

EUR/USD

Harvey Bastidas, Andrés Caicedo, Alexander Alvarado

2024-11-11



# Contents

|  |           |
|--|-----------|
| <b>1 Justificación de elección del Dataset</b>                         | <b>5</b>  |
| 1.1 Información del Dataset . . . . .                                  | 5         |
| 1.2 Justificación de la Elección del Dataset . . . . .                 | 6         |
| 1.3 Importancia de Analizar este Dataset . . . . .                     | 7         |
| <b>2 Estructura de los datos</b>                                       | <b>9</b>  |
| 2.1 Cálculo de Medias Móviles Simples: . . . . .                       | 9         |
| 2.2 Análisis de Rezagos . . . . .                                      | 10        |
| 2.3 Análisis de Estacionalidad . . . . .                               | 11        |
| <b>3 Preprocesamiento y Visualización</b>                              | <b>15</b> |
| 3.1 Descomposición de la Serie de Tiempo . . . . .                     | 15        |
| 3.2 Prueba de Estacionariedad . . . . .                                | 17        |
| 3.3 Diferenciación para Estacionariedad . . . . .                      | 19        |
| 3.4 Verificación de Estacionariedad en la Serie Diferenciada . . . . . | 20        |
| 3.5 Justificación de la Transformación . . . . .                       | 23        |
| 3.6 Análisis de Autocorrelación . . . . .                              | 23        |
| 3.7 Modelo ARIMA . . . . .   | 25        |
| 3.8 Detección de Puntos de Cambio . . . . .                            | 25        |
| 3.9 Media Cero de los Residuos . . . . .                               | 26        |
| 3.10 Independencia de los Residuos . . . . .                           | 27        |
| 3.11 Distribución de los Residuos . . . . .                            | 27        |

|  |           |
|--|-----------|
| <b>4 Análisis de Series de Tiempo con el Método Holt-Winters</b>     | <b>29</b> |
| 4.1 Carga de Bibliotecas y Datos . . . . .                           | 29        |
| 4.2 Normalización de la Columna ‘close’ . . . . .                    | 30        |
| 4.3 Descomposición Estacional . . . . .                              | 30        |
| 4.4 Suavizado Exponencial Simple . . . . .                           | 31        |
| 4.5 Suavizado Exponencial Doble (Aditivo y Multiplicativo) . . . . . | 32        |
| 4.6 Calculo de error . . . . .                                       | 32        |
| 4.7 Conclusiones . . . . .   | 33        |

# Chapter 1

## Justificación de elección del Dataset

**Integrantes:**

Harvey Bastidas, Alexander Alvarado y Andrés Caicedo

**Materia:**

Análisis de series de tiempo

**Profesora:**

Isabel Cristina García

---

### 1.1 Información del Dataset

El dataset seleccionado para este análisis es el siguiente: EUR/USD Forex Trading Data (2003-2021), el cual fue extraído de Forex Trading Charts, propiedad de Barchart Solutions, una empresa especializada en servicios financieros en los Estados Unidos. Barchart Solutions cuenta con reconocidos clientes como el Banco Goldman Sachs y el Bank of Canada, lo que nos lleva a concluir que los datos ofrecidos son confiables y precisos para realizar análisis financieros.

La empresa matriz, **Barchart Solutions**, tiene su sede en:

**222 S. Riverside Plaza, Suite 810,  
Chicago, IL 60606, Estados Unidos**

El dataset presenta las tasas de cambio del par de divisas EUR/USD desde el **5 de mayo de 2003** hasta el **16 de octubre de 2021**, con una periodicidad de **4 horas**. Este conjunto de datos incluye las siguientes 6 columnas:

- **Open:** Precio de apertura para el periodo.
- **High:** Precio máximo durante el periodo.
- **Low:** Precio mínimo durante el periodo.
- **Close:** Precio de cierre para el periodo.
- **Volume:** Volumen de transacciones reportado.

Es importante resaltar que los datos no contienen valores nulos ni perdidos para los días de semana, aunque no se registran valores durante los fines de semana, lo que es normal en los mercados de Forex.

El análisis se centrará en la columna **Close**, que representa el precio de cierre, dado que es la variable más relevante para el pronóstico de tendencias.

## 1.2 Justificación de la Elección del Dataset

La elección de este dataset se fundamenta en la importancia de analizar la tasa de cambio EUR/USD, uno de los pares de divisas más negociados en el mercado Forex. La predicción de la tendencia de la tasa de cambio es crucial para varias estrategias financieras, como el balanceo de portafolios de inversión. En este contexto, se pueden usar tanto la **Teoría Moderna de Portafolios (MPT)** como la **Teoría de Portafolios Post-Moderna (PMPT)**, que son ampliamente utilizadas en la optimización de la distribución de activos. Ambas teorías se benefician de predicciones precisas de la tendencia de los activos subyacentes, como es el caso de las divisas.

Además, este dataset ofrece una excelente relación señal-ruido, lo que mejora la predictibilidad de los modelos basados en series de tiempo. Utilizamos el **coeficiente de variación (CV)** como criterio para seleccionar este dataset, debido a que es una métrica robusta para medir la variabilidad en relación con la media. Un CV bajo indica que la variabilidad en los datos es relativamente baja, lo que es favorable para la predicción de tendencias.

El **coeficiente de variación** calculado para la columna “Close” de este dataset es de aproximadamente **9.5%**, lo que lo convierte en el más bajo entre los datasets que evaluamos. Esto lo hace ideal para el pronóstico, ya que un CV bajo sugiere que la señal en los datos es fuerte en comparación con el ruido.

### 1.2.1 Relación con el Signal-to-Noise Ratio (SNR)

El **coeficiente de variación (CV)** está inversamente relacionado con el **Signal-to-Noise Ratio (SNR)**, una métrica comúnmente utilizada en el procesamiento de señales. El SNR mide la proporción de la potencia de la señal con respecto a la del ruido, lo que ayuda a evaluar la calidad de los datos.

El SNR para este dataset se calcula como el cuadrado del inverso del CV. Para el EUR/USD, con un CV de **0.09**, el SNR es aproximadamente:

$$SNR = \left( \frac{1}{0.09} \right)^2 \approx 123$$

Este valor indica que la señal es mucho más fuerte que el ruido en este dataset. En contraste, otro dataset que probamos, con un CV de **20%**, tenía un SNR mucho menor, alrededor de **25**. La diferencia muestra que el dataset de EUR/USD tiene una mayor preponderancia de señal sobre el ruido, lo que permite una mejor predictibilidad en los modelos de pronóstico.

Dado que sabemos que el **SNR** es aprox **123**, esto significa que el ruido es aprox  $1/123 = 0.008$  que nos da una idea del máximo error absoluto (**MAE**) en los datos Normalizados en el pronóstico del siguiente periodo que podemos obtener, ya que errores por debajo de este valor probablemente incluirían la predicción del ruido y podrían ser indicador de overfitting en el modelo predictivo usado.

### 1.3 Importancia de Analizar este Dataset

La capacidad de predecir con precisión la tasa de cambio EUR/USD tiene múltiples aplicaciones en el sector financiero. En primer lugar, es fundamental para el **trading de divisas (Forex)**, donde una mejor predicción de las tendencias puede resultar en decisiones de inversión más acertadas. Además, es esencial para la gestión de portafolios financieros, ya que permite el uso de herramientas como la **Teoría Moderna de Portafolios (MPT)**, donde la diversificación se realiza teniendo en cuenta tanto la tendencia como la variabilidad de los activos.

Por lo tanto, este dataset fue seleccionado debido a su alta calidad y baja variabilidad, lo que lo hace ideal para el análisis predictivo en el mercado de Forex, así como en la gestión de portafolios. El alto **SNR** nos da confianza en que las predicciones basadas en este dataset serán precisas y útiles en un entorno real, al mismo tiempo que nos alerta sobre los límites predictivos en función del ruido presente en los datos.



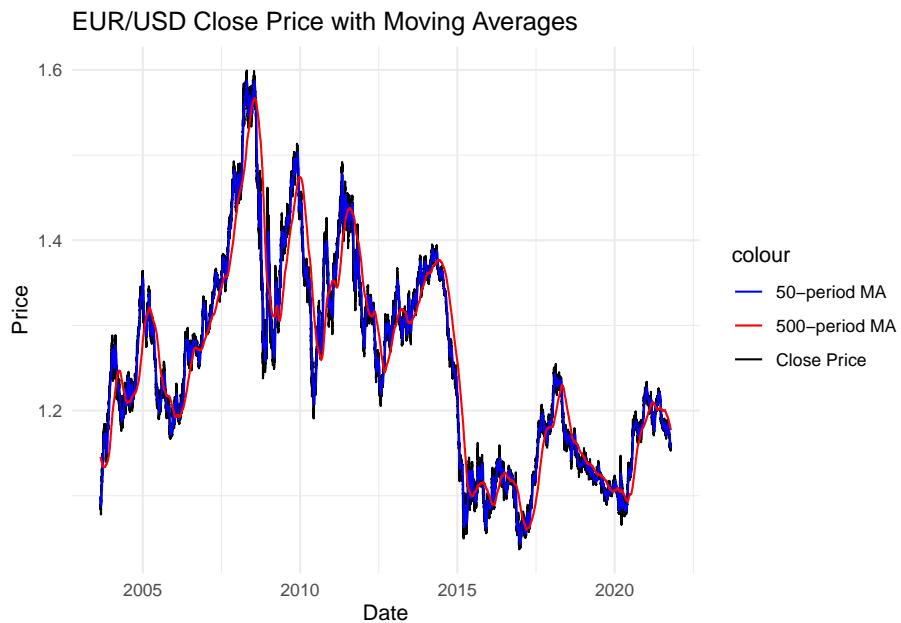
# **Chapter 2**

## **Estructura de los datos**

Con el propósito de observar las tendencias y cambios estructurales en la serie, se realizan pruebas estadísticas para conocer la estructura subyacente de la serie.

### **2.1 Cálculo de Medias Móviles Simples:**

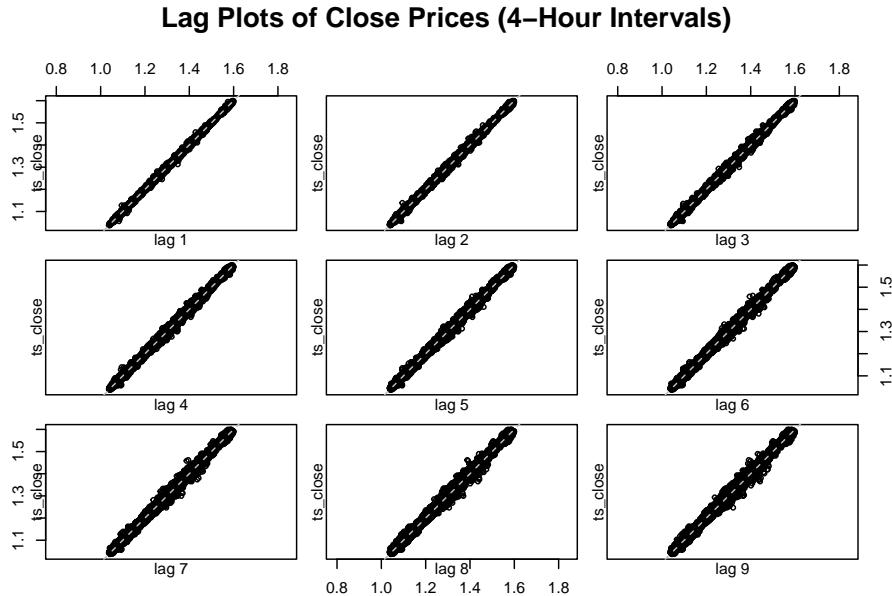
El cálculo de medias móviles es una técnica común en el análisis de series de tiempo utilizada para suavizar las fluctuaciones a corto plazo y destacar las tendencias subyacentes en los datos. En este análisis, se implementan medias móviles de corto y largo plazo para identificar patrones de comportamiento y ayudar en la toma de decisiones basadas en tendencias más claras.



- **Media móvil de 50 períodos (MA corta):**
  - Sigue de cerca las fluctuaciones del precio de cierre, respondiendo rápidamente a los cambios de tendencia.
  - Captura las tendencias a **corto plazo**, pero también refleja mucha volatilidad.
- **Media móvil de 500 períodos (MA larga):**
  - Se mueve de forma más suave, reaccionando más lentamente a los cambios de precios.
  - Indica la **tendencia a largo plazo**, proporcionando una visión más estable del comportamiento del mercado.

## 2.2 Análisis de Rezagos

Cómo se comporta la serie de tiempo con respecto a sus valores pasados, introduciendo rezagos.



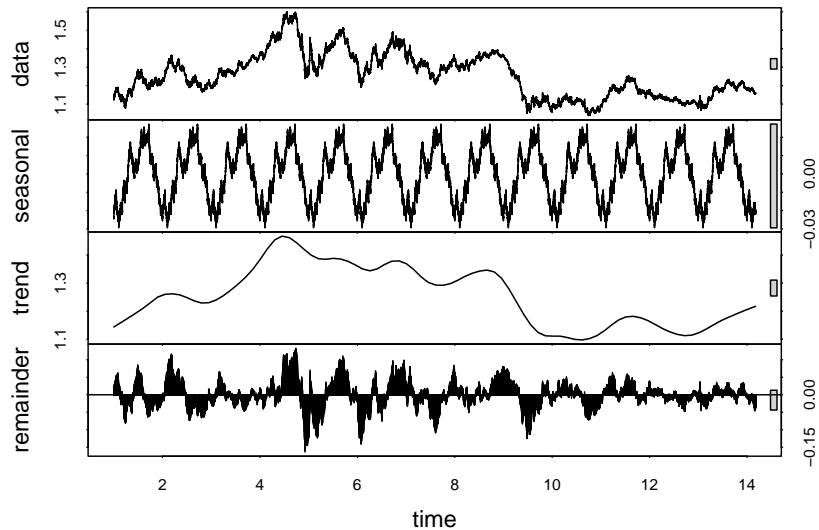
En cada uno de los gráficos, los puntos siguen una línea casi perfectamente recta, sugiriendo una **alta autocorrelación** entre los valores de la serie con sus rezagos cercanos.

La pendiente positiva indica que cuando el valor anterior era alto, el valor actual también tiende a ser alto, y lo mismo sucede para valores bajos. Esto sugiere que **la serie es muy persistente**, es decir, los precios tienden a seguir una dirección similar en el corto plazo.

Dado que no hay patrones dispersos o sin forma definida, se puede inferir que la serie no tiene cambios abruptos o comportamiento caótico entre los puntos cercanos. Esto podría indicar que **no hay mucha volatilidad** en los intervalos de 4 horas.

## 2.3 Análisis de Estacionalidad

Para detectar si existe estacionalidad (patrones repetitivos), utilizaremos **decomposición o test de estacionalidad**.



En la gráfica de **descomposición de series de tiempo** se visualizan los **componentes de la serie**: datos originales, estacionalidad, tendencia y residuales (remainder):

1. **Datos Originales (data):** En la primera gráfica (data), se observan los valores de cierre a lo largo del tiempo. Vemos fluctuaciones en los precios con algunas subidas y bajadas claras, lo que indica la volatilidad normal del mercado Forex.
2. **Componente Estacional (seasonal):** El segundo gráfico muestra un **patrón repetitivo y periódico**. Este patrón sugiere que hay **ciclos regulares** en la serie. La estacionalidad se mantiene constante a lo largo del tiempo, lo que indica que ciertos movimientos en el mercado se repiten con una periodicidad fija (en este caso, podría ser diaria o semanal). Es probable que este componente estacional refleje la actividad cíclica en horarios específicos o días determinados, como mayor volatilidad durante sesiones overlap (como entre Londres y Nueva York).
3. **Componente de Tendencia (trend):** El tercer gráfico muestra una **tendencia suavizada** que sigue la dirección general del mercado. Observamos fases de **alzas y caídas**: primero hay una subida clara, luego una caída, y finalmente otra leve tendencia hacia la estabilidad.
4. **Componente de Residuos o Resto (remainder):** El último gráfico (remainder) muestra los **residuos** o la parte de los datos que no es explicada por la tendencia ni la estacionalidad. Estos residuos parecen ser **ruido blanco**, con fluctuaciones alrededor de cero, lo que indica que no hay patrones significativos

adicionales no capturados por los otros componentes.



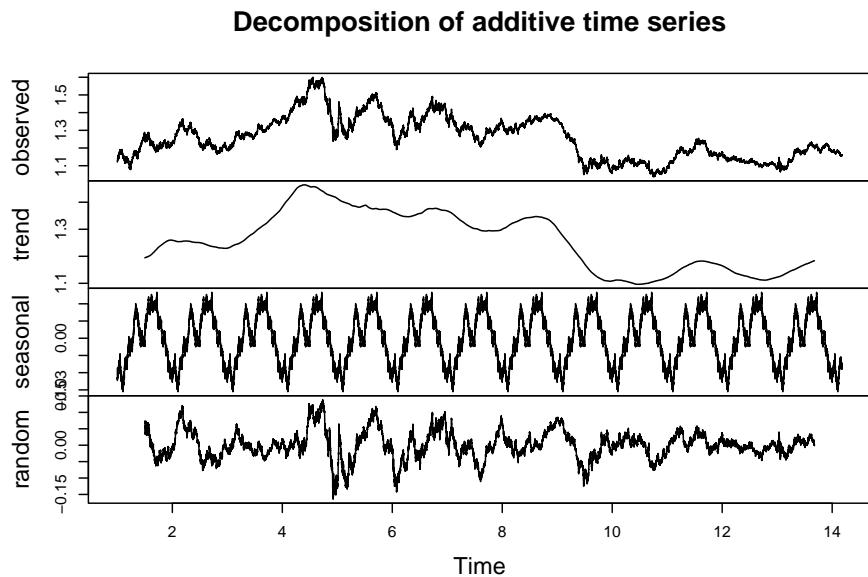
# Chapter 3

## Preprocesamiento y Visualización

En este análisis, trabajaremos con el dataset de tipo de cambio EUR/USD de Forex proporcionado, el cual incluye datos históricos de 2003 a 2021. El objetivo es estudiar las tendencias, estacionalidad y comportamiento estructural de la serie de tiempo. Además, evaluaremos si es necesario realizar transformaciones en la serie para estabilizar la varianza y facilitar la predicción.

### 3.1 Descomposición de la Serie de Tiempo

En esta etapa se busca realizar la descomposición de la serie de tiempo para identificar los componentes de **tendencia, estacionalidad y residuos**.



En la gráfica de **descomposición de series de tiempo** se visualizan los **componentes de la serie**: datos originales, estacionalidad, tendencia y residuales (remainder):

- 1. Datos Originales (data):** En la primera gráfica (data), se observan los valores de cierre a lo largo del tiempo. Vemos fluctuaciones en los precios con algunas subidas y bajadas claras, lo que indica la volatilidad normal del mercado Forex.
- 2. Componente Estacional (seasonal):** El segundo gráfico muestra un **patrón repetitivo y periódico**. Este patrón sugiere que hay **ciclos regulares** en la serie. La estacionalidad se mantiene constante a lo largo del tiempo, lo que indica que ciertos movimientos en el mercado se repiten con una periodicidad fija (en este caso, podría ser diaria o semanal). Es probable que este componente estacional refleje la actividad cíclica en horarios específicos o días determinados, como mayor volatilidad durante sesiones overlap (como entre Londres y Nueva York).
- 3. Componente de Tendencia (trend):** El tercer gráfico muestra una **tendencia suavizada** que sigue la dirección general del mercado. Observamos fases de **alzas y caídas**: primero hay una subida clara, luego una caída, y finalmente otra leve tendencia hacia la estabilidad.
- 4. Componente de Residuos o Resto (remainder):** El último gráfico (remainder) muestra los **residuos** o la parte de los datos que no es explicada por la tendencia ni la estacionalidad. Estos residuos parecen ser **ruido blanco**, con fluctuaciones alrededor de cero, lo que indica que no hay patrones significativos

adicionales no capturados por los otros componentes.

## 3.2 Prueba de Estacionariedad

La estacionariedad es importante en el análisis de series de tiempo porque indica si las propiedades estadísticas de la serie (como la media y la varianza) se mantienen constantes a lo largo del tiempo. Una serie estacionaria es generalmente más fácil de modelar y predecir.

```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##      lag      ADF p.value
## [1,] 0 -0.135  0.606
## [2,] 1 -0.133  0.606
## [3,] 2 -0.135  0.606
## [4,] 3 -0.131  0.607
## [5,] 4 -0.143  0.603
## [6,] 5 -0.146  0.603
## [7,] 6 -0.149  0.602
## [8,] 7 -0.147  0.602
## [9,] 8 -0.155  0.600
## [10,] 9 -0.156  0.600
## [11,] 10 -0.158 0.599
## [12,] 11 -0.174 0.594
## [13,] 12 -0.172 0.595
## [14,] 13 -0.169 0.596
## [15,] 14 -0.166 0.597
## Type 2: with drift no trend
##      lag      ADF p.value
## [1,] 0 -2.20  0.248
## [2,] 1 -2.21  0.244
## [3,] 2 -2.23  0.238
## [4,] 3 -2.19  0.253
## [5,] 4 -2.20  0.250
## [6,] 5 -2.20  0.248
## [7,] 6 -2.21  0.244
## [8,] 7 -2.20  0.248
## [9,] 8 -2.19  0.254
## [10,] 9 -2.18 0.257
## [11,] 10 -2.18 0.255
## [12,] 11 -2.18 0.257
## [13,] 12 -2.18 0.257
```

```

## [14,] 13 -2.16 0.263
## [15,] 14 -2.17 0.258
## Type 3: with drift and trend
##      lag      ADF p.value
## [1,] 0 -2.91 0.192
## [2,] 1 -2.93 0.186
## [3,] 2 -2.94 0.180
## [4,] 3 -2.90 0.196
## [5,] 4 -2.90 0.197
## [6,] 5 -2.90 0.196
## [7,] 6 -2.91 0.193
## [8,] 7 -2.90 0.197
## [9,] 8 -2.88 0.207
## [10,] 9 -2.87 0.210
## [11,] 10 -2.87 0.209
## [12,] 11 -2.85 0.219
## [13,] 12 -2.85 0.218
## [14,] 13 -2.84 0.223
## [15,] 14 -2.85 0.216
## -----
## Note: in fact, p.value = 0.01 means p.value <= 0.01

## $type1
##      lag      ADF p.value
## [1,] 0 -0.1349353 0.6056581
## [2,] 1 -0.1332845 0.6061322
## [3,] 2 -0.1350012 0.6056392
## [4,] 3 -0.1311022 0.6067589
## [5,] 4 -0.1432485 0.6032707
## [6,] 5 -0.1455745 0.6026027
## [7,] 6 -0.1491280 0.6015823
## [8,] 7 -0.1466525 0.6022932
## [9,] 8 -0.1553530 0.5997946
## [10,] 9 -0.1558747 0.5996447
## [11,] 10 -0.1575228 0.5991714
## [12,] 11 -0.1743717 0.5943327
## [13,] 12 -0.1723248 0.5949206
## [14,] 13 -0.1686293 0.5959819
## [15,] 14 -0.1662663 0.5966605
##
## $type2
##      lag      ADF p.value
## [1,] 0 -2.199690 0.2481241
## [2,] 1 -2.210308 0.2438767
## [3,] 2 -2.225752 0.2376991
## [4,] 3 -2.186315 0.2534741

```

```

## [5,]   4 -2.196090 0.2495641
## [6,]   5 -2.199991 0.2480038
## [7,]   6 -2.210932 0.2436273
## [8,]   7 -2.199519 0.2481923
## [9,]   8 -2.185380 0.2538481
## [10,]  9 -2.178336 0.2566657
## [11,] 10 -2.183564 0.2545744
## [12,] 11 -2.178602 0.2565592
## [13,] 12 -2.178490 0.2566041
## [14,] 13 -2.163515 0.2625942
## [15,] 14 -2.174827 0.2580694
##
## $type3
##      lag      ADF  p.value
## [1,]  0 -2.911720 0.1919072
## [2,]  1 -2.925278 0.1861988
## [3,]  2 -2.940306 0.1798712
## [4,]  3 -2.902268 0.1958871
## [5,]  4 -2.899542 0.1970351
## [6,]  5 -2.901407 0.1962497
## [7,]  6 -2.909540 0.1928254
## [8,]  7 -2.900196 0.1967594
## [9,]  8 -2.875536 0.2071427
## [10,] 9 -2.867569 0.2104971
## [11,] 10 -2.871608 0.2087965
## [12,] 11 -2.848029 0.2187245
## [13,] 12 -2.850350 0.2177475
## [14,] 13 -2.838511 0.2227323
## [15,] 14 -2.853430 0.2164506

```

La interpretación es la siguiente:

- **Hipótesis nula (H0):** La serie no es estacionaria (tiene una raíz unitaria).
- **Hipótesis alternativa (H1):** La serie es estacionaria.

En todas las configuraciones (sin tendencia ni drift, con drift, y con drift y tendencia), los p-valores son mayores a 0.05. Esto implica que, bajo ninguna de estas configuraciones, la serie es estacionaria en su forma actual.

### 3.3 Diferenciación para Estacionariedad

Como la serie no es estacionaria, el siguiente paso es aplicar una diferenciación para intentar volverla estacionaria. La diferenciación ayuda a eliminar tenden-

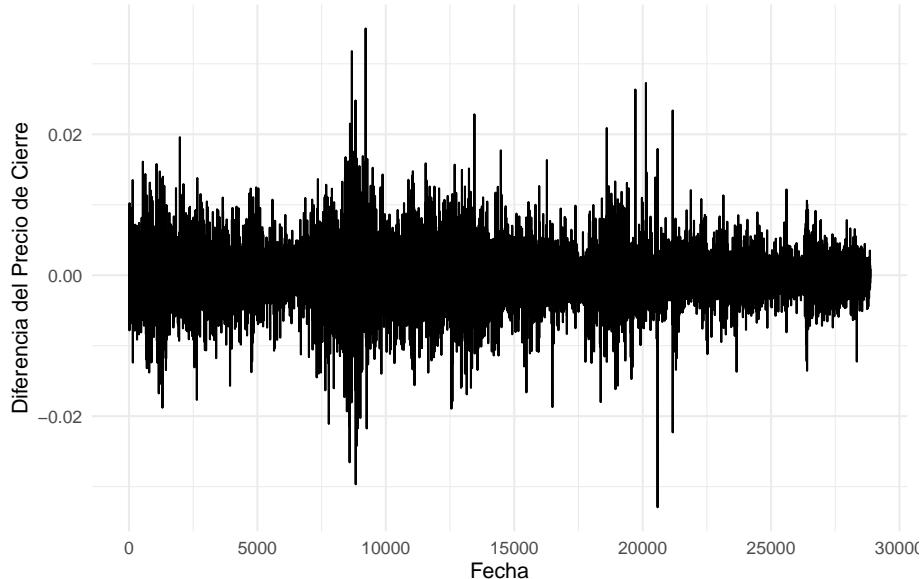
cias y hacer que las propiedades estadísticas de la serie se mantengan constantes a lo largo del tiempo.

Aplicaremos una diferenciación de primer orden y realizaremos nuevamente la prueba ADF para verificar si la serie se ha vuelto estacionaria.

Para este análisis, utilizaremos la columna de cierre (`Close`) del dataset como nuestra serie de tiempo principal. Convertiremos los datos a formato de serie temporal.

Aplicaremos una primera diferenciación.

Serie Temporal Diferenciada del Precio de Cierre EUR/USD



Después de la primera diferenciación, se puede obtener que por cada tick quedó el valor del return ( $x-x_{\text{ant}}$ ) que es el resultado de la aplicación de la diferenciaión de primer orden. Como se aprecia, la serie ya no tiene una tendencia sino que presenta un comportamiento estacionario, sin embargo, esta serie de tiempo parece ser menos predecible que la serie de tiempo antes de la diferenciación, ya que evidentemente tiene mayor desviación estándar y por tanto un mayor Coeficiente de Variación y un menor SNR.

### 3.4 Verificación de Estacionariedad en la Serie Diferenciada

Aplicamos nuevamente la prueba Dickey-Fuller a la serie diferenciada para verificar si ahora es estacionaria.

### 3.4. VERIFICACIÓN DE ESTACIONARIEDAD EN LA SERIE DIFERENCIADA21

```
## Augmented Dickey-Fuller Test
## alternative: stationary
##
## Type 1: no drift no trend
##          lag      ADF p.value
## [1,]    0 -169.3   0.01
## [2,]    1 -119.0   0.01
## [3,]    2 -99.3    0.01
## [4,]    3 -84.8    0.01
## [5,]    4 -75.7    0.01
## [6,]    5 -68.7    0.01
## [7,]    6 -64.1    0.01
## [8,]    7 -60.1    0.01
## [9,]    8 -56.8    0.01
## [10,]   9 -53.7    0.01
## [11,]  10 -50.9    0.01
## [12,]  11 -48.8    0.01
## [13,]  12 -47.3    0.01
## [14,]  13 -45.4    0.01
## [15,]  14 -43.7    0.01
## Type 2: with drift no trend
##          lag      ADF p.value
## [1,]    0 -169.3   0.01
## [2,]    1 -119.0   0.01
## [3,]    2 -99.3    0.01
## [4,]    3 -84.8    0.01
## [5,]    4 -75.7    0.01
## [6,]    5 -68.7    0.01
## [7,]    6 -64.1    0.01
## [8,]    7 -60.1    0.01
## [9,]    8 -56.8    0.01
## [10,]   9 -53.7    0.01
## [11,]  10 -50.9    0.01
## [12,]  11 -48.8    0.01
## [13,]  12 -47.3    0.01
## [14,]  13 -45.4    0.01
## [15,]  14 -43.7    0.01
## Type 3: with drift and trend
##          lag      ADF p.value
## [1,]    0 -169.3   0.01
## [2,]    1 -119.0   0.01
## [3,]    2 -99.3    0.01
## [4,]    3 -84.9    0.01
## [5,]    4 -75.7    0.01
## [6,]    5 -68.7    0.01
## [7,]    6 -64.1    0.01
```

```
## [8,]    7  -60.1   0.01
## [9,]    8  -56.8   0.01
## [10,]   9  -53.7   0.01
## [11,]  10  -50.9   0.01
## [12,]  11  -48.8   0.01
## [13,]  12  -47.3   0.01
## [14,]  13  -45.4   0.01
## [15,]  14  -43.7   0.01
## ----
## Note: in fact, p.value = 0.01 means p.value <= 0.01

## $type1
##      lag      ADF p.value
## [1,]  0 -169.27722  0.01
## [2,]  1 -119.03874  0.01
## [3,]  2  -99.30334  0.01
## [4,]  3  -84.84969  0.01
## [5,]  4  -75.69590  0.01
## [6,]  5  -68.70703  0.01
## [7,]  6  -64.07591  0.01
## [8,]  7  -60.05652  0.01
## [9,]  8  -56.77621  0.01
## [10,] 9  -53.68723  0.01
## [11,] 10 -50.91061  0.01
## [12,] 11 -48.81278  0.01
## [13,] 12 -47.29027  0.01
## [14,] 13 -45.38630  0.01
## [15,] 14 -43.68676  0.01
##
## $type2
##      lag      ADF p.value
## [1,]  0 -169.27433  0.01
## [2,]  1 -119.03671  0.01
## [3,]  2  -99.30166  0.01
## [4,]  3  -84.84825  0.01
## [5,]  4  -75.69462  0.01
## [6,]  5  -68.70587  0.01
## [7,]  6  -64.07484  0.01
## [8,]  7  -60.05550  0.01
## [9,]  8  -56.77525  0.01
## [10,] 9  -53.68632  0.01
## [11,] 10 -50.90973  0.01
## [12,] 11 -48.81194  0.01
## [13,] 12 -47.28946  0.01
## [14,] 13 -45.38553  0.01
## [15,] 14 -43.68601  0.01
```

```

## $type3
##      lag      ADF p.value
## [1,]    0 -169.27489  0.01
## [2,]    1 -119.03832  0.01
## [3,]    2  -99.30402  0.01
## [4,]    3  -84.85094  0.01
## [5,]    4  -75.69772  0.01
## [6,]    5  -68.70929  0.01
## [7,]    6  -64.07866  0.01
## [8,]    7  -60.05942  0.01
## [9,]    8  -56.77944  0.01
## [10,]   9  -53.69075  0.01
## [11,]  10  -50.91391  0.01
## [12,]  11  -48.81642  0.01
## [13,]  12  -47.29426  0.01
## [14,]  13  -45.39063  0.01
## [15,]  14  -43.69120  0.01

```

## 3.5 Justificación de la Transformación

Dado que la serie original no era estacionaria, fue necesario aplicar una diferenciación de primer orden para hacerla estacionaria. Esta transformación es importante para poder aplicar modelos de series de tiempo que asumen estacionariedad y para obtener mejores resultados en el análisis de patrones y predicciones.

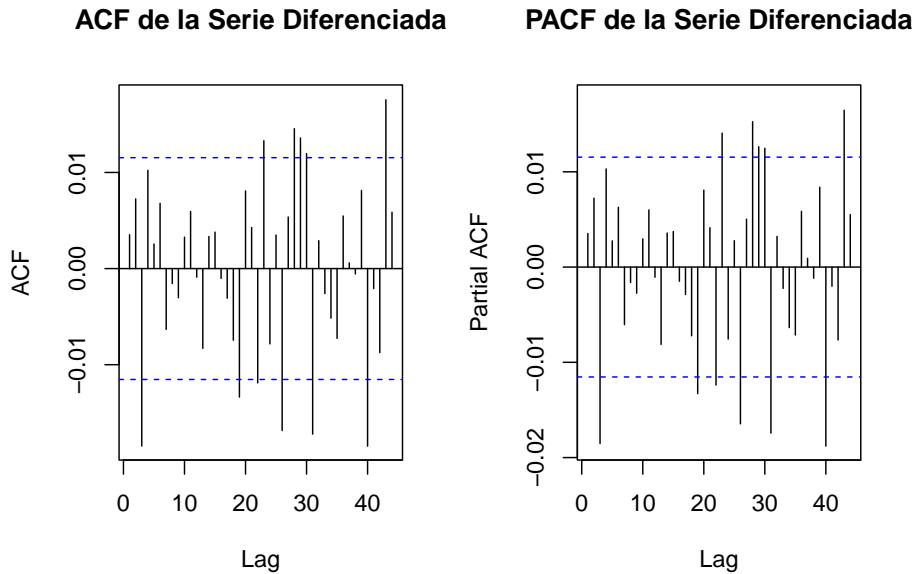
## 3.6 Análisis de Autocorrelación

Graficaremos las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para observar la dependencia temporal en los datos diferenciados.

```

# Graficar ACF y PACF de la serie diferenciada
par(mfrow = c(1, 2))
Acf(forex_diff, main = "ACF de la Serie Diferenciada")
Pacf(forex_diff, main = "PACF de la Serie Diferenciada")

```



```
par(mfrow = c(1,1))
```

La ACF a la izquierda muestra cómo los valores actuales de la serie diferenciada están correlacionados con sus valores en diferentes rezagos (lags). Algunos puntos de la ACF están fuera de las líneas de significancia (líneas punteadas azules), lo cual sugiere que hay correlaciones significativas en esos rezagos específicos. Este patrón puede ser indicativo de que aún existen estructuras autoregresivas o de medias móviles en la serie, incluso después de la diferenciación.

La PACF a la derecha muestra la autocorrelación de la serie diferenciada en cada rezago eliminando el efecto de los rezagos intermedios. Similar a la ACF, algunos valores están fuera de las líneas de significancia, lo que indica correlación significativa en esos rezagos específicos. Este patrón puede sugerir la presencia de efectos autoregresivos en los rezagos correspondientes.

Los picos significativos en la ACF y PACF sugieren que la serie diferenciada podría beneficiarse de un modelo ARIMA para capturar la estructura subyacente. Dependiendo de la cantidad de rezagos significativos en cada gráfico, podría ser apropiado un modelo ARIMA específico (por ejemplo, con ciertos órdenes autoregresivos y de medias móviles).

### 3.7 Modelo ARIMA

Utilizaremos `auto.arima` para identificar el mejor modelo ARIMA para los datos.

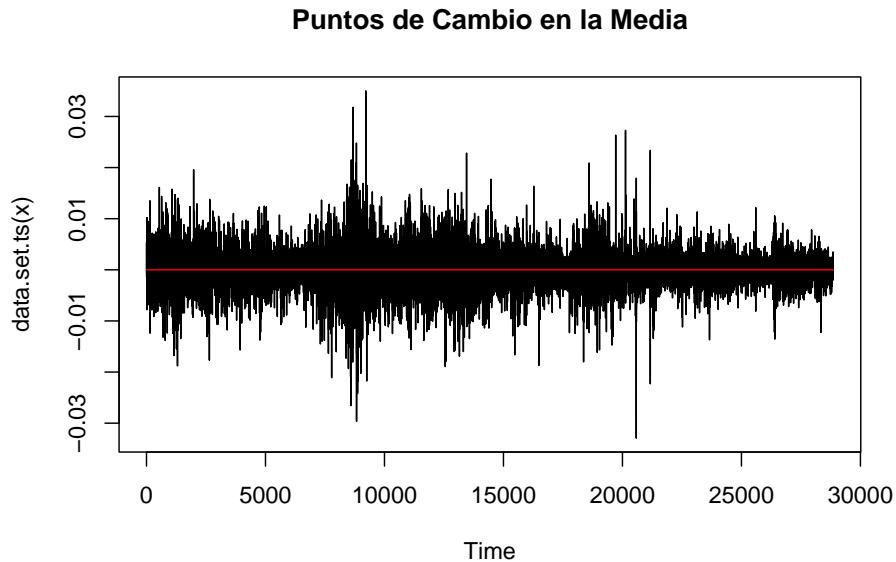
```
# Ajuste del modelo ARIMA
forex_arima <- auto.arima(forex_diff)
summary(forex_arima)

## Series: forex_diff
## ARIMA(0,0,0) with zero mean
##
## sigma^2 = 8.978e-06: log likelihood = 126732.8
## AIC=-253463.6   AICc=-253463.6   BIC=-253455.4
##
## Training set error measures:
##               ME        RMSE       MAE      MPE      MAPE      MASE      ACF1
## Training set 1.304966e-06 0.002996269 0.001979825 100    100 0.675058 0.003518741
```

El modelo ajustado para la serie `forex_diff` es un ARIMA(0,0,0) con media cero, lo que sugiere que la serie no presenta patrones autoregresivos ni de medias móviles significativos, siendo esencialmente ruido blanco. El valor de  $\sigma^2 = 8.978 \times 10^{-6}$  representa la varianza del error, con una alta verosimilitud (log likelihood) de 126732.8. Los criterios de información, AIC y BIC, son de -253463.6 y -253455.4, respectivamente, indicando un buen ajuste para este modelo sencillo. Las medidas de error en el conjunto de entrenamiento muestran un error medio (ME) cercano a cero (1.30e-06) y un RMSE de 0.002996, lo cual refleja una precisión razonable. La autocorrelación en el primer rezago (ACF1) es baja (0.0035), sugiriendo independencia en los residuos.

### 3.8 Detección de Puntos de Cambio

Usaremos la función `cpt.mean` para detectar cambios significativos en la media de la serie.



No se detectaron puntos de cambio, debido a que después de la diferenciación, se convierte básicamente en ruido blanco.

### 3.9 Media Cero de los Residuos

Comprobamos si la media de los residuos es cero.

```
# Prueba t en los residuos
residuals_arima <- residuals(forex_arima)
t_test_residuals <- t.test(residuals_arima)
print(t_test_residuals)

##
## One Sample t-test
##
## data: residuals_arima
## t = 0.073986, df = 28858, p-value = 0.941
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
## -3.326619e-05 3.587612e-05
## sample estimates:
##   mean of x
## 1.304966e-06
```

La prueba t de una muestra realizada sobre los residuos (residuals\_arima) arroja un valor de =0.073986 con 28,858 grados de libertad y un valor p de 0.941. Dado que el valor p es significativamente mayor a 0.05, no rechazamos la hipótesis nula de que la media de los residuos es igual a cero. Esto sugiere que los residuos no presentan un sesgo significativo. El intervalo de confianza del 95% para la media de los residuos y la media estimada muy cercana a cero es consistente con un modelo bien ajustado sin tendencia sistemática en los errores.

## 3.10 Independencia de los Residuos

Evaluamos la independencia de los residuos usando la prueba de Ljung-Box.

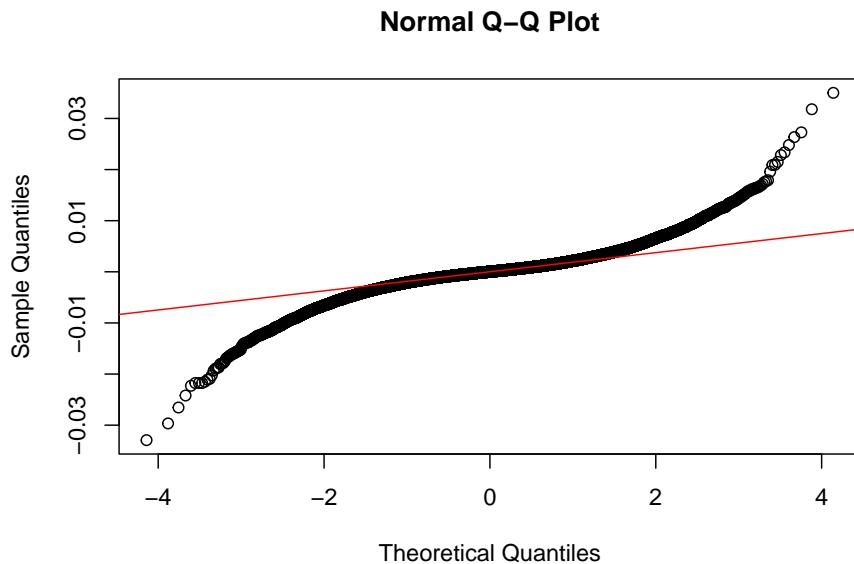
```
# Prueba de independencia
Box.test(residuals_arima, lag = 20, type = "Ljung-Box")
```

```
##
## Box-Ljung test
##
## data: residuals_arima
## X-squared = 30.772, df = 20, p-value = 0.05827
```

Los datos cargados contienen 28,860 filas y 6 columnas de información sobre el tipo de cambio EUR/USD. La prueba de Dickey-Fuller Aumentada (ADF) realizada en tres configuraciones (sin constante ni tendencia, con constante sin tendencia, y con constante y tendencia) muestra valores ADF altamente negativos y p-valores menores o iguales a 0.01, lo que indica que la serie diferenciada es estacionaria. Además, la prueba de Box-Ljung aplicada a los residuos del modelo ARIMA arroja un valor de 30.772 con un valor p de 0.05827, lo cual sugiere que los residuos **no tienen autocorrelación significativa**, indicando independencia en los errores del modelo.

## 3.11 Distribución de los Residuos

Analizaremos la normalidad de los residuos con un gráfico Q-Q.



El gráfico Q-Q muestra que los residuos del modelo se alinean con la normalidad en el centro de la distribución, pero presentan desviaciones significativas en las colas. Esto sugiere que, aunque los residuos se comportan aproximadamente como una distribución normal en el centro, tienen colas más pesadas de lo esperado, lo que indica la presencia de valores extremos.

## Chapter 4

# Análisis de Series de Tiempo con el Método Holt-Winters

Este documento realiza un análisis de series de tiempo utilizando el método de Holt-Winters aplicado exclusivamente a la columna `close` del dataset `EURUSD_ForexTrading_4hrs.csv`. Se utilizarán solo 6000 datos, normalizando la columna `close`, dividiendo en conjunto de entrenamiento y prueba, y calculando la métricas de error MAE en el conjunto de entrenamiento y en el conjunto de prueba.

### 4.1 Carga de Bibliotecas y Datos

El dataset `EURUSD_ForexTrading_4hrs.csv` contiene datos de trading del par de divisas EUR/USD con una frecuencia de 4 horas. Solo se ha seleccionado la columna `close` con los primeros 6000 datos para este análisis.

Table 4.1: Primeras filas del dataset EURUSD ForexTrading 4hrs Columna close

| close   |
|---------|
| 1.12274 |
| 1.12126 |
| 1.12113 |
| 1.12174 |
| 1.12712 |
| 1.12804 |

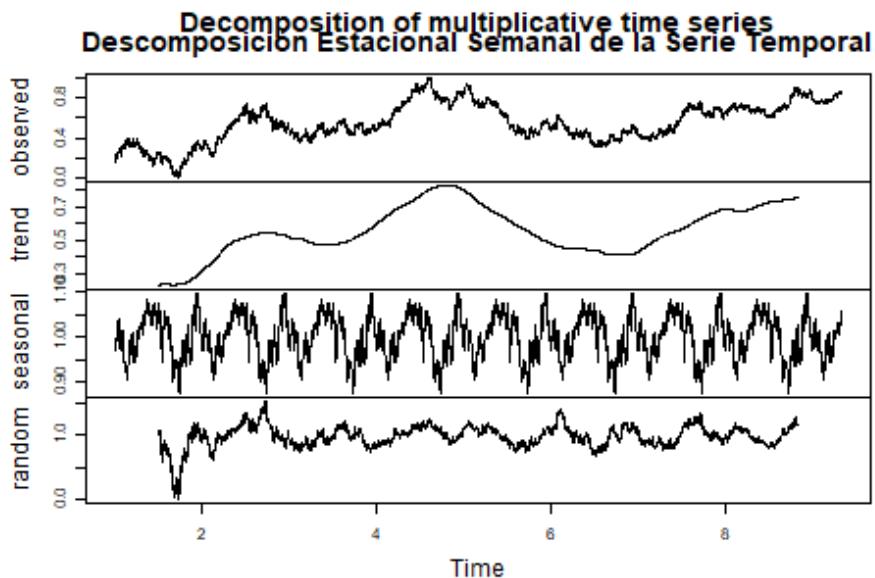
Table 4.2: Columna 'close' Normalizada

| x         |
|-----------|
| 0.1566893 |
| 0.1515141 |
| 0.1510595 |
| 0.1531925 |
| 0.1720050 |
| 0.1752220 |

## 4.2 Normalización de la Columna 'close'

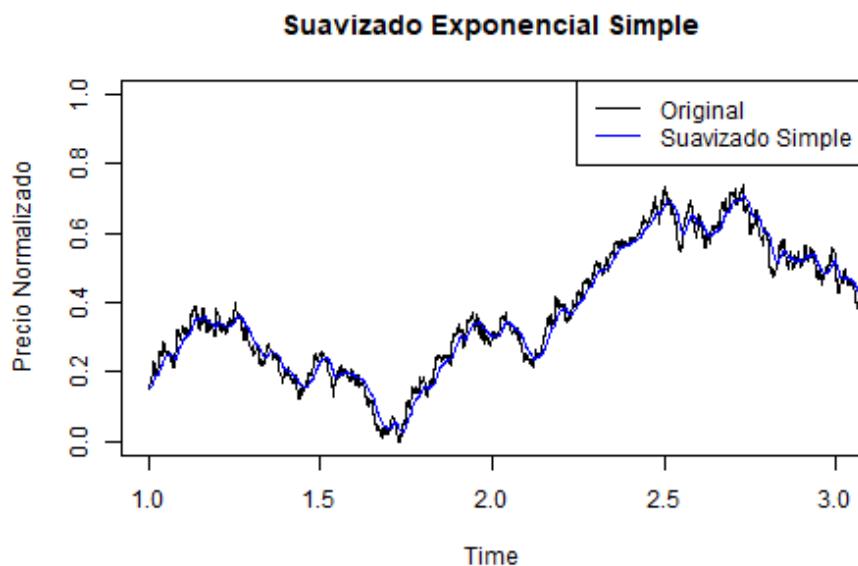
Se ha normalizado la columna `close` utilizando la técnica Min-Max, transformando los valores entre 0 y 1 para mejorar la estabilidad del modelo.

## 4.3 Descomposición Estacional



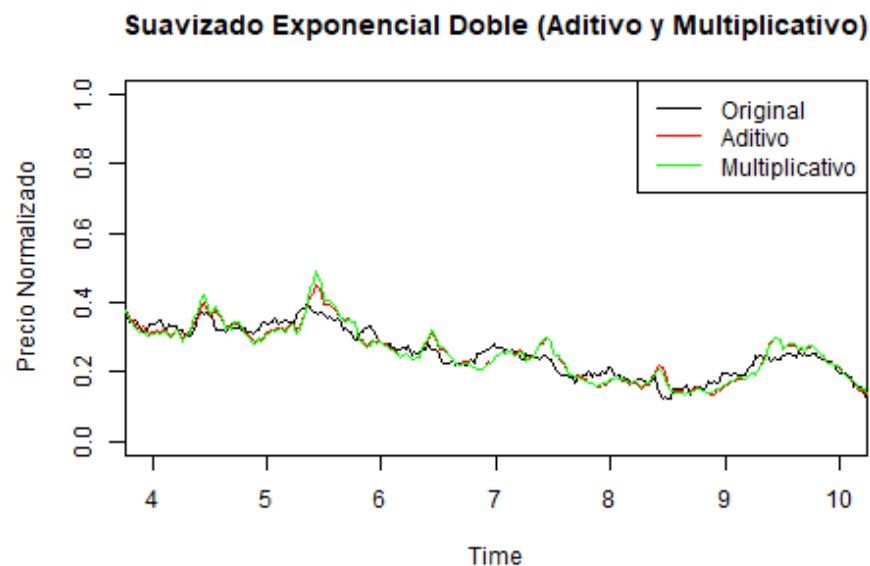
Descomponemos la serie temporal en componentes de tendencia, estacionalidad y ruido para analizar los patrones internos de la serie temporal antes de aplicar el modelo.

#### 4.4 Suavizado Exponencial Simple



Aplicamos el suavizado exponencial simple a la serie `close` para visualizar una versión suavizada de la serie de tiempo. Se pude observar como la señal original contiene mas ruido que la suavizada.

## 4.5 Suavizado Exponencial Doble (Aditivo y Multiplicativo)



Se puede apreciar que tanto el suavizado aditivo como el multiplicativo producen una estimación cercana a los datos originales, aunque estas señales contienen mas ruido que el suavizado simple e incluso al parecer mas que la señal original.

## 4.6 Calculo de error

Se calcularon los errores de los modelos multiplicativo y aditivo en el dataset de training. Se usa un ajuste de 0.001 para el modelo multiplicativo, porque este requiere que todos los datos sean positivos y mayores que cero (no admite ceros), y como los datos fueron nomalizados con min-max, obligatoriamente existe al menos un valor de cero.

```
## MAE en entrenamiento (Aditivo): 0.01972768
```

```
## MAE en entrenamiento (Multiplicativo): 0.05277391
```

Finalmente se calcularon los errores de los modelos en el dataset de validación

```
## MAE en validación (Aditivo): 2.350168
```

```
## MAE en validación (Multiplicativo): 4.704901
```

Este gigantesco error es debido a que el modelo trata de predecir todo el dataset de validación de una sola vez (1238 ticks). En otros modelos predictivos en series de tiempo como redes neuronales, se usa un sliding window usando los últimos 128 ticks como entrada del modelo, se predice el siguiente, y esto se repite para cada tick, luego se promedian todos los errores y esa es la medida de desempeño de la red neuronal.

Para poder comparar el desempeño predictivo del modelo Holt-Winter con otros modelos predictivos en una serie de tiempo larga como la nuestra, probablemente se requiera usar sliding window como en las redes neuronales, se requeriría adaptar el modelo Holt-Winter para que se entrene con una ventana y prediga segmentos cortos que se concatenan y que formarían la señal pronosticada, con la cual se calcularían y promediarían los errores por tick, en lugar de tratar de predecir la serie de tiempo completa de una sola vez.

## 4.7 Conclusiones

El método Holt-Winters aplicado a la columna `close` del conjunto de datos muestra que este modelo es capaz de capturar patrones de tendencia y estacionalidad en los datos de precios de cierre normalizados. Las métricas de evaluación como MAE muestran la precisión del modelo tanto en el conjunto de entrenamiento como en el conjunto de prueba, donde se puede apreciar que la predicción de todo el dataset de validación completo no es una buena forma de evaluar el desempeño de estos modelos, especialmente para compararlos con modelos ampliamente usados como las redes neuronales.