

EUR/USD

Harvey Bastidas, Andrés Caicedo, Alexander Alvarado

2024-11-04



# 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 . . . . .	16
3.3 Diferenciación para Estacionariedad . . . . .	17
3.4 Justificación de la Transformación . . . . .	17



# 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.



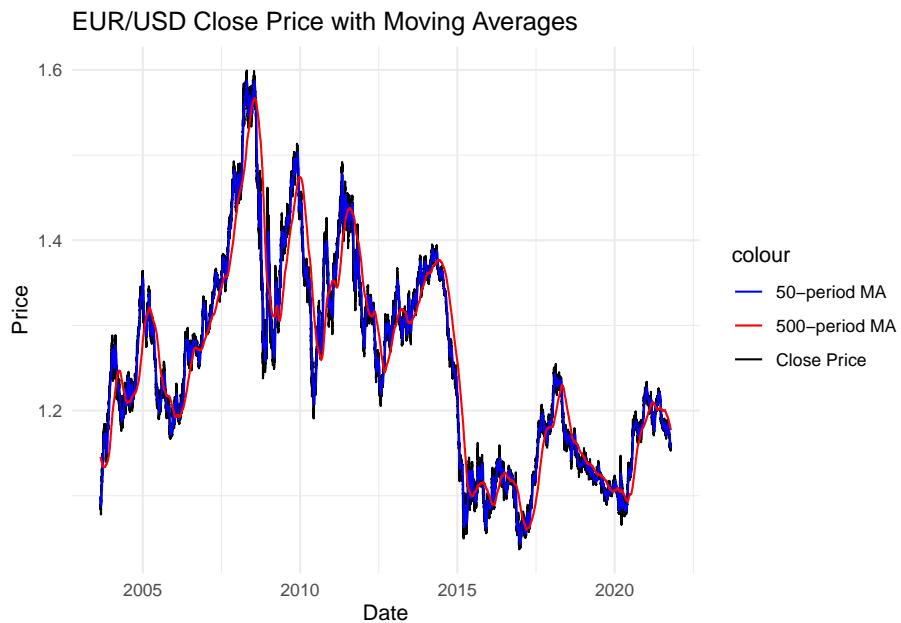
# **Chapter 2**

## **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:**

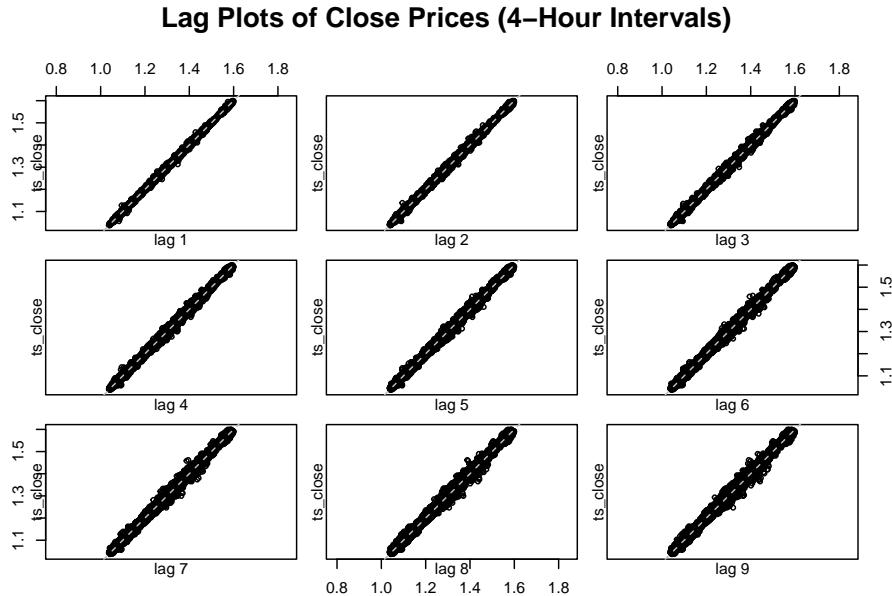
El cálculo de medias móviles es una técnica común en el análisis de series de tiempo utilizada para suavizar las fluctuaciones a corto plazo y destacar las tendencias subyacentes en los datos. En este análisis, se implementan medias móviles de corto y largo plazo para identificar patrones de comportamiento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



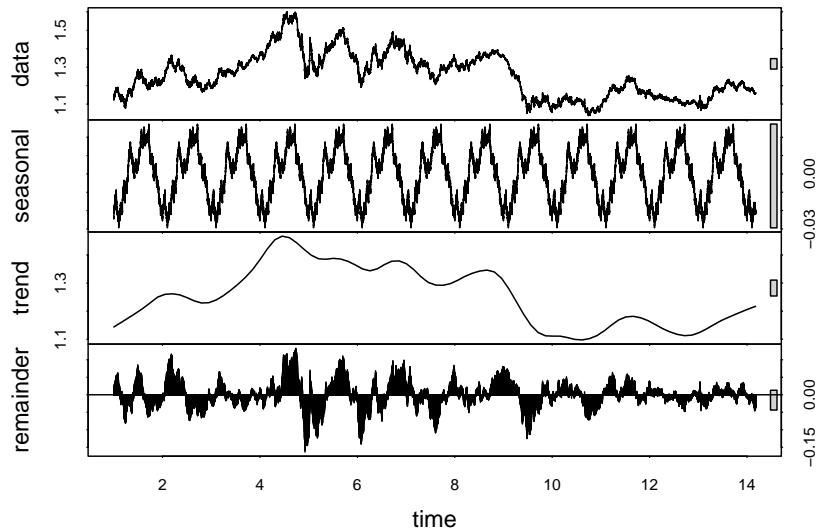
En cada uno de los gráficos, los puntos siguen una línea casi perfectamente recta, sugiriendo una **alta autocorrelación** entre los valores de la serie con sus rezagos cercanos.

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.

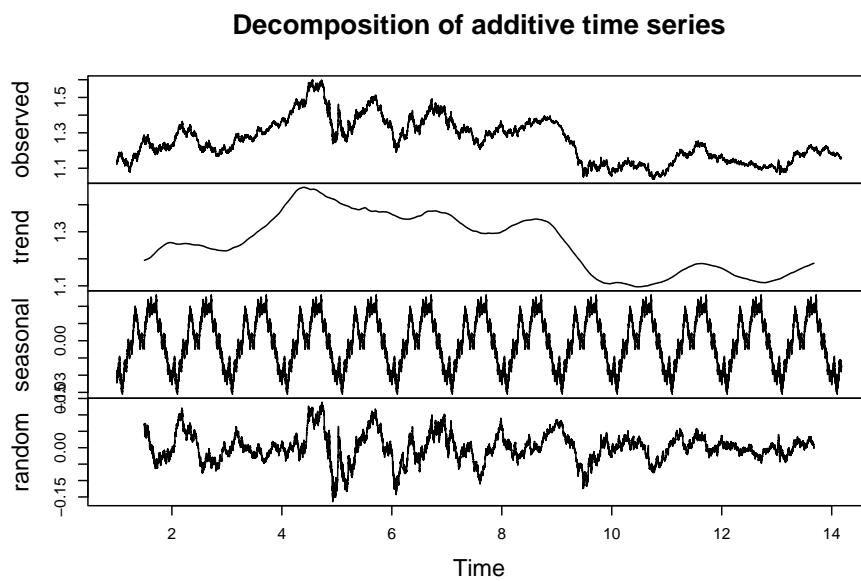


# Chapter 3

## Preprocesamiento y Visualización

### 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 **com-**

**ponentes 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  
##  
## data: ts_data  
## Dickey-Fuller = -2.8283, Lag order = 30, p-value = 0.2269  
## alternative hypothesis: stationary
```

La interpretación es la siguiente:

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

- **Hipótesis alternativa (H1):** La serie es estacionaria.

Con un valor de p (**0.2269**) al ser mayor a 0.05, significa que no podemos rechazar la hipótesis nula, por lo tanto la prueba ADF sugiere que la serie no 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 tendencias 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.

```
##  
## Augmented Dickey-Fuller Test  
##  
## data: ts_data_diff  
## Dickey-Fuller = -30.409, Lag order = 30, p-value = 0.01  
## alternative hypothesis: stationary
```

Con el valor p (0.01) indica que podemos rechazar la hipótesis nula con un nivel de confianza del 95% y por lo tanto la serie diferenciada **es estacionaria** después de la diferenciación de primer orden.

### 3.4 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.