

Universidad Nacional De Colombia
Sede Medellín



Facultad de Ciencias
Departamento de Estadística

Trabajo 3
Pronósticos para series de tiempo univariadas.

Juan David García Zapata
Juan Camilo Miranda Paz

Series de tiempo Univariadas

septiembre, 2024

Índice

1. Punto 1	2
1.1. Introducción	2
1.2. Literal A)	3
1.3. Literal B)	5
1.4. Literal C)	9
1.5. Literal D)	12
2. Punto 2	18
3. Punto 3	21
4. Punto 4)	27

1. Punto 1

Enunciado

Para la serie de tiempo considerada, **separe los datos correspondientes a los últimos 2 años para evaluar pronósticos de la serie** y ajuste un modelo ARIMA estacional, a la parte restante de los datos discutiendo el proceso de ajuste paso a paso, presentando

- A) Gráfica y examen inicial de los datos.
- B) Transformación de los datos en caso de ser necesaria.
- C) Orden de diferenciación e identificación del modelo (grados de los polinomios).
- D) Estimación de parámetros y diagnóstico de residuales del modelo seleccionado.

1.1. Introducción

Librerías Utilizadas

```
library(ggplot2)
library(tseries)
library(car)
library(TSA)
library(foreign)
library(lmtest)
library(quantmod)
library(astsa)
library(MASS)
library(forecast)
library(urca)
library(fANCOVA)
library(fUnitRoots)
```

En este punto se analizarán los datos correspondientes a la base de datos **UnempRate**, dado que la último dígito de la cédula de cada integrante es 0 (En el caso de Juan David) y 2 (En el caso de Juan Camilo) por lo tanto el residuo que se obtiene al dividir la suma entre 4 es $2(2 \% \%4=2)$

Lectura de base de datos.

```
serie_UnempRate <- astsa::UnempRate
```

La base de datos **UnempRate** del paquete **astsa** en R proporciona una serie temporal mensual de la tasa de desempleo en Estados Unidos, expresada en porcentaje. Estos datos abarcan un período prolongado y reflejan las fluctuaciones en la tasa de desempleo, permitiendo identificar tendencias a largo plazo, ciclos económicos, efectos estacionales y posibles eventos externos que influyan en la economía.

```
start(serie_UnempRate)
end(serie_UnempRate)
[1] 1948    1
[1] 2016   11
#longitud de la serie
length(serie_UnempRate)
[1] 827
```

El formato de los datos es una serie temporal de longitud 827, que abarca desde enero de 1948 hasta noviembre de 2016.

```
frequency(UnempRate)
[1] 12
```

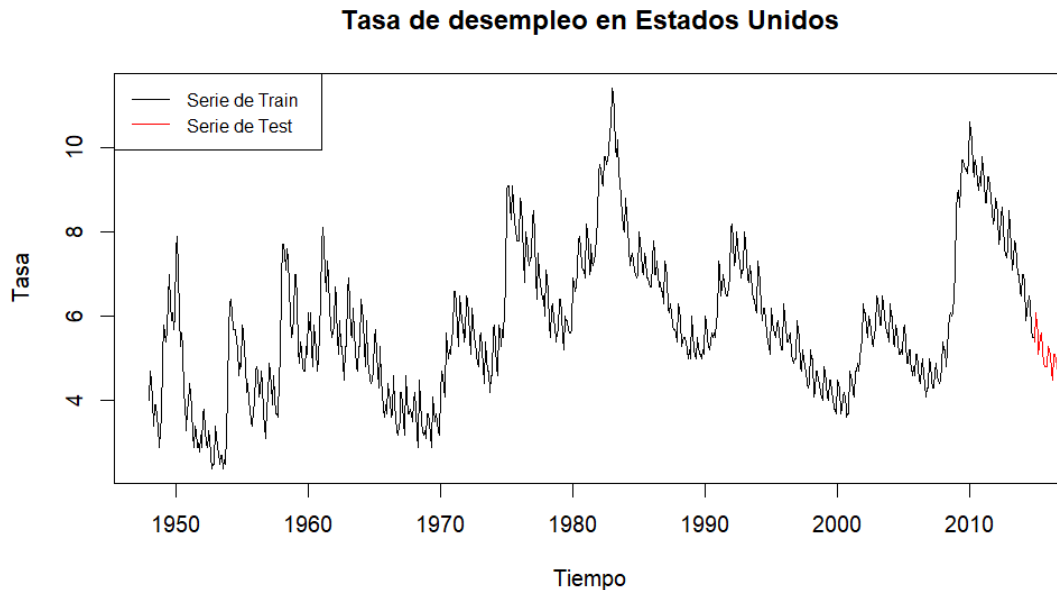
La serie temporal presenta una estacionalidad con un ciclo de 12 periodos, lo que indica que las fluctuaciones recurrentes ocurren anualmente, con patrones que se repiten cada 12 meses.

Separar los datos correspondientes a los últimos 2 años para evaluar pronósticos de la serie

```
t_pronosticos <- 24
n_total <- length(serie_UnempRate)
n_ajuste <- n_total - t_pronosticos
Datos_ajuste <- window(serie_UnempRate, end = c(2014, 11))
Datos_pronostico <- window(serie_UnempRate, start = c(2014, 12),
end = c(2016, 11))
```

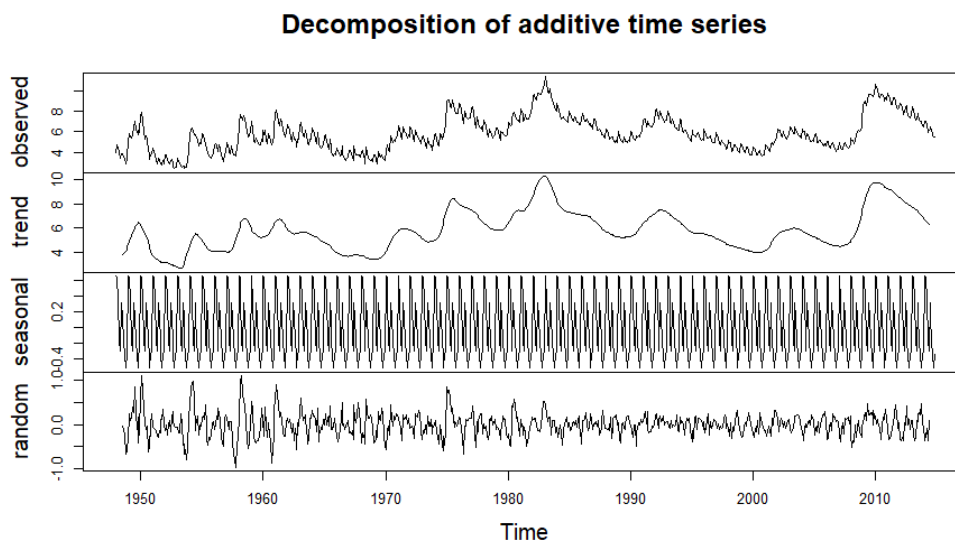
1.2. Literal A)

```
plot(Datos_ajuste, main = "Tasa de desempleo en Estados Unidos",
ylab = "Tasa", xlab = "Tiempo")
lines(Datos_pronostico, col = "red")
legend("topleft", legend = c("Serie de Train", "Serie de Test"),
col = c("black", "red"), lty = 1, cex = 0.8)
```



La tasa de desempleo en Estados Unidos muestra un comportamiento **no estacionario**, ya que no se mantiene alrededor de una media constante ni presenta una varianza fija a lo largo del tiempo. Las fluctuaciones en los niveles de desempleo, con picos importantes en algunos periodos y fases de recuperación posteriores, indican que la serie tiene una tendencia a cambiar y ajustarse en respuesta a eventos económicos significativos. Por tanto, no sigue un patrón estable, sino que refleja episodios de alta variabilidad y cambios en la dinámica económica del país.

```
des.serie <- decompose(Datos_ajuste)
plot(des.serie)
```



la descomposición de la serie de tiempo aditiva en tres componentes: tendencia, estacionalidad y ruido.

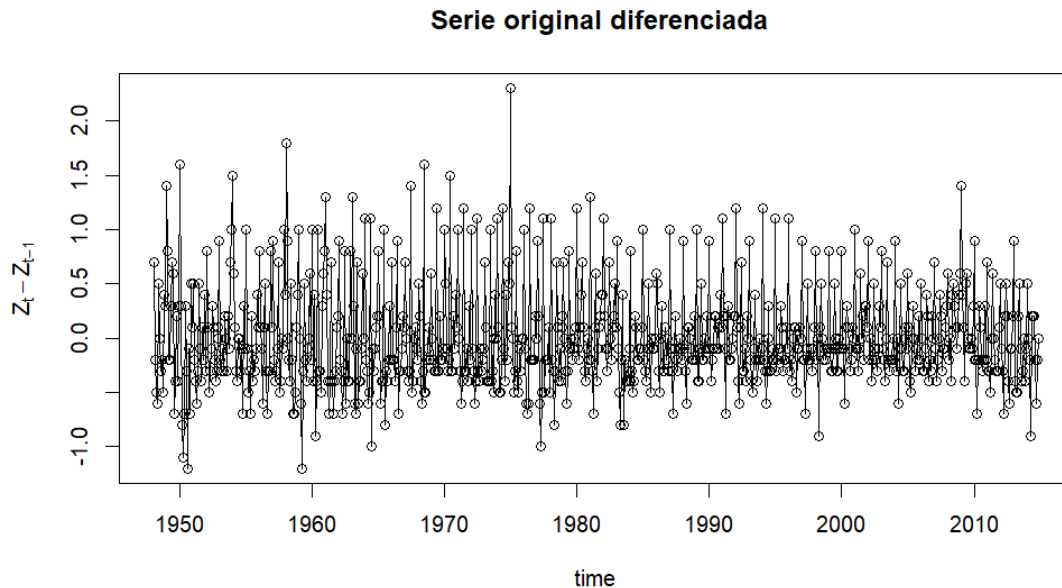
- **Componente observada:** Esta es la serie de tiempo original, donde se observan fluctuaciones notables en ciertos periodos, con picos y caídas que indican cambios significativos a lo largo del tiempo.
- **Tendencia:** La tendencia muestra los cambios de largo plazo en la serie. Se puede apreciar un aumento en algunos periodos, como antes de 1980, seguido de una tendencia a la baja y cierta estabilización después de 2000. Esto indica que la serie no es estacionaria, ya que la tendencia cambia de forma notable a lo largo del tiempo.
- **Componente estacional:** La estacionalidad es constante a lo largo del tiempo, lo que sugiere que los patrones estacionales tienen un comportamiento regular. La amplitud de la componente estacional no parece cambiar significativamente, lo que indica que este patrón repetitivo no está afectado por las fluctuaciones de la serie observada.
- **Componente aleatoria (ruido):** El componente aleatorio presenta variabilidad en ciertos periodos, lo que podría indicar la presencia de factores no explicados por la tendencia ni por la estacionalidad. Este componente tiene mayor variabilidad en ciertos momentos, como alrededor de 1980, lo que sugiere que hubo perturbaciones adicionales no capturadas por los otros componentes.

La descomposición de esta serie de tiempo indica que la serie tiene una tendencia no estacionaria, ya que los patrones de largo plazo varían significativamente con el tiempo. La estacionalidad, por otro lado, es constante, lo que sugiere un patrón repetitivo estable. El componente aleatorio introduce variabilidad que puede estar relacionada con eventos externos o factores no modelados directamente.

1.3. Literal B)

Como se discutió de el literal anterior hay evidencia de que la varianza no es constante y habría necesidad de hacer una transformación box-cox por que diferenciaríamos la serie temporal para ver mejor la varianza.

```
plot.ts(diff(Datos_ajuste), main = " Serie original diferenciada",  
ylab = expression(Z[t]-Z[t-1]), xlab = "time", col='black',  
lwd = 0.1, type = "o")
```



La gráfica muestra fluctuaciones de mayor magnitud en ciertos periodos en comparación con otros, lo que indica una variabilidad no homogénea. Por lo tanto, es recomendable aplicar una transformación Box-Cox para estabilizar la varianza y lograr una mayor homogeneidad en la serie.

Metodo 1:

```
tBoxCox <- powerTransform(Datos_ajuste)
summary(tBoxCox)
```

bcPower Transformation to Normality

	Est	Power	Rounded	Pwr	Wald	Lwr	Bnd	Wald	Upr	Bnd
Datos_ajuste	0.2678		0.33		0.0659		0.4697			

Likelihood ratio test that transformation parameter is equal to 0
(log transformation)

Likelihood ratio test that no transformation is needed

Metodo 2:

```
BoxCox.lambda(Datos_ajuste, method=c("loglik"), lower=-2, upper=2)
[1] 0.15
```

Metodo 3:

```
BoxCox.lambda(Datos_ajuste, method=c("guerrero"))
[1] 0.5649325
```

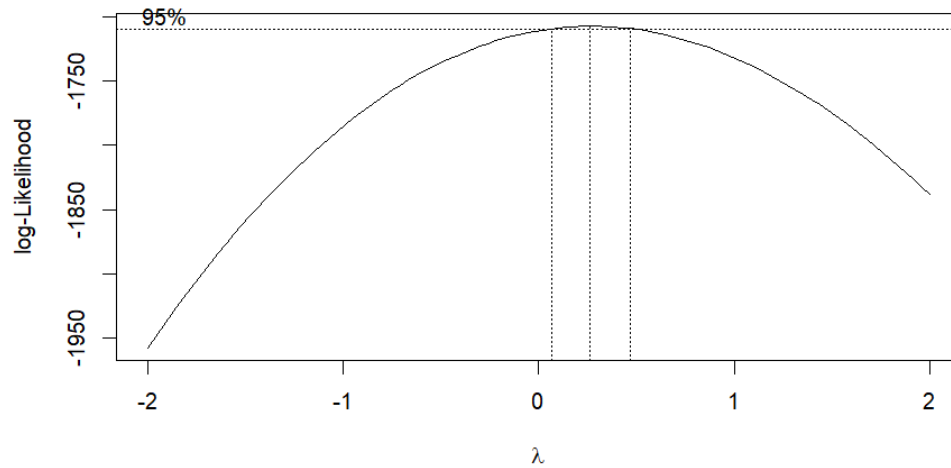
Resumen:

Método	Valor
Método 1	$0.4697 \approx 0.47$
Método 2	0.15
Método 3	$0.5649325 \approx 0.56$

Tabla 1: Valores obtenidos por los diferentes valores de λ

En este caso el logaritmo no se puede considerar ya que:

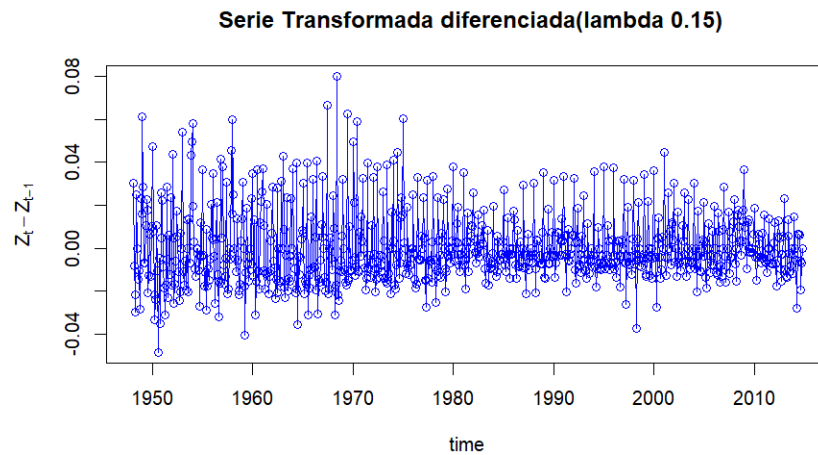
```
bc <- boxcox(lm(Datos_ajuste~1))
```



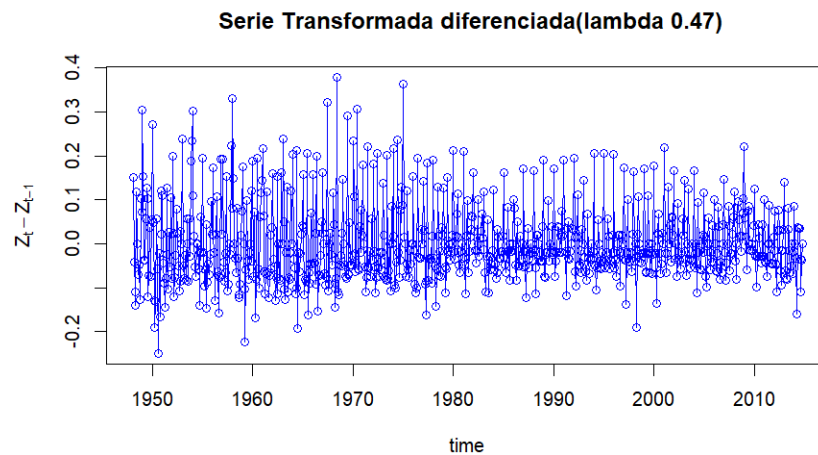
El valor 0 no se encuentra dentro del intervalo, por lo que solo se considerarán los tres valores previamente descritos.

Nota: A continuación, solo se mostrarán las gráficas de la serie diferenciada, ya que permiten observar con mayor claridad la variabilidad en la varianza.

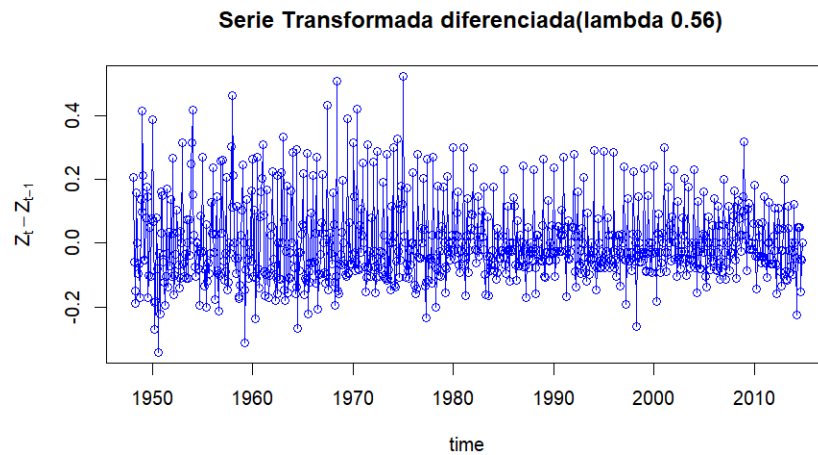
```
plot.ts(diff(Datos_ajuste^{0.15}),  
main = "Serie Transformada diferenciada(lambda 0.15)",  
ylab = expression(Z[t]-Z[t-1]), xlab = "time",col='blue',  
lwd = 0.1, type = "o")
```

```
plot.ts(diff(Datos_ajuste^{0.47}),
main = "Serie Transformada diferenciada(lambda 0.47)",
ylab = expression(Z[t]-Z[t-1]), xlab = "time",col='blue',
lwd = 0.1, type = "o")
```



```
plot.ts(diff(Datos_ajuste^{0.56}),
main = "Serie Transformada diferenciada(lambda 0.56)",
ylab = expression(Z[t]-Z[t-1]), xlab = "time",col='blue',
lwd = 0.1, type = "o")
```



Los distintos valores de λ no logran completamente estabilizar la varianza. Sin embargo, con $\lambda = 0.56$, se observa un mejor ajuste del comportamiento de la serie. Por lo tanto, a partir de este punto, se trabajará con la serie utilizando esta transformación.

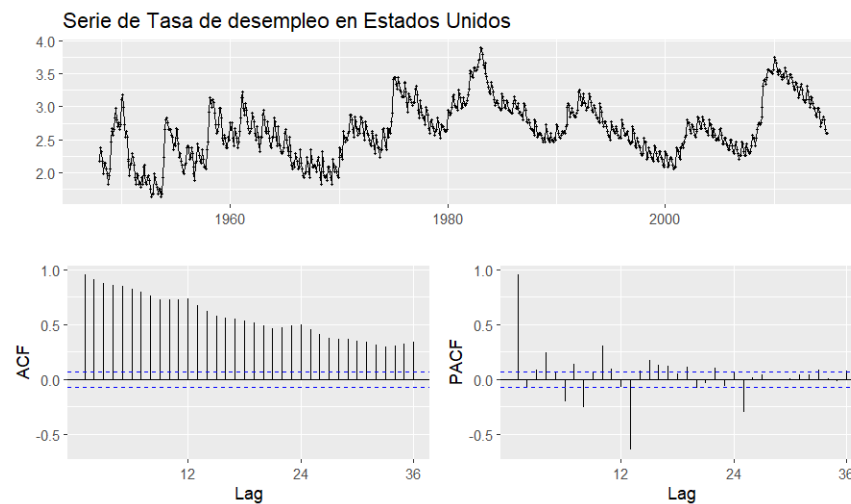
```
Z1 <- Datos_ajuste^{0.56}
```

1.4. Literal C)

Para la identificación del orden de diferenciación, lo más recomendable sería realizar una prueba de HEGY. Sin embargo, para agilizar el desarrollo de este ejercicio, en ocasiones dicha prueba. En su lugar, nos basaremos en los análisis de las funciones de autocorrelación (**ACF**) y autocorrelación parcial (**PACF**) para identificar el orden de diferenciación.

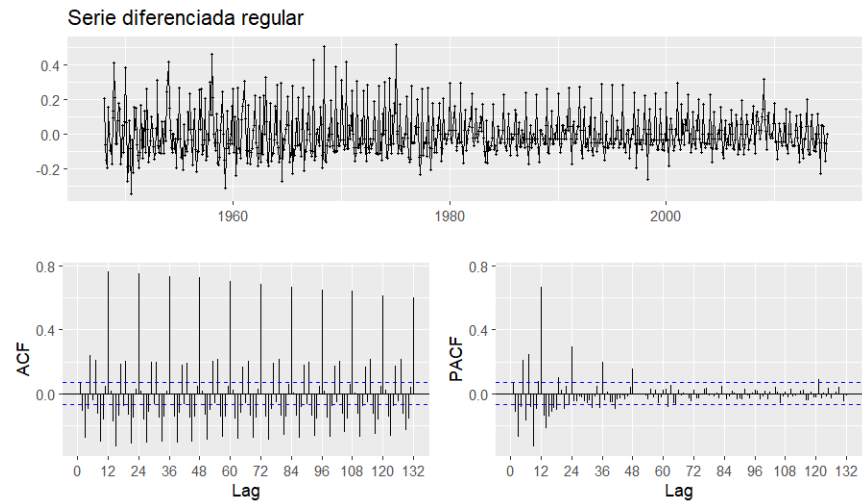
ACF y **PACF**

```
ggtsdisplay(Z1, main = "Serie de Tasa de desempleo en Estados Unidos")
```



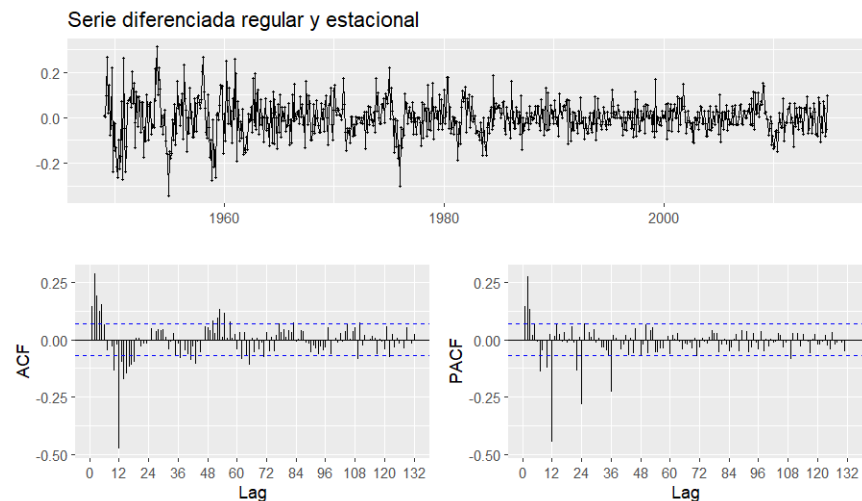
Como se puede observar en el gráfico de la **ACF**, esta no decae de forma exponencial, sino geométricamente. Esto evidencia la necesidad de aplicar una diferenciación. Comenzaremos realizando una diferenciación **regular**.

```
ggtsdisplay(diff(Z1),lag.max=132, main = "Serie diferenciada regular")
```



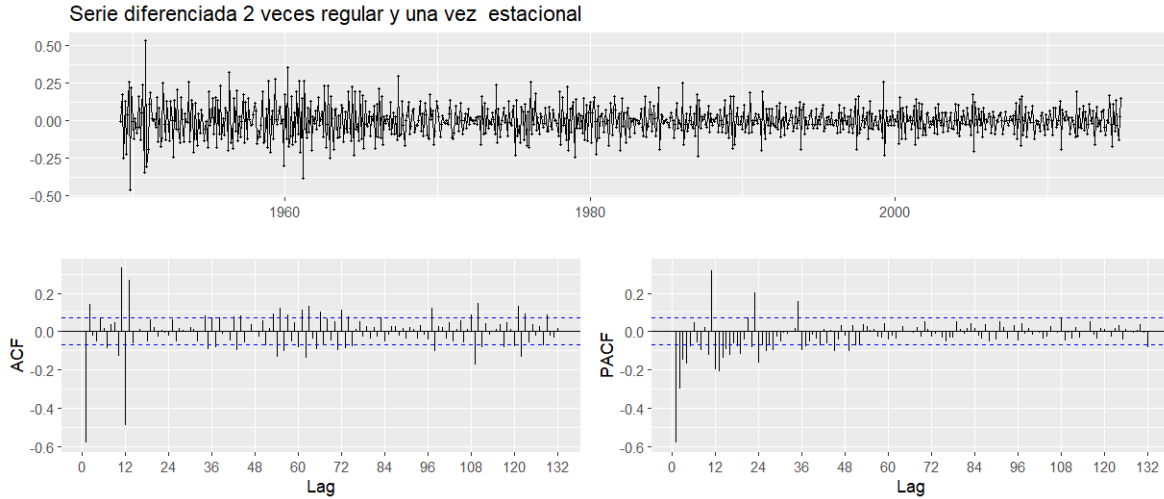
Como se observa en la **ACF**, tras la primera diferenciación, la serie aparenta comportarse de manera estacionaria. Sin embargo, se nota que el patrón estacional no decae exponencialmente, sino de forma geométrica. Por lo tanto, es necesario aplicar una diferenciación estacional.

```
ggtsdisplay(diff(diff(Z1),lag = 12), lag.max=132,
main = "Serie diferenciada regular y estacional")
```



Después de aplicar tanto la diferenciación regular como la estacional, se observa una mejora en la estructura estacionaria de la serie. No obstante, en la gráfica de la serie aún se aprecia que la media no es constante en algunas secciones. Por lo tanto, podría ser necesario aplicar una diferenciación regular adicional para estabilizar completamente la media.

```
ggtsdisplay(diff(diff(diff(Z1),lag = 12)), lag.max=132, main = "Serie diferenciada 2 v
```



Después de aplicar tanto la segunda diferenciación regular como la estacional, se concluye que el orden de diferenciación apropiado es 2 para la parte regular y 1 para la estacional. Ahora, es necesario determinar el orden correspondiente a los términos autorregresivos (AR) y de medias móviles (MA) tanto para la parte regular como para la estacional, con el fin de completar el modelo ARIMA. Este proceso se realizara mediante la inspección de las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para identificar los valores más adecuados para los órdenes de p y q en ambos componentes.

Parte regular

Se observa que la **PACF** presenta un patrón de cola, es decir, un decaimiento exponencial, mientras que la **ACF** muestra cortes claros en los rezagos 1, 2. Esto sugiere la presencia de un proceso **MA(2)** en la serie, lo que indica que un modelo media móvil de orden 2 es adecuado para capturar la estructura de la serie en su parte regular.

Parte estacional

Se observa que la **PACF** presenta un patrón de cola, es decir, un decaimiento exponencial, mientras que la **ACF** muestra un corte claro en el rezago 12, correspondiente al primer rezago estacional, dado que la serie tiene una frecuencia de 12 meses. Esto sugiere la presencia de un proceso **MA(1)** en la parte estacional, indicando que un modelo de media móvil de orden 1 es adecuado para capturar la dependencia estacional en la serie.

Modelo propuesto

$$ARIMA(0, 2, 2)X(0, 1, 1)[12]$$

$$(1 - B)^2(1 - B^{12})X_t = (1 + \theta_1 B + \theta_2 B^2)(1 + \Theta_1 B^{12})\varepsilon_t$$

Donde:

- B es el operador de rezago.
- θ_1, θ_2 son los parámetros de la parte MA (Media Móvil) regular.
- Θ_1 es el parámetro de la parte MA estacional.
- ε_t es el término de error o ruido blanco.

¿Que modelo propone `auto.arima`?

```
auto.arima(Z1,max.p=5,max.q=5)
Series: Z1
ARIMA(3,0,1)(2,1,2)[12]
Coefficients:
          ar1          ar2          ar3          ma1          sar1          sar2          sma1
          1.6815   -0.5501   -0.1435   -0.6393   -0.3227   -0.0173   -0.4173
s.e.      0.0714    0.1081    0.0430    0.0671    0.3685    0.0582    0.3680
          sma2
          -0.2061
s.e.      0.2795
sigma^2 = 0.003784:  log likelihood = 1081.26
AIC=-2144.53   AICc=-2144.29   BIC=-2102.47
```

La función `auto.arima` sugiere un modelo **ARIMA(3,0,1)(2,1,2)[12]**. Esto puede parecer contradictorio, ya que no considera un orden de diferenciación en la parte regular. Sin embargo, como observamos anteriormente en la **ACF**, se evidenciaba un decaimiento geométrico, lo que indicaba la necesidad de aplicar una diferenciación regular. Por lo tanto, seguiremos utilizando el modelo que hemos propuesto previamente, que incorpora dicha diferenciación para asegurar una mejor representación de la serie.

1.5. Literal D)

Ajuste del modelo.

```
modelo <- Arima(Z1, order = c(0, 2, 2),
seasonal = list(order = c(0, 1, 1), period = 12))
summary(modelo)
Series: Z1
ARIMA(0,2,2)(0,1,1)[12]

Coefficients:
          ma1          ma2          sma1
          -0.8604   0.0573   -0.7501
s.e.      0.0338   0.0400   0.0250
```

```
sigma^2 = 0.004156: log likelihood = 1036.66  
AIC=-2065.32 AICc=-2065.27 BIC=-2046.64
```

Training set error measures:

	ME	RMSE	MAE
Training set	-2.247342e-05	0.06377924	0.04793485

	MPE	MAPE	MASE
Training set	0.04029685	1.875582	0.2106495

	ACF1
Training set	-0.004797343

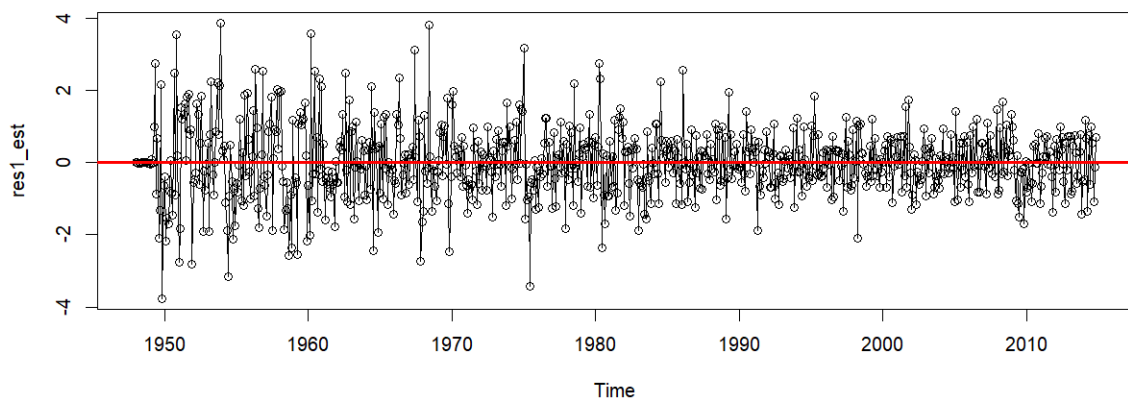
El modelo ARIMA(0,2,2)(0,1,1)[12] ajustado muestra un buen desempeño en la captura de la estructura de la serie temporal, con coeficientes significativos, baja varianza de los residuos ($\sigma^2 = 0.004156$) y valores aceptables en los criterios de información (AIC, AICc, BIC). Las medidas de error (RMSE, MAE, MAPE) son bajas, indicando una precisión adecuada en las predicciones.

No obstante, es importante realizar pruebas adicionales para verificar que los residuos del modelo cumplen con el supuesto de ser ruido blanco, lo que implica que no deberían presentar autocorrelación significativa, deben tener media cero y varianza constante.

Evaluación de supuestos

Es fundamental evaluar si los errores cumplen con los supuestos de ruido blanco, lo que implica que no deben presentar autocorrelación significativa, deben tener una media cero, varianza constante, y además seguir una distribución aproximadamente normal. Estos supuestos aseguran que el modelo es adecuado para hacer inferencias confiables y que ha capturado correctamente la estructura de la serie temporal.

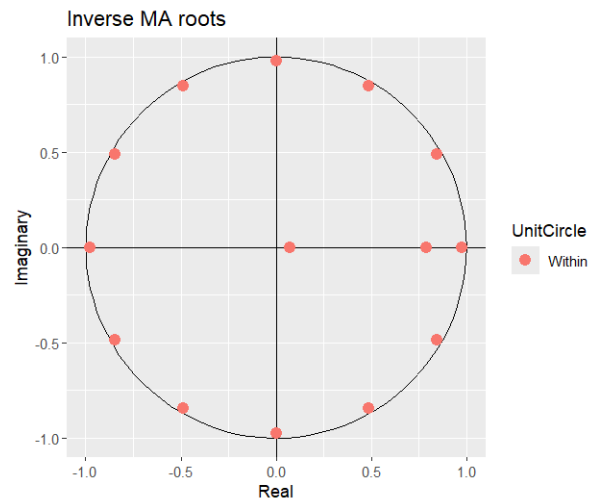
```
res1_CSS_ML_ML=residuals(modelo)  
res1_est=res1_CSS_ML_ML/(modelo$sigma2^.5)  
plot.ts(res1_est, type="o")  
abline(a=0,b=0,lwd=3, col="red")
```



El análisis de los residuos del modelo **ARIMA(0,2,2) × (0,1,1)**[12] muestra que los errores están distribuidos alrededor de cero, sin evidencia de patrones claros de tendencia, lo que es un buen indicio de un ajuste adecuado. Sin embargo, en cuanto a la varianza, los residuos parece no cumplir con el supuesto de homocedasticidad, ya que se observa que la varianza no es constante. Esto es consistente con lo que mencionamos previamente: la transformación de Box-Cox, aunque mejora la estabilización, no garantiza completamente la constancia de la varianza.

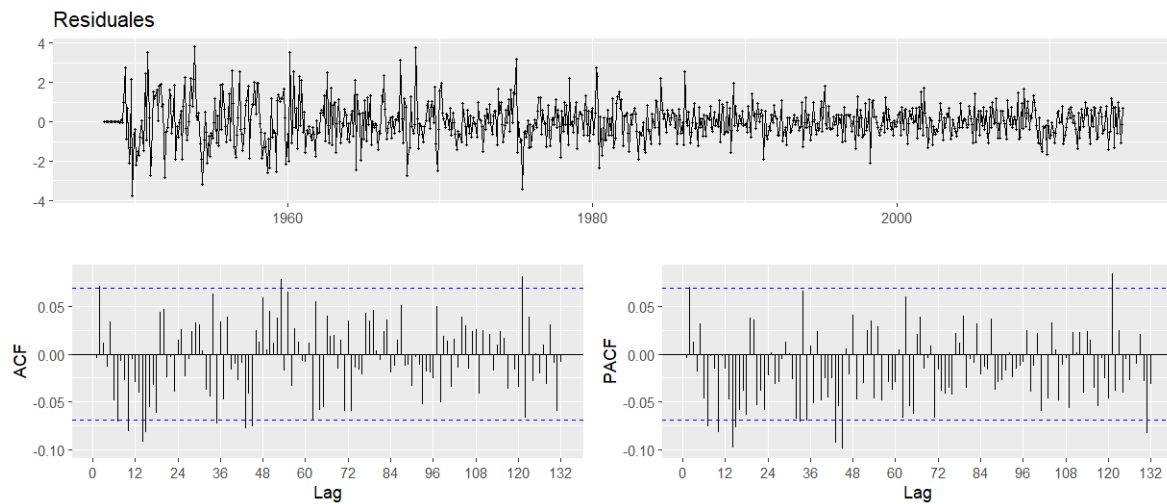
También podemos ver que se comporta de forma estacionario, para validar esta afirmación, evaluamos las raíces del polinomio utilizando la función autoplot en R.

```
autoplot(modelo)
```

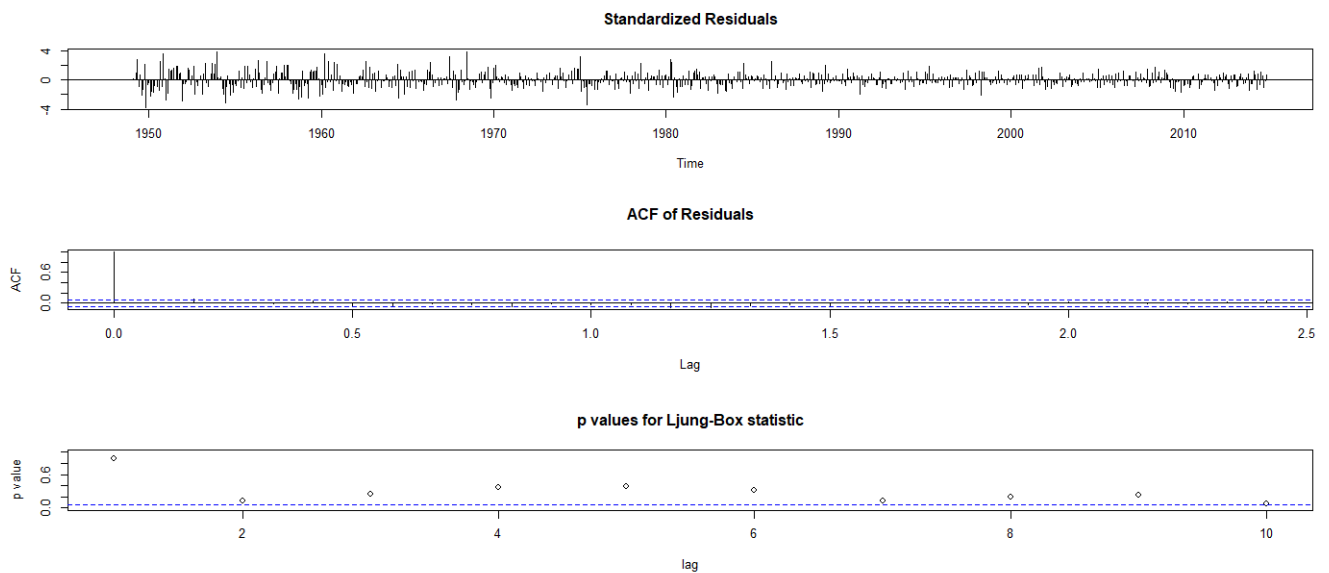


La gráfica de las raíces inversas del modelo MA muestra que todas se encuentran dentro del círculo unitario, lo que confirma que el modelo ARIMA es invertible y estable. Esto garantiza que las predicciones generadas por el modelo son confiables y que el proceso subyacente es adecuado. Sin embargo, es importante complementar esta validación verificando que los residuos del modelo se comporten como ruido blanco, lo cual es esencial para asegurar que el modelo ha capturado toda la estructura de la serie.

```
ggtsdisplay(res1_est, lag.max=132, main = "Residuales")
```



`tsdiag(modelo)`



Aunque los residuos del modelo $ARIMA(0,2,2)(0,1,1)[12]$ parecen estar centrados en cero, la ACF muestra algunos rezagos con autocorrelación significativa, lo que indica que el modelo no ha capturado completamente todos los patrones de autocorrelación en los residuos. Esto sugiere que podrían ser necesarios ajustes adicionales para mejorar el modelo.

Además, los residuos estandarizados presentan algunos valores fuera del rango de $(-3, 3)$, lo que podría señalar la presencia de datos atípicos.

Datos atípicos

```
(Nobs_Esp=round(length(Datos_ajuste)*2*pnorm(-3, mean = 0, sd = 1,
lower.tail = TRUE)))
[1] 2
```

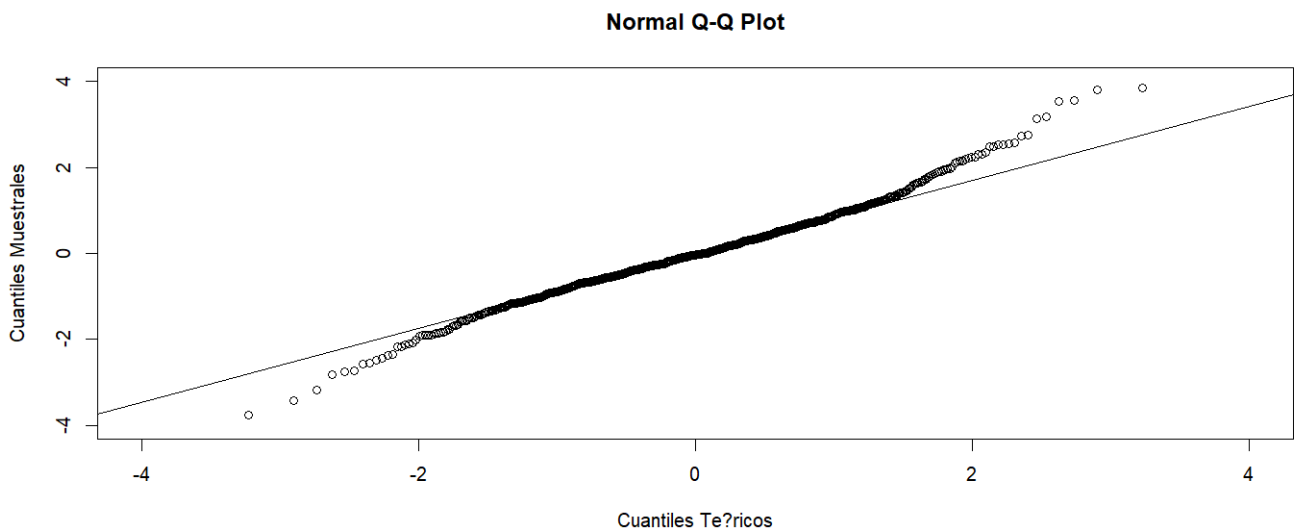


```
ind=(abs(res1_est)>3.0)
sum(ind)
[1] 9
```

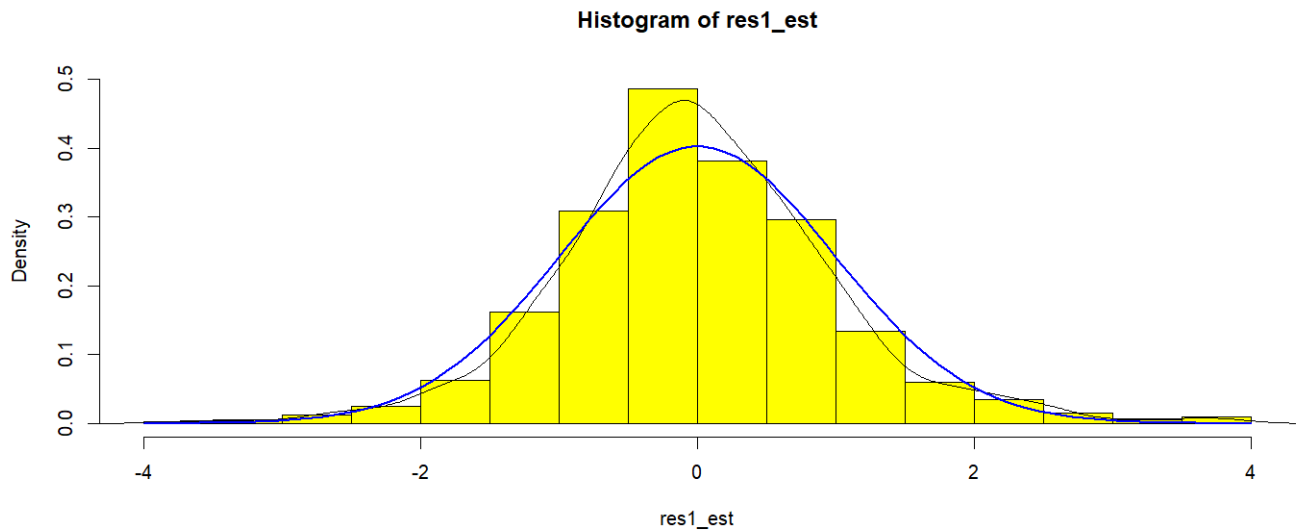
Se observaron 9 datos atípicos, lo que podría afectar la suposición de normalidad en los residuos. Aunque este problema podría abordarse más a fondo, el enfoque principal de este trabajo es la predicción, por lo que la presencia de estos valores atípicos no representa necesariamente un problema crítico(a priori) para la capacidad predictiva del modelo.

¿Se distribuye aproximadamente normal?

```
qqnorm(res1_est, xlab = "Cuantiles Teóricos", ylab = "Cuantiles Muestrales",
xlim=c(-4,4), ylim=c(-4,4))
qqline(res1_est)
```



```
mu<-mean(res1_est)
sigm<-sd(res1_est)
x<-seq(-4,4,length=100)
y<-dnorm(x,mu,sigm)
hist(res1_est,prob=T,ylim=c(0,0.5),xlim=c(-4,4),col="yellow")
lines(density(res1_est))
lines(x,y,lwd=2,col="blue")
```



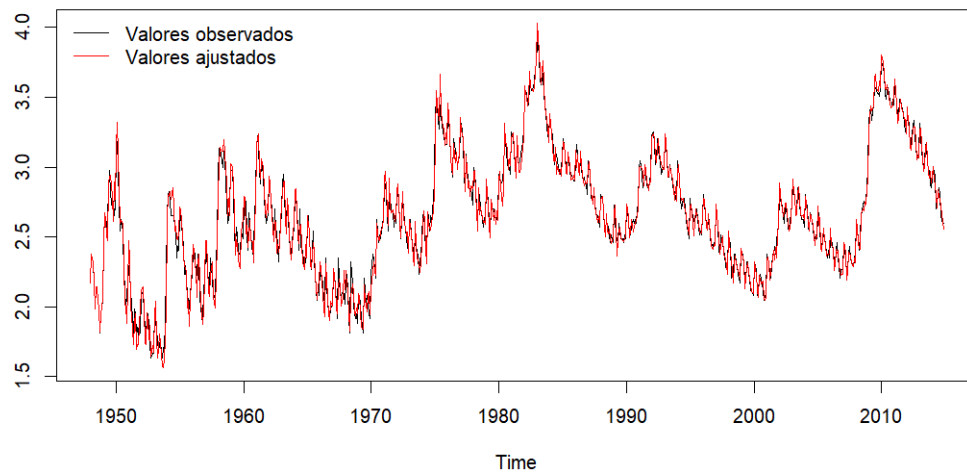
Shapiro-Wilk normality test

```
data: res1_est
W = 0.98463, p-value = 1.851e-07
```

Los residuos no cumplen con la suposición de normalidad, como se evidencia en el Q-Q plot, el histograma y la prueba de Shapiro-Wilk. Aunque esto puede ser un problema si el objetivo es hacer inferencias estadísticamente rigurosas, para fines de predicción, esta desviación de la normalidad no necesariamente afectará la calidad predictiva del modelo, pero debes tener en cuenta la presencia de valores atípicos y la posible falta de ajuste de normalidad en los residuos.

Valores observados Vs valores ajustados

```
res1_CSS_MLn=residuals(modelo)
ajustn=Z1-residuals(modelo)
ts.plot(Z1,ajustn)
lines(Z1, col="black")
lines(ajustn, col="red")
legend("topleft", legend=c("Valores observados", "Valores ajustados"),
col=c("black", "red"), lty=1, bty="n")
```



El modelo ajusta correctamente la serie temporal, capturando las principales tendencias y fluctuaciones. Aunque se detectaron algunos valores atípicos que afectan la normalidad de los residuos, no se encontró autocorrelación significativa, lo que sugiere que los residuos se comportan como ruido blanco. Para fines de predicción, el modelo es adecuado, pero la falta de normalidad podría ser un problema si se busca realizar inferencias.

2. Punto 2

Enunciado

Utilice el modelo estimado en el ítem anterior para realizar un pronóstico para los últimos 24 meses.

Con ayuda de la función `forecast()` se calculan los pronósticos al 95 % correspondientes al último año de la serie como se ve a continuación

```
forecast_data <- forecast(modelo, h=24, level=95)
```

Pero recordemos que el modelo se le realizó una transformación box-cox por lo cual los anteriores pronósticos no son adecuados ya que no tiene en cuenta el Factor de corrección de sesgo el cual está expresado de la siguiente forma teniendo en cuenta que $\lambda \neq 0$:

$$\hat{Z}_t(h) = T^{-1}[\hat{T}(Z_t)(h)] * \hat{c}_{t,\lambda}(h)$$

donde $\hat{c}_{t,\lambda}(h)$ es un factor de corrección de sesgos dado por,

Month	Point Forecast	Lo 95	Hi 95
Dec 2014	2.594604	2.46825394	2.720953
Jan 2015	2.743749	2.55218461	2.935314
Feb 2015	2.697539	2.44216876	2.952908
Mar 2015	2.620581	2.30003120	2.941131
Apr 2015	2.457655	2.06964550	2.845664
May 2015	2.475159	2.01706781	2.933251
Jun 2015	2.556065	2.02514092	3.086990
Jul 2015	2.560923	1.95438513	3.167460
Aug 2015	2.484104	1.79919027	3.169018
Sep 2015	2.376495	1.61048219	3.142508
Oct 2015	2.328789	1.47900461	3.178573
Nov 2015	2.297223	1.36105100	3.233394
Dec 2015	2.291813	1.25343313	3.330193
Jan 2016	2.436801	1.29346400	3.580138
Feb 2016	2.386433	1.13466025	3.638205
Mar 2016	2.305318	0.94175244	3.668884
Apr 2016	2.138234	0.65962507	3.616842
May 2016	2.151581	0.55477594	3.748386
Jun 2016	2.228329	0.51026287	3.946396
Jul 2016	2.229029	0.38671629	4.071341
Aug 2016	2.148053	0.17858294	4.117522
Sep 2016	2.036286	-0.06318307	4.135756
Oct 2016	1.984423	-0.24782580	4.216671
Nov 2016	1.948699	-0.41904866	4.316446

$$\hat{c}_{t,\lambda}(h) = \begin{cases} \left\{ \left\{ 0.5 + 0.5 \left[1 + 2 (\lambda^{-1} - 1) \frac{\hat{Var}_t[e_t(h)]}{[\lambda^{-1} + \hat{T}(Z_t)(h)]^2} \right]^{1/2} \right\}^{1/\lambda} & \text{si } \lambda \neq 0 \\ \exp \left(\hat{Var}_t[e_t(h)/2] \right) & \text{si } \lambda = 0 \end{cases}$$

Nuestro caso $\lambda \neq 0$, en R teniendo en cuenta el factor de correccion de sesgo los pronosticos nos da como resultado:

```
forecast_data <- forecast(modelo, h=24, level=95)
r <- ts(as.data.frame(forecast_data), freq=12,
        start=start(Datos_pronostico))^(1/0.56)
lami <- 1/0.56
ses <- 0.5 + 0.5 * (1 + 2*(lami-1)*((summary(modelo)$sigma2) /
        ((lami + forecast_data$mean)^2)))^(1/2)
sesgo <- ses^(1/0.56)
z_t <- r*sesgo
```

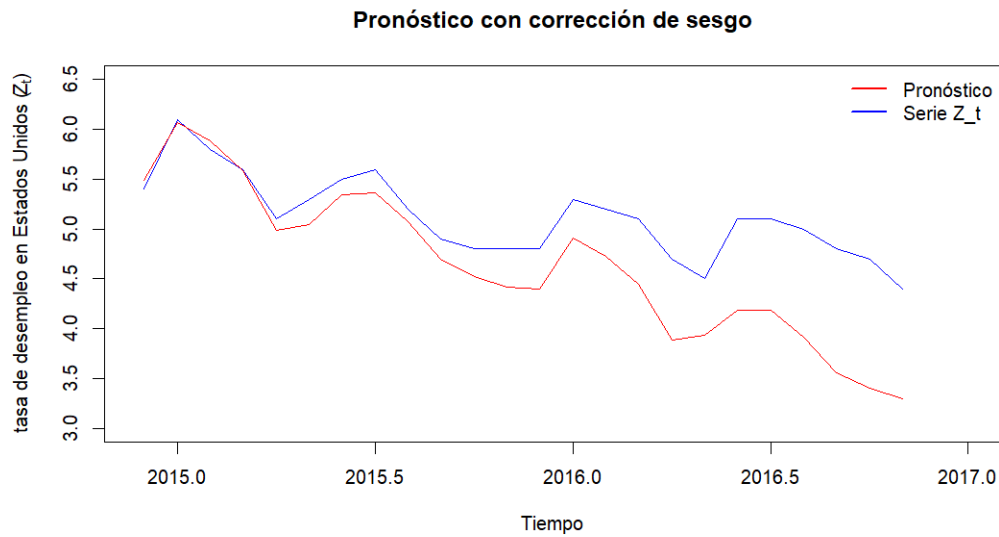
Month	Point Forecast	Lo 95	Hi 95
Dec 2014	5.488814	5.02067515	5.975216
Jan 2015	6.064843	5.32954865	6.841622
Feb 2015	5.883667	4.92628510	6.915046
Mar 2015	5.587324	4.42607013	6.865910
Apr 2015	4.982290	3.66581245	6.473102
May 2015	5.045829	3.50117225	6.833164
Jun 2015	5.344095	3.52621406	7.485781
Jul 2015	5.362241	3.30923789	7.837797
Aug 2015	5.078433	2.85472482	7.844725
Sep 2015	4.692336	2.34227149	7.727993
Oct 2015	4.525477	2.01183239	7.887113
Nov 2015	4.416533	1.73435250	8.131688
Dec 2015	4.397980	1.49712064	8.571507
Jan 2016	4.907057	1.58355202	9.753888
Feb 2016	4.727427	1.25327808	10.038227
Mar 2016	4.444361	0.89853320	10.189947
Apr 2016	3.885675	0.47577165	9.933432
May 2016	3.929088	0.34925999	10.587748
Jun 2016	4.182830	0.30080295	11.607068
Jul 2016	4.185176	0.18334935	12.271441
Aug 2016	3.917592	0.04614044	12.521203
Sep 2016	3.561104	NaN	12.620529
Oct 2016	3.400781	NaN	13.064911
Nov 2016	3.292244	NaN	13.622127

Tabla 2: Point Forecasts with 95 % Confidence Intervals

Así la gráfica de los pronósticos nos da:

```
plot(Datos_pronostico,
     xlab = "Tiempo",
     ylab = expression(paste("tasa de desempleo en Estados Unidos (", Z[t], ")")),
     main = "Pronóstico con corrección de sesgo",
     xlim = c(2014.9, 2017), ylim=c(3,6.5), col="blue")

lines(z_t[,1], col="red")
# Agregar leyenda
legend("topright", legend = c("Pronóstico", "Serie Z_t"),
      col = c("red", "blue"), lwd = 2, lty = 1, bty = "n")
```



El Pronostio utilizando el modelo ARIMA, corregido por el factor de sesgo, muestra un buen desempeño en la predicción de los primeros meses, ajustándose de manera aceptable a los valores reales. Sin embargo, a medida que se avanza en el tiempo, su capacidad predictiva disminuye, lo que indica una pérdida de eficiencia en las predicciones a largo plazo. Por lo tanto, se puede concluir que este modelo es más adecuado para realizar pronósticos a corto plazo, mientras que para horizontes temporales más largos sería necesario explorar otros modelos o ajustes adicionales para mejorar su rendimiento.

3. Punto 3

Enunciado

Explore la función `HoltWinters()` para realizar un pronóstico de las mismas observaciones, utilizando EWMA, es decir una media móvil exponencialmente ponderada o más breve suavizamiento exponencial o el mismo método de `HoltWinters` si es el caso. Que valor de los parámetros de suavizamiento α , β y γ .

Encuentra Ud que son los más adecuados para su conjunto de datos?

El método Holt-Winters es una extensión del suavizamiento exponencial, también conocido como Exponential Weighted Moving Average (EWMA). Es útil para hacer pronósticos en series de tiempo que muestran tanto tendencia como estacionalidad. Holt-Winters tiene tres parámetros de suavizamiento: α , β y γ , que controlan diferentes aspectos del modelo:

- α : El parámetro de nivel o suavizamiento. Controla el grado de suavizamiento aplicado a la serie de tiempo original. Valores altos de α asignan más peso a las observaciones recientes.

- β : El parámetro de tendencia. Controla el grado de suavizamiento de la tendencia. Un valor alto de β permitirá que la tendencia reaccione rápidamente a los cambios.
- γ : El parámetro estacional. Controla la suavización del componente estacional. Un valor alto de γ hará que la estacionalidad cambie rápidamente.

El proceso para determinar los mejores valores de α , β y γ generalmente implica una combinación de pruebas con distintos valores y la evaluación del modelo resultante mediante un criterio de ajuste (como el error cuadrático medio, AIC, BIC o alguna otra métrica de error).

El modelo es:

$$x_t = \mu_t + w_t, \quad (1)$$

donde μ_t es la media no estacionaria del proceso en el tiempo t y las w_t son variables aleatorias independientes con una media de 0 y una desviación estándar σ . Dado que no hay una tendencia sistemática, una estimación intuitivamente razonable de la media en el tiempo t viene dado por un promedio ponderado de la observación en el tiempo t , (x_t) y la estimación de la media en el tiempo $t - 1$ (a_{t-1}):

$$a_t = \alpha x_t + (1 - \alpha)a_{t-1}, \quad 0 < \alpha < 1. \quad (2)$$

El método de Holt-Winters generaliza la ecuación de recursión del suavizamiento exponencial:

$$a_t = \lambda x_t + (1 - \lambda)a_{t-1} \quad (3)$$

La ecuación de pronóstico es

$$\hat{x}_{n+k|n} = a_n, \quad k = 1, 2, \dots \quad (4)$$

De forma tal que:

$$\begin{aligned} a_t &= \alpha(x_t - s_{t-p}) + (1 - \alpha)(a_{t-1} + b_{t-1}) \\ b_t &= \beta(a_t - a_{t-1}) + (1 - \beta)b_{t-1} \\ s_t &= \gamma(x_t - a_t) + (1 - \gamma)s_{t-p}. \end{aligned} \quad (5)$$

donde a_t , b_t y s_t son el nivel estimado, la pendiente y la componente estacional, en el instante t respectivamente.

Ajuste del modelo

```
modelo_HW <- HoltWinters(Datos_ajuste, seasonal = "additive")
```

Holt-Winters exponential smoothing with trend and additive seasonal component.

Call:

```
HoltWinters(x = Datos_ajuste, seasonal = "additive")
```

Smoothing parameters:

alpha: 0.8490578

beta : 0.09371791

gamma: 1

Coefficients:

[,1]

a 5.97926958

b -0.07465667

s1 -0.26976858

s2 0.50755872

s3 0.42725737

s4 0.22554356

s5 -0.27875314

s6 -0.08441898

s7 0.26806368

s8 0.26788572

s9 -0.09495196

s10 -0.47295861

s11 -0.52003027

s12 -0.47926958

El modelo de suavizamiento exponencial Holt-Winters con estacionalidad aditiva ajustado a la serie se puede concluir lo siguiente

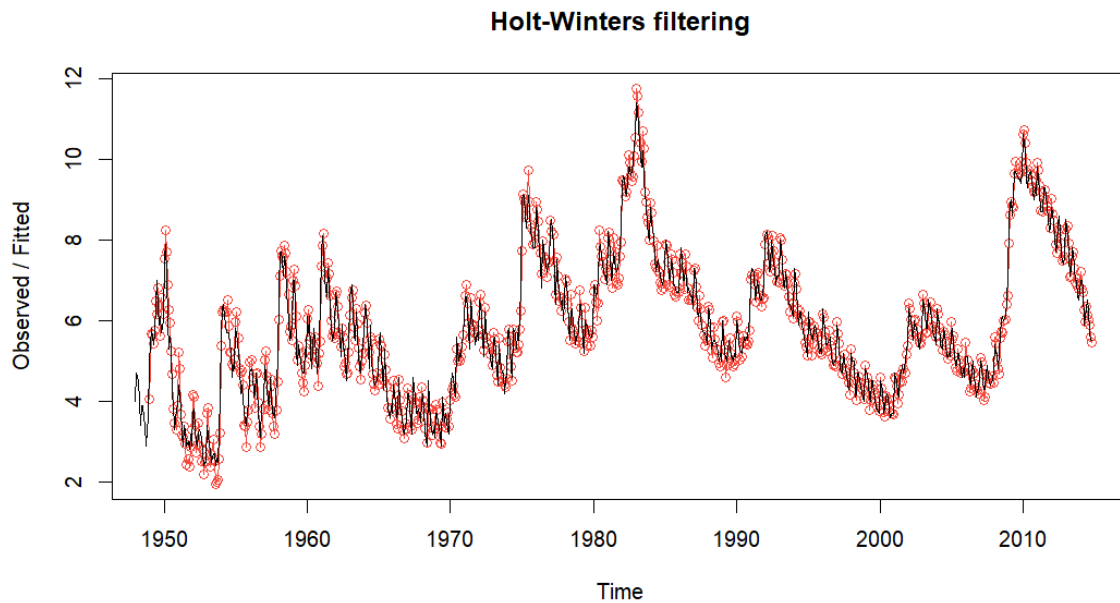
- $\alpha = 0.849$: Un valor alto de α indica que el modelo está dando más peso a las observaciones recientes, lo cual es apropiado si los datos recientes son más representativos para el futuro pronóstico.
- $\beta = 0.094$: Un valor bajo de β indica que la tendencia en tu serie temporal es relativamente estable, con pocos cambios bruscos.
- $\gamma = 1$: Un valor máximo de γ muestra que el modelo ajusta completamente a los cambios estacionales sin suavización, lo que implica una fuerte y constante estacionalidad en los datos.
- **Tendencia negativa**: La serie muestra una ligera tendencia descendente con un valor de $b = -0.075$, lo que sugiere que, en general, el nivel de la serie podría estar disminuyendo de forma suave con el tiempo.

Implicaciones

El modelo ajusta correctamente la estacionalidad y refleja que la tendencia es estable y ligeramente descendente. Los componentes estacionales muestran fluctuaciones

significativas entre periodos, lo que sugiere que el componente estacional es un factor importante para el pronóstico.

```
plot(modelo_HW,type = "o")
```



Caption

El modelo Holt-Winters ajusta bien la serie temporal, capturando tanto la estacionalidad como la tendencia general. Aunque sigue de cerca las fluctuaciones, puede haber pequeñas discrepancias en los picos más extremos, pero en general, ofrece un ajuste adecuado para realizar pronósticos futuros.

Predicciones

```
predicciones <- stats::predict(modelo_HW,n.ahead=length(Datos_pronostico),  
interval="prediction",level=0.95)
```

Año/Mes	2014	2015	2016
Jan		6.337515	5.441635
Feb		6.182557	5.286677
Mar		5.906186	5.010306
Apr		5.327233	4.431353
May		5.446911	4.551030
Jun		5.724737	4.828856
Jul		5.649902	4.754022
Aug		5.212408	4.316527
Sep		4.759744	3.863864
Oct		4.638016	3.742136
Nov		4.604120	3.708240
Dec	5.634844	4.738964	

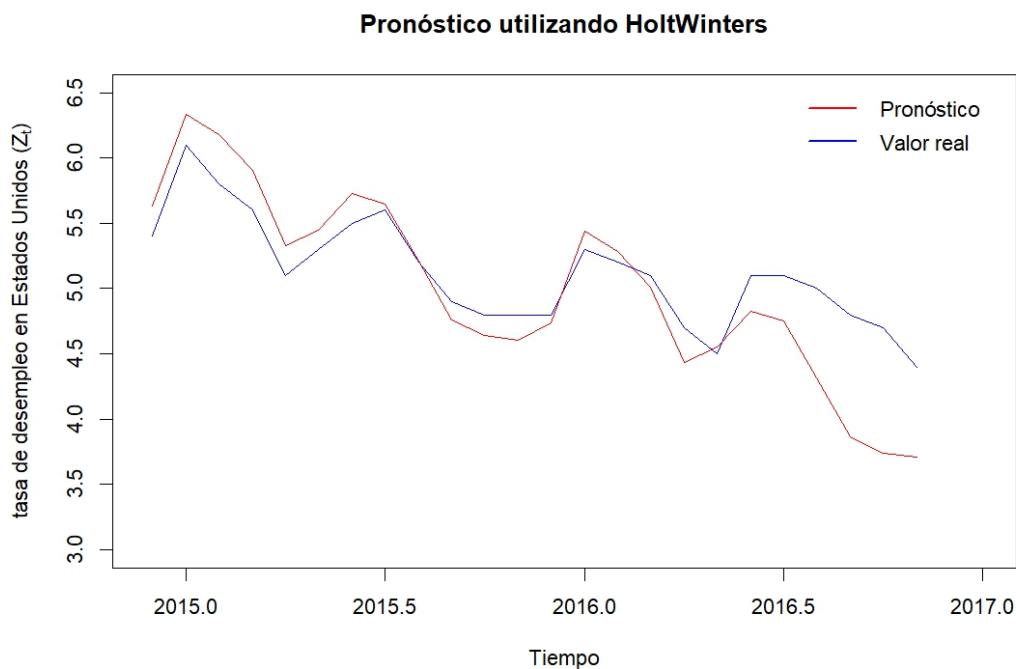
Tabla 3: Comparación de valores observados y ajustados utilizando HoltWinters()

mes	Observados	Ajustados
2014-12	5.4	5.634844
2015-1	6.1	6.337515
2015-2	5.8	6.182557
2015-3	5.6	5.906186
2015-4	5.1	5.327233
2015-5	5.3	5.446911
2015-6	5.5	5.724737
2015-7	5.6	5.649902
2015-8	5.2	5.212408
2015-9	4.9	4.759744
2015-10	4.8	4.638016
2015-11	4.8	4.604120
2015-12	4.8	4.738964
2016-1	5.3	5.441635
2016-2	5.2	5.286677
2016-3	5.1	5.010306
2016-4	4.7	4.431353
2016-5	4.5	4.551030
2016-6	5.1	4.828856
2016-7	5.1	4.754022
2016-8	5.0	4.316527
2016-9	4.8	3.863864
2016-10	4.7	3.742136
2016-11	4.4	3.708240

En la tabla se muestra comparación entre los valores observados y ajustados por el modelo. En general, los valores ajustados siguen de cerca los valores observados, lo que

sugiere que el modelo tiene un buen desempeño en la captura de la tendencia de los datos. Sin embargo, en algunos casos, como en los últimos periodos (por ejemplo, 5.1 frente a 4.316527), se observan pequeñas discrepancias, lo que indica que puede haber ligeras imprecisiones en el ajuste. A pesar de estas diferencias, el modelo parece ser razonablemente preciso en la mayoría de los puntos.

```
plot(Datos_pronostico,
     xlab = "Tiempo",
     ylab = expression(paste("tasa de desempleo en Estados Unidos (", Z[t], ")")),
     main = "Pronóstico utilizando HoltWinters",
     xlim = c(2014.9, 2017), ylim=c(3,6.5), col="blue")
lines(predicciones, col="red")
# Agregar leyenda
legend("topright", legend = c("Pronóstico", "Valor real"),
     col = c("red", "blue"), lwd = 2, lty=1, bty="n")
```



Encuentra Ud que son los más adecuados para su conjunto de datos?

Después de ajustar el modelo a los datos y evaluar los resultados, considero que los parámetros obtenidos son adecuados para el conjunto de datos. Esto se basa en la comparación entre los valores observados y los ajustados, donde las discrepancias son mínimas en la mayoría de los casos. Además, el modelo captura correctamente tanto la tendencia como los patrones estacionales presentes en los datos. Aunque existen pequeñas diferencias en algunos puntos, el modelo es capaz de ofrecer predicciones

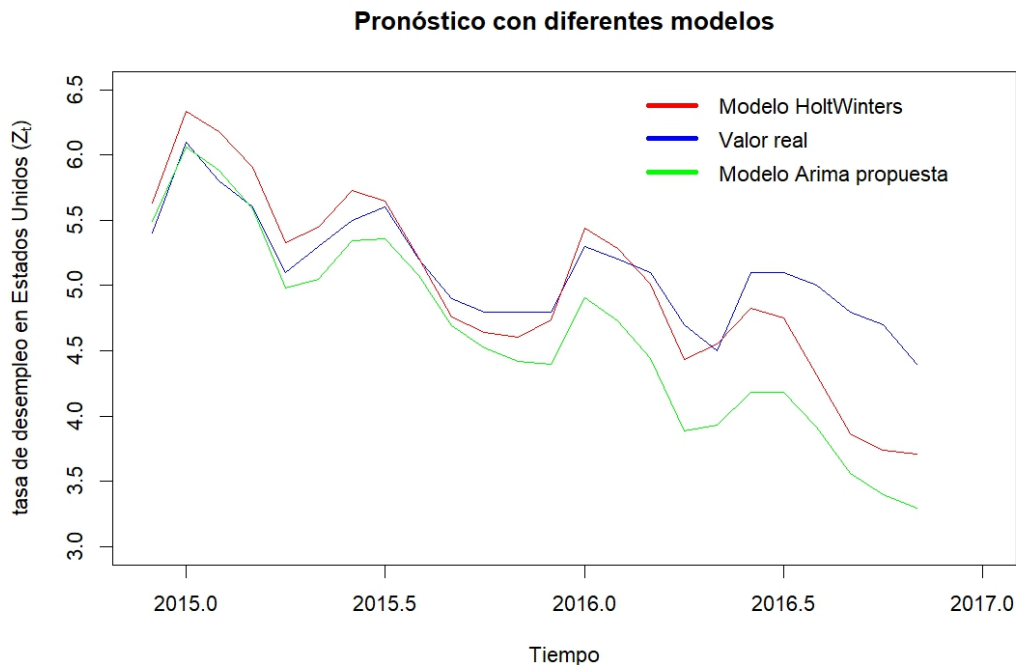
razonablemente precisas, lo que sugiere que los valores de suavizamiento seleccionados (β, α, γ) son apropiados.

4. Punto 4)

Enunciado

Compare los 2 pronósticos obtenidos en los ítems anteriores evaluando para cada uno medidas de calidad de pronóstico RMSE, MAE, MAPE en la pag 72 de las notas (2DA parte). En base a los pronósticos cuál modelo prefiere y por qué? anote sus propias conclusiones.

```
plot(Datos_pronostico,
     xlab = "Tiempo",
     ylab = expression(paste("tasa de desempleo en Estados Unidos (", Z[t], ")")),
     main = "Pronóstico con diferentes modelos",
     xlim = c(2014.9, 2017), ylim=c(3,6.5), col="blue")
lines(predicciones, col="red")
lines(z_t[,1], col="green")
# Agregar leyenda
legend("topright", legend = c("Modelo HoltWinters", "Valor real",
                              "Modelo Arima propuesta"),
      col = c("red", "blue", "green"), lwd = 4, lty=1, bty="n")
```



El Modelo Holt-Winters (rojo) es el que mejor captura la tendencia de la tasa de desempleo en comparación con el Modelo ARIMA propuesto (verde). Se ajusta más de cerca a los valores reales (azul) a lo largo de todo el periodo, especialmente en los picos y valles. El Modelo ARIMA presenta un mayor sesgo en ciertos tramos, especialmente hacia el final del periodo, donde no sigue tan de cerca la tendencia descendente. Por tanto, el Modelo Holt-Winters es el más adecuado para estos datos.

A continuación, se compararán los pronósticos obtenidos por ambos métodos. Para evaluar la calidad del ajuste, se calcularán las métricas RMSE, MAE y MAPE.

```
medidas_HoltWinters <- accuracy(predicciones,Datos_pronostico)
medidas_modelo <- accuracy(z_t,Datos_pronostico)
```

Tabla 4: Medidas de calidad de pronóstico

	Holt-Winters	Modelo
RMSE	5.777552	10.09597
MAE	0.2877286	0.4934069
MAPE	2.592336	9.838697

Basado en las medidas de calidad de pronóstico, el Modelo Holt-Winters ofrece un mejor desempeño que el otro modelo. Presenta un RMSE significativamente menor (5.78 frente a 10.10), lo que indica un ajuste más preciso en términos de errores cuadráticos. Así mismo, el MAE y el MAPE son más bajos para Holt-Winters, lo que refleja una menor desviación absoluta y porcentual respecto a los valores reales. Por lo tanto, el Modelo Holt-Winters es notablemente mejor en términos de precisión de pronóstico, por lo que sería el más adecuado para utilizar.