INTRODUCCIÓN A LA **CIENCIA DE DATOS**

UNIVERSIDAD DE MONTEVIDEO

Data Science Work Flow

UM
UNIVERSIDAD DE MONTEVIDEO

1. Definición de la Pregunta
2. Obtención de los Datos
3. Preparación de los Datos
4. Análisis Exploratorio
5. **Creación y Evaluación de Modelos**
6. Comunicación de Resultados

SERIES DE TIEMPO

INTRODUCCIÓN

Series de Tiempo

Introducción

- Las **series de tiempo** son datos en **forma de secuencia** de observaciones **cuantitativas** sobre un sistema o proceso realizadas en **puntos sucesivos en el tiempo**.

- Normalmente, los puntos en el tiempo están **igualmente espaciados**.

- **Ejemplos**: producto interno bruto, los volúmenes de ventas, los precios de las acciones, los atributos climáticos cuando se registran en un intervalo de tiempo de varios años, meses, días, horas, etc.

- La **frecuencia** de observación depende de la naturaleza de la variable y sus aplicaciones.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Objetivos

Introducción

El análisis de series de tiempo tiene como **objetivo** utilizar los datos para varios propósitos que se pueden clasificar en términos generales como:

- **Comprender e interpretar** las fuerzas subyacentes que producen el estado observado de un sistema o proceso a lo largo del tiempo.
- Para **pronosticar** el estado futuro del sistema o proceso en términos de características observables.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Objetivos

Introducción

El análisis de series de tiempo tiene como **objetivo** utilizar los datos para varios propósitos que se pueden clasificar en términos generales como:

- **Comprender e interpretar** las fuerzas subyacentes que producen el estado observado de un sistema o proceso a lo largo del tiempo.

- Para **pronosticar** el estado futuro del sistema o proceso en términos de características observables.

Para lograr los objetivos, se aplican diferentes **métodos estadísticos** que exploraran y modelan las **estructuras internas** de los datos de las series de tiempo, tales como tendencias, fluctuaciones estacionales, comportamiento cíclico y cambios irregulares.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Ejemplo

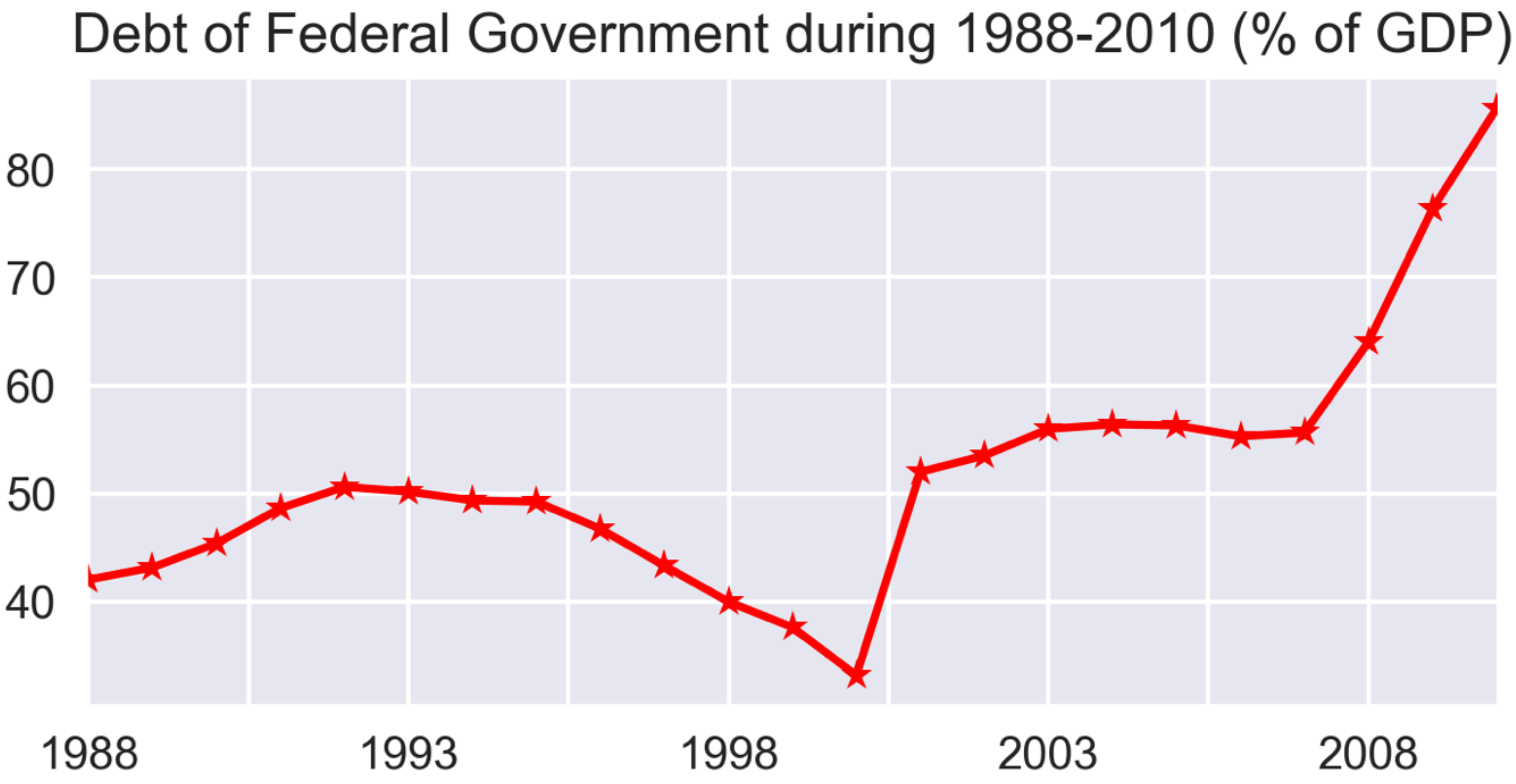
Introducción

	Federal Military Expenditure	Debt of Federal Government
1988	57993	0258
1989	37472	1439
1990	12025	3772
1991	53985	633
1992	66626	6016
1993	32693	1657
1994	94129	3475
1995	63849	2366
1996	35074	7174
1997	2099	2997

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Ejemplo

Introducción



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Estructura Interna

Introducción

- Una serie de tiempo se puede expresar como **$x_t = f_t + s_t + c_t + e_t$**, que es una **suma de los componentes tendenciales, estacionales, cíclicos e irregulares** en ese orden. Aquí, t es el índice de tiempo en el que se han realizado observaciones sobre la serie en t = 1,2,3 ... N puntos sucesivos e igualmente espaciados en el tiempo.

- El **objetivo** del análisis de series de tiempo es **descomponer una serie de tiempo en sus características constituyentes y desarrollar modelos matemáticos para cada una**. Estos modelos se utilizan luego para comprender qué causa el comportamiento observado de la serie temporal y predecir la serie para puntos futuros en el tiempo.

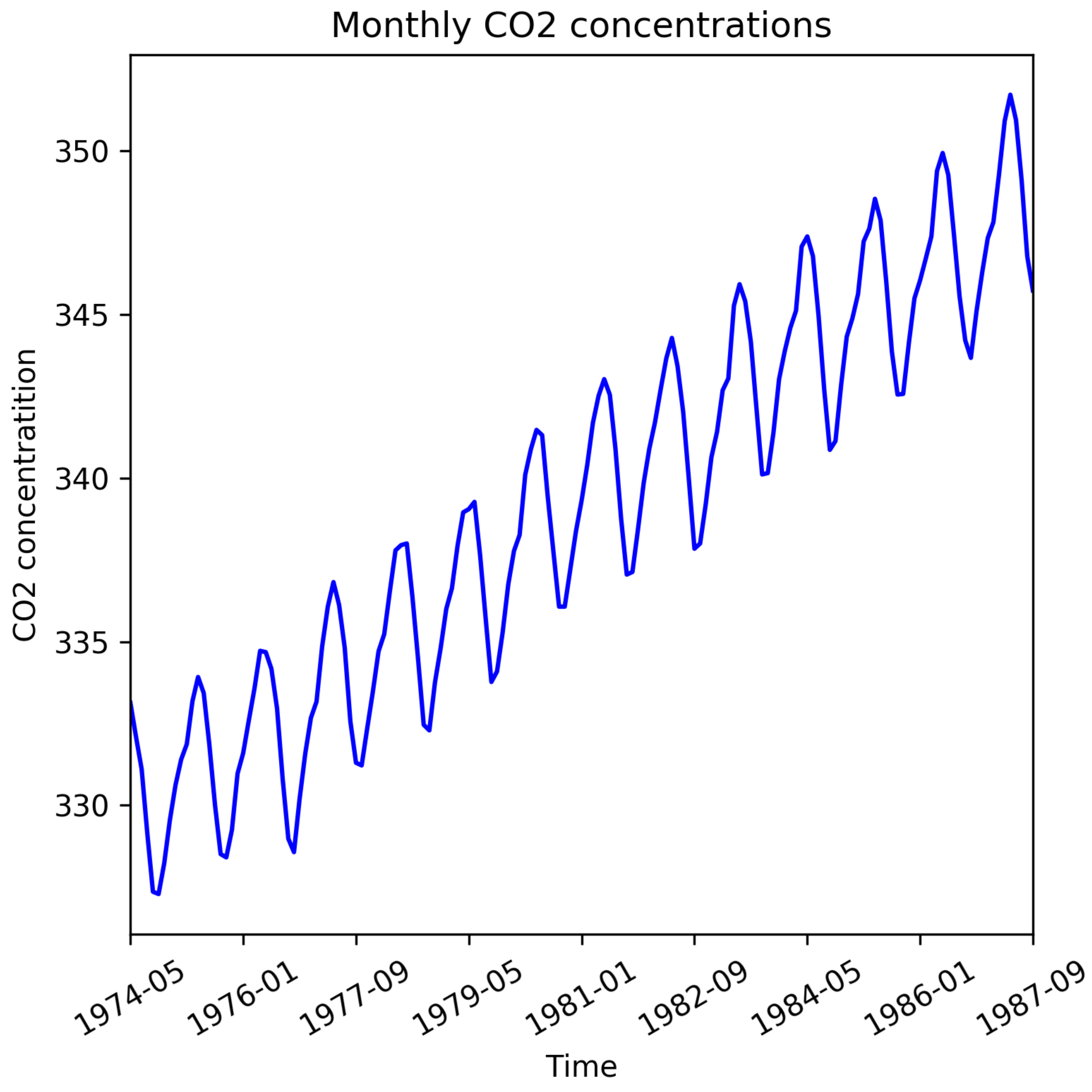
Fuente: *Practical Time Series Analysis*, Pal & Prakash

Tendencia

Introducción

- Cuando una serie de tiempo muestra un **movimiento ascendente o descendente a largo plazo**, se dice que tiene una **tendencia general.**

- Una forma rápida de verificar la presencia de una tendencia general es **graficar la serie de tiempo** como en la siguiente figura, que muestra las concentraciones de CO_2 en el aire medidas durante 1974 a 1987:

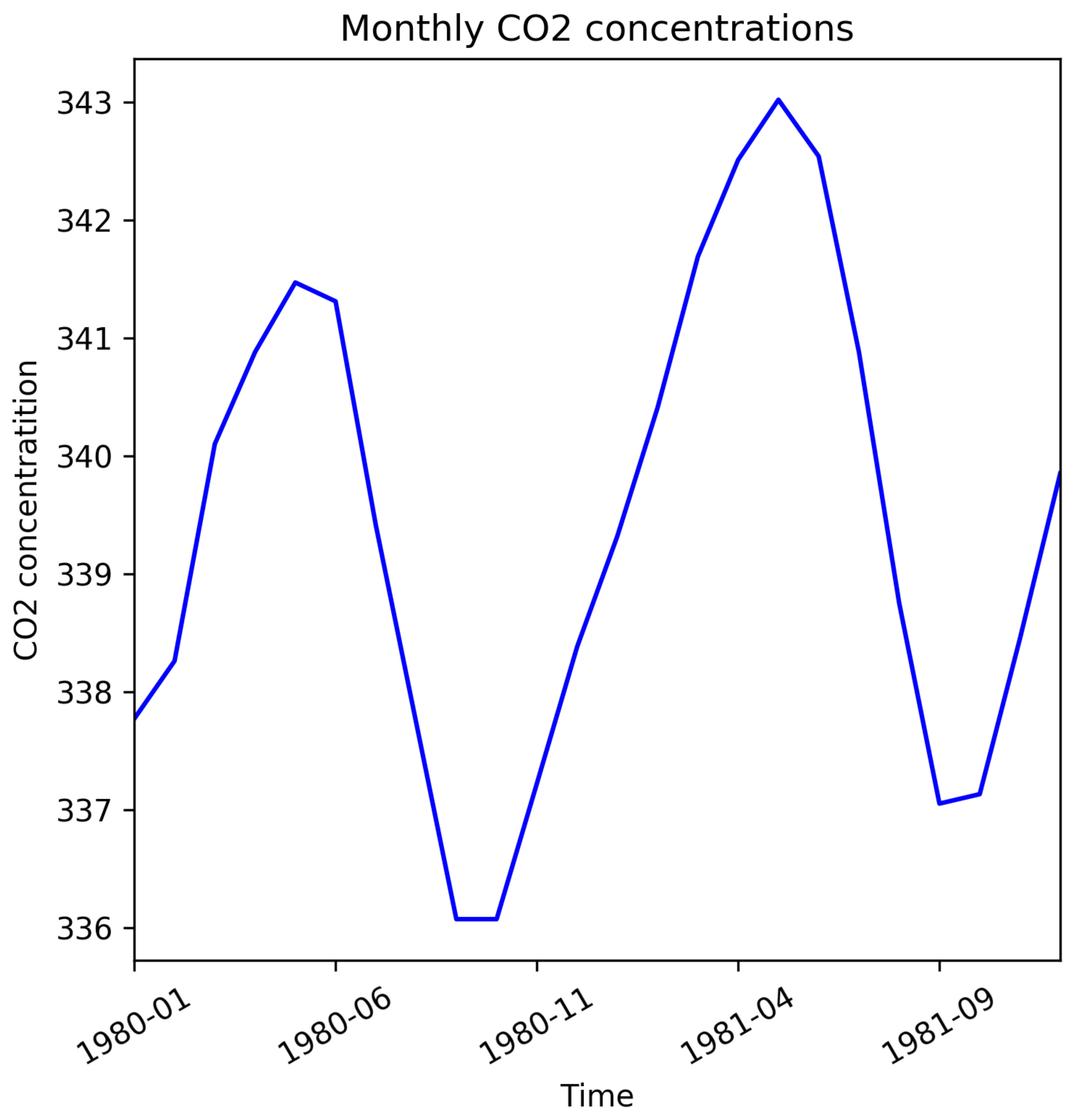


Fuente: *Practical Time Series Analysis*, Pal & Prakash

Tendencia

Introducción

- Sin embargo, **es posible que la tendencia general no sea evidente a corto plazo de la serie.** Los efectos a corto plazo, como las fluctuaciones estacionales y las variaciones irregulares, hacen que la serie de tiempo vuelva a visitar los valores más bajos o más altos observados en el pasado y, por lo tanto, pueden ofuscar temporalmente cualquier tendencia general.



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Tendencia

Introducción

- La **tendencia general en la serie de tiempo se debe a cambios fundamentales o cambios sistémicos del proceso o sistema que representa.** Por ejemplo, el movimiento ascendente de las concentraciones de CO2 durante 1974 a 1987 puede atribuirse al aumento gradual de los automóviles y la industrialización durante estos años.

- Una tendencia general **se modela comúnmente estableciendo la serie de tiempo como una regresión contra el tiempo** y otros factores conocidos como variables explicativas. La regresión o línea de tendencia se puede utilizar como una predicción del movimiento a largo plazo de la serie de tiempo.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Estacionalidad

Introducción

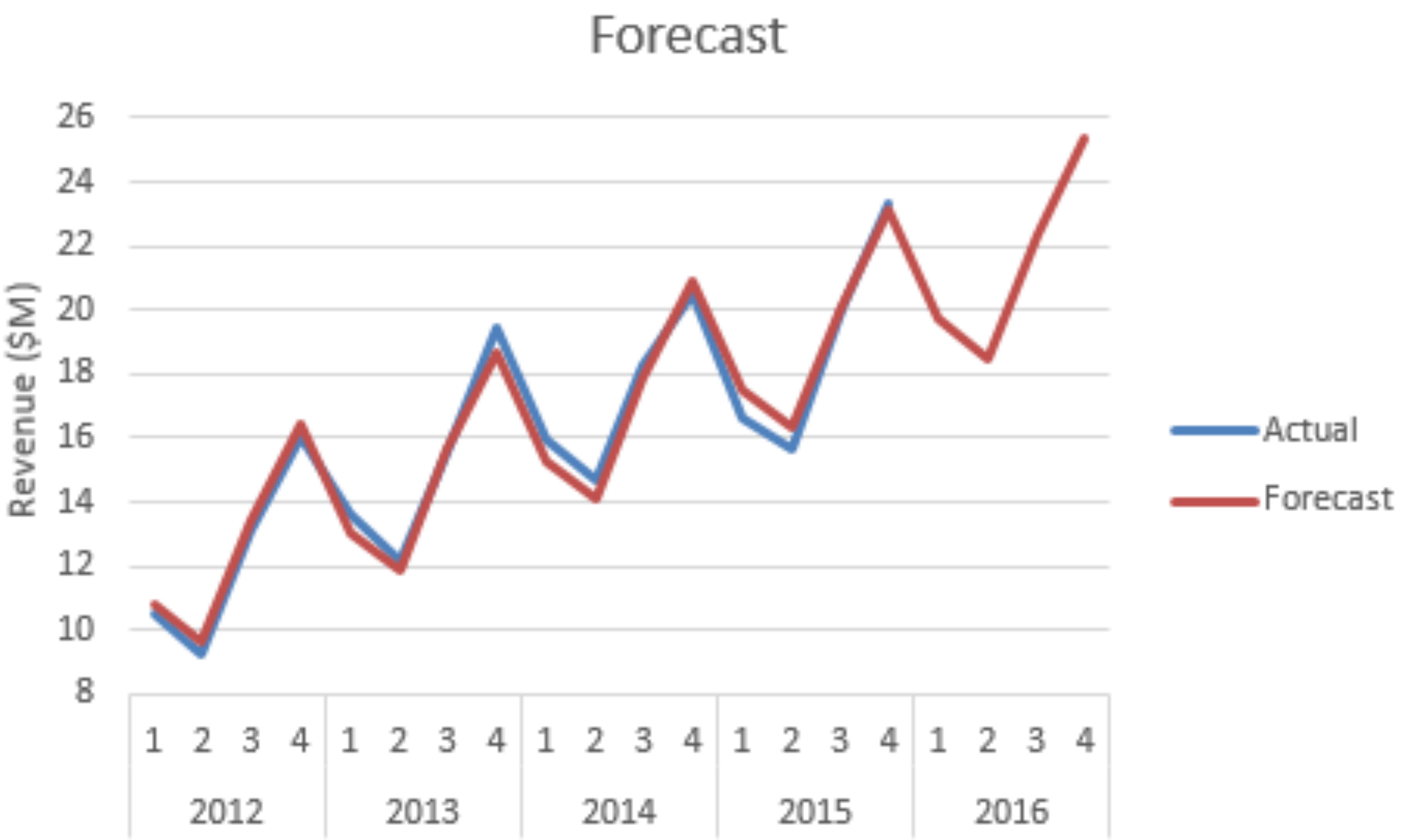
- La **estacionalidad** se manifiesta como **variaciones repetitivas y periódicas** en una serie de tiempo. En la mayoría de los casos, el análisis de datos exploratorios revela la presencia de estacionalidad.

- Una **técnica práctica para determinar la estacionalidad** es mediante el análisis exploratorio de datos a través de los siguientes gráficos:
  - Run sequence plot
  - Seasonal sub series plot
  - Multiple box plots

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Estacionalidad

Introducción

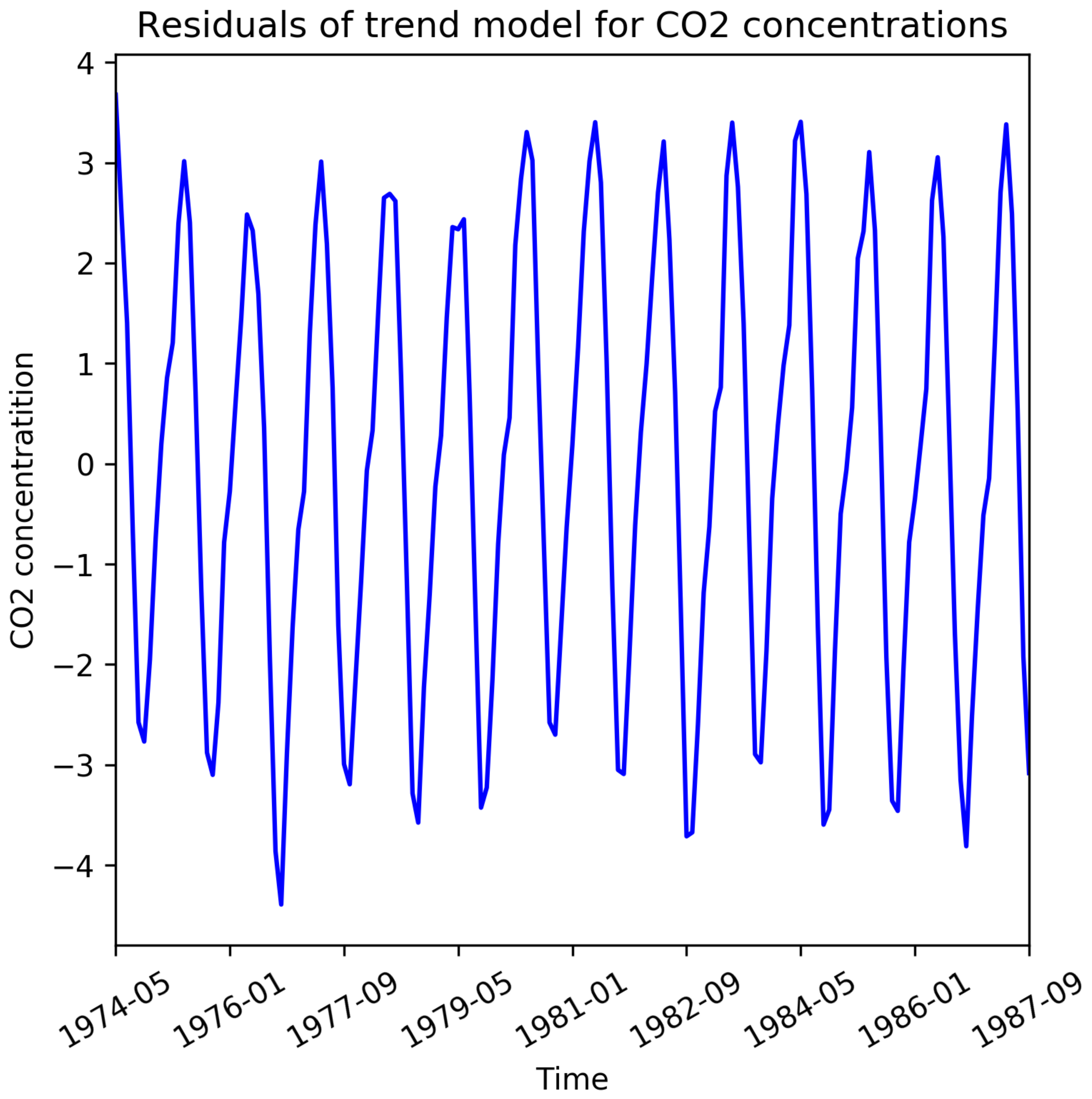


Fuente: *Practical Time Series Analysis*, Pal & Prakash

Sequence plot

Un diagrama de secuencia simple de la serie de tiempo original con el tiempo en el eje *x* y la variable en el eje *y* es bueno para indicar las siguientes propiedades de la serie de tiempo:

- Movimientos en la media de la serie
- Cambios en la varianza
- Presencia de valores atípicos



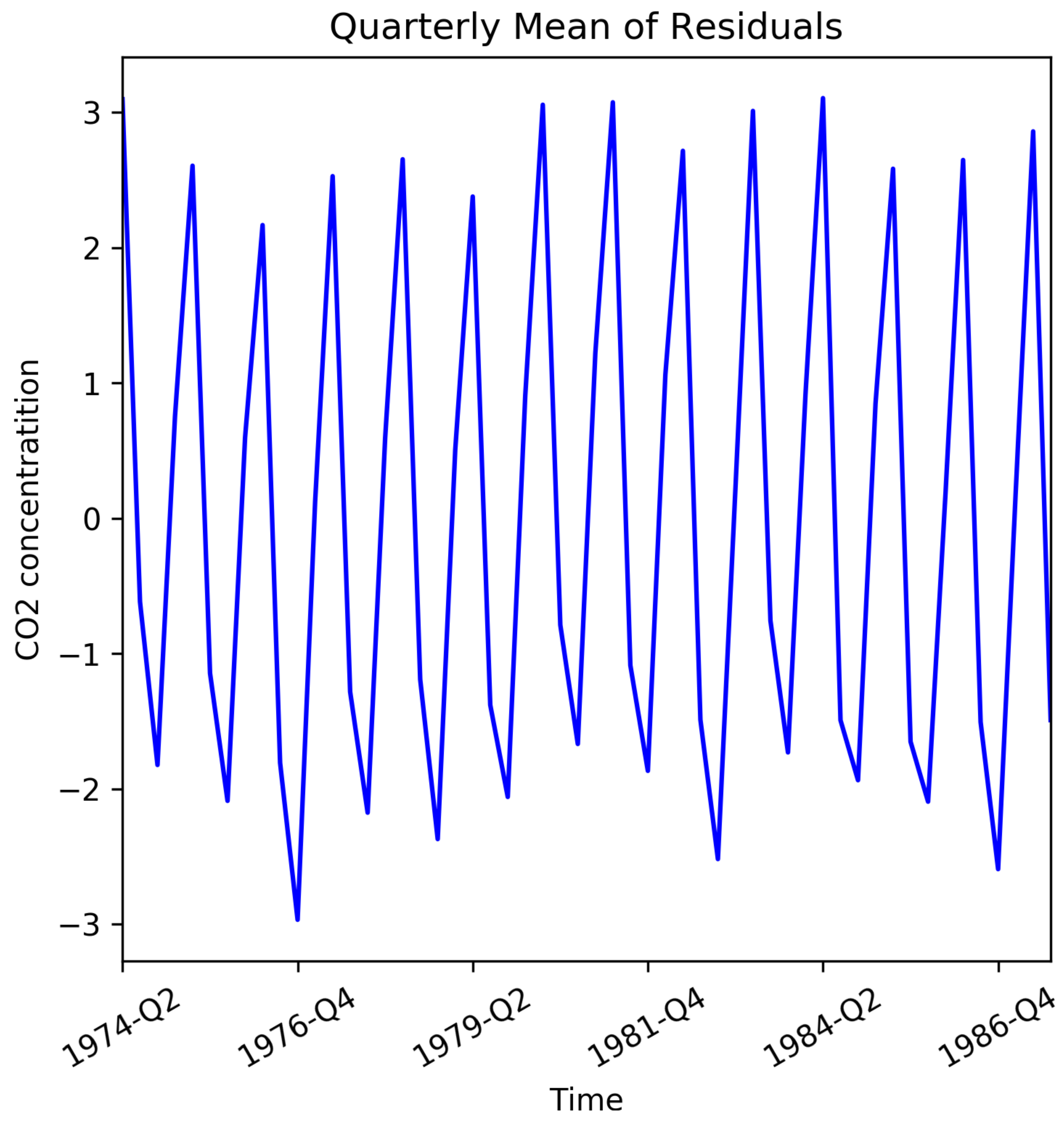
Fuente: *Practical Time Series Analysis*, Pal & Prakash

Estacionalidad

Introducción

Seasonal sub series plot

- Para una periodicidad conocida de variaciones estacionales, la subserie estacional vuelve a dibujar la serie original en lotes de períodos de tiempo sucesivos.
- La gráfica de subserie estacional puede ser más informativa cuando se vuelve a dibujar con diagramas de caja estacionales como se muestra en la siguiente figura.



Fuente: *Practical Time Series Analysis*, Pal & Prakash

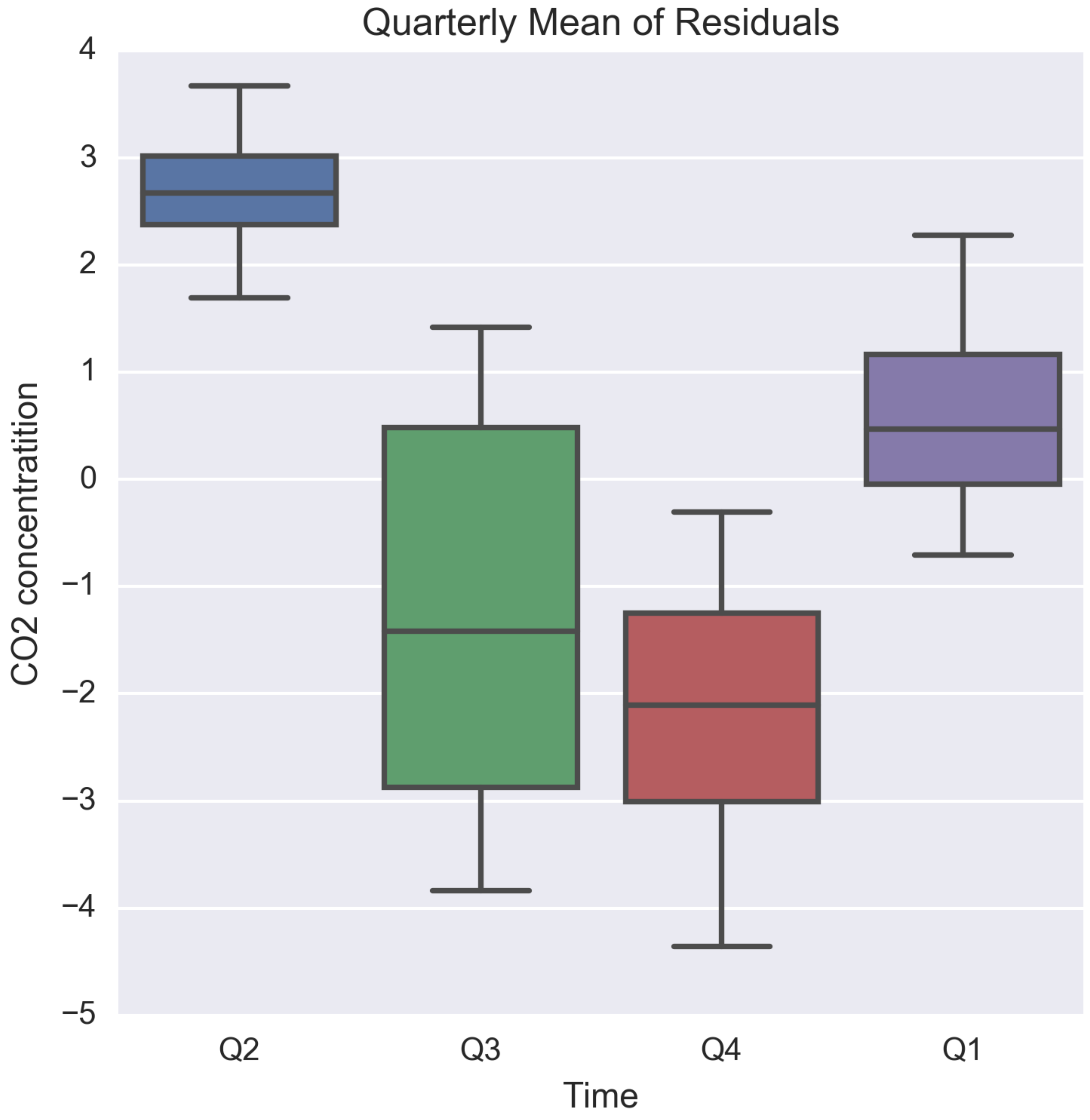
Estacionalidad

Introducción

Multiple box plots

- Un diagrama de caja muestra tanto la **tendencia central** como la **dispersión** dentro de los datos estacionales en un lote de unidades de tiempo.

- Además, la separación entre dos parcelas de caja adyacentes revela las variaciones dentro de la temporada.



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Cambios cíclicos

Introducción

- Los **cambios cíclicos** son **movimientos observados cada pocas unidades de tiempo**, pero ocurren con **menos frecuencia que las fluctuaciones estacionales**.

- A diferencia de la estacionalidad, los cambios cíclicos **pueden no tener un período fijo** de variaciones.

- Además, la **periodicidad promedio de los cambios cíclicos sería mayor** (más comúnmente en años), mientras que las variaciones estacionales se observan dentro del mismo año y corresponden a divisiones de tiempo anuales como estaciones, trimestres y períodos de festividad y feriados, etc.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Cambios cíclicos

Introducción

- Se requiere una **gráfica de largo plazo** de la serie de tiempo para identificar los **cambios cíclicos** que pueden ocurrir, por ejemplo, cada pocos años y **se manifiestan como crestas y valles repetidos**.

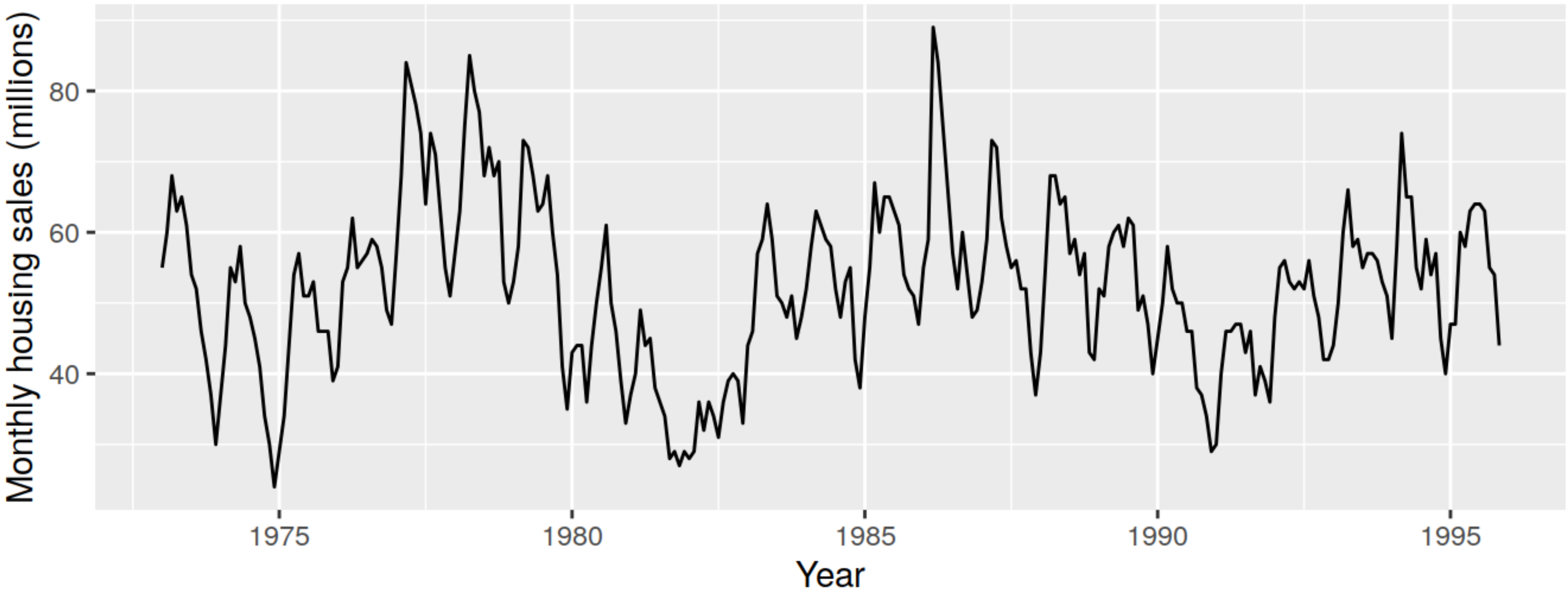
- En este sentido, las series de tiempo relacionadas con la economía y los negocios a menudo muestran cambios cíclicos que corresponden a los **ciclos comerciales y macroeconómicos habituales,** como los períodos de recesión seguidos por los auges, pero están separados por un lapso de tiempo de pocos años.

- De manera similar a las tendencias generales, la identificación de movimientos cíclicos puede requerir datos que se remontan significativamente al pasado.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Cambios cíclicos

Introducción



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Cambios irregulares

Introducción

- Refiriéndonos a nuestro modelo que expresa una serie de tiempo como una suma de cuatro componentes, es de notar que a pesar de poder dar cuenta de los otros tres componentes, todavía podríamos quedarnos con un **componente de error irreductible que es aleatorio y no exhibe dependencia sistemática del índice de tiempo**.

- Este cuarto componente **refleja variaciones inesperadas en la serie temporal**. Las variaciones inesperadas son estocásticas y no se pueden enmarcar en un modelo matemático para una predicción futura definitiva. Este tipo de error se debe a la falta de información sobre las variables explicativas que puedan modelar estas variaciones o debido a la presencia de un ruido aleatorio.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

MODELOS

Introducción

Modelos

- El propósito del análisis de series de tiempo es desarrollar un **modelo matemático** que pueda **explicar el comportamiento observado** de una serie de tiempo y posiblemente **pronosticar** el estado futuro de la serie.

- El modelo elegido debe poder dar cuenta de una o más de las **estructuras internas** que podrían estar presentes. Los siguientes modelos generales se utilizan a menudo como **bloques de construcción del análisis** de series de tiempo:
    - Zero mean models
    - Random walk
    - Trend models
    - Seasonality models

Fuente: *Practical Time Series Analysis,* Pal & Prakash

Zero Mean Models

Modelos

UNIVERSIDAD DE MONTEVIDEO

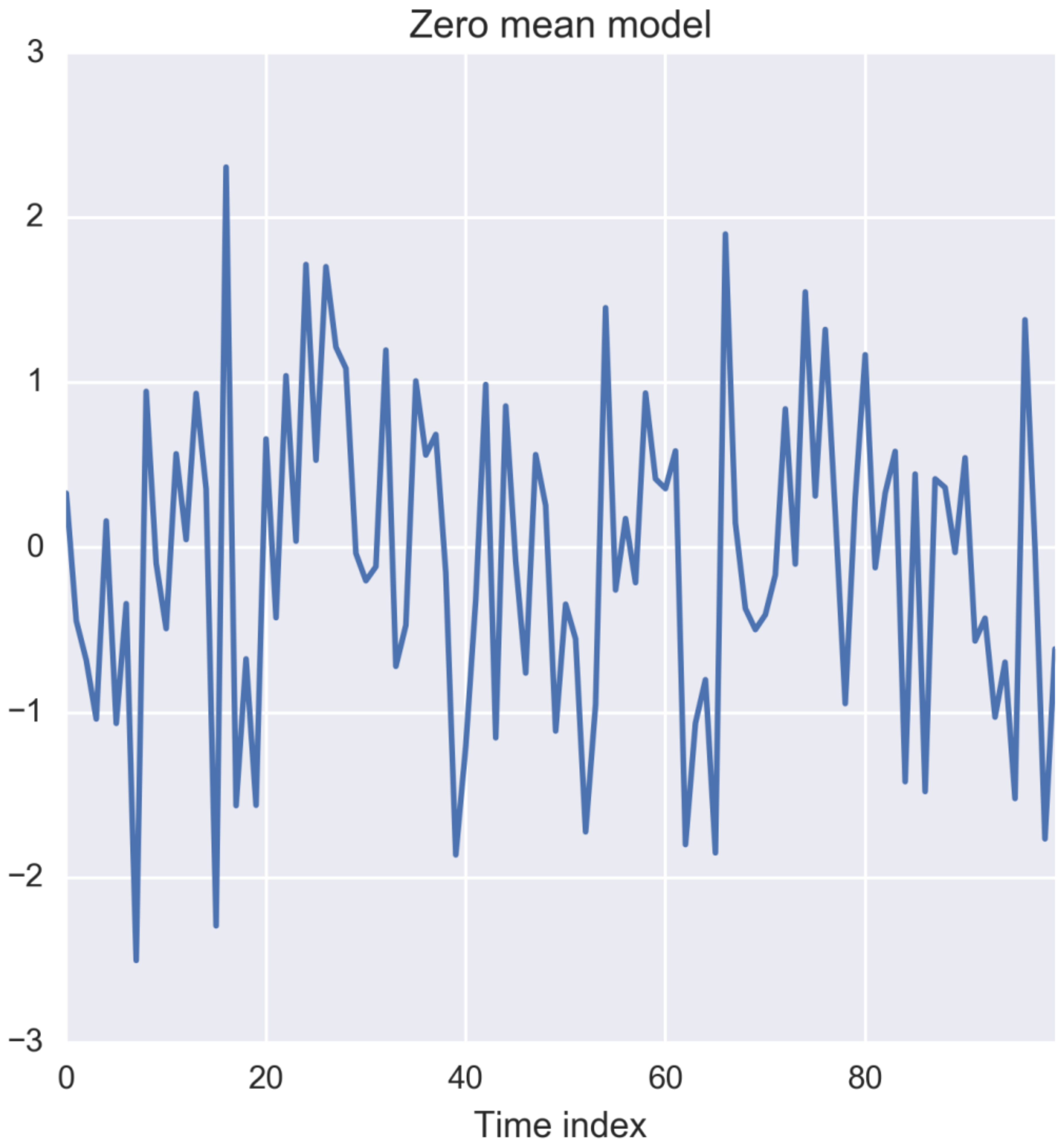
- Los **modelos de media cero** tienen una **media constante y una varianza constante** y **no muestran tendencias o estacionalidad predecibles**.

- Se supone que las **observaciones** de un modelo de media cero son **independientes e idénticamente distribuidas (iid)** y **representan el ruido aleatorio alrededor de una media fija**, que se ha deducido de la serie de tiempo como un término constante.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Zero Mean Models

Modelos



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Zero Mean Models

Modelos

¿Por qué son importantes?

- **Previsibilidad:** si la serie temporal es ruido blanco, entonces, por definición, es aleatoria. No se puede modelar razonablemente y hacer predicciones.
- **Diagnóstico del modelo:** la serie de errores de un modelo de pronóstico de series de tiempo idealmente debería ser ruido blanco.

Se espera que los datos de series de tiempo contengan algún componente de ruido blanco además de la señal generada por el proceso subyacente.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Zero Mean Models

Modelos

¿Por qué son importantes?

- Una vez que se han realizado las predicciones mediante un modelo de predicción de series de tiempo, se pueden recopilar y analizar. Idealmente, **la serie de errores de pronóstico debería ser ruido blanco.**

- Cuando los errores de pronóstico son ruido blanco, **significa que el modelo ha aprovechado toda la información de la señal de la serie temporal para realizar predicciones**. Todo lo que queda son las fluctuaciones aleatorias que no se pueden modelar.

- Una señal de que las predicciones del modelo **no son ruido blanco** es una indicación de que **es posible que se realicen más mejoras** en el modelo de predicción.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Random Walk

Modelos

- Un random walk es otro modelo de series de tiempo donde la observación actual es igual a la observación anterior con un paso aleatorio hacia arriba o hacia abajo.

> **Random Walk**
>
> A *random walk* is a time series model x_t such that $x_t = x_{t-1} + w_t$, where w_t is a discrete white noise series.

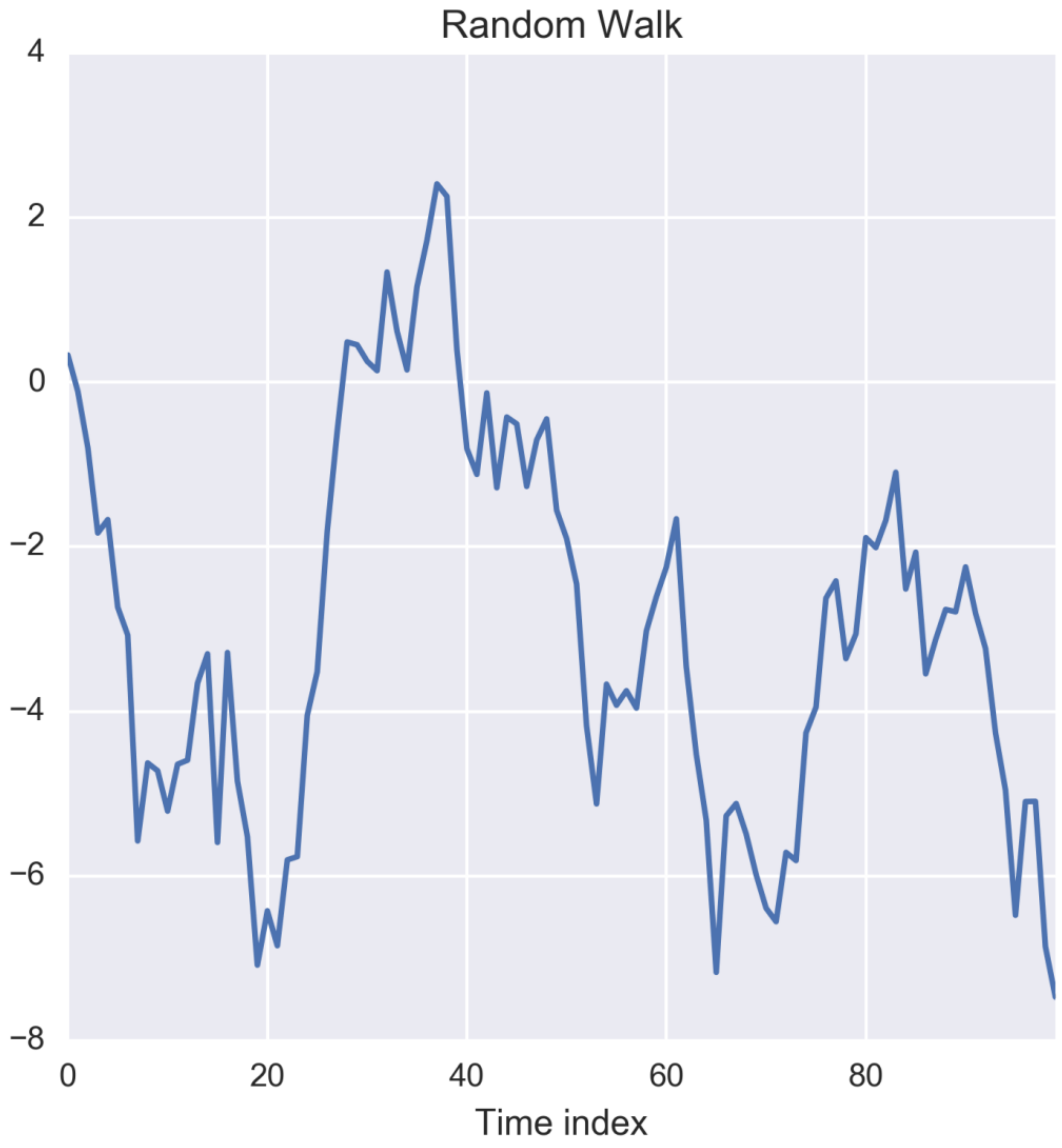
- Imagínese parado en el punto 0 y tiene dos direcciones posibles para el movimiento: hacia adelante o hacia atrás. Lanza una moneda para elegir la dirección, la cara hacia adelante, la cruz hacia atrás. Después de cada lanzamiento de la moneda, das un solo paso.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Random Walk

Modelos

- El random walk es importante porque si tal comportamiento se encuentra en una serie de tiempo, **se puede reducir fácilmente a un modelo de media cero tomando diferencias de las observaciones de dos índices de tiempo consecutivos** como: St - St-1 = xt que es un iid con media cero y varianza constante.



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Trend Models

Modelos

- Este tipo de modelo tiene como **objetivo capturar la tendencia a largo plazo** en la serie de tiempo que se puede ajustar como regresión lineal del índice de tiempo.

- Cuando la serie de tiempo no muestra fluctuaciones periódicas o estacionales, se puede expresar simplemente como la **suma de la tendencia y el modelo de media cero** como **xt = μ(t) + yt**, donde μ(t) es la tendencia de la serie a largo plazo dependiente del tiempo.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Trend Models

Modelos

UNIVERSIDAD DE MONTEVIDEO

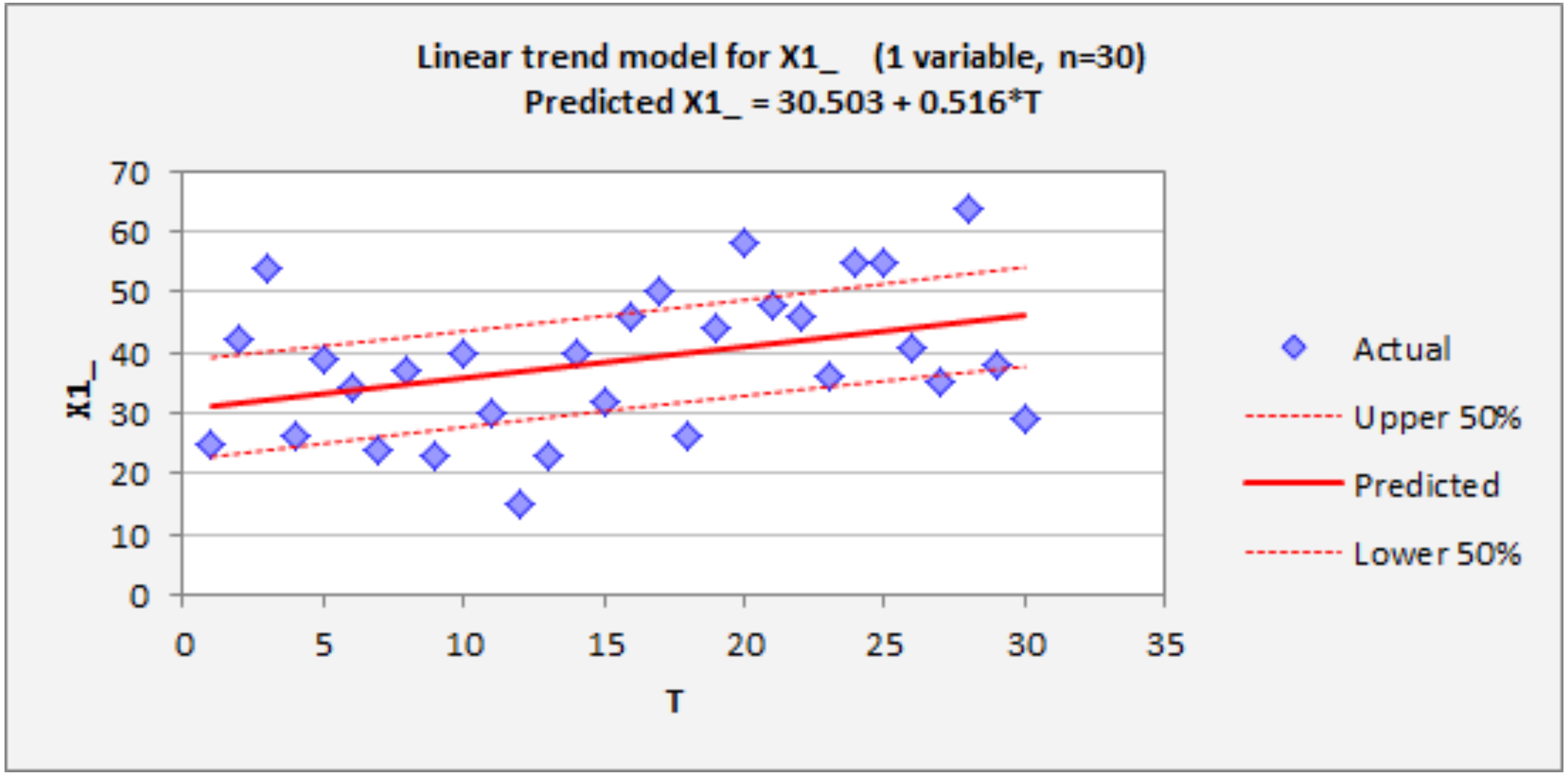
- **La elección del modelo de tendencia μ(t) es fundamental para capturar correctamente el comportamiento de la serie temporal.** El análisis de datos exploratorios a menudo proporciona pistas para formular hipótesis sobre si el modelo debería ser **lineal o no lineal en t**. Un modelo lineal es simplemente $\mu(t) = wt + b$, mientras que el modelo cuadrático es $\mu(t) = w1t + w2t2 + b$. A veces, la tendencia puede formularse como hipótesis mediante una relación más compleja en términos del índice de tiempo, como $\mu(t) = w0tp + b$.

- Los **residuos** $xt - \mu(t)$ del modelo de tendencia se consideran para el **ruido irreducible** y como realización del componente media cero yt.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Trend Models

Modelos



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Seasonality Models

Modelos

- La **estacionalidad** se manifiesta como **fluctuaciones periódicas y repetitivas** en una serie de tiempo y, por lo tanto, se modela como la suma de la suma ponderada de ondas sinusoidales de periodicidad conocida.

- Suponiendo que la tendencia a largo plazo ha sido eliminada por una línea de tendencia, el modelo de estacionalidad se puede expresar como xt = st + yt, donde la variación estacional con periodicidad conocida es α.

$$s_t \sum \frac{L}{k=1} (w_k cos\alpha t + v_k sin\alpha t) + b$$

Los modelos de estacionalidad también se conocen como **modelos de regresión armónica**, ya que intentan ajustar la suma de múltiples ondas sinusoidales.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Seasonality Models

Modelos

- La **estacionalidad** se manifiesta como **fluctuaciones periódicas y repetitivas** en una serie de tiempo y, por lo tanto, se modela como la suma de la suma ponderada de ondas sinusoidales de periodicidad conocida.

- Suponiendo que la tendencia a largo plazo ha sido eliminada por una línea de tendencia, el modelo de estacionalidad se puede expresar como xt = st + yt, donde la variación estacional con periodicidad conocida es α.

$$s_t \sum \frac{L}{k=1}(w_k cos\alpha t + v_k sin\alpha t) + b$$

Los modelos de estacionalidad también se conocen como **modelos de regresión armónica**, ya que intentan ajustar la suma de múltiples ondas sinusoidales.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Resumen

Modelos

- Los cuatro modelos descritos aquí son **bloques de construcción** de un modelo de serie de tiempo completo.
- Un **modelo de suma cero** representa un **error irreductible** del sistema y los **otros tres modelos** tienen como **objetivo transformar una serie de tiempo dada en modelos de suma cero** a través de transformaciones matemáticas adecuadas.
- Para obtener pronósticos en términos de la serie de tiempo original, se aplican transformaciones inversas relevantes.

Fuente: *Practical Time Series Analysis,* Pal & Prakash

Steps

Modelos

UNIVERSIDAD DE MONTEVIDEO

1. **Visualice los datos** en diferentes granularidades del índice de tiempo para revelar tendencias a largo plazo y fluctuaciones estacionales.
2. Ajustar la **línea de tendencia**, capturar las tendencias a largo plazo y trazar los residuos para verificar la estacionalidad o el error irreducible
3. Ajustar un modelo de regresión armónica para capturar la **estacionalidad**
4. Grafique los residuos que deja el modelo de estacionalidad para comprobar si hay **errores irreductibles.**

Estos pasos suelen ser suficientes para desarrollar modelos matemáticos para la mayoría de las series de tiempo. Los modelos individuales de tendencia y estacionalidad pueden ser simples o complejos según la serie de tiempo original y la aplicación.

Fuente: *Practical Time Series Analysis,* Pal & Prakash

ESTACIONARIEDAD

Introducción

Estacionariedad

- Una **serie de tiempo estacionaria** es aquella cuyas **propiedades estadísticas** como la media, la varianza y la autocorrelación **son todas constantes en el tiempo**.

- No significa que la serie no cambie con el tiempo, solo que **la forma en que cambia no cambia en sí misma con el tiempo.**

- La mayoría de los métodos de pronóstico estadístico se basan en el supuesto de que la serie de tiempo puede volverse aproximadamente estacionaria (es decir, "estacionaria") mediante el uso de transformaciones matemáticas.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Importancia

Estacionariedad

- Otra razón para tratar de **estacionarizar una serie de tiempo** es poder **obtener estadísticas muestrales significativas**, como medias, varianzas y correlaciones con otras variables.

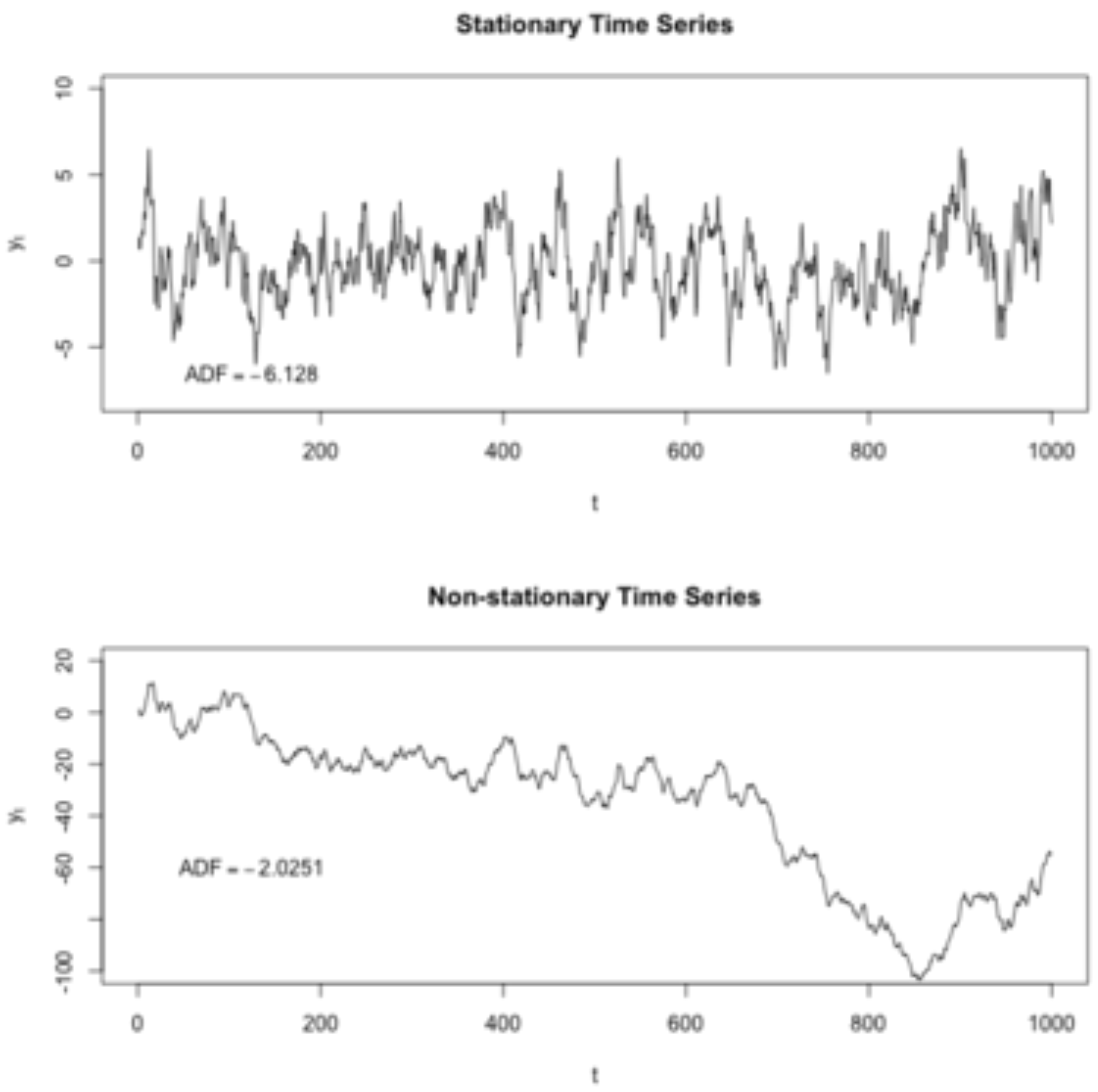
- Tales **estadísticas son útiles** como descriptores del comportamiento futuro **solo si la serie es estacionaria.**

- Por ejemplo, si la serie aumenta constantemente con el tiempo, la media y la varianza de la muestra crecerán con el tamaño de la muestra y siempre subestimarán la media y la varianza en períodos futuros. Y si la media y la varianza de una serie no están bien definidas, tampoco lo están sus correlaciones con otras variables. Por esta razón, **debe tener cuidado al intentar extrapolar modelos de regresión ajustados a datos no estacionarios.**

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Ejemplo

Modelos



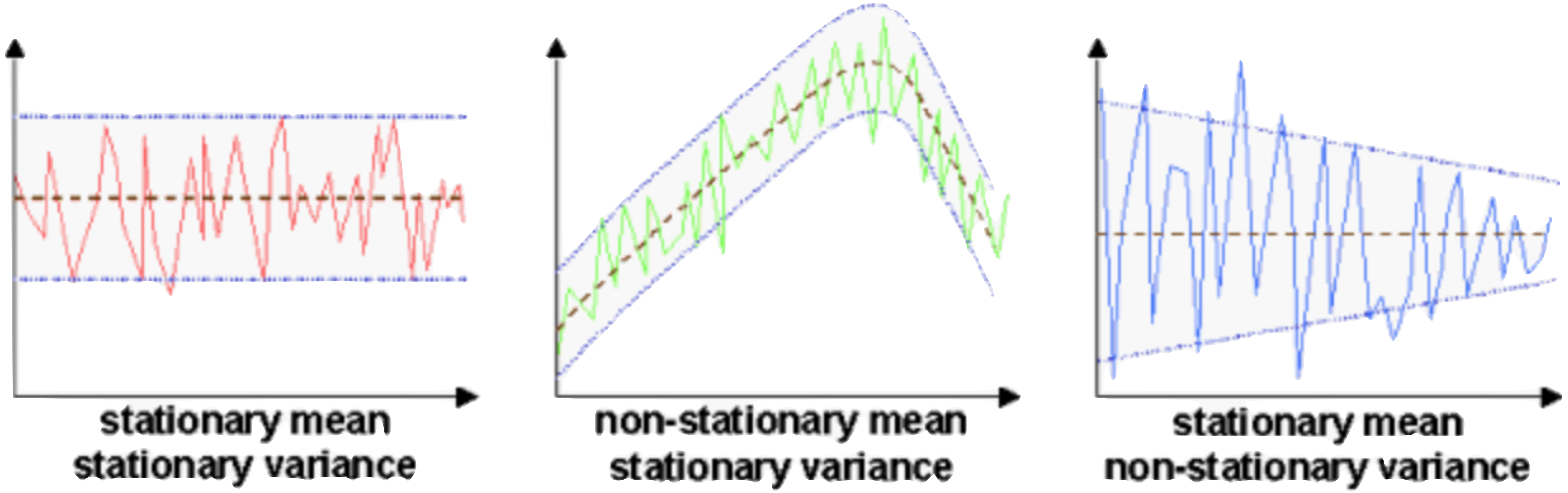
Fuente: *Practical Time Series Analysis*, Pal & Prakash

Test

Modelos

Visualizaciones

- Los métodos más básicos para la detección de estacionariedad se basan en graficar los datos, o sus funciones, y determinar visualmente si presentan alguna propiedad conocida de los datos estacionarios (o no estacionarios).

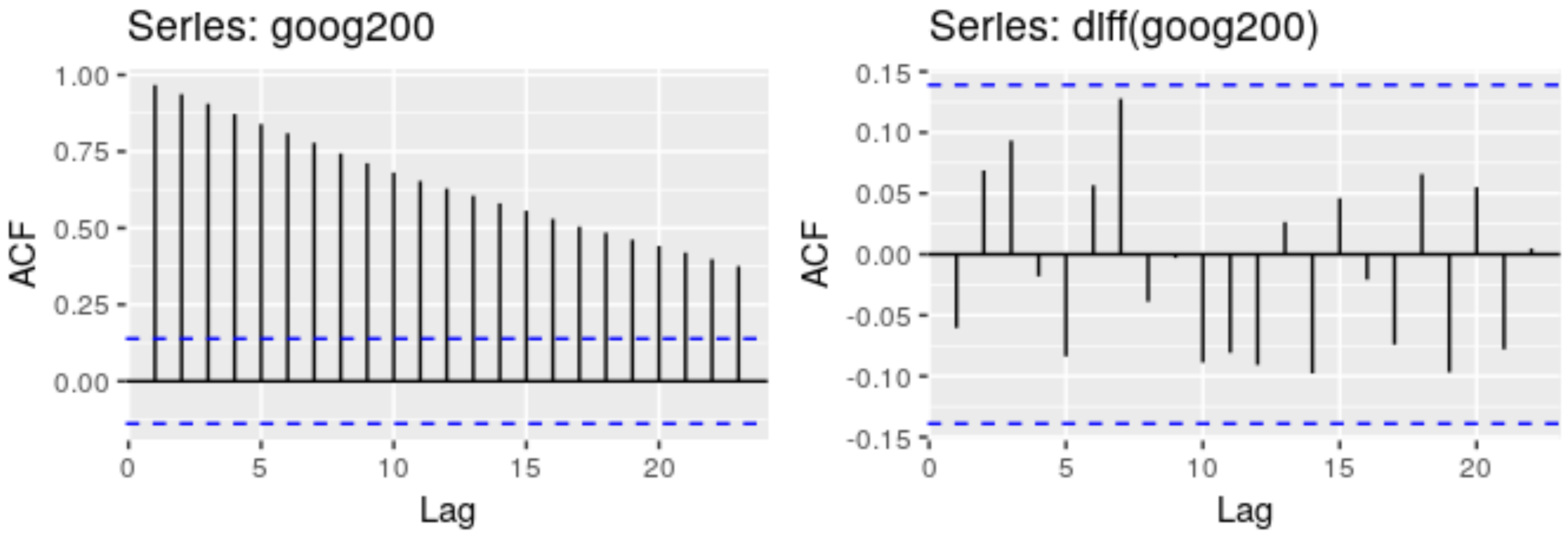


Fuente: *Practical Time Series Analysis*, Pal & Prakash

UNIVERSIDAD DE MONTEVIDEO

Mirando los gráficos de la función de autocorrelación (ACF)

- La autocorrelación es la correlación de una señal con una copia retrasada de sí misma en función del retraso. Al trazar el valor del ACF para retrasos crecientes (un diagrama llamado correlograma), **los valores tienden a degradarse a cero rápidamente para series de tiempo estacionarias** (ver figura 1, derecha), mientras que para datos no estacionarios la degradación ocurrirá más lentamente (ver figura 1, izquierda).



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Test

Modelos

UNIVERSIDAD DE MONTEVIDEO

Pruebas paramétricas

- La prueba **Augmented Dickey-Fuller** es un tipo de prueba estadística denominada prueba de raíz unitaria.

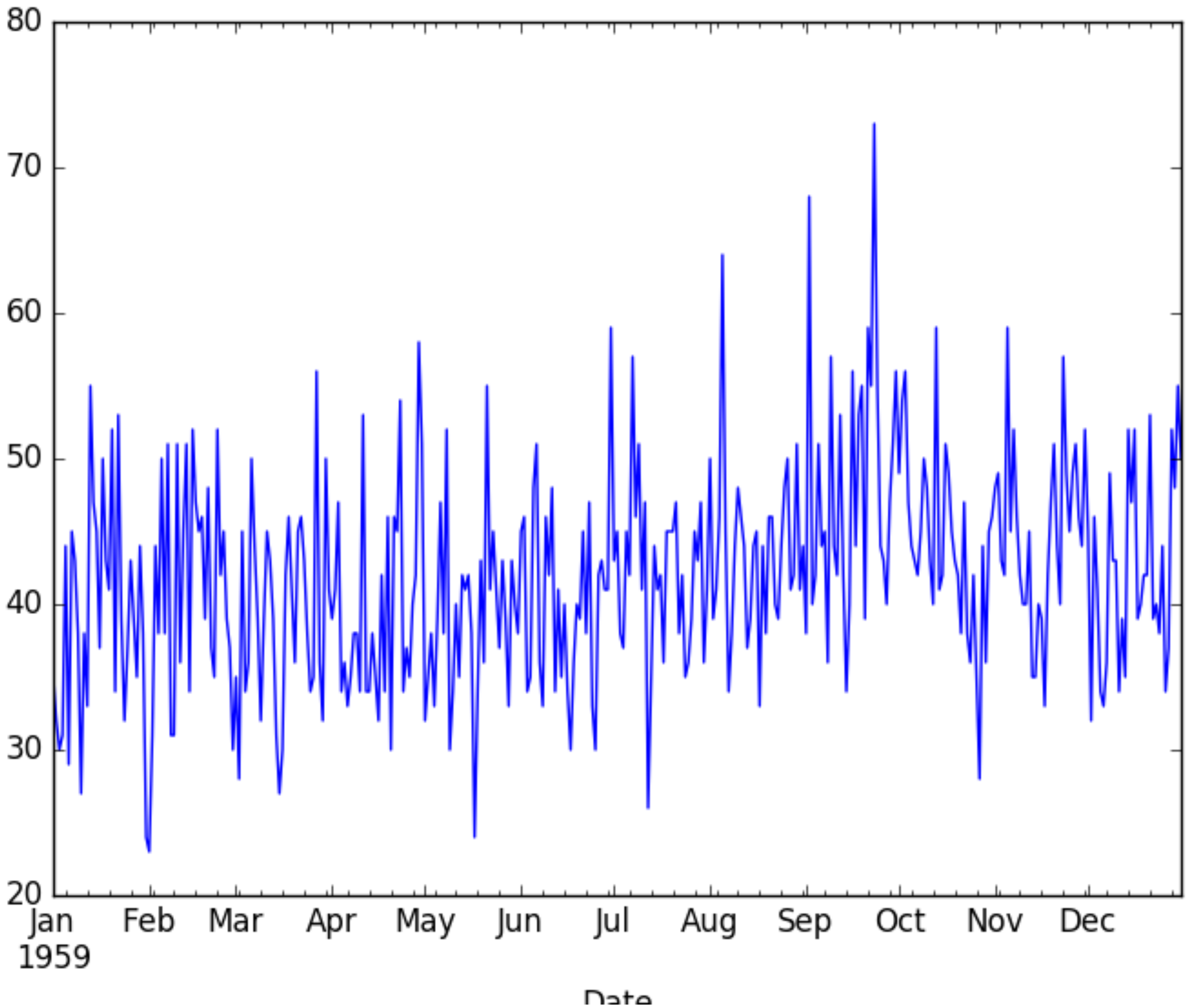
- La intuición detrás de una prueba de raíz unitaria es que **determina la fuerza con la que una serie de tiempo está definida por una tendencia**.

- La **hipótesis nula** de la prueba es que la serie de tiempo se puede representar mediante una raíz unitaria, **que no es estacionaria** (tiene alguna estructura dependiente del tiempo). La hipótesis alternativa (rechazando la hipótesis nula) es que la serie de tiempo es estacionaria.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Test

Modelos



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Test

Modelos

UNIVERSIDAD DE MONTEVIDEO

- **Rechazar la hipótesis nula** significa que el proceso no tiene raíz unitaria y, a su vez, **la serie de tiempo es estacionaria** o no tiene una estructura dependiente del tiempo.

```
from pandas import read_csv
from statsmodels.tsa.stattools import adfuller
series = read_csv('daily-total-female-births.csv', header=0, index_col=0, squeeze=
X = series.values
result = adfuller(X)
print('ADF Statistic: %f' % result[0])
print('p-value: %f' % result[1])
print('Critical Values:')
for key, value in result[4].items():
    print('\t%s: %.3f' % (key, value))
```

```
ADF Statistic: -4.808291
p-value: 0.000052
Critical Values:
    5%: -2.870
    1%: -3.449
    10%: -2.571
```


Fuente: *Practical Time Series Analysis*, Pal & Prakash

Order of Integration I(d)

Modelos

- Si tiene raíces unitarias en su serie de tiempo, una serie de diferencias sucesivas, d, puede transformar la serie de tiempo en una con estacionariedad.

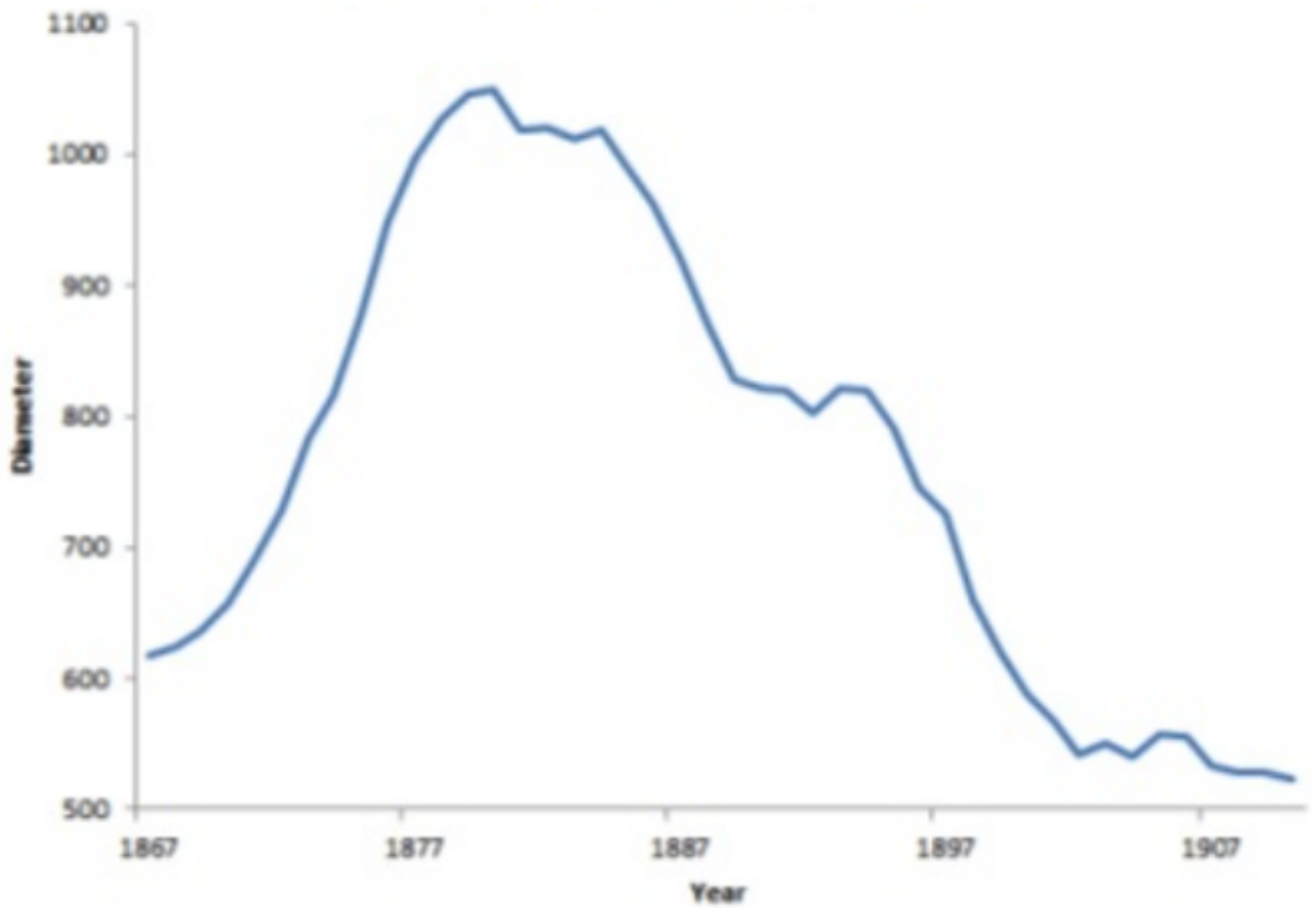
- Las diferencias se indican mediante I(d), donde d es el orden de integración.

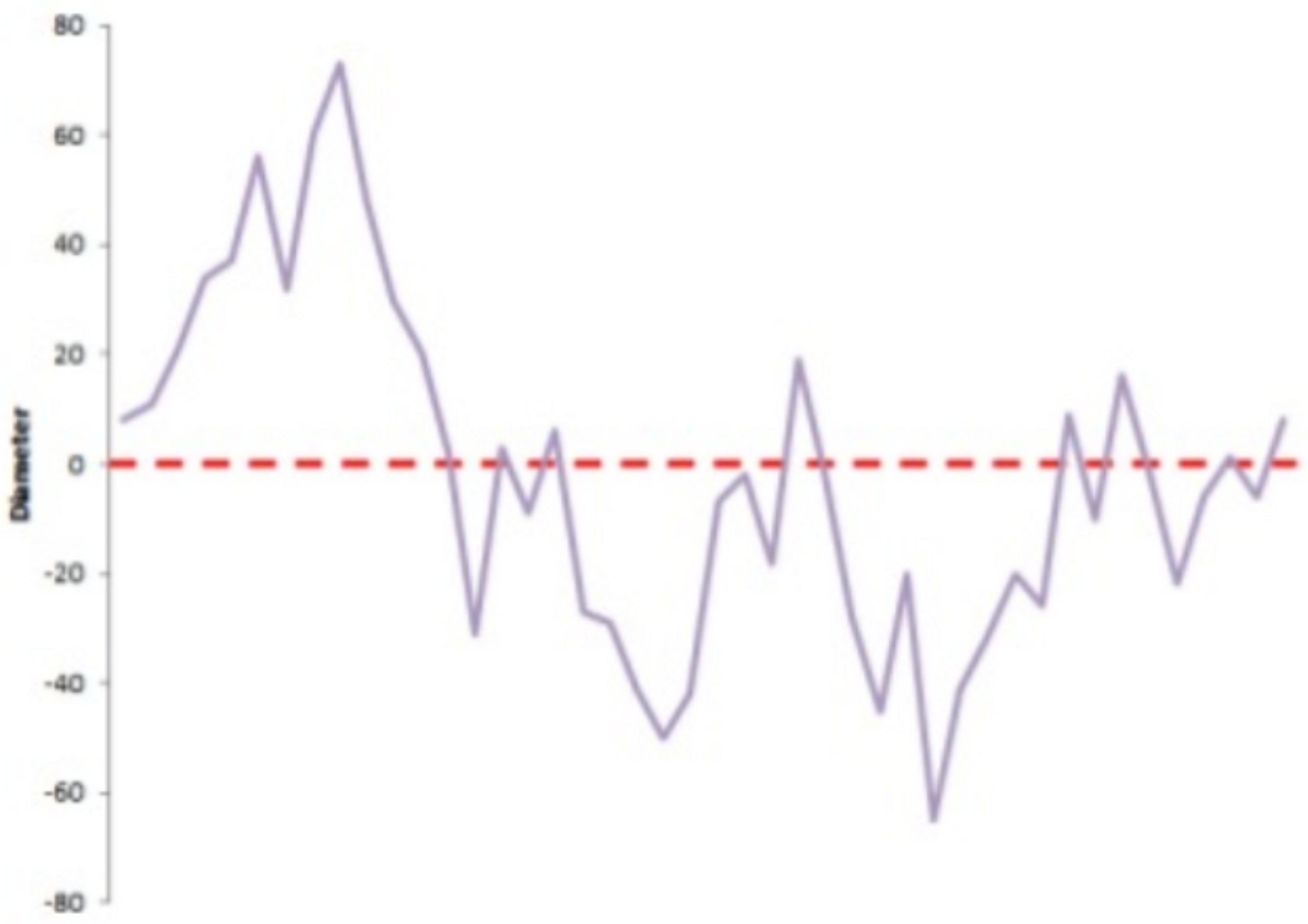
- Las series de tiempo no estacionarias que pueden transformarse de esta manera se denominan series integradas de orden k. Generalmente, el orden de integración es I(0) o I (1); Es raro ver valores de d que sean 2 o más

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Order of Integration I(d)

Modelos





Fuente: *Practical Time Series Analysis*, Pal & Prakash

AUTOCORRELACIÓN

Introducción

Autocorrelación

- Cuando tiene una serie de tiempo y existe un patrón tal que **los valores de la serie se pueden predecir en función de los valores anteriores de la serie**, se dice que la serie de tiempo exhibe **autocorrelación**.

- Esto también se conoce como correlación serial y dependencia serial.

- La existencia de **autocorrelación en los residuos** de un modelo es una **señal** de que el **modelo puede no ser sólido.**

- La autocorrelación se diagnostica mediante un **correlograma** (gráfico ACF) y se puede probar con la prueba de **Durbin-Watson**.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Correlograma/ACF

Autocorrelación

- Un correlograma muestra la correlación de una serie de datos consigo mismo; también se conoce como gráfico de autocorrelación y gráfico ACF.

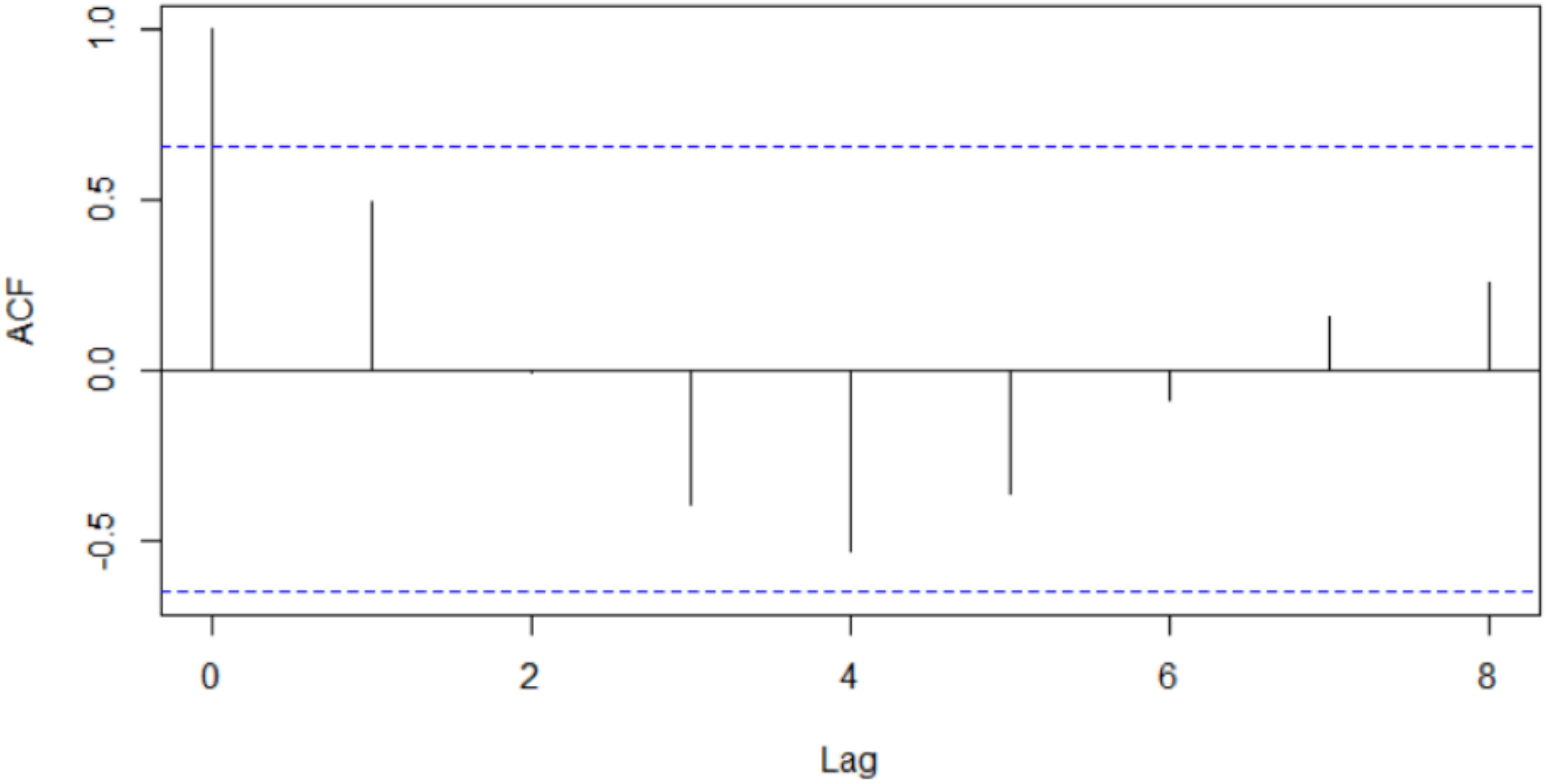
- El *lag* se refiere al orden de correlación. Podemos ver en el gráfico que en el rezago 0, la correlación es 1, ya que los datos están correlacionados consigo mismos.

- Con un rezago de 1, la correlación se muestra alrededor de 0,5. También podemos ver que tenemos correlaciones negativas cuando los puntos están separados por 3, 4 y 5.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Correlograma/ACF

Autocorrelación



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Correlograma/ACF

Autocorrelación

- Una **autocorrelación positiva** indica que los **valores presentes y futuros de la serie temporal se mueven en la misma dirección**, mientras que los valores negativos significan que los valores presentes y futuros se mueven en la dirección opuesta.

- Si la autocorrelación es cercana a cero, las dependencias temporales dentro de la serie pueden ser difíciles de encontrar.

- Debido a esta propiedad, **la autocorrelación es útil para predecir el estado futuro de una serie** de tiempo en h pasos de tiempo por delante.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Autocorrelación Parcial

Autocorrelación

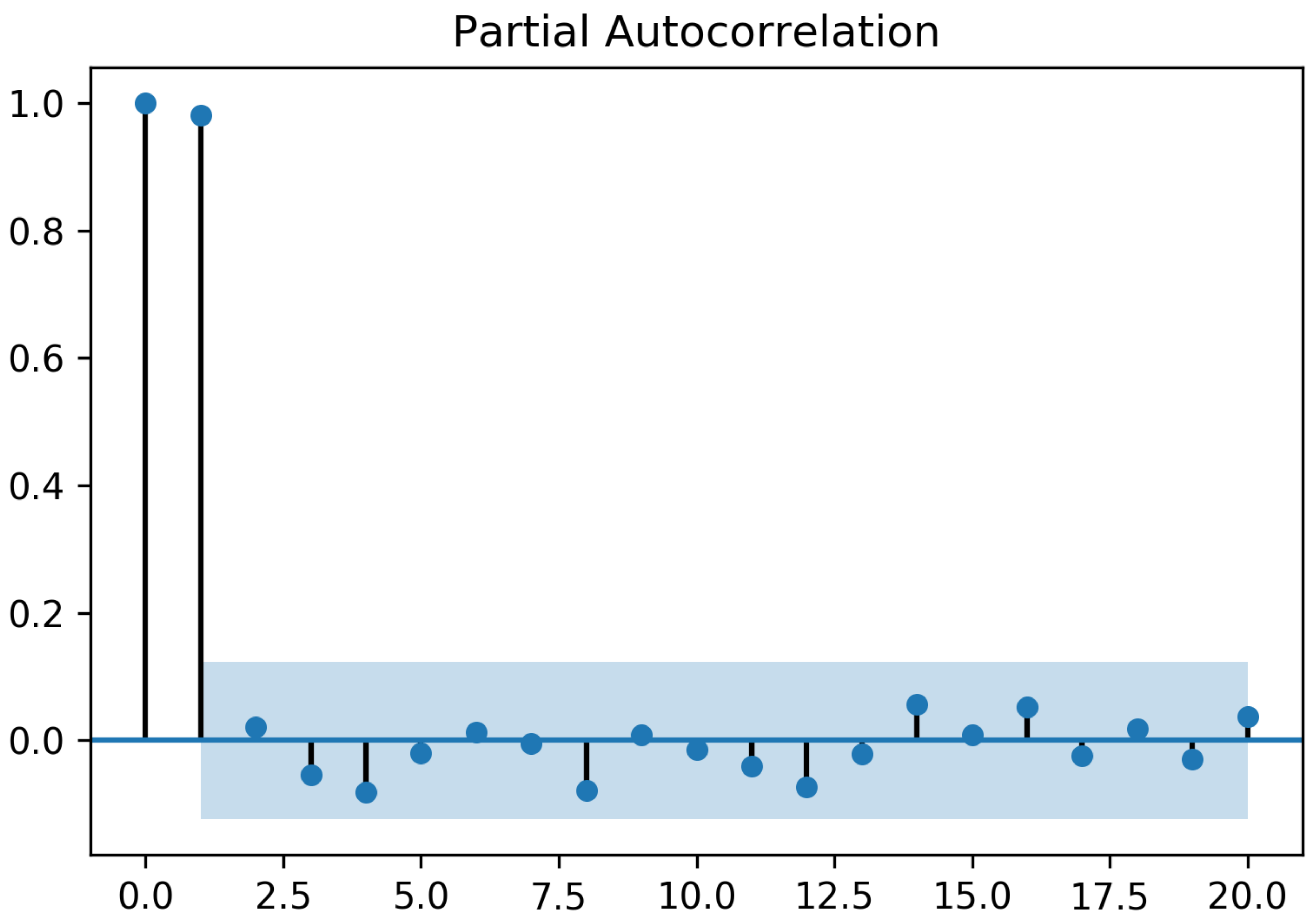
- Las series de tiempo tienden a transportar información y estructuras de dependencia en pasos y, por lo tanto, la autocorrelación en el rezago h también está influenciada por las variables intermedias x_t, x_ (t + 1)… x_ (t + h-1).

- Por tanto, la autocorrelación no es la medida correcta de la correlación mutua entre x_t y x_ (t + h) en presencia de las variables intermedias. Por tanto, sería erróneo elegir h en modelos AR basados en la autocorrelación.

- La autocorrelación parcial resuelve este problema midiendo la correlación entre x_t y x_ (t + h) cuando se ha eliminado la influencia de las variables intermedias. Por tanto, la autocorrelación parcial en el análisis de series de tiempo define la correlación entre x_t y x_ (t + h) que no se explica por los rezagos t + 1 a t + h-1. La autocorrelación parcial ayuda a identificar el orden h de un modelo AR (h).

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Autocorrelación Parcial



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Durbin-Watson test

Autocorrelación

- En estadística, la **autocorrelación de un proceso aleatorio es la correlación de Pearson entre valores del proceso en diferentes momentos**, en función de los dos tiempos o del desfase temporal.

- La **prueba de Durbin-Watson** es una medida de autocorrelación en los residuos del análisis de regresión.

- Cabe señalar que la prueba DW se utiliza para probar un tipo específico de autorregresión (AR1), es decir, una autorregresión de primer orden. La variable de resultado en un proceso de autorregresión de primer orden en algún momento t solo está relacionada con períodos de tiempo que están separados por un período

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Durbin-Watson test

Autocorrelación

UNIVERSIDAD DE MONTEVIDEO

La **prueba de Durbin-Watson** da valores que se encuentran **entre 0 y 4** con el siguiente significado:

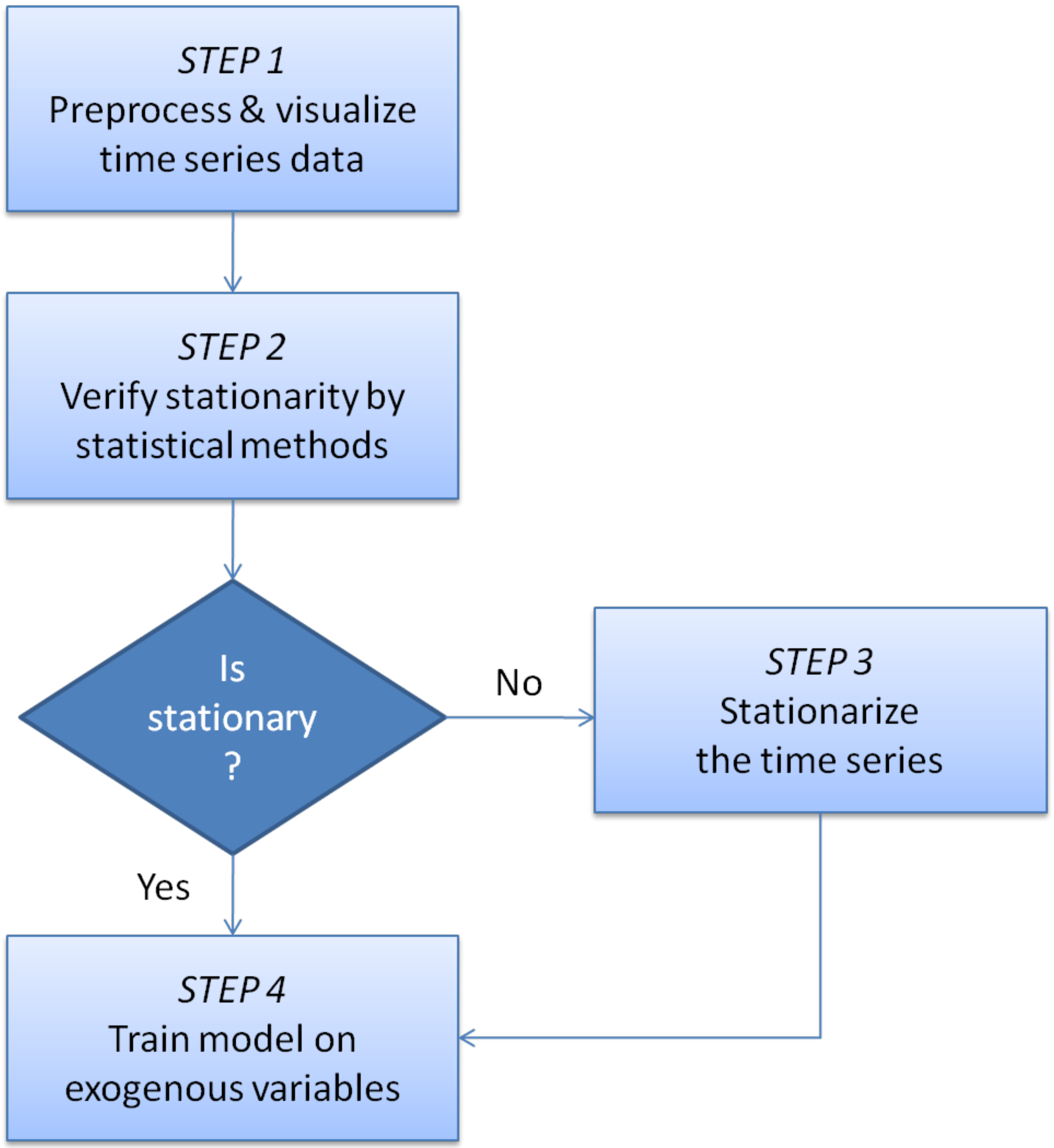
- 2 no es una autocorrelación.
- 0 a <2 es una autocorrelación positiva (común en los datos de series de tiempo).
- 2 a 4 es una autocorrelación negativa (menos común en los datos de series de tiempo).

Fuente: *Practical Time Series Analysis*, Pal & Prakash

STEPS

Series de Tiempo

Steps

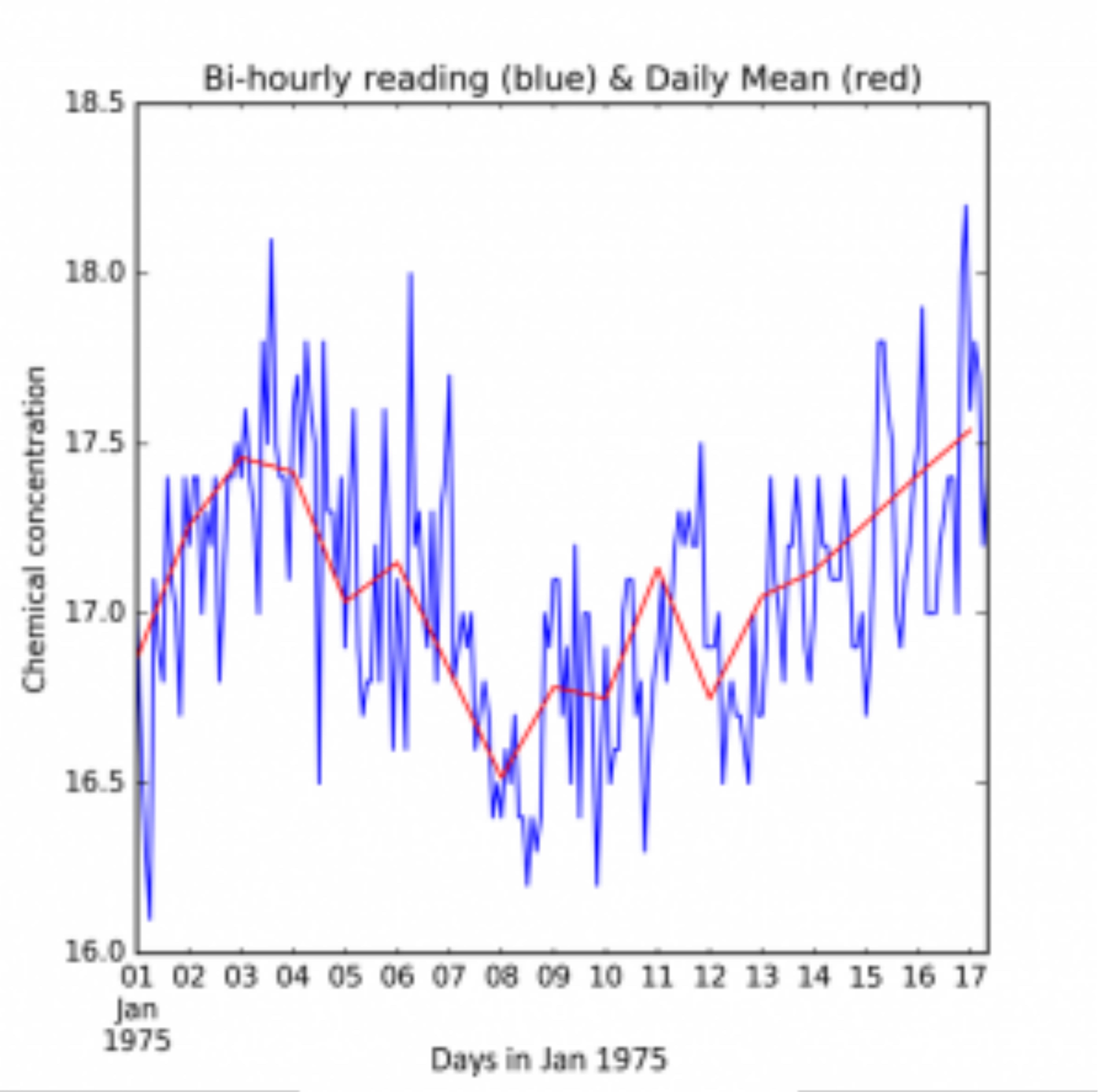


Fuente: *Practical Time Series Analysis*, Pal & Prakash

ADVANCED PROCESSING

Resampling

Advanced Processing



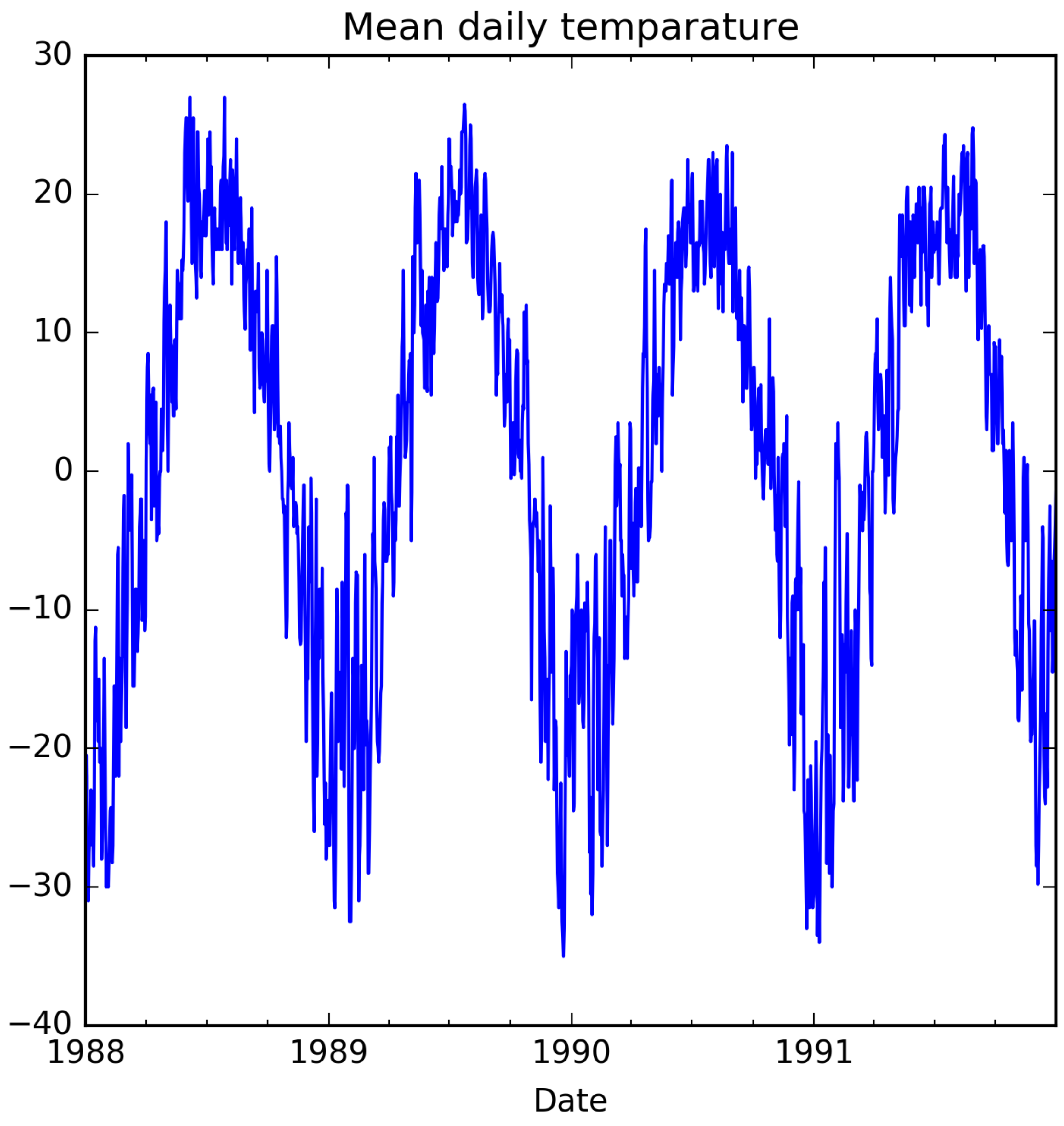
Fuente: *Practical Time Series Analysis*, Pal & Prakash

Group wise aggregation

Advanced Processing

- La serie de tiempo original parece tener **patrones mensuales que se repiten todos los años** y se pueden verificar calculando promedios mensuales.

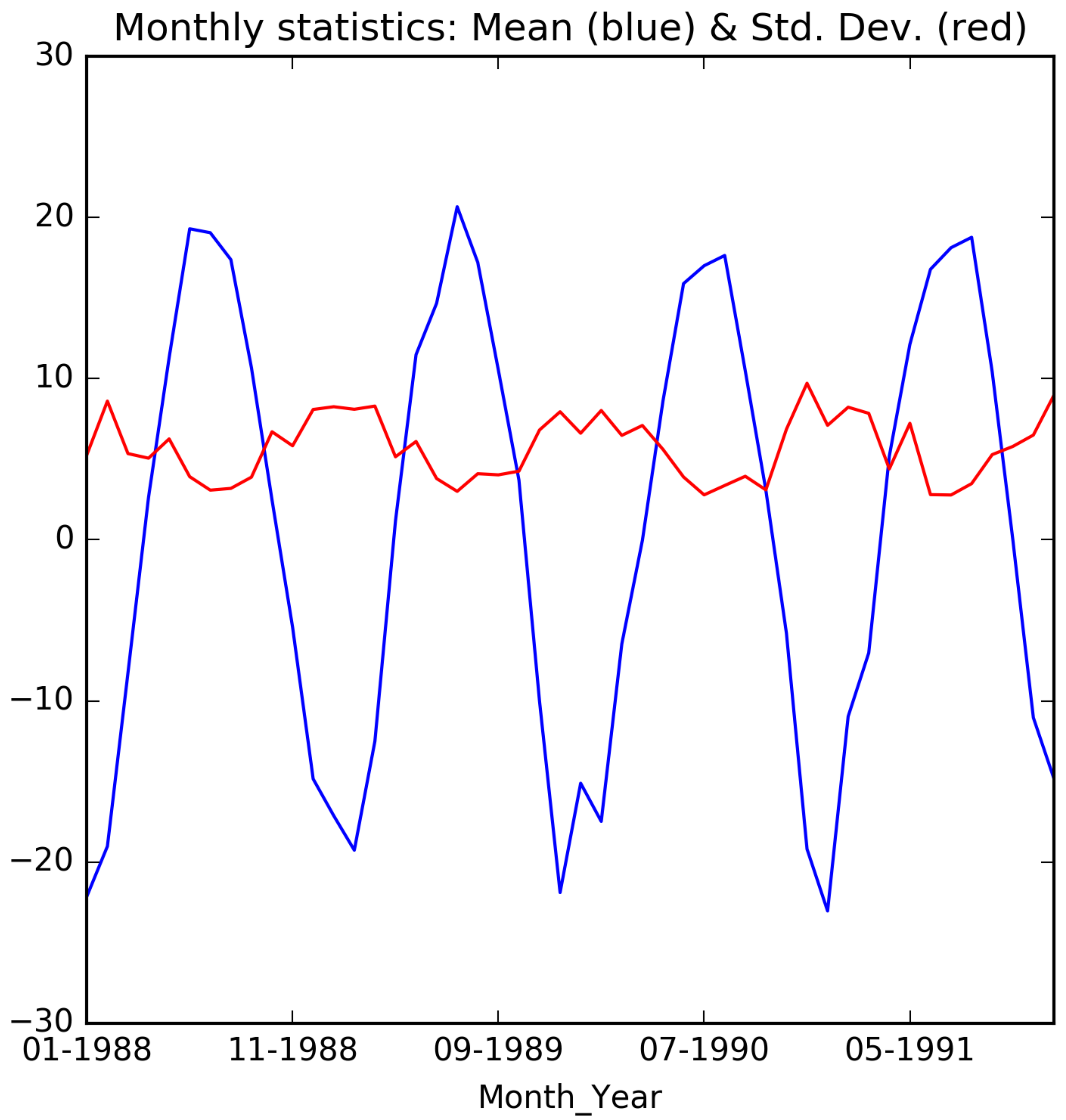
- Esto se hace **agrupando los datos** en 12 meses y luego calculando los promedios de cada mes.



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Group wise aggregation

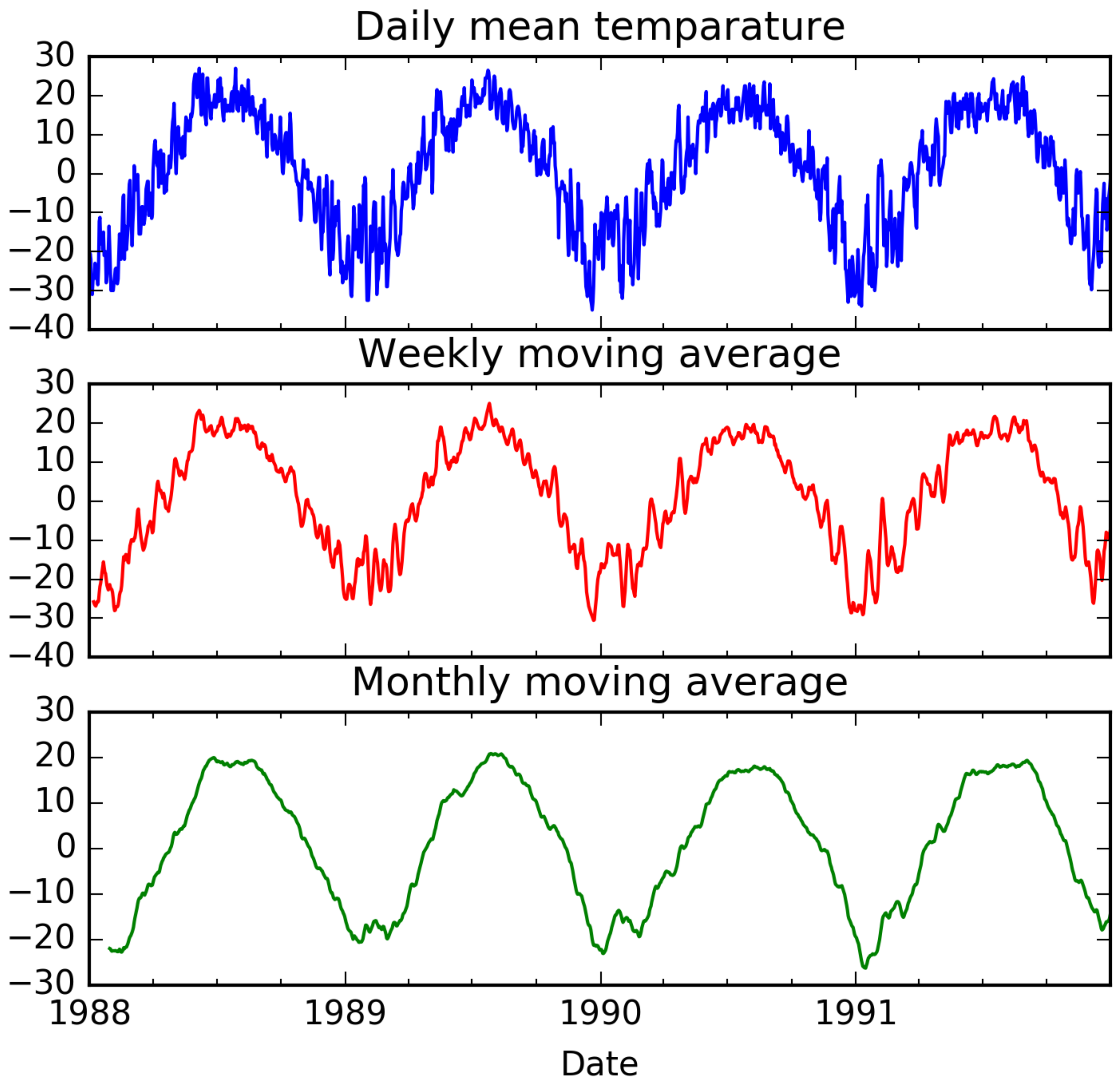
Advanced Processing



Fuente: *Practical Time Series Analysis*, Pal & Prakash

Moving statistics

Advanced Processing



Fuente: *Practical Time Series Analysis*, Pal & Prakash

TIME SERIES DECOMPOSITION

El **objetivo** de la descomposición de series de tiempo es **modelar la tendencia a largo plazo y la estacionalidad** y estimar la serie de tiempo general como una combinación de ellas. Dos modelos populares para la descomposición de series de tiempo son:

- Modelo aditivo
- Modelo multiplicativo

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Modelo Aditivo

Time Series Decomposition

El **modelo aditivo** formula la serie temporal original (xt) como la **suma del ciclo de tendencia** (Ft) y los **componentes estacionales** (St) de la siguiente manera:

- $xt = Ft + St + \epsilon t$

- **Los residuos Ꞓt** obtenidos tras ajustar los componentes tendenciales y estacionales **son las variaciones irregulares**. El modelo aditivo generalmente se aplica cuando hay un componente de ciclo de tendencia dependiente del tiempo, pero una estacionalidad independiente que no cambia con el tiempo.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Modelo Aditivo

Time Series Decomposition

- El **modelo de descomposición multiplicativa**, que da la serie de tiempo como producto de los componentes tendenciales, estacionales e irregulares, es útil cuando existe una **estacionalidad variable en el tiempo:**

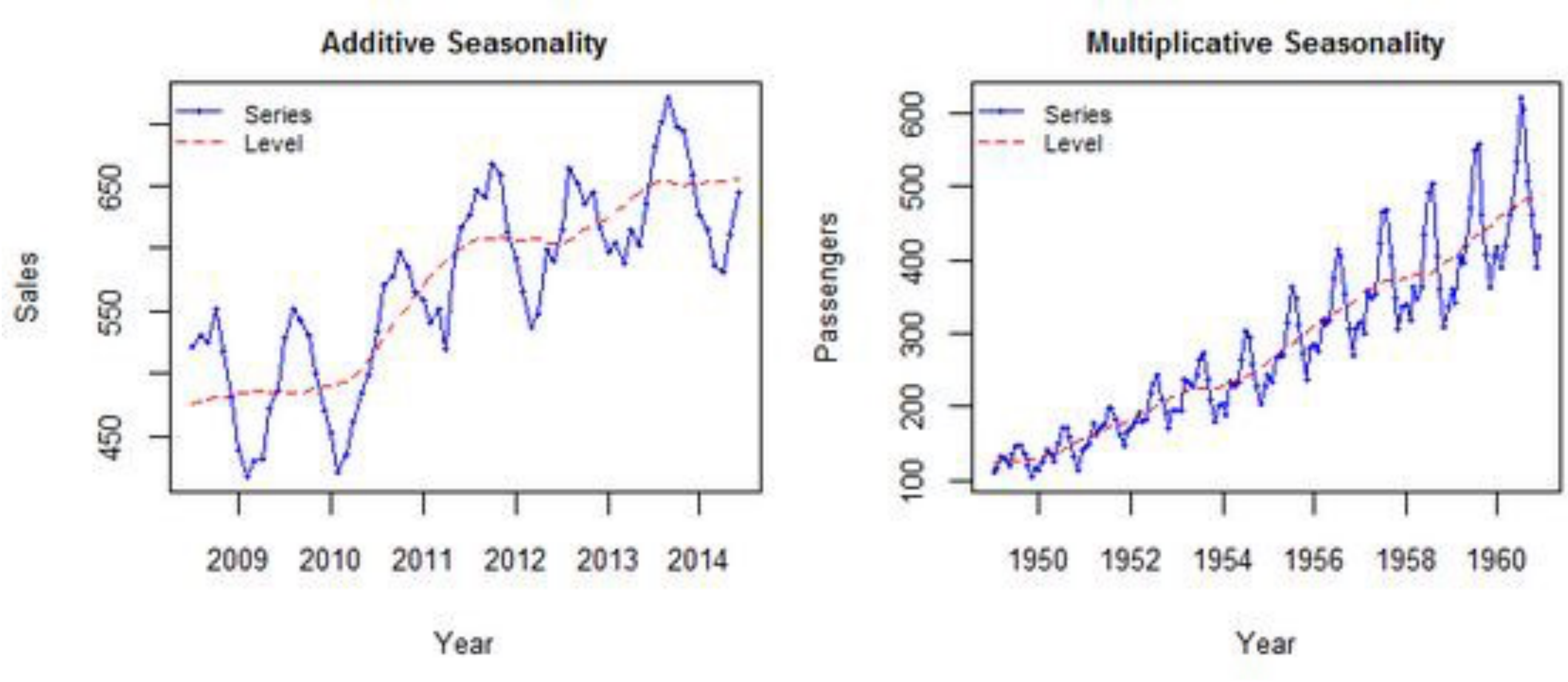
- xt = Ft x St x Єt

- Al tomar el logaritmo, el modelo multiplicativo se convierte en un modelo aditivo de logaritmo de los componentes individuales.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Aditivo vs. Multiplicativo

Time Series Decomposition



Fuente: *Practical Time Series Analysis*, Pal & Prakash

EXPONENTIAL SMOOTHING

Introducción

Exponential smoothing

El **exponential smoothing** es un método de pronóstico de series de tiempo para datos univariados.

- La predicción es una suma ponderada de observaciones pasadas, pero el modelo usa explícitamente una ponderación exponencialmente decreciente para las observaciones pasadas.

> *Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent the observation the higher the associated weight.*

— Page 171, Forecasting: principles and practice, 2013.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Types

Exponential smoothing

Hay **tres tipos principales de métodos** de pronóstico de series de tiempo de exponential smoothing.

- Un método simple que no asume una estructura sistemática, una extensión que maneja explícitamente las tendencias y el enfoque más avanzado que agrega soporte para la estacionalidad.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Single Exponential Smoothing

Exponential smoothing

- Método de pronóstico de series de tiempo para datos **univariados sin una tendencia o estacionalidad.**

- Requiere **un solo parámetro**, llamado alfa (a), también llamado **factor de suavizado** o coeficiente de suavizado.

- **Este parámetro controla la velocidad a la que la influencia de las observaciones en los pasos de tiempo anteriores decae exponencialmente**. Alpha se establece a menudo en un valor entre 0 y 1. Los valores grandes significan que el modelo presta atención principalmente a las observaciones pasadas más recientes, mientras que los valores más pequeños significan que se tiene en cuenta una mayor parte del historial al realizar una predicción.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Double Exponential Smoothing

Exponential smoothing

- Es una extensión que **agrega** explícitamente **soporte para las tendencias** en la serie de tiempo univariante.

- Se agrega un factor de suavizado adicional para controlar la disminución de la influencia del cambio de tendencia llamado beta (b).

- El método admite tendencias que cambian de diferentes formas: una aditiva y una multiplicativa, dependiendo de si la tendencia es lineal o exponencial respectivamente.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Double Exponential Smoothing

Exponential smoothing

- Para pronósticos de mayor alcance, la tendencia puede continuar de manera poco realista. Como tal, puede ser útil **amortiguar la tendencia** a lo largo del tiempo.

- Amortiguación significa **reducir el tamaño de la tendencia en tiempos futuros** hacia una línea recta (sin tendencia).

- Al igual que con el modelado de la tendencia en sí, podemos usar los mismos principios para amortiguar la tendencia, específicamente de forma aditiva o multiplicativa para obtener un efecto de amortiguación lineal o exponencial. Se utiliza un **coeficiente de amortiguación** Phi (p) para controlar la tasa de amortiguación.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

Triple Exponential Smoothing

Exponential smoothing

Extensión del que **agrega** explícitamente **soporte para la estacionalidad** a la serie de tiempo univariante.

- Además de los factores de suavizado alfa y beta, se agrega un **nuevo parámetro** llamado gamma (g) que controla la **influencia sobre el componente estacional**.
- Al igual que con la tendencia, la estacionalidad puede modelarse como un **proceso aditivo o multiplicativo** para un cambio lineal o exponencial en la estacionalidad.
- Además, para garantizar que la estacionalidad se modele correctamente, **se debe especificar el número de períodos de tiempo en un período estacional** (Período). Por ejemplo, si la serie es de datos mensuales y el período estacional se repite cada año, entonces el período = 12.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

AUTO-REGRESSIVE MODELS

MA Model

Auto-Regressive Models

- El modelo de media móvil es probablemente el enfoque más ingenuo para el modelado de series de tiempo. Este modelo simplemente establece que **la siguiente observación es la media de todas las observaciones pasadas.**

- Aunque simple, este modelo puede ser sorprendentemente bueno y representa un buen punto de partida.

- De lo contrario, la media móvil se puede utilizar para identificar tendencias interesantes en los datos. Podemos definir una ventana para aplicar el modelo de media móvil para suavizar la serie de tiempo y resaltar diferentes tendencias.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

MA Model

Auto-Regressive Models

- **Asume que la serie de tiempo es estacionaria o no muestra tendencias obvias** (movimiento creciente o decreciente a largo plazo) o **estacionalidad** (estructura periódica consistente).

- Existen muchos métodos para eliminar las tendencias y la estacionalidad de un conjunto de datos de series de tiempo al realizar pronósticos. Dos buenos métodos para cada uno son utilizar el método de diferenciación y modelar el comportamiento y restarlo explícitamente de la serie.

- Los valores de promedio móvil se pueden usar de varias maneras cuando se usan algoritmos de aprendizaje automático en problemas de series de tiempo.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

MA Model

Auto-Regressive Models

```
obs(t) = 1/3 * (t-2 + t-1 + t)
```

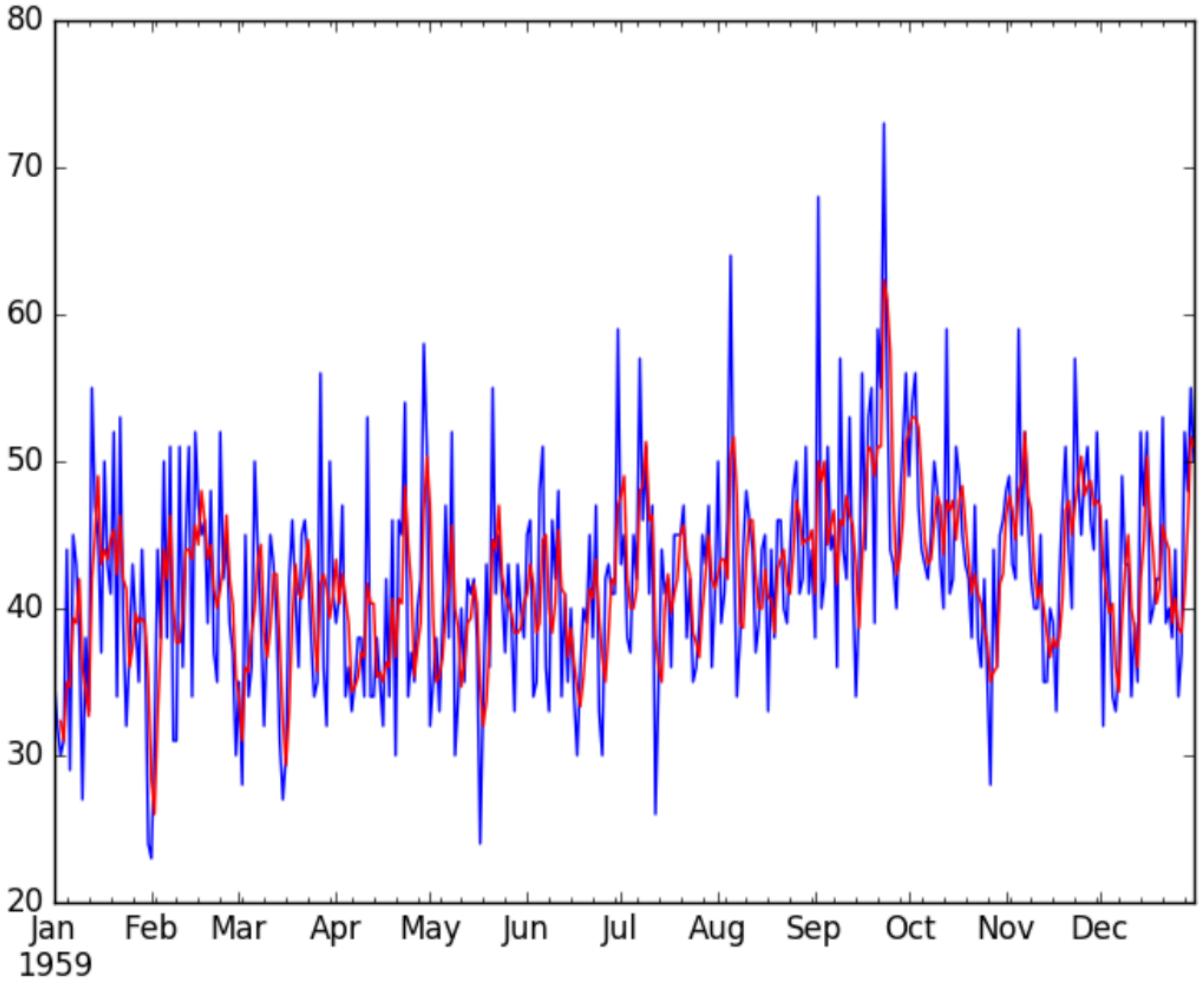
```
obs(t) = 1/3 * (t-2 + t-1 + t)
obs(t) = 1/3 * (35 + 32 + 30)
obs(t) = 32.333
```

```python
from pandas import read_csv
from matplotlib import pyplot
series = read_csv('daily-total-female-births.csv', header=0, index_col=0)
# Tail-rolling average transform
rolling = series.rolling(window=3)
rolling_mean = rolling.mean()
print(rolling_mean.head(10))
# plot original and transformed dataset
series.plot()
rolling_mean.plot(color='red')
pyplot.show()
```


Fuente: *Practical Time Series Analysis*, Pal & Prakash

MA Model

Auto-Regressive Models



Fuente: *Practical Time Series Analysis,* Pal & Prakash

AR Model

Auto-Regressive Models

- Un **modelo de regresión**, como la regresión lineal, modela un valor de salida basado en una combinación lineal de valores de entrada. Esta técnica se puede utilizar en series de tiempo en las que las **variables de entrada se toman como observaciones en pasos de tiempo anteriores**, llamadas variables rezagadas.

- Por ejemplo, podemos predecir el valor para el siguiente paso de tiempo (t+1) dadas las observaciones en los dos últimos pasos de tiempo (t-1 y t-2). Como modelo de regresión, se vería de la siguiente manera:

- $X(t + 1) = b0 + b1 * X(t-1) + b2 * X(t-2)$

- Debido a que el modelo de regresión usa datos de la misma variable de entrada en los pasos de tiempo anteriores, se lo conoce como una autorregresión.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

AR Model

Auto-Regressive Models

UNIVERSIDAD DE MONTEVIDEO

- Un modelo de autorregresión s**upone que las observaciones en los pasos de tiempo anteriores son útiles para predecir el valor en el siguiente paso de tiempo**.

- Las **estadísticas de correlación** también pueden **ayudar a elegir qué variables de rezago** serán útiles en un modelo y cuáles no.

- Si todas las variables de rezago muestran una correlación baja o nula con la variable de salida, entonces sugiere que el problema de la serie de tiempo puede no ser predecible. Esto puede resultar muy útil al comenzar con un nuevo conjunto de datos.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

AR Model

Auto-Regressive Models

- Un **modelo de autorregresión** es un modelo de regresión lineal que utiliza **variables rezagadas como variables de entrada.**

- Podríamos calcular el modelo de regresión lineal manualmente usando la clase **LinearRegession** en scikit-learn y especificar manualmente las variables de entrada de retraso a usar.

- Alternativamente, la biblioteca statsmodels proporciona un modelo de autorregresión en el que debe especificar un valor de retraso apropiado y entrena un modelo de regresión lineal. Se proporciona en la clase **AutoReg**.

- Toma un **parámetro p que representa el retraso máximo**. Para encontrarlo, miramos el gráfico de autocorrelación parcial e identificamos el retraso después del cual la mayoría de los retrasos no son significativos.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

ARIMA Model

Auto-Regressive Models

Un modelo **ARIMA** es una clase de modelos estadísticos para analizar y pronosticar datos de series de tiempo. ARIMA es un acrónimo de AutoRegressive Integrated Moving Average.

Los parámetros del modelo ARIMA se definen de la siguiente manera:

- p: El número de observaciones de retraso incluidas en el modelo, también llamado orden de retraso.
- d: El número de veces que se diferencian las observaciones sin procesar, también llamado grado de diferenciación.
- q: El tamaño de la ventana de media móvil, también llamado orden de media móvil.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

ARIMA Model

Auto-Regressive Models

UNIVERSIDAD DE MONTEVIDEO

- **AR: Autoregresión**. Un modelo que utiliza la relación dependiente entre una observación y cierto número de observaciones retrasadas.

- **I: Integrado.** El uso de la diferenciación de observaciones en bruto (por ejemplo, restar una observación de una observación en el paso de tiempo anterior) para hacer estacionaria la serie de tiempo.

- **MA: Media móvil.** Modelo que utiliza la dependencia entre una observación y un error residual de un modelo de promedio móvil aplicado a observaciones retrasadas.

Cada uno de estos **componentes se especifica explícitamente en el modelo** como parámetro. Se utiliza una notación estándar de ARIMA (p, d, q) donde los parámetros se sustituyen por valores enteros para indicar rápidamente el modelo ARIMA específico que se está utilizando.

Fuente: *Practical Time Series Analysis*, Pal & Prakash

ARIMA Model

- Se construye un modelo de regresión lineal que incluye el número y el tipo de términos especificados, y los datos se preparan mediante un grado de diferenciación para hacerlos estacionarios, es decir, para eliminar las estructuras de tendencia y estacionales que afectan negativamente al modelo de regresión.

- Se puede usar un valor de 0 para un parámetro, lo que indica no usar ese elemento del modelo. De esta manera, el modelo ARIMA se puede configurar para realizar la función de un modelo ARMA, e incluso un modelo simple AR, I o MA.

Fuente: *Practical Time Series Analysis*, Pal & Prakash