obtenemos las ecuaciones generales de Yule-Walker.

$$\begin{aligned} \rho_1 &= \phi_1 + \phi_2 \rho_1 + \phi_3 \rho_2 + \dots + \phi_p \rho_{p-1} \\ \rho_2 &= \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_1 + \dots + \phi_p \rho_{p-2} \\ &\vdots \\ \rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \phi_3 \rho_{p-3} + \dots + \phi_p. \end{aligned} \quad (1.15)$$

Dados valores numéricos para $\phi_1, \phi_2, \dots, \phi_p$, estas ecuaciones lineales se pueden resolver como obtener valores numéricos para $\rho_1, \rho_2, \dots, \rho_p$ entonces la ecuación (1.14) se puede utilizar para obtener valores numéricos para ρ_k en cualquier número de rezagos más altos. Señalando que:

$$E(e_t Y_t) = E[e_t (\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + e_t)] = E(e_t^2) = \sigma_e^2,$$

podemos multiplicar la ecuación (1.10) por Y_t , tome esperanzas y se encuentra:

$$\gamma_0 = \phi_1 \gamma_1 + \phi_2 \gamma_2 + \dots + \phi_p \gamma_p + \sigma_e^2,$$

que, usando $\rho_k = \gamma_k / \gamma_0$, se puede escribir como:

$$\gamma_0 = \frac{\sigma_e^2}{1 - \phi_1 \gamma_1 - \phi_2 \gamma_2 - \dots - \phi_p \gamma_p} \quad (1.16)$$

y expresar la varianza del proceso γ_0 en términos de los parámetros, $\sigma_e^2, \phi_1, \phi_2, \dots, \phi_p$, y los valores ahora conocidos de $\gamma_1, \gamma_2, \dots, \gamma_p$. Por supuesto, las soluciones explícitas para ρ_k son esencialmente imposible en esta generalidad, pero podemos decir que γ_k será una combinación lineal de términos que decaen exponencialmente (correspondientes a las raíces reales de la ecuación característica) y términos de onda sinusoidal amortiguada (correspondientes a las raíces complejas de la ecuación característica).

1.6 Procesos de media móvil

Se dice que Y_t tiene una estructura MA(q) si se puede expresar según la ecuación:

$$Y_t = e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}. \quad (1.17)$$

Llamamos a esta serie una media móvil de orden q. La terminología media móvil surge del hecho de que Y_t se obtiene aplicando los pesos $1, -\theta_1, -\theta_2, \dots, -\theta_q$ a las variables $e_t, e_{t-1}, e_{t-2}, \dots, e_{t-q}$ y luego los pesos móviles y aplicando $e_{t+1}, e_t, e_{t-1}, \dots, e_{t-q+1}$ para obtener Y_{t+1} y así sucesivamente.

El proceso de media móvil de primer orden

Consideramos en detalle el simple pero importante proceso de media móvil de orden 1, es decir, la serie MA(1). El modelo es $Y_t = \mu + e_t - \theta e_{t-1}$. Ya que solo una θ está involucrada, descartamos el subíndice redundante 1. Claramente $E(Y_t) = \mu$ y $Var(Y_t) = \sigma_e^2(1 + \theta^2)$ ahora

$$\begin{aligned} Cov(Y_t, Y_{t-1}) &= Cov(e_t - \theta e_{t-1}, e_{t-1} - \theta e_{t-2}) = \\ &= Cov(-\theta e_{t-1}, e_{t-1}) = -\theta \sigma_e^2, \end{aligned}$$

Y

$$Cov(Y_t, Y_{t-2}) = Cov(e_t - \theta e_{t-1}, e_{t-2} - \theta e_{t-3}) = 0,$$

ya que no hay e 's con subíndices en común entre Y_t y Y_{t-2} . Del mismo modo, cuando $k \geq 2$ será $Cov(Y_t, Y_{t-k}) = 0$, es decir, el proceso no tiene correlación más allá del retraso 1. Este hecho será importante cuando necesitemos elegir modelos con datos reales. En resumen, para un modelo MA(1) $Y_t = \mu + e_t - \theta e_{t-1}$,

$$E(Y_t) = \mu$$

$$\begin{aligned}\gamma_0 &= \text{Var}(Y_t) = \sigma_e^2(1 + \theta^2) \\ \gamma_1 &= -\theta\sigma_e^2 \\ \rho_1 &= (-\theta) / (1 + \theta^2) \\ \gamma_k &= \rho_k = 0 \text{ para } k \geq 2.\end{aligned}\tag{1.18}$$

La Fig. 3, presenta la función de autocorrelación muestral de una serie simulada con estructura MA(1)

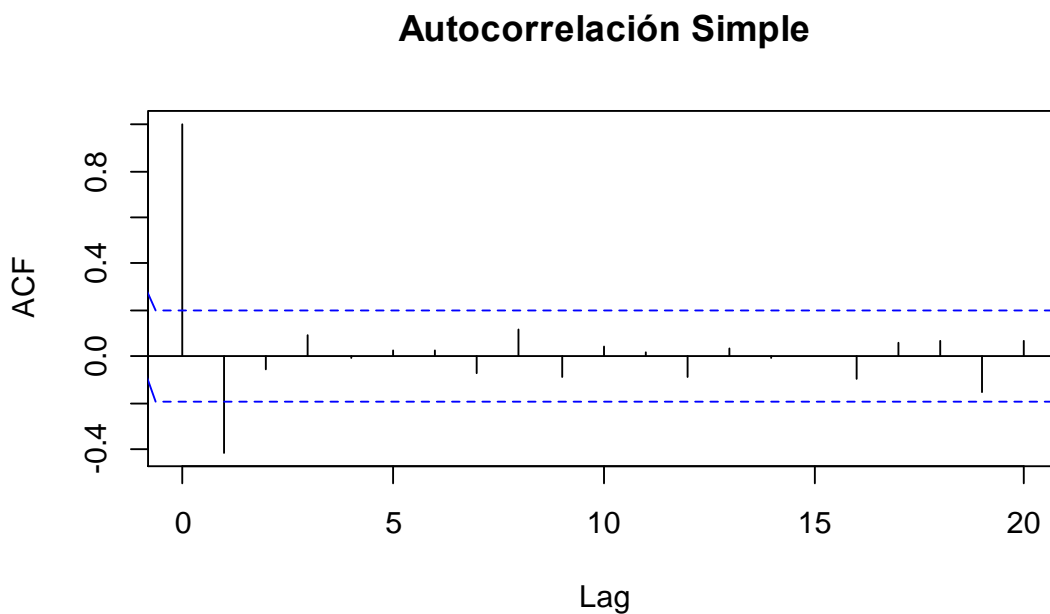


Fig. 3 Función de autocorrelación simple, MA(1) “simulado”.

El proceso de media móvil de segundo orden

Considere el proceso de promedio móvil de orden 2:

$$Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}$$

Aquí

$$\begin{aligned}\gamma_0 &= \text{Var}(Y_t) = \text{Var}(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}) = (1 + \theta_1^2 + \theta_2^2)\sigma_e^2 \\ \gamma_1 &= \text{Cov}(Y_t, Y_{t-1}) = \text{Cov}(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, e_{t-1} - \theta_1 e_{t-2} - \theta_2 e_{t-3}) \\ &= \text{Cov}(-\theta_1 e_{t-1}, e_{t-1}) + \text{Cov}(-\theta_1 e_{t-2}, -\theta_2 e_{t-2}) \\ &= [-\theta_1 + (-\theta_1)(-\theta_2)]\sigma_e^2 \\ &= (-\theta_1 + \theta_1\theta_2)\sigma_e^2,\end{aligned}$$

Y

$$\begin{aligned}\gamma_2 &= \text{Cov}(Y_t, Y_{t-2}) = \text{Cov}(e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2}, e_{t-2} - \theta_1 e_{t-3} - \theta_2 e_{t-4}) \\ &= \text{Cov}(-\theta_2 e_{t-2}, e_{t-2}) \\ &= -\theta_2 \sigma_e^2.\end{aligned}$$

Así, para un proceso MA(2):

$$\rho_1 = \frac{-\theta_1 + \theta_1 \theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_2 = \frac{-\theta_2}{1 + \theta_1^2 + \theta_2^2}$$

$$\rho_k = 0 \text{ para } k = 3, 4, \dots$$

La Fig. 4, presenta la función de autocorrelación muestral de una serie simulada con estructura MA(2)

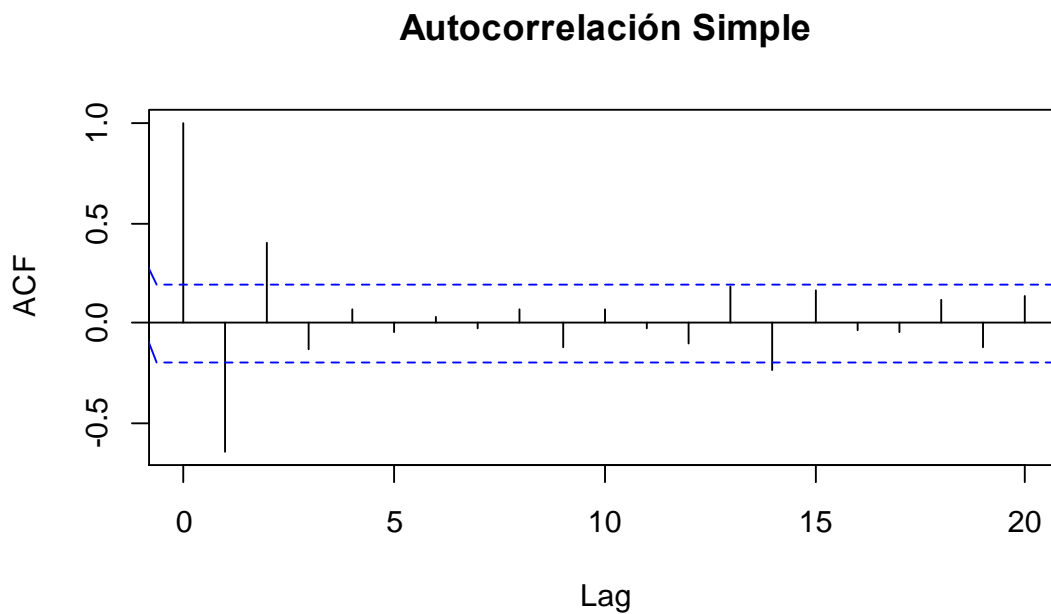


Fig. 4 Función de autocorrelación simple, MA(2) "simulado".

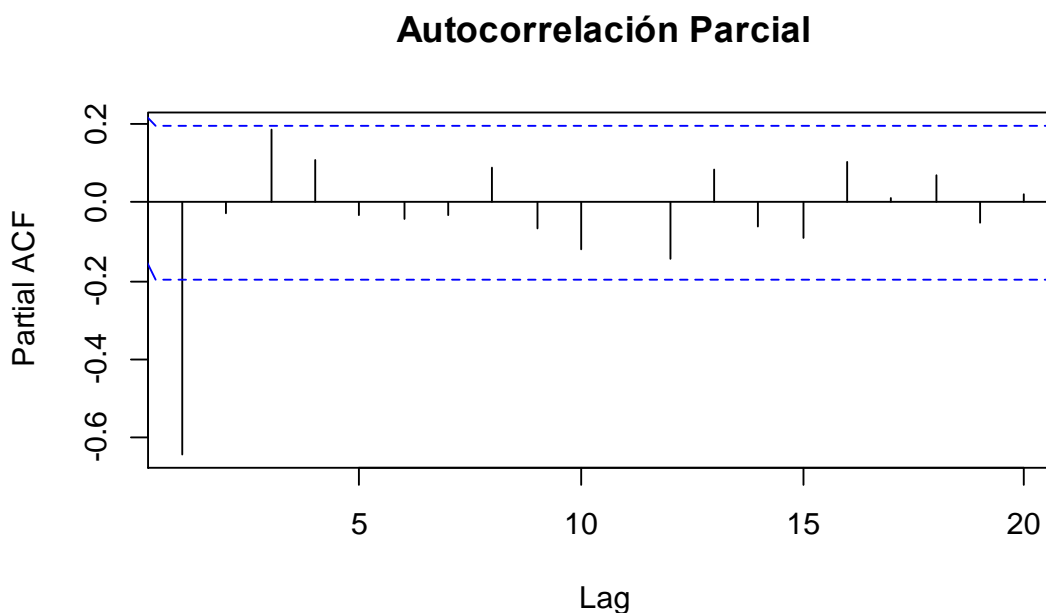


Fig. 5 Función de autocorrelación parcial, MA(2) "simulado".

El proceso general MA(q)

Para el proceso general MA $Y_t = \mu + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}$, cálculos similares muestran que:

$$\gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2) \sigma_e^2, \quad (1.19)$$

y

$$\rho_k = \begin{cases} \frac{-\theta_k + \theta_1 \theta_{k+1} + \theta_1 \theta_{k+2} + \dots + \theta_{q-k} \theta_q}{1 + \theta_1^2 + \theta_2^2 + \dots + \theta_q^2} & \text{para } k = 1, 2, \dots, q \\ 0 & \text{para } k > q \end{cases} \quad (1.20)$$

Donde el numerador de ρ_q es simplemente $-\theta_q$. La función de autocorrelación se "corta" después del retraso q es decir, es cero. Su forma puede ser casi cualquier cosa para los rezagos anteriores.

1.7 Propiedades y funciones de la autocorrelación

Propiedades de la función de autocorrelación muestral

La definición de la muestra o función de autocorrelación muestral estimada, para la serie observada Y_1, Y_2, \dots, Y_n es

$$r_k = \frac{\sum_{t=k+1}^n (Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^n (Y_t - \bar{Y})^2} \quad \text{para } k = 1, 2, \dots \quad (1.21)$$

Nuestro objetivo es reconocer, en la medida de lo posible, patrones en r_k que son característicos de los patrones conocidos en ρ_k para modelos ARMA comunes. Por ejemplo, sabemos que $\rho_k = 0$ para $k > q$ en un modelo MA (q). Sin embargo, como r_k son solo estimaciones de ρ_k necesitamos investigar sus propiedades de muestreo para facilitar la comparación de estimaciones correlaciones con correlaciones teóricas.

A partir de la definición de r_k , una relación de funciones cuadráticas de variables posiblemente dependientes, debería ser evidente que las propiedades de muestreo de r_k no se obtendrán fácilmente para tamaños muestrales pequeños. Incluso el valor esperado de r_k es difícil de determinar; recuerde que el valor esperado de una razón no es la razón de los respectivos valores esperados. Aceptaremos un resultado general para muestra grande y sus implicaciones en casos especiales. Bartlett (1946), realizó el trabajo original. Tomaremos un resultado más general de Anderson (1971). Una discusión reciente de estos resultados se puede encontrar en Shumway y Stoffer (2006, p. 519)

Suponemos que

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j e_{t-j},$$

Donde el e_t son independientes e idénticamente distribuidos con medias cero y varianzas comunes finitas, distintas de cero. Suponemos además que

$$\sum_{j=0}^{\infty} |\psi_j| < \infty \quad \text{y} \quad \sum_{j=0}^{\infty} j \psi_j^2 < \infty.$$

(Estos serán satisfechos por cualquier modelo ARMA estacionario). Entonces, para cualquier m fijo, la distribución conjunta de

$$\sqrt{n}(r_1 - \rho_1), \sqrt{n}(r_2 - \rho_2), \dots, \sqrt{n}(r_m - \rho_m),$$

Se aproxima, a una distribución normal conjunta con medias cero, varianzas c_{ij} , y covarianzas c_{ij} , donde

$$c_{ij} = \sum_{k=-\infty}^{\infty} (\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} + 2\rho_i\rho_j\rho_k)^2. \quad (1.22)$$

Para n grande, diríamos que r_k se distribuye aproximadamente normalmente con una media ρ_k

y varianza c_{kk}/n . Además, $Corr(r_k, r_j) \approx c_{kj}/\sqrt{c_{kk}c_{jj}}$. Observe que la varianza aproximada de r_k es inversamente proporcional al tamaño de la muestra, pero $Corr(r_k, r_j)$ es aproximadamente constante para n grande.

Dado que la ecuación (1.22) es claramente difícil de interpretar en su generalidad actual, considerará algunos casos especiales importantes y simplificaciones. Supongamos primero que

$\{Y_t\}$ es ruido blanco. Entonces la ecuación (1.22) se reduce considerablemente y obtenemos:

$$Var(r_k) \approx \frac{1}{n} \quad y \quad Corr(r_k, r_j) \approx 0 \quad para \quad k \neq j. \quad (1.23)$$

A continuación, suponga que $\{Y_t\}$ es generado por un proceso AR (1) con $\rho_k = \Phi^k$ para $k > 0$.

Luego, de considerar el álgebra y de sumar varias series geométricas, la ecuación

(1.22) con $i = j$ produce

$$Var(r_k) \approx \frac{1}{n} \left[\frac{(1+\phi^2)(1-\phi^{2k})}{1-\phi^2} - 2k\phi^{2k} \right], \quad (1.24)$$

en particular,

$$Var(r_1) \approx \frac{1-\phi^2}{n}. \quad (1.25)$$

Observe que cuanto más cerca esté ϕ de ± 1 , más precisa será nuestra estimación de ρ_1 ($= \phi$).

Para grandes rezagos, los términos de la ecuación (1.24) que involucran ϕ^k puede ser ignorado, y tenemos

$$Var(r_k) \approx \frac{1}{n} \left[\frac{1+\phi^2}{1-\phi^2} \right] \quad para \quad k \quad grande. \quad (1.26)$$

Observe que aquí, en contraste con la ecuación (1.25), los valores de ϕ cercanos a ± 1 implican grandes variaciones para r_k . Por lo tanto, no deberíamos esperar estimaciones

tan precisas de $\rho_k = \phi^k \approx 0$ para k grande como lo hacemos de $\rho_k = \phi^k$ para k pequeños.

Para el modelo AR (1), la ecuación (1.22) también se puede simplificar (después de mucha álgebra) por lo general $0 < i < j$ como

$$c_{ij} = \frac{(\phi^{j-i} - \phi^{j+1})(1 + \phi^2)}{1 - \phi^2} + (j - 1)\phi^{j-i} - (j + i)\phi^{j+1}. \quad (1.27)$$

En particular encontramos

$$Corr(r_1, r_2) \approx 2\phi \sqrt{\frac{1 - \phi^2}{1 + 2\phi^2 - 3\phi^4}}. \quad (1.28)$$

Las funciones de autocorrelación parcial

Dado que para los modelos MA (q) la función de autocorrelación es cero para rezagos más allá de q, la autocorrelación de la muestra es un buen indicador del orden del proceso. Sin embargo, las relaciones de autocorrelación de un modelo AR (p) no se vuelven cero después de un cierto número de retardos. Entonces se necesita una función diferente para ayudar a determinar el orden de modelos autorregresivos. Tal función puede definirse como la correlación entre Y_t y Y_{t-k} después de eliminar el efecto de las variables intermedias $Y_{t-1}, Y_{t-2}, Y_{t-3}, \dots, Y_{t-k+1}$. Este coeficiente se llama autocorrelación parcial en el rezago k y se denotará por ϕ_{kk} .

Hay varias formas de precisar esta definición. Si Y_t es una serie de tiempo distribuida normalmente, podemos dejar

$$\phi_{kk} = Corr(Y_t, Y_{t-k} | Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}). \quad (1.29)$$

Es decir, ϕ_{kk} es la correlación en la distribución bivariada de Y_t y Y_{t-k} condicionado a $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$.

Un enfoque alternativo, no basado en la normalidad, se puede desarrollar en lo siguiente. Considere la posibilidad de predecir Y_t basado en una función lineal de las variables intervinientes $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$. Es decir $\beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_{k-1} Y_{t-k+1}$ con los β elegidos para minimizar el error cuadrático medio de predicción.

Si asumimos que los β han sido tan elegido y luego pensar hacia atrás en el tiempo, se deduce de la estacionariedad que el mejor "predicador" de Y_{t-k} basado en el mismo $Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1}$. será $\beta_1 Y_{t-k+1} + \beta_2 Y_{t-k+2} + \dots + \beta_{k-1} Y_{t-1}$. La función de autocorrelación parcial en el rezago k es entonces definido como la correlación entre los errores de predicción; es decir:

$$\phi_{kk} = \text{Corr}(Y_t - \beta_1 Y_{t-1} - \beta_2 Y_{t-2} - \dots - \beta_{k-1} Y_{t-k+1}, Y_{t-k} - \beta_1 Y_{t-k+1} - \beta_2 Y_{t-k+2} - \dots - \beta_{k-1} Y_{t-1}). \quad (1.30)$$

(Para series distribuidas normalmente, se puede demostrar que las dos definiciones coinciden). Por convención, tomamos $\phi_{11} = \rho_1$.

Veremos que para el caso AR (1), $\phi_{kk} = 0$ para todo $k > 1$. Por lo tanto, la relación de autocorrelación parcial es distinta de cero para el rezago 1, el orden del proceso AR (1), pero es cero para todo retraso mayor que 1. Mostraremos que este es generalmente el caso de los modelos AR (p). Algunas veces decimos que la función de autocorrelación parcial para un proceso AR (p) se corta después del retraso excede el orden del proceso.

Considere un caso AR (p) general, que la mejor lineal predictor de Y_t basado en una función lineal de las variables $Y_{t-1}, Y_{t-2}, \dots, Y_p, \dots, Y_{t-k+1}$ para $k > p$ es $\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p}$. Además, el mejor predictor lineal de Y_{t-k} es alguna función de $Y_{t-1}, Y_{t-2}, \dots, Y_p, \dots, Y_{t-k+1}$, llámelo $h(Y_{t-1}, Y_{t-2}, \dots, Y_p, \dots, Y_{t-k+1})$.

Entonces, la covarianza entre los dos errores de predicción es

$$\begin{aligned} \text{Cov} \left(Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \dots - \phi_p Y_{t-p}, Y_{t-k} - h(Y_{t-k+1}, Y_{t-k+2}, \dots, Y_{t-1}) \right) \\ = \text{Cov}(e_t, Y_{t-k} - h(Y_{t-k+1}, Y_{t-k+2}, \dots, Y_{t-1})) = 0, \end{aligned}$$

ya que e_t es independiente de $Y_{t-k}, Y_{t-k+1}, Y_{t-k+2}, \dots, Y_{t-1}$.

Por lo tanto, hemos establecido el hecho clave de que, para un modelo AR (p)

$$\phi_{kk} = 0 \text{ para } k > p. \quad (1.31)$$

Para un modelo MA (1), la ecuación $\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$ produce rápidamente

$$\phi_{22} = \frac{-\theta^2}{1 + \theta^2 + \theta^4}. \quad (1.32)$$

Además, para el caso MA (1), se puede demostrar que

$$\phi_{kk} = -\frac{\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}} \text{ para } k \geq 1. \quad (1.33)$$

Observe que la autocorrelación parcial de un modelo MA(1) nunca es igual a cero, pero esencialmente decae a cero exponencialmente rápido a medida que aumenta el retardo, al igual que la función de autocorrelación del proceso AR (1). De manera más general, se puede demostrar que, la autocorrelación de un modelo MA (q) se comporta de manera muy parecida a la autocorrelación de un modelo AR (q).

Un método general para encontrar la función de autocorrelación parcial para cualquier proceso estacionario con función de autocorrelación ρ_k es el siguiente. Para un rezago k dado, se puede demostrar que ϕ_{kk} satisface las ecuaciones de Yule-Walker.

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \phi_{k3}\rho_{j-3} + \dots + \phi_{kk}\rho_{j-k} \text{ para } j = 1, 2, \dots, k \quad (1.34)$$

Más explícitamente, podemos escribir estas k ecuaciones lineales como

$$\begin{aligned} \phi_{k1} + \rho_1\phi_{k2} + \rho_2\phi_{k3} + \dots + \rho_{k-1}\phi_{kk} &= \rho_1 \\ \rho_1\phi_{k1} + \phi_{k2} + \rho_1\phi_{k3} + \dots + \rho_{k-2}\phi_{kk} &= \rho_2 \\ &\vdots \\ \rho_{k-1}\phi_{k1} + \rho_{k-2}\phi_{k2} + \rho_{k-3}\phi_{k3} + \dots + \phi_{kk} &= \rho_k \end{aligned} \quad (1.35)$$

Aquí estamos tratando $\rho_1, \rho_2, \dots, \rho_k$ y deseamos resolver para $\phi_{k1}, \phi_{k2}, \dots, \phi_{kk}$ (descartando todos menos ϕ_{kk}).

Estas ecuaciones producen ϕ_{kk} para cualquier proceso estacionario. Sin embargo, si el proceso está hecho AR (p), entonces, dado que para $k = p$, las ecuaciones (1.35) son solo las ecuaciones de Yule-Walker que se sabe que satisface el modelo AR (p), debemos tener $\phi_{pp} = \phi_p$. Además, como ya hemos visto mediante una derivación alternativa, $\phi_{kk} = 0$ para $k > p$. Así, la autocorrelación parcial muestra efectivamente el orden correcto p de un proceso autorregresivo, ya que el mayor retraso k antes de ϕ_{kk} se convierte en cero.

La función de autocorrelación parcial muestral

Para una serie de tiempo observada, necesitamos poder estimar la autocorrelación parcial en una variedad de retrasos, el método es estimar la autocorrelaciones muestrales y usar la ecuación (1.35), $k = 1, 2, 3, \dots$ para obtener estimaciones de ϕ_{kk} . Nosotros llamamos la función estimada la función de autocorrelación parcial muestral (muestra PACF) y denotarlo por.

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j} \rho_j} \quad (1.36)$$

Donde

$$\phi_{k,j} = \phi_{k-1,j} - \phi_{kk} \phi_{k-1,k-j} \quad \text{para } j=1,2,\dots,k-1,$$

Por ejemplo $\phi_{11} = \rho_1$, para empezar tenemos lo siguiente:

$$\phi_{22} = \frac{\rho_2 - \phi_{11}\rho_1}{1 - \phi_{11}\rho_1} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}.$$

Como antes con $\phi_{21} = \phi_{11} - \phi_{22}\phi_{11}$ que se necesita para el siguiente paso.

$$\phi_{33} = \frac{\rho_3 - \phi_{21}\rho_2 - \phi_{22}\rho_1}{1 - \phi_{21}\rho_1 - \phi_{22}\rho_2}.$$

Por tanto, podemos calcular numéricamente tantos valores para ϕ_{kk} como deseemos. Como se dijo, Estas ecuaciones recursivas nos dan las autocorrelaciones parciales teóricas, pero al reemplazar ρ por r , obtenemos las autocorrelaciones parciales estimadas o muestrales. Este algoritmo recibe el nombre de Durbin-Levinson.

Para evaluar la posible magnitud de las autocorrelaciones parciales muestrales, ha demostrado que, bajo la hipótesis de que un modelo AR (p) es correcto, la muestra las autocorrelaciones parciales en rezagos mayores que p tienen una distribución aproximadamente normal con medias cero y varianzas $1/n$. Por tanto, para $k > p$, $\pm 1.96 / \sqrt{n}$, se pueden utilizar como límites críticos $\hat{\phi}_{kk}$ para probar la hipótesis nula de que un modelo AR(p) es correcto.

La función de autocorrelación extendida

El método EACF utiliza el hecho de que, si la parte AR de un modelo ARMA mixto es conocido, "filtrar" la autorregresión de la serie de tiempo observada da como resultado un proceso MA que goza de la propiedad de corte en su ACF. Los coeficientes AR pueden estimarse mediante una secuencia finita de regresiones. Ilustramos el procedimiento para el caso en que el verdadero modelo es un modelo ARMA(1,1):

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}.$$

En este caso, una regresión lineal simple de Y_t en Y_{t-1} da como resultado un estimador inconsistente de ϕ , incluso con una cantidad infinita de datos. De hecho, el coeficiente de regresión teórico es igual a $\rho_1 = (\phi - \theta)(1 - \phi\theta) / (1 - 2\phi\theta + \theta^2)$ no ϕ . Pero los residuos de esta regresión contienen información sobre el proceso de error $\{e_t\}$. Se realiza una segunda regresión múltiple que consiste en retroceder Y_t en Y_{t-1} y en el rezago 1 de los residuos de la

primera regresión. El coeficiente de Y_{t-1} en la segunda regresión, denotado por $\tilde{\phi}$, resulta ser un estimador consistente de ϕ . Define $W_t = Y_t - \tilde{\phi}Y_{t-1}$, entonces es aproximadamente un proceso MA(1). Para un modelo ARMA(1,2), una tercera regresión que regresa Y_t en su rezago 1, el rezago 1 de los residuos de la segunda regresión, y el rezago 2 de los residuos de la primera regresión conducen a que el coeficiente de Y_{t-1} sea un estimador consistente de ϕ . De manera similar, los coeficientes AR de un modelo ARMA(p, q) pueden ser consistentemente estimado a través de una secuencia de q regresiones.

Como se desconocen los órdenes AR y MA, se requiere un procedimiento iterativo.

$$W_{t,k,j} = Y_t - \tilde{\phi}Y_{t-1} - \dots - \tilde{\phi}_k Y_{t-k}. \quad (1.37)$$

Sean los residuos autorregresivos definidos con los coeficientes AR estimados iterativamente

suponiendo que el orden AR es k y el orden MA es j . Las autocorrelaciones muestrales de $W_{t,k,j}$ se denominan autocorrelaciones extendida de la muestra. Para $k = p$ y $j \geq q$, $\{W_{t,k,j}\}$ es aproximadamente un modelo MA(q), de modo que sus autocorrelaciones

teóricas de retraso $q + 1$ o superior son iguales a cero. Para $k > p$, ocurre un problema de sobreajuste, y esto aumenta la orden MA para el proceso W por el mínimo de $k - p$ y $j - q$, sugirió resumir la información en la muestra EACF por una tabla con el elemento en la k -ésima fila y la j -ésima columna igual al símbolo X si el rezago $j + 1$ correlación muestral de $W_{t,k,j}$ es significativamente diferente de 0 (es decir, si su magnitud es mayor que $1.96 / \sqrt{n - j - k}$ ya que la autocorrelación muestral es asintóticamente $N(0,1) / (n - k - j)$) si las W 's son aproximadamente un proceso $MA(j)$) y 0 en caso contrario. En la tabla 1, la región triangular de ceros que se muestra en la muestra EACF, indica claramente que un modelo mixto con $q = 1$ y con $p = 1$ o 2 podría ser más apropiado.

Tabla 1. Muestra simulada para la serie ARMA(1,1)

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	o	o	o	o	o	o	o	o	o	o
1	x	o	o	o	o	o	o	o	o	o	o	o	o	o
2	x	o	o	o	o	o	o	o	o	o	o	o	o	o
3	x	x	o	o	o	o	o	o	o	o	o	o	o	o
4	x	o	x	o	o	o	o	o	o	o	o	o	o	o
5	x	o	o	o	o	o	o	o	o	o	o	o	o	o
6	x	o	o	o	x	o	o	o	o	o	o	o	o	o
7	x	o	o	o	x	o	o	o	o	o	o	o	o	o

1.8 Modelos ARMA, ARIMA Y SARIMA

La serie temporal $\{X_t\}$ es un proceso ARMA (p,q) si $\{X_t\}$ es estacionaria y si por cada t ,

$$X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p} = Z_t + \theta_1 Z_{t-1} + \dots + \theta_q Z_{t-q}, \tag{1.38}$$

Donde $(Z_t) \sim WN(0, \sigma^2)$ y los polinomios $(1 - \phi_1 z - \dots - \phi_p z^p)$ y $(1 + \theta_1 z + \dots + \theta_q z^q)$ no tiene factores comunes. Se dice que el proceso $\{X_t\}$ es un proceso ARMA (p, q) con media μ si $\{X_t - \mu\}$ es un proceso ARMA (p, q). Es conveniente utilizar la forma más concisa la ecuación anterior (1.38)

$$\phi(B)X_t = \theta(B)Z_t, \tag{1.39}$$

Donde $\phi(\cdot)$ y $\theta(\cdot)$ son la p y q grado de polinomios de grado p y q respectivamente

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p,$$

y

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q,$$

B es el operador de desplazamiento hacia atrás ($B^j X_t = X_{t-j}, j = 1, 2, \dots$). Se dice que la serie de tiempo $\{X_t\}$ es un proceso autoregresivo de orden p (o AR (p)) si $\theta(z) = 1$, y un proceso de promedio móvil de orden q (o MA (q)) si $\phi(z) = 1$.

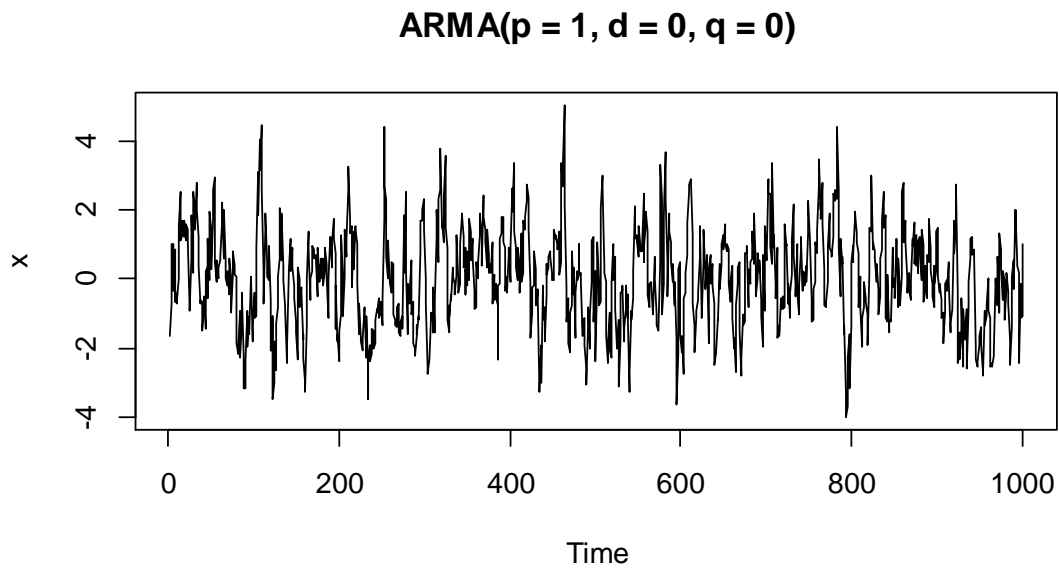


Fig. 6 Modelo AR(1) “simulado”.

Serie x

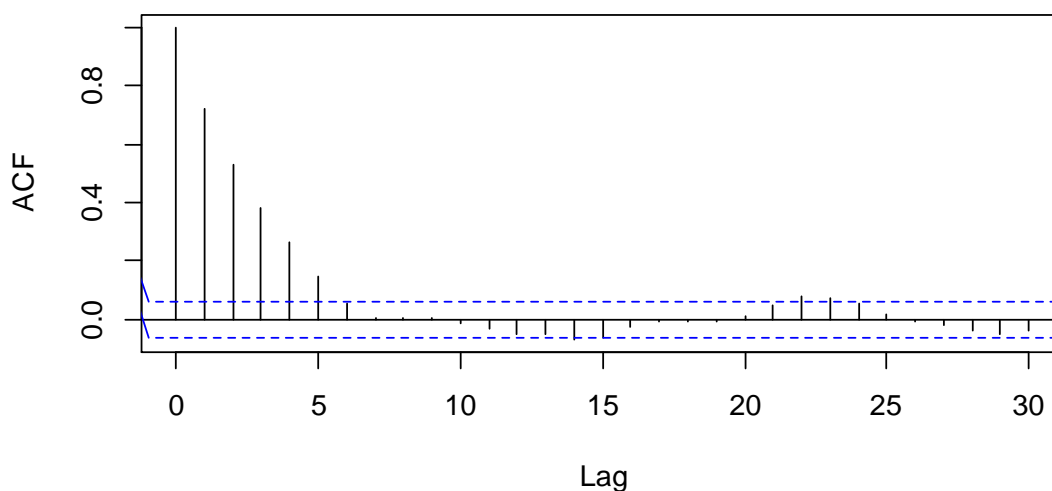


Fig. 7 Función de autocorrelación del modelo AR(1) “simulado”.

Serie x

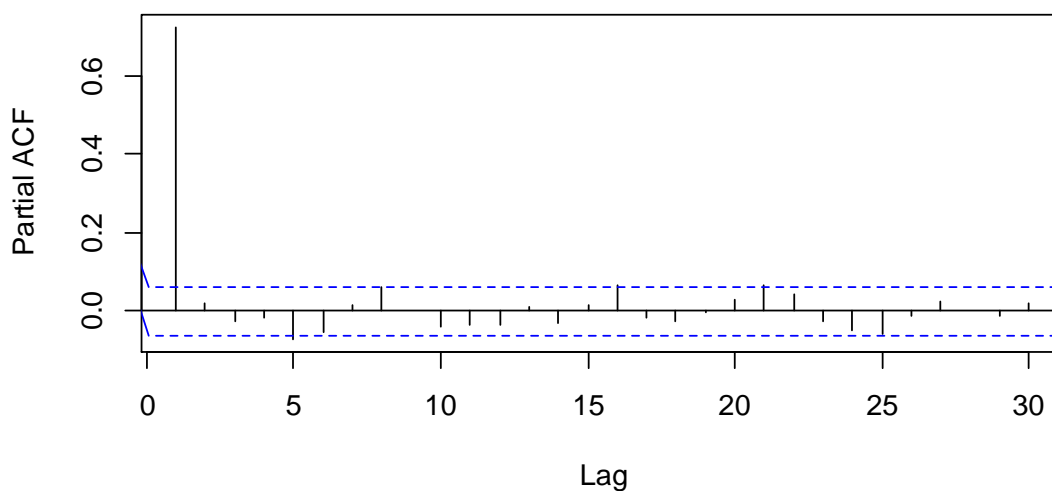


Fig. 8 Función de autocorrelación parcial del modelo AR(1) “simulado”

Existencia y singularidad

Existe una solución estacionaria $\{X_t\}$ de la ecuación (1.38) (y también es la única solución estacionaria) si y solo si

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0 \text{ para todo } |z| = 1. \quad (1.40)$$

Causalidad

Un proceso ARMA (p, q) $\{X_t\}$ es causal, o una función causal de $\{Z_t\}$ si hay constantes existen $\{\psi_j\}$ tal que $\sum_{j=0}^{\infty} |\psi_j| < \infty$ y

$$X_t = \sum_{j=0}^{\infty} \psi_j Z_{t-j} \text{ para todo } t, \quad (1.41)$$

La causalidad es equivalente a la condición

$$\phi(z) = 1 - \phi_1 z - \dots - \phi_p z^p \neq 0 \text{ para todo } |z| \leq 1. \quad (1.42)$$

Invertibilidad

Un ARMA (p, q) proceso $\{X_t\}$ es invertible si hay constantes $\{\pi_j\}$ tal que $\sum_{j=0}^{\infty} |\pi_j| < \infty$ y

$$Z_t = \sum_{j=0}^{\infty} \pi_j X_{t-j} \text{ para todo } t$$

La invertibilidad es equivalente a la condición

$$\theta(z) = 1 + \theta_1 z + \dots + \theta_q z^q \neq 0 \text{ para todo } |z| \leq 1. \quad (1.43)$$

Intercambiando los roles de los polinomios AR y MA, encontramos que la secuencia $\{\pi_j\}$ está determinada por las ecuaciones:

$$\pi_j + \sum_{k=1}^q \theta_k \pi_{j-k} = -\phi_j, \quad j = 0, 1, \dots, \quad (1.44)$$

donde $\phi_0 = -1$, $\phi_j = 0$ para $j > p$, y $\pi_j = 0$ para $j < 0$.

Modelo ARIMA

Una serie temporal $\{X_t\}$ se dice que sigue un modelo autorregresivo integrado media móvil, si la diferencia dth , $W_t = \bar{\nabla}^d Y_t$ es un proceso ARMA estacionario si W_t , sigue un modelo ARMA (p,q) decimos que $\{Y_t\}$ es un proceso ARIMA (p,d,q).

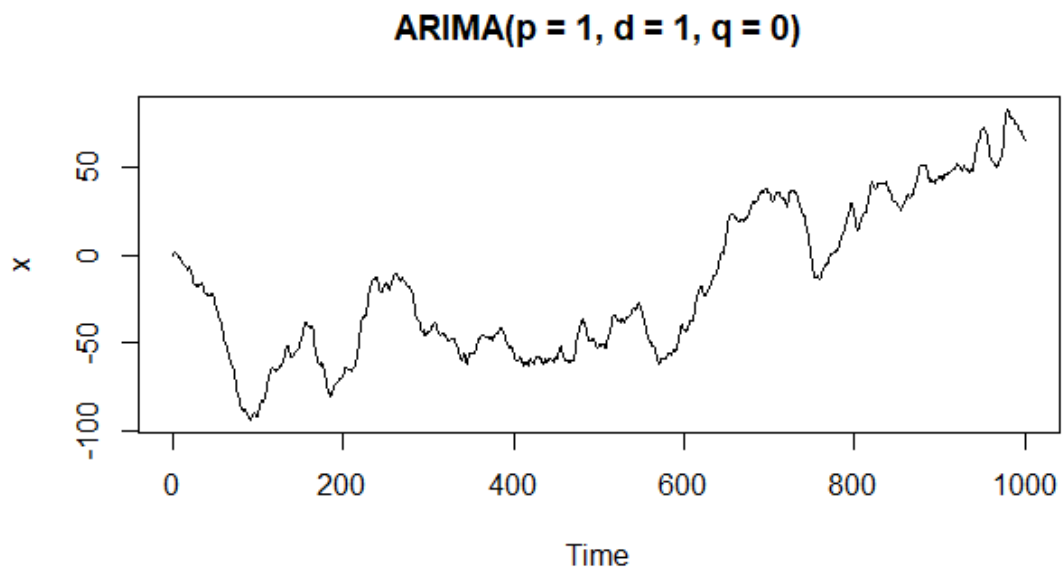


Fig. 9 Modelo ARIMA(1,1,0) “simulado”.

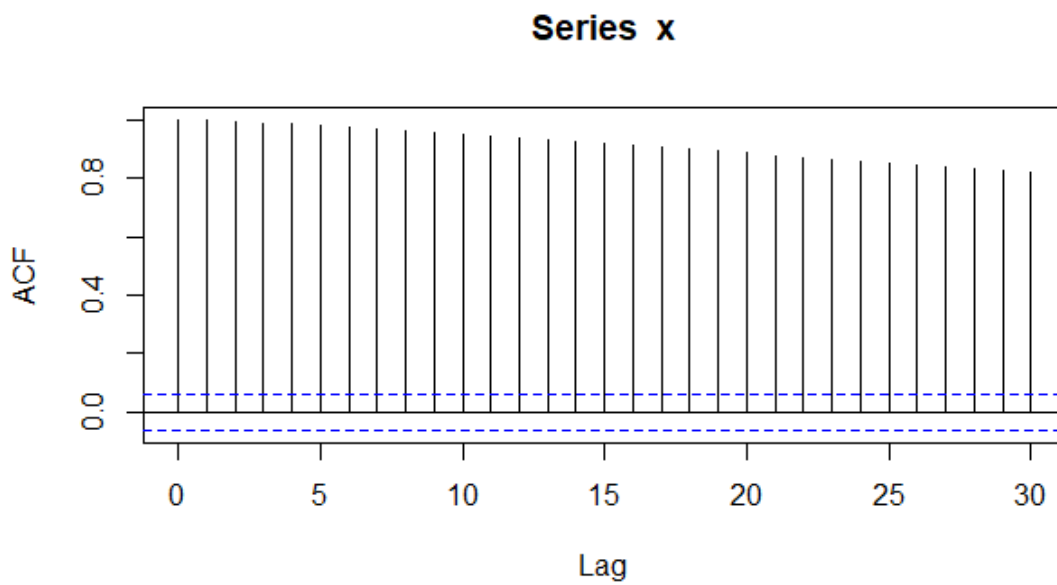


Fig. 10 Función de autocorrelación del modelo ARMA(1,1,0) “simulado”.

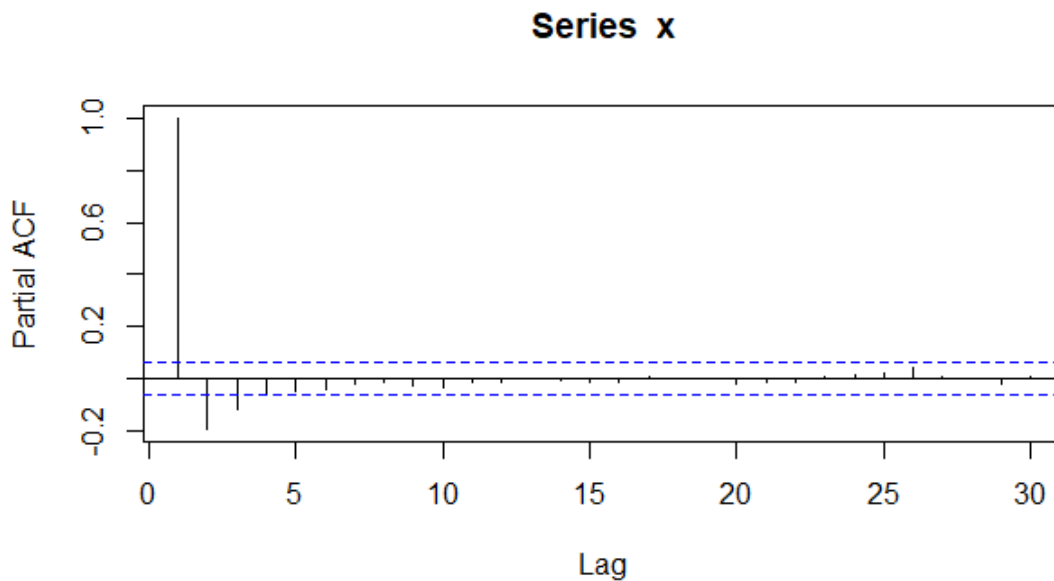


Fig. 11 Función de autocorrelación parcial del modelo ARIMA(1,1,0) “simulado”

La prueba de raíz unitaria de Dickey-Fuller

Si bien el decrecimiento lineal de la muestra de ACF a menudo se toma como un síntoma de que la serie de tiempo subyacente no es estacionaria y requiere diferenciación, también es útil para cuantificar la evidencia de no estacionariedad en el mecanismo de generación de datos. Esto puede ser hecho a través de pruebas de hipótesis. Considere el modelo.

$$Y_t = \alpha Y_{t-1} + e_t \text{ para } t = 1, 2, \dots$$

Donde $\{e_t\}$ es un proceso estacionario. El proceso $\{Y_t\}$ es no estacionario si el coeficiente $\alpha = 1$, pero es estacionario si $|\alpha| < 1$. Suponga que $\{X_t\}$ es un proceso AR(k): $X_t = \phi_1 X_{t-1} + \dots + \phi_k X_{t-k} + e_t$. Bajo la hipótesis nula de que $\alpha = 1$, $X_t = Y_t - Y_{t-1}$. Dejando $a = \alpha - 1$, tenemos.

$$\begin{aligned} Y_t - Y_{t-1} &= (\alpha - 1)Y_{t-1} + X_t \\ &= aY_{t-1} + \phi_1 X_{t-1} + \dots + \phi_k X_{t-k} + e_t \\ &= aY_{t-1} + \phi_1(Y_{t-1} - Y_{t-2}) + \dots + \phi_k(Y_{t-k} - Y_{t-k-1}) + e_t. \end{aligned} \quad (1.45)$$

Donde $a = 0$ bajo la hipótesis de que Y_t es la diferencia no estacionaria. Por otro lado, si $\{Y_t\}$ es estacionario de modo que $-1 < \alpha$, entonces se puede verificar que Y_t todavía satisface una ecuación similar a la ecuación anterior, pero con coeficientes diferentes;

por ejemplo, $a = (1 - \phi_1 - \dots - \phi_k)(1 - \alpha) < 0$. De hecho, $\{Y_t\}$ es un proceso AR $(k + 1)$ cuya ecuación característica AR está dada por $\Phi(x)(1 - \alpha x) = 0$, donde $\Phi(x) = 1 - \phi_1 x - \dots - \phi_k x^k$. Entonces la hipótesis nula corresponde al caso donde el polinomio característico AR tiene una raíz unitaria y la hipótesis alternativa establece que no tiene raíces unitarias. En consecuencia, la prueba para diferenciar equivale a probar una raíz unitaria en el polinomio característico AR de $\{Y_t\}$.

Según el análisis anterior, la hipótesis nula de que $\alpha = 1$ (equivalentemente $a = 0$) puede ser probado haciendo una regresión de la primera diferencia de la serie de tiempo observada en el rezago 1 de la serie observada y en el pasado k rezagos de la primera diferencia de la serie observada. El test es equivalente a comprobar, si el coeficiente $a = 0$, siendo la hipótesis nula que el proceso es una diferencia no estacionaria. Es decir, el proceso no es estacionario, pero se vuelve estacionario después de primera diferenciación. La hipótesis alternativa es que ($a < 0$) y, por tanto, $\{Y_t\}$ es estacionario.

El estadístico de prueba de Dickey-Fuller aumentado (ADF) es el estadístico t -statistic del coeficiente estimado de a a partir del método de regresión por mínimos cuadrados. Sin embargo, la estadística de prueba ADF no tiene una distribución t aproximadamente bajo la hipótesis nula. Se han tabulado los puntos porcentuales de esta distribución nula.

Criterios de información Akaike y criterio de información bayesiano

Los dos métodos principales comúnmente aplicados para elegir entre los modelos apropiados son los criterios de información de Akaike (AIC) y el criterio de información bayesiano de Schwarz (BIC). Además, la fórmula matemática para AIC y BIC se muestran a continuación:

$$AIC = -2\log(L) + 2k = -2\log(L) + 2(p + q + P + Q). \quad (1.46)$$

$$BIC = -2\log(L) + k \ln(n) = -2\log(L) + (p + q + P + Q) \ln(n). \quad (1.47)$$

Donde n se refiere al tamaño de la serie y k representa el número de los parámetros del método ARIMA. L es la verosimilitud, el modelo es más eficiente si el valor de AIC es menor. Entonces, el óptimo El modelo de pronóstico se selecciona en función del mejor ajuste que tiene el valor AIC mínimo del grupo.

Términos constantes en modelos ARIMA

Para un modelo ARIMA (p, d, q) , $\nabla^d Y_t$ es un proceso ARMA (p, q) estacionario. Nuestro supuesto estándar es que los modelos estacionarios tienen una media cero; es decir, en realidad se trabajando con desviaciones de media constante. Una media constante distinta de cero, μ , en un modelo ARMA estacionario $\{W_t\}$ se puede acomodar de dos formas. Podemos asumir que:

$$W_t - \mu = \phi_1(W_{t-1} - \mu) + \phi_2(W_{t-2} - \mu) + \dots + \phi_p(W_{t-p} - \mu) + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} \quad (1.48)$$

Alternativamente podemos introducir un término constante θ_0 en el modelo de la siguiente manera:

$$W_t = \theta_0 + \phi_1 W_{t-1} + \phi_2 W_{t-2} + \dots + \phi_p W_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q}. \quad (1.49)$$

Cambios porcentuales y logaritmos

Suponga que Y_t tiende a tener cambios porcentuales relativamente estables de un período de tiempo a el siguiente. Específicamente, suponga que

$$Y_t = (1 + X_t)Y_{t-1}. \quad (1.50)$$

Donde $100X_t$ es el cambio porcentual (posiblemente negativo) de Y_{t-1} a Y_t . Entonces

$$\begin{aligned} \log(Y_t) - \log(Y_{t-1}) &= \log\left(\frac{Y_t}{Y_{t-1}}\right) \\ &= \log(1 + X_t). \end{aligned} \quad (1.51)$$

Si X_t está restringido digamos, $|X_t| < 0.2$ (es decir, los cambios porcentuales son como máximo $\pm 20\%$), entonces, con una buena aproximación, $\log(1 + X_t) \approx X_t$. Como consecuencia,

$$\nabla[\log(Y_t)] \approx X_t. \quad (1.52)$$

Será relativamente estable y quizás bien modelado por un proceso estacionario. Darse cuenta si primero tomamos los registros y luego calculamos las primeras diferencias; el orden sí importa. En la literatura financiera, las diferencias de los logaritmos (naturales) generalmente se denominan retornos.

Modelo SARIMA

El modelo SARIMA que refleja la característica de variación estacional en series de tiempo. Generalmente, la serie de tiempo original $\{Y_t\}$ utiliza un operador de retardo B para procesar SARIMA $(p, d, q) \times (P, D, Q)_S$. Un ARIMA estacional el modelo puede escribirse como:

$$\phi(B)\Phi_p(B^S)(1-B)^d(1-B^S)^DY_t = \theta_q(B)\Theta_Q(B^S)\epsilon_t. \quad (1.53)$$

En la formula (1.53). B es el operador de retraso (definido como $B^k Y_t = Y_{t-k}$)

$$\begin{aligned} \phi_p(B) &= 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \\ \Phi_p(B^S) &= 1 - \Phi_s B^S - \Phi_{s2} B^{2S} - \dots - \Phi_{ps} B^{Ps} \\ \theta_q(B) &= 1 - \theta_1 B - \theta_1 B^2 - \dots - \theta_q B^q \\ \Theta_Q(B^S) &= 1 - \Theta_s B^S - \Theta_{2s} B^{2S} - \dots - \Theta_{Qs} B^{Qs}. \end{aligned} \quad (1.54)$$

Donde $\phi(B)$ y $\theta(B)$ son polinomios de orden p y q , $\Phi(B^S)$ y $\Theta(B^S)$ son polinomios en B^S de grados P y Q , respectivamente p es el orden de autorregresión no estacional; d es el número de diferencias regulares; q es el orden de la media móvil no estacional; P es el orden de autorregresión estacional; D es el número de diferencias estacionales; Q es el orden de la media móvil estacional; y S es la duración de la temporada.

2. Metodología Box – Jenkins

La metodología de Box & Jenkins fue descrita en 1970 por los estadísticos George Edward Pelham Box y Gwilym Meirion Jenkins, para intentar hacer más fácil el trabajo a los analistas a la hora de crear un modelo para una serie de tiempo, ayudando a explicar su estructura y a realizar una predicción a cerca del comportamiento futuro de la serie. Esta metodología, que se explicó por primera vez en el libro “Time Series Analysis: Forecasting and Control” se trata de un procedimiento de análisis estadístico que sirve para ajustar un tipo especial de modelos autorregresivos ARMA, siglas en inglés de (Autorregressive Moving Average) además de los modelos autorregresivos integrados de medias móviles, también llamados ARIMA (Autorregresive Integrated Moving Average). descrita

La idea principal de la metodología Box - Jenkins es proponer un conjunto de procedimientos para escoger entre los modelos, ARMA ARIMA y SARIMA estacional el que se ajuste a los datos de una serie temporal observada y con ello realizar pronósticos sobre ésta. Este procedimiento plantea cuatro pasos:

1. Identificación: Este primer paso busca establecer los valores apropiados para p, d, q, P, D, Q .
2. Estimación: Luego se deben estimar los parámetros incluidos en el modelo.
3. Verificación de diagnóstico: Al seleccionar el modelo ARIMA particular se debe comprobar si el modelo se ajusta a los datos (puede existir otro modelo ARIMA que presente mejor ajuste). La prueba más simple de ajuste es comprobar si los residuos obtenidos son ruido blanco.
4. Predicción: El último plazo consiste en la elaboración de los pronósticos de la serie temporal en particular.

Como desarrollo de la temática para el presente capítulo se procede a estudiar, de forma más precisa, cada uno de los pasos que comprende la metodología Box – Jenkins.

2.1 Identificación

Este paso comprende identificar la estructura si es no estacionaria, realizar las transformaciones a que haya lugar para obtener varianza y media constantes y finalmente determinar los órdenes del modelo.

Este análisis nos permite identificar en la serie temporal características como la alta frecuencia (característica intrínseca de la serie que no es corregible), el comportamiento no estacionario y la presencia de estacionalidad en los datos.

Esta familia de transformaciones, se emplea para la estabilización de la serie en algunas situaciones de no estacionariedad.

Transformaciones de Box - Cox

Una familia flexible de transformaciones, fue introducida por Box y Cox (1964). Para un valor dado del parámetro λ , la transformación se define por:

$$g(x) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \text{para } \lambda \neq 0 \\ \log x & \text{para } \lambda = 0 \end{cases} .$$

(2.1)

El término x^λ es la parte importante de la primera expresión, pero restando 1 y dividiendo por λ hace que $g(x)$ cambie suavemente a medida que λ se acerca a cero. De hecho, un argumento de cálculo muestra que como $\lambda \rightarrow 0$, $(x^\lambda - 1)/\lambda \rightarrow \log(x)$. Observe que $\lambda = 1/2$ produce una raíz cuadrada, transformación útil con datos de tipo Poisson, y $\lambda = -1$ corresponde a una transformación recíproca.

La transformación de potencia se aplica solo a valores de datos positivos. Si algunos de los valores son negativos o cero, se puede agregar una constante positiva a todos los valores para hacerlos todo positivo antes de hacer la transformación. El cambio a menudo se determina subjetivamente. Por ejemplo, para datos de captura no negativos en biología, la aparición de ceros es a menudo se trata agregando una constante igual al valor de datos positivo más pequeño a todos los valores de los datos. Un enfoque alternativo consiste en utilizar transformaciones aplicables a cualquier dato, positivo o no. Un inconveniente de este enfoque alternativo es que las interpretaciones de tales transformaciones son a menudo menos sencillas que las interpretaciones de la transformación de potencia.

Podemos considerar λ como un parámetro adicional en el modelo a estimar a partir de los datos observados. Sin embargo, normalmente no se justifica una estimación precisa de λ . Evaluación de un rango de transformaciones basadas en una cuadrícula de valores de λ , digamos $\pm 1, \pm 1/2, \pm 1/3, \pm 1/4$ y 0, normalmente será suficiente y puede tener algún significado intuitivo.

Diferenciación

Este procedimiento implica identificar si la serie temporal tiene un centro de gravedad o si carece de éste, es decir si se presentan tendencias (para lo cual se usará, sobre todo, el gráfico de la serie) y además se busca identificar en la función de autocorrelación muestral un decaimiento lento.

Para estabilizar la media se toman diferenciaciones del tipo:

$$\nabla_s X_t = (1 - B^s)X_t = X_t - X_{t-s} \quad t = s + 1, \dots, T.$$

En este caso ($s = 1$) los instrumentos gráficos son muy útiles, se utilizan el gráfico de la serie, y el gráfico de autocorrelación con ellos se puede realizar una diferenciación regular o una diferenciación en la parte estacional D . También se puede usar técnicas analíticas como el test de Dickey-Fuller.

En la diferenciación en d el gráfico presentará una tendencia clara, la función de autocorrelación muestral decaerá de forma lenta y lineal, y la función de autocorrelación parcial muestral presentará un coeficiente de primer retardo cercano a 1. Se debe tener presente que las tendencias lineales se eliminan con $d = 1$ y las tendencias cuadráticas con $d = 2 \dots$. Para la diferenciación en B^s se mostrará un gráfico con pautas repetidas de periodo s y se observará que la función de autocorrelación simple muestral mostrará coeficientes altos que decrecen de manera lenta en los retardos múltiplos de periodo s .

2.1.2 Identificación de órdenes del proceso

Se proceden a estimar los órdenes p, q, P y Q para ello se compara las funciones estimadas de autocorrelación simple y parcial con sus respectivas funciones teóricas, se debe seleccionar un conjunto de modelos que se supongan adecuados. Los coeficientes de autocorrelación muestrales se estiman mediante la ecuación:

$$\hat{\rho}_k = \frac{\sum_{t=d+sD+1}^{T-k} (\omega_t - \bar{\omega})(\omega_{t+k} - \bar{\omega})}{\sum_{t=d+sD+1}^T (\omega_t - \bar{\omega})^2}, \quad k = 1, 2, \dots \quad (2.2)$$

Donde $\omega_t = \nabla^d \nabla_s^D X_t$, que representa la serie estacionaria. Para la ecuación (2.2) se debe calcular la covarianza muestral del retardo k , $\hat{\Omega}_k$ y la varianza muestral $\hat{\Omega}_0$ definidas:

$$\hat{\Omega}_k = \frac{\sum (\omega_t - \bar{\omega})(\omega_{t+k} - \bar{\omega})}{n}.$$

$$\hat{\Omega}_0 = \frac{\sum (\omega_t - \bar{\omega})^2}{n}.$$

Donde $\bar{\omega}$ es la media muestral y n el tamaño de la muestra, por tanto $\hat{\rho}_k = \frac{\hat{\Omega}_k}{\hat{\Omega}_0}$ Para identificar los órdenes del modelo es necesario apoyarse en la función de autocorrelación parcial ya que ρ_k (coeficiente de correlación parcial de orden k) mide el grado de asociación lineal existente entre variables habiendo ajustado el efecto lineal

de todas las variables intermedias, el cálculo se realiza mediante la regresión lineal entre variables, representada seria.

$$X_t = \mu + p_1 X_{t-1} + \dots + p_k X_{t-k} + e_t. \quad (2.3)$$

La función de autocorrelación parcial se estima basándose en los datos de la serie y como función de $\hat{\rho}_k$. Es claro que si $\hat{\rho}_k = 0$, la serie X_t es ruido blanco, si por el contrario $\hat{\rho}_k$, es significativo no es ruido blanco.

Para identificar órdenes en un proceso estacional se sigue un proceso similar con la diferencia que se deberán observar los coeficientes en los retardos específicos que muestren estacionalidad ya que indicarán en la función de autocorrelación y la función de autocorrelación simple las órdenes P y Q .

2.2 Estimación

2.2.1 Estimación de máxima verosimilitud

Para cualquier conjunto de observaciones, Y_1, Y_2, \dots, Y_n , serie de tiempo, la función de verosimilitud L es definida como la densidad de probabilidad conjunta de obtener los datos realmente observados. Sin embargo, se considera en función de los parámetros desconocidos en el modelo con los datos observados se mantienen fijos. Para los modelos ARIMA, L será una función de las, $\phi's, \theta's, \mu$ y σ_e^2 dadas las observaciones Y_1, Y_2, \dots, Y_n . Los estimadores de máxima verosimilitud son entonces definidos como aquellos valores de los parámetros para los cuales los datos realmente observados son más probables, es decir, los valores que maximizan la función de verosimilitud.

Comenzamos mirando en detalle el modelo AR (1). La suposición más común es que los términos de ruido blanco son variables aleatorias independientes, normalmente distribuidas con media cero y desviación estándar común. La función de densidad de probabilidad (pdf) de cada e_t es entonces:

$$(2\pi\sigma_e^2)^{-1/2} \exp\left(-\frac{e_t^2}{2\sigma_e^2}\right) \text{ para } -\infty < e_t < \infty,$$

Y por la independencia, del pdf para e_2, e_3, \dots, e_n es

$$(2\pi\sigma_e^2)^{-(n-1)/2} \exp\left(-\frac{1}{2\sigma_e^2} \sum_{t=2}^n e_t^2\right). \quad (2.4)$$

Ahora consideremos

$$\begin{aligned}
 Y_2 - \mu &= \phi(Y_1 - \mu) + e_2 \\
 Y_3 - \mu &= \phi(Y_2 - \mu) + e_3 \\
 &\vdots \\
 Y_n - \mu &= \phi(Y_{n-1} - \mu) + e_n
 \end{aligned} \tag{2.5}$$

Si condicionamos a $Y_1 = y_1$, la ecuación (2.5) define una transformación lineal entre e_2, e_3, \dots, e_n y Y_2, Y_3, \dots, Y_n (con jacobiano igual a 1). Así, el pdf conjunto de Y_2, Y_3, \dots, Y_n dado $Y_1 = y_1$ se puede obtener utilizando la ecuación (2.4) para sustituir las e 's en términos de las Y 's en la ecuación (2.4). Así obtenemos:

$$f(y_2, y_3, \dots, y_n | y_1) = (2\pi\sigma_e^2)^{-(n-1)/2} \times \exp\left\{-\frac{1}{2\sigma_e^2} \sum_{t=2}^n [(y_t - \mu) - \phi(y_{t-1} - \mu)]^2\right\}. \tag{2.6}$$

Ahora considere la distribución (marginal) de Y_1 . De la representación del proceso lineal del proceso AR (1), se deduce que Y_1 tendrá una distribución normal con media μ y varianza $\sigma_e^2 / (1 - \phi^2)$. Multiplicando el pdf condicional, la ecuación (2.6) por la función de densidad de probabilidad marginal de Y_1 nos da la función de densidad de probabilidad conjunta de Y_1, Y_2, \dots, Y_n que exigir. Interpretado en función de los parámetros ϕ, μ y σ_e^2 la función de verosimilitud para un modelo AR(1) viene dado por:

$$L(\phi, \mu, \sigma_e^2) = (2\pi\sigma_e^2)^{-n/2} (1 - \phi^2)^{1/2} \exp\left[-\frac{1}{2\sigma_e^2} S(\phi, \mu)\right], \tag{2.7}$$

Donde

$$S(\phi, \mu) = \sum_{t=2}^n [(Y_t - \mu) - \phi(Y_{t-1} - \mu)]^2 + (1 - \phi^2) + (1 + \phi^2)(Y_1 - \mu). \tag{2.8}$$

La función $S(\phi, \mu)$ se llama función de suma de cuadrados incondicional. Como regla general, el logaritmo de la función de verosimilitud es más conveniente para trabajar con la probabilidad misma. Para el caso AR(1), **la función logarítmica de verosimilitud**, denotado, $\ell(\phi, \mu, \sigma_e^2)$ está dado por:

$$\ell(\phi, \mu, \sigma_e^2) = -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log(\sigma_e^2) + \frac{1}{2} \log(1 - \phi^2) - \frac{1}{2\sigma_e^2} S(\phi, \mu). \tag{2.9}$$

Para valores dados de ϕ y μ , $\ell(\phi, \mu, \sigma_e^2)$ se puede maximizar analíticamente con respecto a σ_e^2 en términos de los estimadores aún por determinar de ϕ y μ . Obtenemos:

$$\hat{\sigma}_e^2 = \frac{S(\hat{\phi}, \hat{\mu})}{n}. \quad (2.10)$$

Como en muchos otros contextos similares, generalmente dividimos por $n - 2$ en lugar de n (ya que estamos estimando dos parámetros, ϕ y μ) para obtener un estimador con menos sesgo. Para típicos tamaños de muestra de series de tiempo, habrá muy poca diferencia.

2.2.2 Método de los mínimos cuadrados condicionales

Con esta técnica se busca minimizar la suma de cuadrados condicionales $S_c(\varphi_p, \theta_q) = \sum_{t=p+1}^n e_t^2$, suma que se obtiene de la suma de cuadrados no condicionales, donde $e_p, e_{p-1}, \dots, e_{p+1-q}$ son la base de los cálculos de e_t igualados a cero que corresponde a su valor esperado.

Para un modelo AR(p) la estimación viene dada por la ecuación:

$$S_c(\varphi_p) = \sum_{t=p+1}^n (X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p})^2. \quad (2.11)$$

De donde se plantea las ecuaciones normales $Y'Y\hat{\varphi}_p = Y'W$ siendo:

$$Y = \begin{bmatrix} X_p & X_{p-1} & \dots & X_1 \\ X_{p+1} & X_p & \dots & X_2 \\ \vdots & \vdots & \ddots & \vdots \\ X_{n-1} & X_{n-2} & & X_{n-p} \end{bmatrix}$$

$$\hat{\varphi}_p = (\hat{\phi}_1, \dots, \hat{\phi}_p)' \text{ y } W = (X_{p+1}, \dots, X_n)' \text{ y por tanto } \hat{\varphi}_p = (Y'Y)^{-1} Y'W.$$

2.2.3 Mínimos cuadrados incondicionales

Como compromiso entre las estimaciones de mínimos cuadrados condicionales y las estimaciones de máxima verosimilitud, podríamos considerar obtener estimaciones de mínimos cuadrados incondicionales; estima minimizando $S(\phi, \mu)$. Desafortunadamente, el término $(1 - \phi^2)(Y_1 - \mu)^2$ es causa de las ecuaciones $\partial S / \partial \phi = 0$ y $\partial S / \partial \mu = 0$ sea no lineal en ϕ y μ y la parametrización a un término constante $\theta_0 = \mu(1 - \phi)$ no mejora sustancialmente la situación, por lo tanto, la minimización debe llevarse a cabo numéricamente. Las estimaciones resultantes se denominan estimaciones de mínimos cuadrados incondicionales.

2.2.4 Métodos de optimización no lineal

Los métodos de suma de cuadrados condicional y máxima verosimilitud exacta, al utilizarse en modelos que contiene términos de medias móviles, generan funciones no cuadráticas lo que obliga a realizar estimaciones no lineales en la maximización en la ecuación (1.59) y la minimización de la suma de cuadrados condicionales $S_c(\varphi_p, \theta_q)$, principalmente se utiliza el algoritmo de Gauss-Newton, el cual resulta de una variación del método de optimización de Newton sin el uso de segundas derivadas, este procedimiento es iterativo por lo cual parte de una estimación del parámetro.

2.3 Diagnóstico del modelo

En este paso se busca verificar que tan adecuado es el modelo, es decir se debe comprobar que:

- Las condiciones de estacionariedad e invertibilidad de los coeficientes estimados del modelo se cumplan, así como determinar que estos parámetros estimados sean significativos.
- Los residuos se comporten como ruido blanco.

2.3.1 Diagnóstico de los coeficientes estimados

Para los modelos $ARMA(p, q)$ se plantean los siguientes contrastes de hipótesis:

$$\begin{aligned} H_0: \mu &= 0 \text{ vs } H_1: \mu \neq 0 \\ H_0: \phi_i &= 0 \text{ vs } H_1: \phi_i \neq 0 \\ H_0: \theta_i &= 0 \text{ vs } H_1: \theta_i \neq 0. \end{aligned} \quad (2.12)$$

Siendo μ la constante media. Para los coeficientes $\beta = (\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)$ los cuales presentan distribución asintótica $\hat{\beta}_i \sim N(\beta_i, V(\hat{\beta}_i))$ esto $\forall i$ donde la inversa de la matriz de información permite estimar la varianza, el estadístico t de contraste con distribución normal viene dado por:

$$t = \frac{\hat{\beta}_i - 0}{\sqrt{V(\hat{\beta}_i)}} \sim N.$$

La hipótesis nula con $\alpha = 0.05$ se rechaza cuando: $\left| \frac{\hat{\beta}_i - 0}{\sqrt{V(\hat{\beta}_i)}} \right| > N_{\frac{\alpha}{2}}(0, 1) \approx 1.96$.

Las condiciones de estacionariedad e invertibilidad se comprueban calculando las raíces del polinomio autorregresivo $\hat{\phi}(B) = 0$ y las raíces del polinomio de medias móviles

$\hat{\theta}(B) = 0$, si alguna se encuentra cercana a 1 se puede presumir falta de estacionariedad o invertibilidad. Por otra parte, la matriz de covarianzas permite detectar presencia de factores comunes al modelo valiéndose de los niveles de correlación entre los modelos.

Correlación entre estimadores

Por otro lado, además se puede examinar la matriz de \mathbf{R} de correlaciones estimadas entre los estimadores de los parámetros, que puede calcularse a partir de \mathbf{V} como $R = D^{-1} \sum V D^{-1/2}$, donde D es una matriz diagonal cuyos elementos son las raíces cuadradas de los elementos en la diagonal principal de V . La presencia de correlaciones elevadas suele indicar que el modelo este sobre parametrizado.

2.3.2 Diagnóstico de los residuos

Después de que se ha estimado un modelo, el siguiente paso para verificar el modelo es realizar un análisis residual. Específicamente, los residuos son denotados por $\hat{Z}_t = Y_t - \hat{Y}_t$ Luego se realice los siguiente:

- Hacer una gráfica de series de tiempo de \hat{Z}_t
- Trazar la ACF y PACF de \hat{Z}_t

Bajo la hipótesis nula de que $Z_t \sim WN(0,1)$, se puede demostrar que r_j (de $Z_t) \sim N(0,1/n)$ como regla general, se puede usar $\pm 2/\sqrt{n}$ para determinar si el r_j observado es significativamente diferente de cero. Pero recuerde que solo los rezagos cortos de r_j son importantes. Para determinar la no existencia de correlaciones entre los residuos se utiliza también el estadístico de Ljung-Box.

Test de Ljung – Box

Además de observar las correlaciones residuales en los rezagos individuales, es útil tener una prueba que tenga en cuenta sus magnitudes como grupo. Por ejemplo, puede ser que la mayoría de las autocorrelaciones residuales son moderadas, algunas incluso cercanas a sus valores críticos, pero, en conjunto, parecen excesivos. Box y Pierce (1970) propusieron el estadístico

$$Q = n(\hat{r}_1^2 + \hat{r}_2^2 + \dots + \hat{r}_k^2) \quad (2.13)$$

para abordar esta posibilidad. Demostraron que si se estima el modelo ARMA (p, q) correcto, entonces, para n grande, Q tiene una distribución de chi-cuadrado aproximada con $K - p - q$ grados de libertad. Ajustar un modelo erróneo tendería a inflar Q. Por lo tanto, el test rechazaría el modelo ARMA (p, q) si el valor observado de Q excedió un valor crítico apropiado en una distribución de chi-cuadrado con $K - p - q$ grados de libertad. El retardo máximo K se selecciona algo arbitrariamente, pero lo suficientemente grande que los pesos ψ sean despreciables para $j > K$.

La distribución de chi-cuadrado para Q se basa en un teorema del límite, Ljung y Box (1978) posteriormente descubrieron que incluso para $n = 100$, la aproximación no es satisfactoria. Al modificar ligeramente el estadístico Q, definieron un estadístico de prueba cuyo valor nulo. La distribución es mucho más cercana a la chi-cuadrado para tamaños de muestra típicos. El modificado. El estadístico Ljung-Box, viene dado por

$$Q_* = n(n + 2) \left(\frac{\hat{r}_1^2}{n-1} + \frac{\hat{r}_2^2}{n-2} + \dots + \frac{\hat{r}_k^2}{n-k} \right). \quad (2.14)$$

Observe que dado que $(n + 2) / (n - k) > 1$ para cada $k \geq 1$, tenemos $Q_* > Q$, que en parte explica por qué el estadístico Q original tendía a pasar por alto modelos inadecuados.

2.4 Pronóstico

Después de que se ha encontrado un modelo adecuado, se pueden llevar a cabo los pronósticos para un período, o varios en el futuro. También puede construirse intervalos de predicción con base en los pronósticos, en general para un nivel de confianza determinado, mientras más largo sea el tiempo guía de pronóstico, mayor será el intervalo de predicción. Esto es sensato porque se espera que la incertidumbre para el pronóstico de un valor lejano sea más grande que, para la observación próxima. A medida que se tiene más datos disponibles, se puede usar el mismo modelo, para generar pronósticos revisados que procedan de otro origen de tiempo.

Los pronósticos de los procesos SARIMA es completamente análoga a los pronósticos de proceso ARIMA. Expandiendo el operador $(1 - B)^d(1 - B^s)^D$ en potencias de B, reordenando la ecuación:

$$(1 - B)^d(1 - B^s)^D X_t = Y_t. \quad (2.15)$$

y establecer $t = n + h$ da el análogo

$$X_{n+h} = Y_{n+h} + \sum_{j=1}^{d+Ds} a_j X_{n+h-j} \quad (2.16)$$

Bajo el supuesto de que las primeras $d + D \times s$ observaciones $X_{-d-Ds+1}, \dots, X_0$ no están correlacionados con $\{Y_t, t \geq 1\}$, podemos determinar los mejores predictores lineales $P_n X_{n+h}$ de X_{n+h} basado en $\{1, X_{-d-Ds+1}, \dots, X_n\}$ aplicando P_n a cada lado, para obtener.

$$P_n X_{n+h} = P_n Y_{n+h} + \sum_{j=1}^{d+Ds} a_j P_n X_{n+h-j}. \quad (2.17)$$

El primer término de la derecha es simplemente el mejor predictor lineal del proceso ARMA (posiblemente una media distinta de cero) $\{Y_t\}$ en términos de $\{1, Y_1, \dots, Y_n\}$ que se puede calcular. Los predictores $P_n X_{n+h}$, se pueden calcular de forma recursiva para $h = 1, 2, \dots$ de (2.16), si observamos que $P_n X_{n+1-j} = X_{n+1-j}$ para cada $j \geq 1$. Un argumento análogo al que conduce a error cuadrado medio de la predicción como:

$$\sigma_n^2(h) = E(X_{n+h} - P_n X_{n+h})^2 = \sum_{j=0}^{h-1} \left(\sum_{r=0}^j \chi_r \theta_{n+h-r-1, j-r} \right)^2 v_{n+h-j-1}. \quad (2.18)$$

Donde θ_{nj} y v_n se obtienen aplicando el algoritmo de innovaciones a la diferencia serie $\{Y_t\}$ y

$$\chi(z) = \sum_{r=0}^{\infty} \chi_r z^r = [\phi(z)\Phi(z^s)(1-z)^d(1-z^s)^D]^{-1}, \quad |z| < 1. \quad (2.19)$$

Para n grande podemos aproximar (2.19), si $\theta(z)\Theta(z^s)$ es distinto de cero para todos $|z| \leq 1$, por

$$\sigma_n^2(h) = \sum_{j=0}^{h-1} \psi_j^2 \sigma^2, \quad (2.20)$$

donde

$$\psi(z) = \sum_{j=0}^{\infty} \psi_j z^j = \frac{\theta(z)\Theta(z^s)}{\phi(z)\Phi(z^s)(1-z)^d(1-z^s)^D}, \quad |z| < 1.$$

Los errores cuadráticos medios se calculan a partir de la aproximación de muestra grande (2.20) si el modelo ajustado es invertible. Si el modelo ajustado no es invertible, calcula el error cuadrado medio al convertir el modelo al equivalente (en términos de Gauss probabilidad) modelo invertible y luego usando la ecuación (2.20).

3. Aplicación usando el software R

3.1 Datos

El proyecto CRUTEM es un conjunto de datos derivados de las temperaturas del aire cerca de la superficie terrestre registradas en las estaciones meteorológicas de todos los continentes de la tierra, ha sido desarrollado y mantenido por la Climatic Research Unit desde principios de los años 80, con financiación proporcionada principalmente por el Departamento de Energía de los Estados Unidos. La última versión de CRUTEM se llama CRUTEM4 y está disponible en formato de texto y netCDF en la Climatic Research Unit en el Met Office Hadley Centre. Pero también es posible acceder a CRUTEM4 a través de Google Earth para ello solamente se debe descargar un archivo KML y abrirlo en Google Earth, y para realizar una consulta o descarga en cualquier parte del mundo, tan solo basta con hacer clic en el lugar deseado. Los datos utilizados en este estudio es de la variable temperatura ambiente en °C de la ciudad de New York, recopilados de la estación meteorológica por el proyecto CRUTEM. Los datos son mensuales y se extiende desde 2000 hasta 2019, que incluyen 240 observaciones.

La temperatura

Según la OMM (organización mundial de meteorología) define la temperatura como una cantidad física que caracteriza la media aleatoria movimiento de moléculas en un cuerpo físico. La temperatura se caracteriza por el comportamiento por el cual dos cuerpos en contacto térmico tienden a tener la misma temperatura. Por tanto, la temperatura representa el estado termodinámico de un cuerpo, y su valor está determinado por la dirección del flujo neto de calor entre dos cuerpos. En tal sistema, el cuerpo que en general pierde calor al otro se dice que está a la temperatura más alta. Definición de la temperatura de la cantidad física en relación al "estado de un cuerpo" sin embargo es difícil. Se encuentra una solución definiendo una escala de temperatura aprobada basada en congelación universal y puntos triples. La escala internacional de temperatura de 1990 (ITS-90) en el que se expresa la temperatura como t_{90} (temperatura Celsius) o T_{90} (temperatura kelvin) .

Para el rango meteorológico (-95°C a $+60^{\circ}\text{C}$), t_{90} se define mediante un conjunto bien especificado de termómetros de resistencia de platino calibrado en una serie de

definir puntos fijos y utilizando procedimientos de interpolación especificados (BIPM, 1989, 1990).

Requisitos y desempeño

La mayoría de los sistemas de radiosondas modernos (introducidos desde 2000) miden la temperatura en la tropósfera y en la estratosfera hasta una altura de unos 31 km con una incertidumbre ($k = 2$) de entre 0,4 y 1 K. Este rendimiento suele estar cerca del rendimiento óptimo para la predicción numérica del tiempo. Sin embargo, la incertidumbre muy por encima de $k = 2$ todavía se encuentra en algunas redes nacionales de radiosondas en regiones tropicales. Mediciones con errores tan grandes dañan pronósticos numéricos de predicción meteorológica, si se utilizan.

En la estratósfera, las incertidumbres de la temperatura de la radiosonda pueden estar cerca del objetivo de pronóstico del tiempo, pero requieren algunas mejoras en las condiciones diurnas para optimizar el clima.

Sensor de temperatura

Los mejores sensores de temperatura modernos tienen una velocidad de respuesta a los cambios de temperatura que es rápida suficiente para asegurar que el sesgo sistemático del retraso térmico durante un ascenso, tasa típica de ascenso 5 a 6 ms^{-1} , permanece menos de 0,1 K a través de cualquier capa de 1 km de profundidad en la troposfera y menos de 0,2 K a través de cualquier capa de profundidad similar en la estratósfera. A velocidades de ascenso típicas de radiosonda de 5-6 ms^{-1} , este es logrado en la mayoría de las ubicaciones utilizando un sensor con una constante de tiempo de respuesta más rápida que 1 s en la primera parte del ascenso. Además, los sensores de temperatura deben diseñarse para que estén lo más libres posible de errores de radiación introducidos por radiación solar directa o retro dispersada. Debe haber una variación tan pequeña como posible en el área de la sección transversal para la calefacción solar, ya que el sensor gira con respecto al sol durante ascenso. El intercambio de calor en el infrarrojo, debe evitarse mediante el uso de recubrimientos de sensor que tienen baja emisividad en el infrarrojo. En el pasado, los recubrimientos de sensor blancos más utilizados tenían una alta emisividad en el infrarrojo. Las mediciones de estos sensores estaban limitadas, especialmente en los niveles

superiores, por errores relativamente grandes del intercambio de calor infrarrojo, los errores dependieron de las temperaturas medido, el efecto de la nube en el fondo infrarrojo y el albedo de la superficie y no es fácil de corregir de forma fiable.

Los sensores de temperatura también deben ser lo suficientemente robustos para resistir golpes durante el lanzamiento y lo suficientemente estable como para mantener una calibración precisa durante varios años. Los principales tipos de sensores de temperatura, en uso rutinario son termistores (semiconductores resistivos cerámicos), sensores capacitivos, sensores bimetálicos y termopares.

Errores de temperatura

Los errores en la calibración del sensor de temperatura durante un ascenso pueden resultar de:

- Errores en la calibración de fábrica. Esto puede ocurrir de vez en cuando y es una de las razones por las que las mediciones de la radiosonda deben comprobarse en tierra antes del lanzamiento.
- Pequeños cambios en el sensor, la capacitancia parásita asociada con un sensor capacitativo o en las conexiones eléctricas al sensor.
- Inestabilidades en el sistema transductor de radiosonda y referencias. Esto es posible durante el almacenamiento. o durante el ascenso. La desviación del sensor o del transductor durante el almacenamiento generalmente se puede corregir parcialmente durante el procesamiento de datos, utilizando ajustes basados en verificaciones en tierra previas al vuelo.

3.2 Identificación del modelo

Para iniciar el estudio de la serie se procede a ejecutar la lectura de los datos y realizar la gráfica respectiva utilizando el software R.

Temperatura en New York

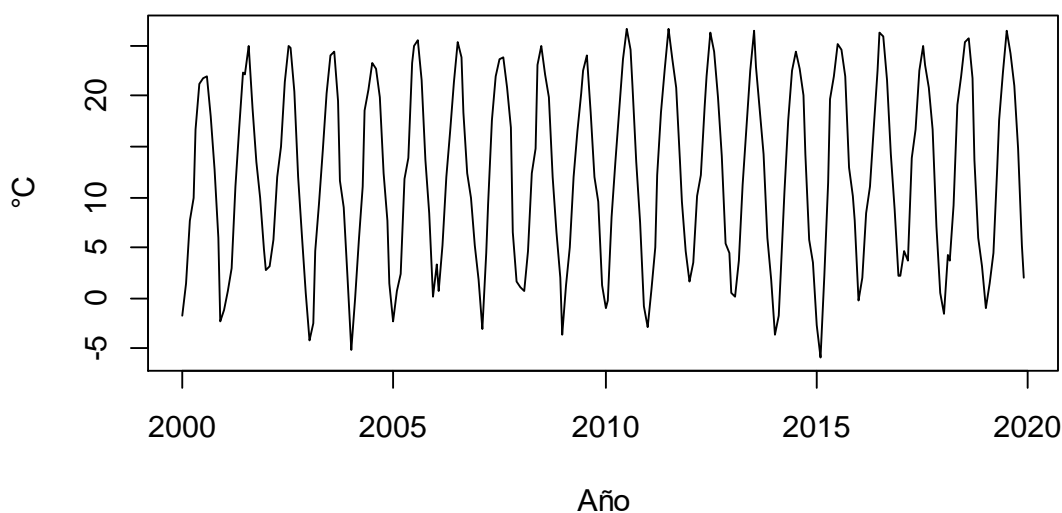


Fig. 12 Temperatura en New York

La Fig.12, muestra la temperatura media como varían a lo largo del tiempo en la ciudad de New York, parece mostrar una ligera tendencia, además para poder determinar los patrones de comportamiento es necesario el método de descomposición de series de tiempo estacional es adecuado para nuestros datos. En la fig. 13, se tiene la descomposición de la serie de tiempo.

Decomposition of additive time series

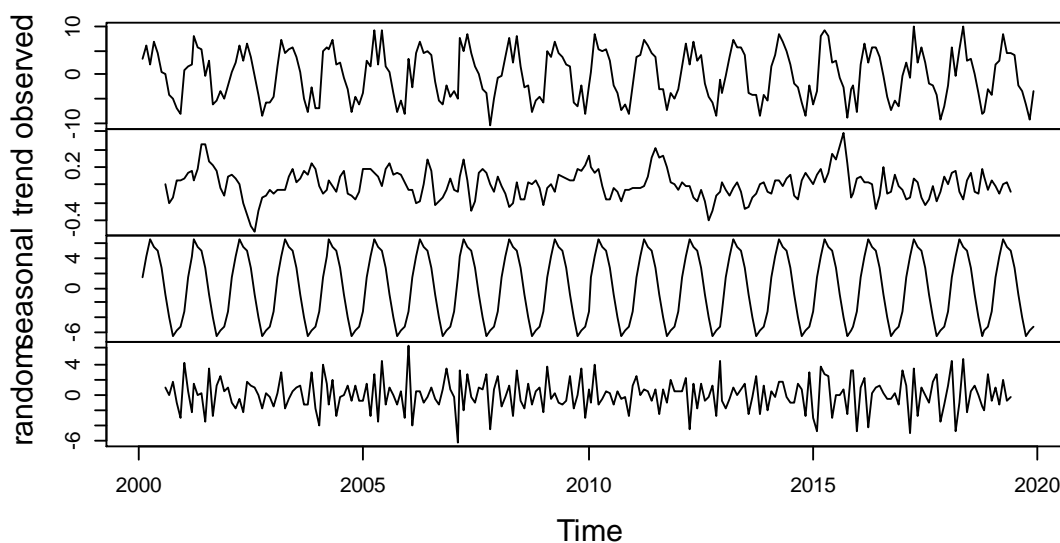


Fig. 13 Gráfica de descomposición de la temperatura mensual.

En la Fig.13 podemos apreciar el patrón de los datos, existe la componente estacional que se repite cada año, Por otra parte, la tendencia primeramente disminuye luego tiende a aumentar muy ligeramente, la parte aleatoria representa la configuración de los residuos que va dejando el modelo, cuando disminuye y cuando aumenta.

Aplicamos el test de Dickey – Fuller a los datos originales, para verificar si la serie es estacionaria.

Augmented Dickey-Fuller Test

data: q

Dickey-Fuller = -3.134, Lag order = 12, p-value = 0.09958

alternative hypothesis: stationary

Según el valor $p = 0.09958 > 0.05$ por tanto, se acepta la hipótesis nula, es decir la serie no es estacionaria. Para poder eliminar la existencia de la componente estacional, se realizó una diferencia en la serie de tiempo el resultado a continuación:

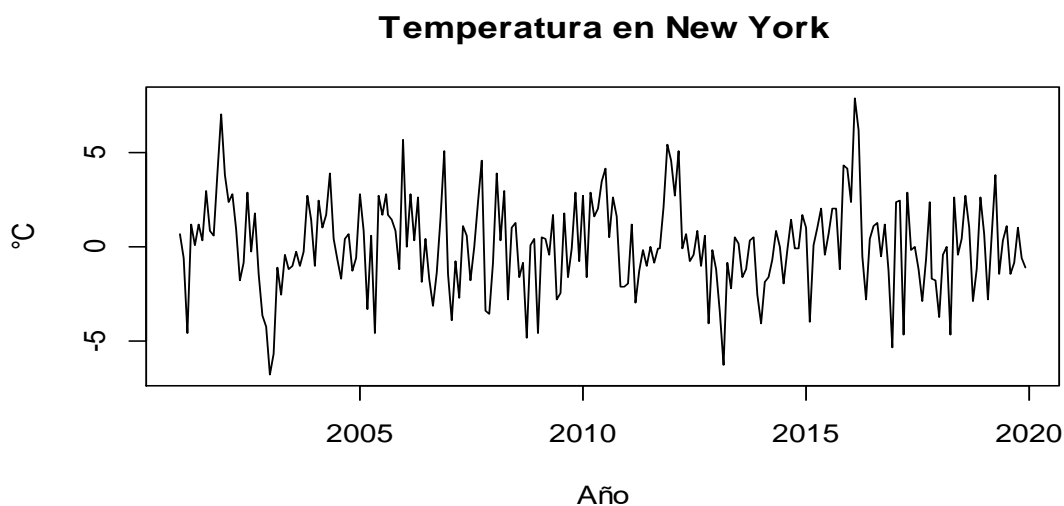


Fig. 14 Grafica de serie de tiempo primera diferencia.

En la Fig. 14, podemos observar que la serie se parece a una serie estacionaria, para asegurarnos aplicamos el test de ADF (Dickey Fuller)

Augmented Dickey-Fuller Test

data: di1

Dickey-Fuller = -5.9355, Lag order = 12, p-value = 0.01

alternative hypothesis: stationary

Según el valor $p = 0.01 < 0.05$ por tanto, se acepta la hipótesis alternativa es decir la serie es estacionaria, con el fin de seleccionar las ordenes de (p) y (q) del modelo examinamos las gráficas del ACF y PACF (autocorrelación simple y parcial).

Autocorrelación de temperatura en New York

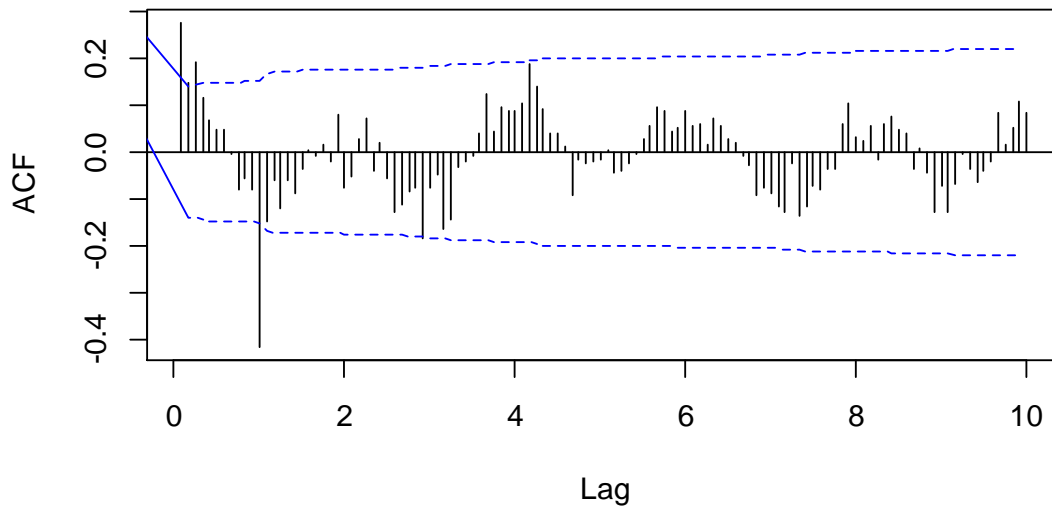


Fig. 15 Gráfica ACF, primera diferencia.

Autocorrelación parcial de temperatura en New York

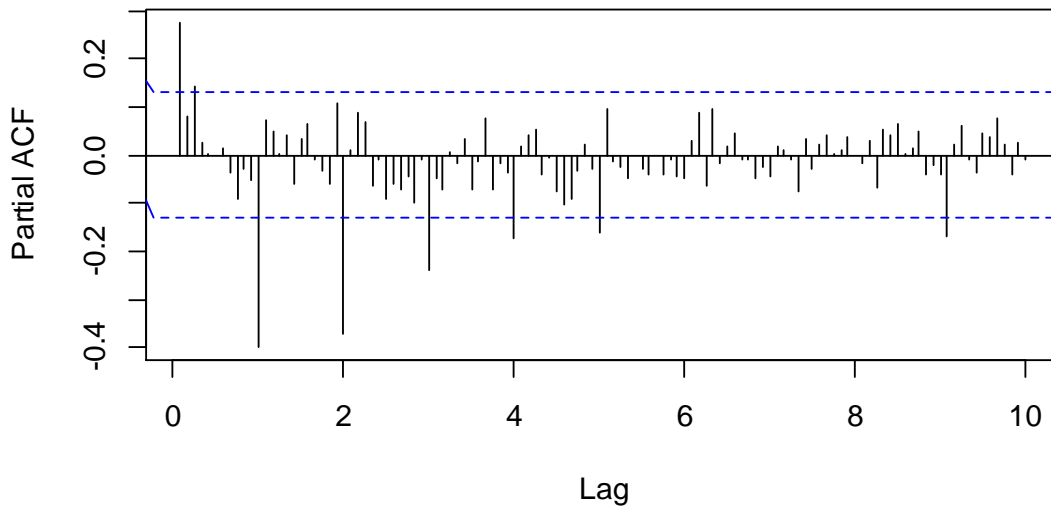


Fig. 16 Gráfica PACF, primera diferencia.

Ahora podemos determinar los polinomios AR y MA. Obsérvese los gráficos de la función de autocorrelación parcial PACF en la Fig. 16, tiene un patrón ondulado y se reduce a cero, lo que sugiere un proceso AR(3) o más. En la Fig.15, la función de autocorrelación ACF, tiene un retraso significativo en una primera conjetura, lo que sugiere un MA(1).

3.3 Estimación

Iniciamos con el ajuste del modelo SARIMA (1,0,1)x(0,1,3)[12] a continuación el resultado:

SARIMA(1,0,1)x(0,1,3)[12]

Call:

```
arima(x = q, order = c(1, 0, 1), seasonal = list(order = c(0, 1, 3)))
```

Coefficients:

	ar1	ma1	sma1	sma2	sma3
	0.7658	-0.5485	-0.9193	-0.1109	0.1614
s.e.	0.1170	0.1510	0.0732	0.0839	0.0810

sigma^2 estimated as 2.93: log likelihood = -456.07, aic = 922.14

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	0.7658	0.1170	6.54	<.0001
ma1	-0.5485	0.1510	-3.63	<.0001
sma1	-0.9193	0.0732	-12.56	<.0001
sma2	-0.1109	0.0839	-1.32	0.1857
sma3	0.1614	0.0810	1.99	0.0468

```

ar1  0.765784  0.116979  6.5464  5.896e-11 ***
ma1  -0.548451  0.151022  -3.6316  0.0002817 ***
sma1 -0.919331  0.073188  -12.5612 < 2.2e-16 ***
sma2 -0.110908  0.083866  -1.3224  0.1860209
sma3  0.161387  0.081005  1.9923  0.0463382 *

```

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

	2.5 %	97.5 %
ar1	0.536510268	0.99505847
ma1	-0.844447440	-0.25245376
sma1	-1.062777762	-0.77588447
sma2	-0.275282558	0.05346644
sma3	0.002619384	0.32015486

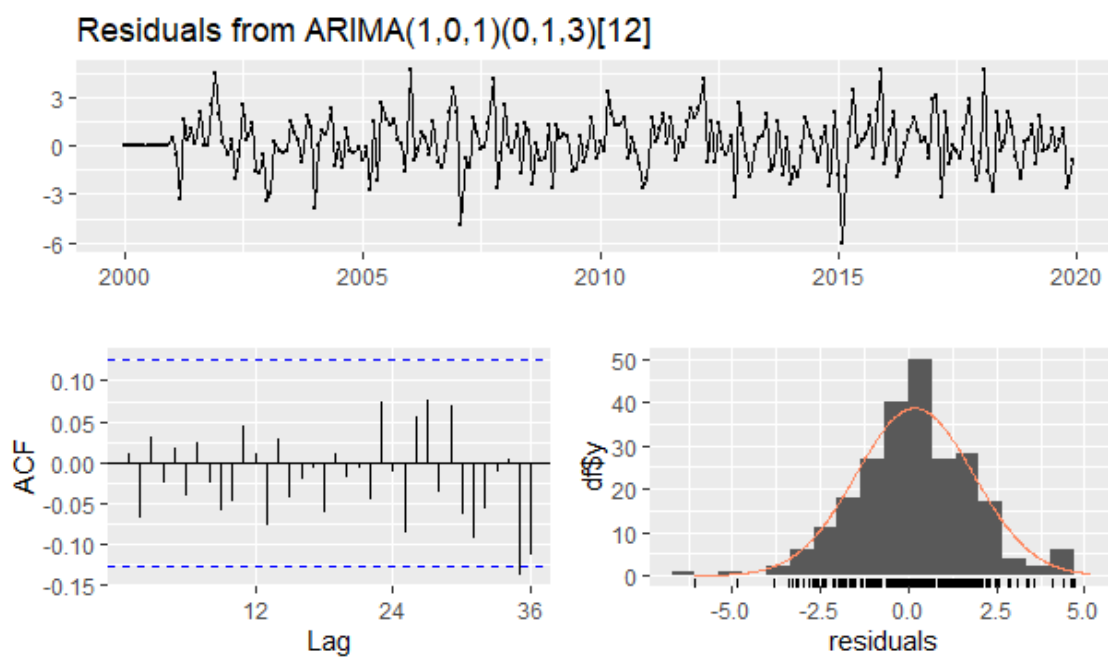


Fig. 17 Gráfico de residuos

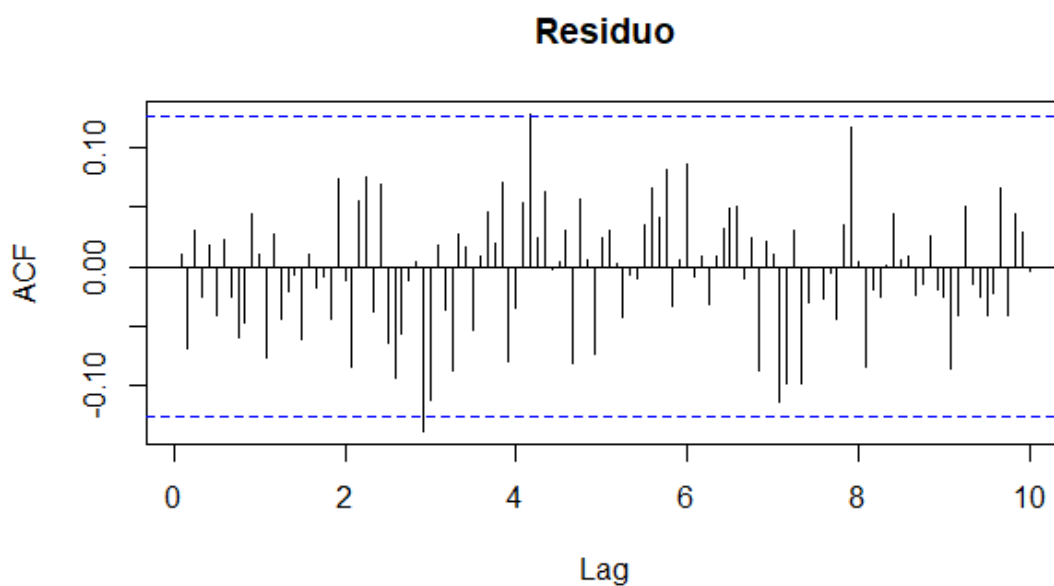


Fig. 18 ACF de Residuo

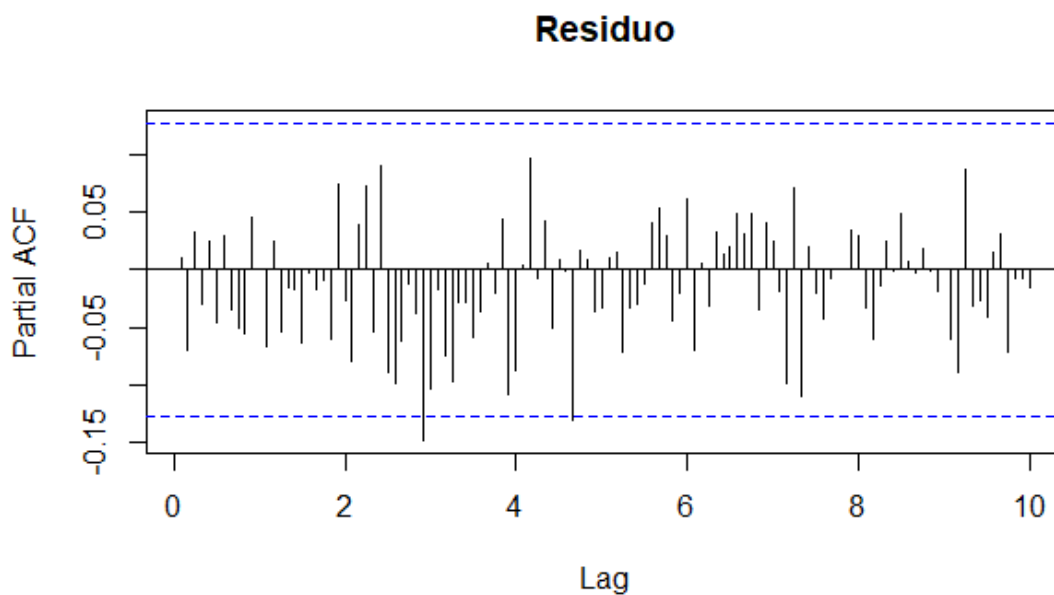


Fig. 19 PACF de Residuo

Para el contraste de hipótesis si los datos se distribuyen de forma independiente, según el P-valor = $0.87 > 0.05$, evidentemente, los datos de la serie de tiempo se distribuyen de forma independiente.

En la Fig. 20, observamos que los datos siguen una distribución normal, para una mayor seguridad comprobamos la normalidad mediante el test de Shapiro Wilk. El p-valor obtenido es $0.059 > 0.05$, por lo que no rechazamos nuestra hipótesis nula, es decir que los residuos siguen una distribución normal.

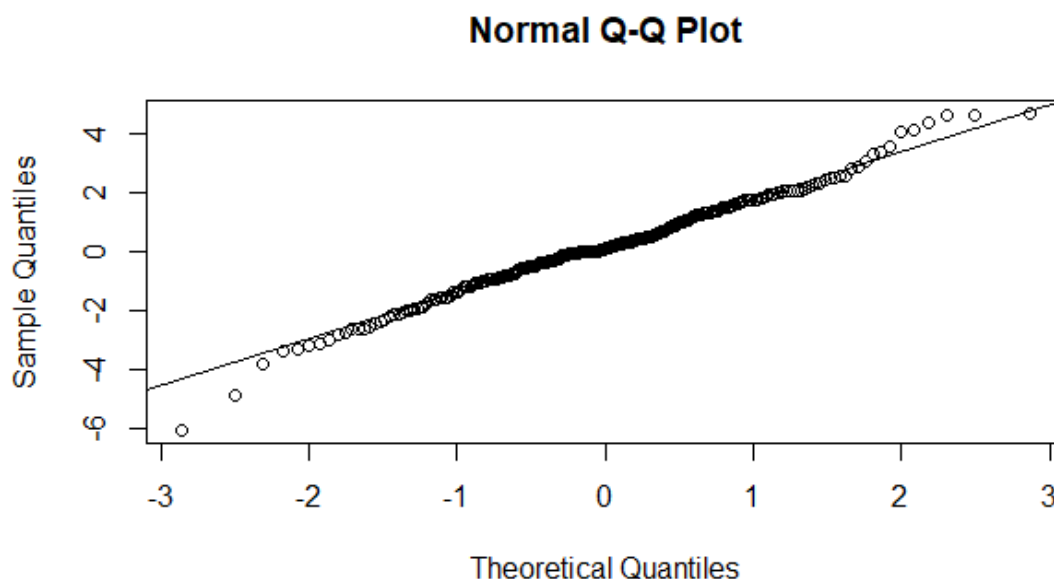


Fig. 20 Gráfica de Cuantiles teóricos

Sin embargo, el modelo SARIMA (1,0,1)x(0,1,3,) [12], posee coeficientes no significativos sma2 y sma3, cumple con la independencia y normalidad, en la fig. 18 y 19 (autocorrelación y autocorrelación parcial), hay retardos significativos en los residuos, por lo que no es ruido blanco, es decir el modelo no es válido. Ahora realizaremos un ajuste haciendo $sma2 = 0$.

SARIMA (1,0,1)x(0,1,3) [12]

Call:

```
arima(x = q, order = c(1, 0, 1), seasonal = list(order = c(0, 1, 3), period = 12), fixed = c(NA, NA, NA, 0, NA))
```

Coefficients:

	ar1	ma1	sma1	sma2	sma3
	0.7757	-0.5654	-0.975	0	0.1037
s.e.	0.1211	0.1578	0.061	0	0.0657

sigma^2 estimated as 2.952: log likelihood = -456.9, aic = 921.8

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	0.775705	0.121129	6.4040	1.514e-10 ***
ma1	-0.565412	0.157773	-3.5837	0.0003388 ***
sma1	-0.974952	0.060957	-15.9942	< 2.2e-16 ***
sma3	0.103725	0.065691	1.5790	0.1143409

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

	2.5 %	97.5 %
ar1	0.53829650	1.0131136
ma1	-0.87464147	-0.2561824
sma1	-1.09442438	-0.8554789
sma2	NA	NA
sma3	-0.02502743	0.2324779

En el modelo ajustado con restricciones, el coeficiente sma3 no es significativo nuevamente. Así que modificamos el modelo eliminando este coeficiente y ajustamos un **SARIMA (1,0,1)x(0,1,1) [12]**

Call:

arima(x = q, order = c(1, 0, 1), seasonal = list(order = c(0, 1, 1), period = 12))

Coefficients:

	ar1	ma1	sma1
	0.7955	-0.5918	-0.9229
s.e.	0.1088	0.1435	0.0679

sigma^2 estimated as 2.949: log likelihood = -458.05, aic = 922.1

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)
ar1	0.795546	0.108761	7.3146	2.581e-13 ***
ma1	-0.591829	0.143472	-4.1250	3.707e-05 ***

sma1 -0.922948 0.067911 -13.5905 < 2.2e-16 ***

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

2.5 % 97.5 %

ar1 0.5823776 1.0087134

ma1 -0.8730286 -0.3106284

sma1 -1.0560518 -0.7898441

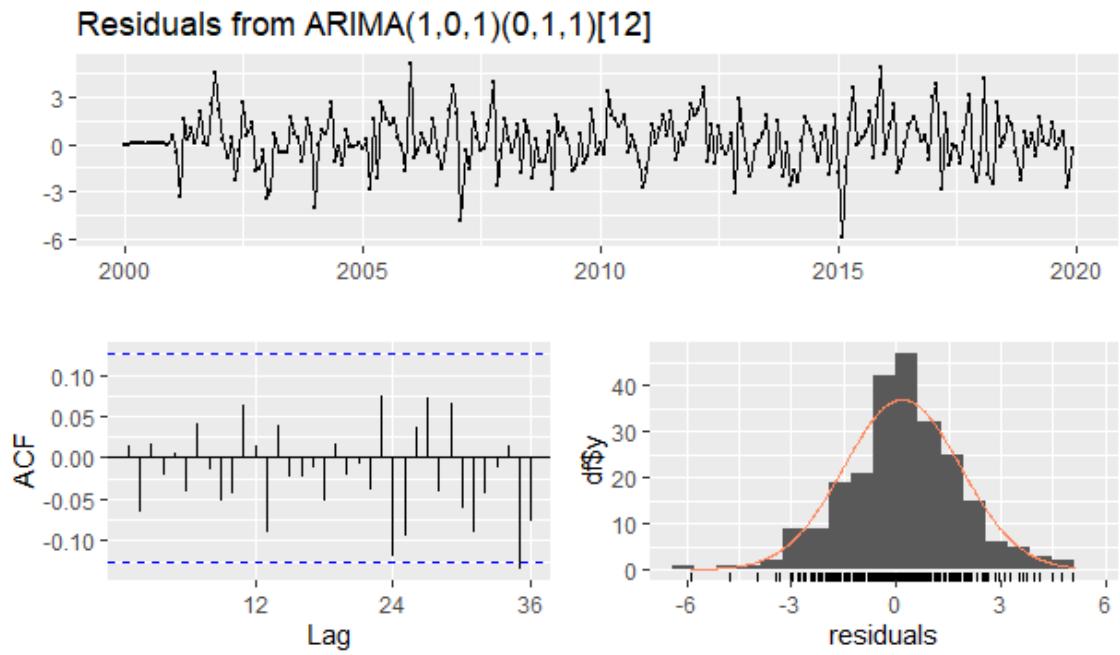


Fig. 21 Gráfico de residuos

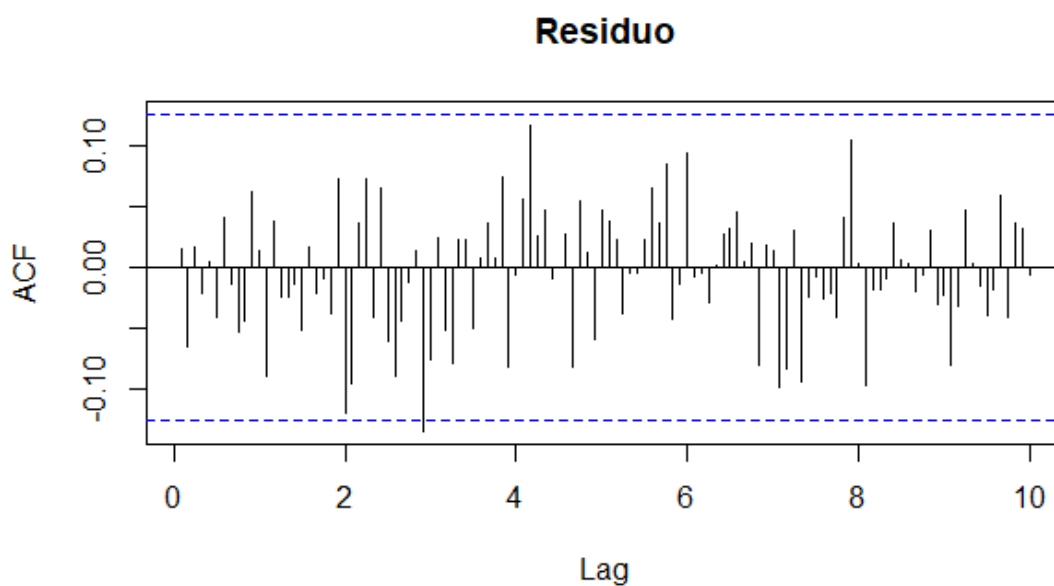


Fig. 22 Gráfico ACF de residuo

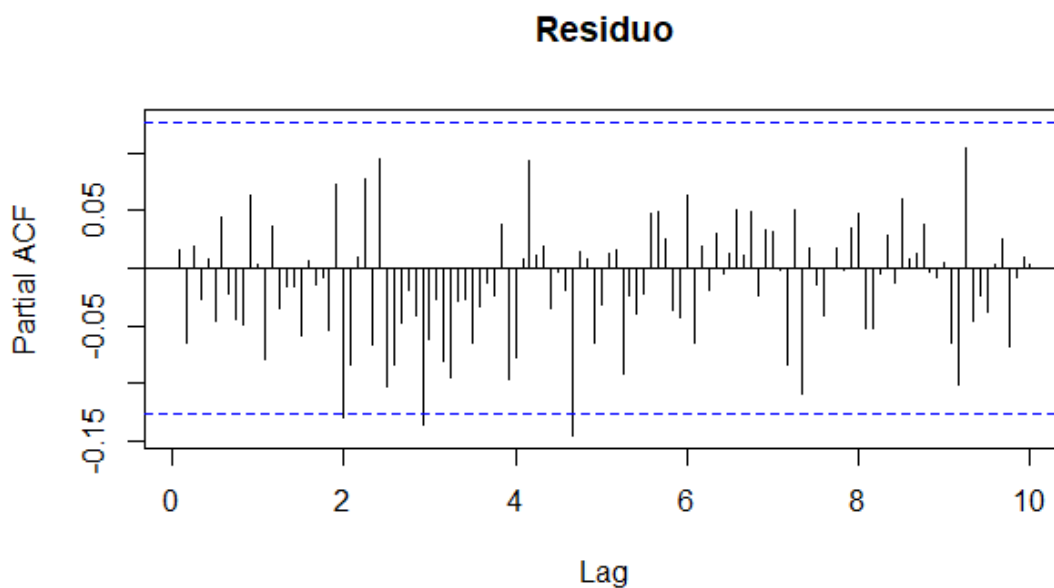


Fig. 23 Gráfico PACF de residuo

Se puede apreciar los gráficos de residuo en la fig. 22, y en la fig. 23, no existe retardo significativo, podemos decir que los residuos son ruido blanco.

Ahora realizaremos el contraste de hipótesis si los datos se distribuyen de forma independiente, el P-valor = $0.81 > 0.05$, evidentemente, los datos de la serie de tiempo se distribuyen de forma

independiente.

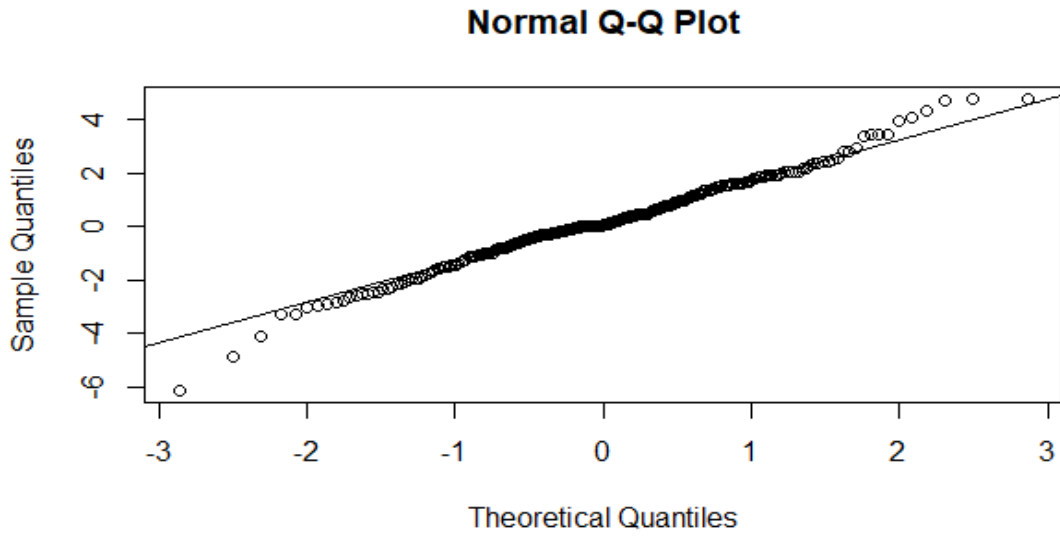


Fig. 24 Gráfica de Cuantiles teóricos

En la Fig. 24. observamos que los datos siguen una distribución normal, para una mayor seguridad comprobamos la normalidad mediante el test de Shapiro Wilk. Se tiene que el valor p-valor del test es de $0.15 > 0.05$, por lo que no rechazamos nuestra hipótesis nula, es decir que los residuos siguen una distribución normal, el modelo cumple con la independencia, normalidad y los residuos son ruido blanco, por lo que el modelo es válido.

3.5 Pronóstico

	Point Forecast	Lo 80	Hí 80	Lo 95	Hí 95
Jan 2020	-0.4467978	-2.6626923	1.769097	-3.835716	2.942120
Feb 2020	0.8933430	-1.3532656	3.139952	-2.542548	4.329234
Mar 2020	5.1507925	2.8738842	7.427701	1.668562	8.633023
Apr 2020	11.4410744	9.1342642	13.747884	7.913113	14.969036
May 2020	17.1912857	14.8549564	19.527615	13.618178	20.764393
Jun 2020	22.1434867	19.7780066	24.508967	18.525797	25.761176
Ju1 2020	24.9285163	22.5342403	27.322792	21.266787	28.590246
Aug 2020	24.2181061	21.7953764	26.640836	20.512861	27.923351
Sep 2020	20.3720192	17.9211662	22.822872	16.623763	24.120276
Oct 2020	13.6198482	11.1411909	16.098506	9.829069	17.410628
Nov 2020	7.5911705	5.0850173	10.097324	3.758340	11.424001
Dec 2020	2.3504191	-0.1829640	4.883802	-1.524056	6.224894
Jan 2021	-0.4467978	-3.0070891	2.113493	-4.362425	3.468830
Feb 2021	0.8933430	-1.6935766	3.480263	-3.063009	4.849695
Mar 2021	5.1507925	2.5375160	7.764069	1.154131	9.147454
Apr 2021	11.4410744	8.8017040	14.080445	7.404506	15.477643
May 2021	17.1912857	14.5260769	19.856494	13.115201	21.267371
Jun 2021	22.1434867	19.4526877	24.834286	18.028265	26.258709
Ju1 2021	24.9285163	22.2123682	27.644664	20.774526	29.082506
Aug 2021	24.2181061	21.4768432	26.959369	20.025706	28.410506
Sep 2021	20.3720192	17.6058696	23.138169	16.141558	24.602480
Oct 2021	13.6198482	10.8290338	16.410663	9.351666	17.888031
Nov 2021	7.5911705	4.7759073	10.406434	3.285597	11.896744
Dec 2021	2.3504191	-0.4891113	5.189950	-1.992268	6.693106

Pronóstico de temperatura en New York

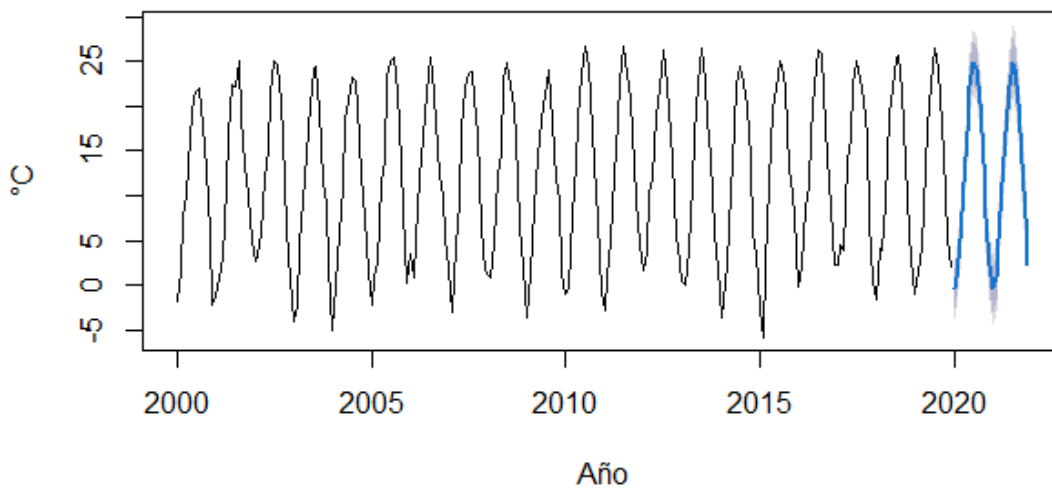


Fig. 26 Gráfica de pronóstico de Temperatura en New York

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	0.03026281	1.677883	1.331042	Inf	Inf	0.709889	0.1006234

El pronóstico se muestra con una línea azul, entre los intervalos de predicción del 80% y del 95%, el error promedio en el pronóstico (ME) es de 0.03 y el error absoluto medio del pronóstico (MAE) es de 1.33 esto significa que es un mejor modelo de predicción.

Conclusión

La metodología de Box & Jenkins nos otorga una manera sencilla y eficaz de ajustar series de tiempo de manera rápida y relativamente sencilla para realizar predicciones fiables para el futuro.

El modelo SARIMA (1,0,1) (0,1,1) [12] que se obtuvo, a pesar de ser un modelo apropiado para series de tiempo, es útil en corto plazo, existe problemas debido a su falla en la predicción, especialmente si en el pasado la secuencia de series de tiempo tiene cambios anormales. En la predicción el error promedio (ME) es de 0.03 y el error absoluto medio (MAE) es de 1.33 esto significa que es un buen modelo de pronóstico.

Bibliografía

- Anderson, T. W. (1971). *The Statistical Analysis of Time Series*. New York: John Wiley.
- Bartlett, M. S. (1946). *Journal of the Royal Statistical Society*, 27- 41.
- Broom, T. A. (2017). Seasonal Time-Series Modelling and Forecasting of Monthly Mean Temperature for Decision Making in the Kurdistan Region of Iraq. *City Research Online*, 1- 35. doi: 10.1080/15598608.2017.1292484
- Chan, N. H. (2002). *Time Series Applications to Finance*. Canada: A John Wiley & Sons Inc Publication.
- Damador N.Gujarati, D. C. (2004). *Econometría*. México DF: Mc.Graw Hill.
- Hellesnes, C. (2017). Forecast evaluation of dynamic regression and sarima models applied to Electricity Spot Prices, Time Series Analysis. *Matematisk statistik*, 1-35. Retrieved from www.math.su.se
- Hyndman, P. R. (2011). Forecasting Time Series Using R. *Monash University*, 1-152.
- Jonathan D. Cryer, K.-S. C. (2008). *Time Series Analysis with Applications in R*. Springer.
- Nash, J. (2015). Measurement of upper - Air Pressure, Temperature and Humidity. *World Meteorological Organization*, 1- 89. Retrieved from https://library.wmo.int/doc_num.php?explnum_id=7366
- Paul S.P.Cowpertwait, A. V. (2008). *Introductory Time Series With R*. New York: Springer.
- Pelgrin, F. (2011). Box-Jenkins Methodology. *University of Lausanne, Ecole des HEC*, 1-32. Retrieved from <https://math.unice.fr> › Chapitre_5_IMEA_1
- Peter J. Brockwell, R. A. (2002). *Introduction to Time Series and Forecasting, Second Edition*. New York: Springer.
- Rao, S. S. (2021). A course in Time Series Analysis. Retrieved from https://web.stat.tamu.edu/~suhasini/teaching673/time_series.pdf
- Robert H. Shumway, D. S. (2016). *Time series analysis and its applications with R examples*. Springer.
- Roberto S. Pindyck, D. L. (2001). *Econometría Modelos y Pronósticos*. México, D.F.: McGRAW HILL.
- Shengwei Wang, J. F. (2013). Application of Seasonal Time Series Model in the Precipitation Forecast. *Elsevier*, 677-683. Retrieved from www.elsevier.com/locate/mcm
- Shumway, R. H. (2006). *Time Series Analysis and Its Applications (with R Example)*. New York: Springer.
- Yule, G. U. (1926). "Why do we sometimes get nonsense-correlations between time series - A study in sampling and the nature of time series". *Journal of the Royal Statistical*, 1-63.

ANEXO 1: Código R

```
library(readxl)
dato <- read_excel("C:/Users/Usuario/Desktop/dato.xlsx")
View(dato)
library(astsa)
library(forecast)
library(quantmod)
library(tseries)
library(lmtest)
library(TSA)
#trimestral#
q<-ts(dato$TP,start =c(2000,1),frequency = 12,end=c(2019,12))
q
#cuantas diferencias se necesita,parte estacional#
plot(q,main="Temperatura en New York",xlab="Año",ylab=""C")
adf.test(q,alternative = "stationary", k=12)
plot(decompose(q))
#autocorrelacion#
ndiffs(q)
nsdiffs(q)
#diferencia parte estacional#
di1<-diff(q,lag = 12)
di1
nsdiffs(di1)
adf.test(di1,alternative = "stationary", k=12)
plot(di1,main="Temperatura en New York",xlab="Año",ylab=""C")
acf(di1,main = "Autocorrelación de temperatura en New York",lag=120,ci.type="ma")
pacf(di1,main="Autocorrelación parcial de temperatura en New York",lag=120)
#estimacion#
modelo1<-arima(q,order =c(0,0,1),seasonal=list(order=c(0,1,1),period=12))
modelo1
coeftest(modelo1)
confint(modelo1)
#residuos#
checkresiduals(modelo1)
acf(modelo1$residuals,main="Residuo",lag.max=120,ci.type="ma")
pacf(modelo1$residuals, main="Residuo",lag.max=120)
#comprobacio de supuestos de Modelo#
independencia <- Box.test(modelo1$residuals,type="Ljung-Box") # Test de Ljung-Box
independencia$p.value
qqnorm(modelo1$residuals)
qqline(modelo1$residuals)
normalidad<-shapiro.test(modelo1$residuals) # Test de Shapiro-Wilk
normalidad$p.value
#pronostico #
pro<-forecast(modelo1,h=24)
pro
plot(pro,main = "Pronóstico de temperatura en New York",xlab="Año",ylab=""C")
accuracy(pro)
```

ANEXO 2: Datos de temperatura ambiente de New York.

YR	TP
2000	-1,8
2000	1,4
2000	7,6
2000	9,9
2000	16,7
2000	21,1
2000	21,8
2000	22
2000	17,9
2000	12,8
2000	6
2000	-2,2
2001	-1,1
2001	0,8
2001	3
2001	11,1
2001	16,8
2001	22,3
2001	22,1
2001	25
2001	18,7
2001	13,4
2001	9,9
2001	4,9
2002	2,7
2002	3,2
2002	5,8
2002	12
2002	15
2002	21,4
2002	25
2002	24,7
2002	20,5
2002	11,9
2002	6,2
2002	0,6
2003	-4,1
2003	-2,5
2003	4,7
2003	9,4
2003	14,6
2003	20,2

2003	24
2003	24,4
2003	19,5
2003	11,6
2003	8,9
2003	2
2004	-5,1
2004	0
2004	5,7
2004	11,1
2004	18,5
2004	20,6
2004	23,3
2004	22,7
2004	19,9
2004	12,3
2004	7,6
2004	1,4
2005	-2,3
2005	0,8
2005	2,4
2005	11,7
2005	13,9
2005	23,3
2005	25
2005	25,5
2005	21,6
2005	13,7
2005	8,4
2005	0,2
2006	3,4
2006	0,8
2006	5,2
2006	12
2006	16,5
2006	21,4
2006	25,4
2006	23,8
2006	18,4
2006	12,3
2006	9,9
2006	5,3
2007	1,9
2007	-3,1

2007	4,4
2007	9,3
2007	17,6
2007	22
2007	23,6
2007	23,8
2007	20,8
2007	16,9
2007	6,5
2007	1,7
2008	1
2008	0,8
2008	4,7
2008	12,3
2008	14,8
2008	23
2008	24,9
2008	22,2
2008	19,9
2008	12
2008	6,6
2008	2,1
2009	-3,6
2009	1,3
2009	5,1
2009	11,9
2009	16,5
2009	20,2
2009	22,4
2009	24
2009	18,3
2009	11,9
2009	9,5
2009	1,3
2010	-0,9
2010	-0,3
2010	8
2010	13,5
2010	18,5
2010	23,7
2010	26,6
2010	24,5
2010	20,9
2010	13,5

2010	7,4
2010	-0,8
2011	-2,9
2011	0,9
2011	5
2011	12,2
2011	18,3
2011	22,7
2011	26,6
2011	23,6
2011	20,8
2011	13,4
2011	9,5
2011	4,6
2012	1,7
2012	3,6
2012	10,1
2012	12,1
2012	19
2012	21,9
2012	26,2
2012	24,4
2012	19,8
2012	14
2012	5,4
2012	4,4
2013	0,5
2013	0,1
2013	3,8
2013	11,2
2013	16,8
2013	22,4
2013	26,4
2013	22,8
2013	18,6
2013	14,3
2013	5,9
2013	1,8
2014	-3,6
2014	-1,8
2014	2,2
2014	10,5
2014	17,6
2014	22,4

2014	24,4
2014	22,6
2014	20
2014	14,2
2014	5,8
2014	3,5
2015	-2,6
2015	-5,8
2015	2,3
2015	11,5
2015	19,6
2015	22
2015	25,1
2015	24,6
2015	22
2015	13
2015	10,1
2015	7,7
2016	-0,2
2016	2,1
2016	8,5
2016	11
2016	16,8
2016	22,4
2016	26,2
2016	25,9
2016	21,5
2016	14,2
2016	8,8
2016	2,3
2017	2,2
2017	4,6
2017	3,8
2017	13,9
2017	16,6
2017	22,4
2017	25
2017	23
2017	20,8
2017	16,6
2017	7,1
2017	0,5
2018	-1,6
2018	4,2

2018	3,8
2018	9,2
2018	19,2
2018	22
2018	25,4
2018	25,7
2018	21,8
2018	13,7
2018	5,9
2018	3,1
2019	-0,9
2019	1,4
2019	4,4
2019	13
2019	17,7
2019	22,3
2019	26,5
2019	24,2
2019	20,9
2019	14,7
2019	5,3
2019	2