

Sexta Clase de Análisis de Datos

Prof: Boris Panes
Universidad Del Desarrollo

Octubre 05, 2024

*Durante la clase tendremos la oportunidad de
conversar sobre el avance de su proyecto T2
y algunos aspectos generales de T3*

Asistencia y Evaluaciones

Asistencia por ZOOM: 28 alumnos promedio con asistencia sobre 120 minutos por clase

Asistencia presencial: 5 alumnos como mínimo todas las clases

33 alumnos efectivos por clase de un total de 38

Evaluación online de preparación de datos: 10 alumnos han completado la evaluación

Evaluación online de test estadístico: 5 alumnos han completado la evaluación

Ahora existen tres evaluaciones online, probablemente habrán dos mas

La evaluación del proyecto T1 tuvo una componente cualitativa y otra más cuantitativa

El lado cualitativo se refiere al cumplimiento con el formato sugerido y la coherencia del trabajo en general. La parte mas cuantitativa se refiere al cumplimiento de los puntos del algoritmo mostrado en clases. Ambas contribuciones contribuyen a la nota final, las cuales fueron comunicadas a cada grupo junto con algunos comentarios respectivos

La evaluación del proyecto T2 será bastante similar en términos de protocolo

Bases para Proyecto T2

Objetivos: Ejemplo directo de modelamiento de datos usando Regresión Lineal Simple. Aplicar el flujo de análisis de datos, desde la selección y limpieza de datos hasta el calculo y visualización de una Regresión Lineal Simple.

Se aplican las mismas reglas generales de T1. En particular, la presentación debería tener entre 10 y 15 laminas y durar entre 15 y 30 minutos. El algoritmo del notebook, hilo conductor de la presentación y esquema del video debería considerar los siguientes pasos

- Seleccionar datos (csv de kaggle u otros)

- Preparación de datos (opcional si continua con datos T1)

- Selección de columnas X e Y considerando contexto causal entre las variables

 - Chequeo del contenido y distribución de las variables

 - Gráfico de dispersión

 - Planteamiento del modelo de Regresión Lineal para ajustar tendencia de X e Y

 - Calcular explícitamente los valores de los coeficientes

- Discusión de resultados

Entrega T2: Octubre 6, 2024

Calendario y Evaluaciones

TRIM.	FECHA	HORA
TRIM.2	sábado, 24 de agosto de 2024	11.20 - 12.30 12.30 - 13.40
TRIM.2	sábado, 31 de agosto de 2024	11.20 - 12.30 12.30 - 13.40
TRIM.2	sábado, 7 de septiembre de 2024	11.20 - 12.30 12.30 - 13.40
TRIM.2	sábado, 14 de septiembre de 2024	11.20 - 12.30 12.30 - 13.40
TRIM.2	sábado, 28 de septiembre de 2024	11.20 - 12.30 12.30 - 13.40
TRIM.2	sábado, 5 de octubre de 2024	11.20 - 12.30 12.30 - 13.40
TRIM.2	sábado, 19 de octubre de 2024	11.20 - 12.30 12.30 - 13.40

Más ejercicios con múltiples alternativas

Entrega T1: Limpieza y Estructura de Datos

Entrega T2: Regresión Lineal

Entrega T3: Series Temporales

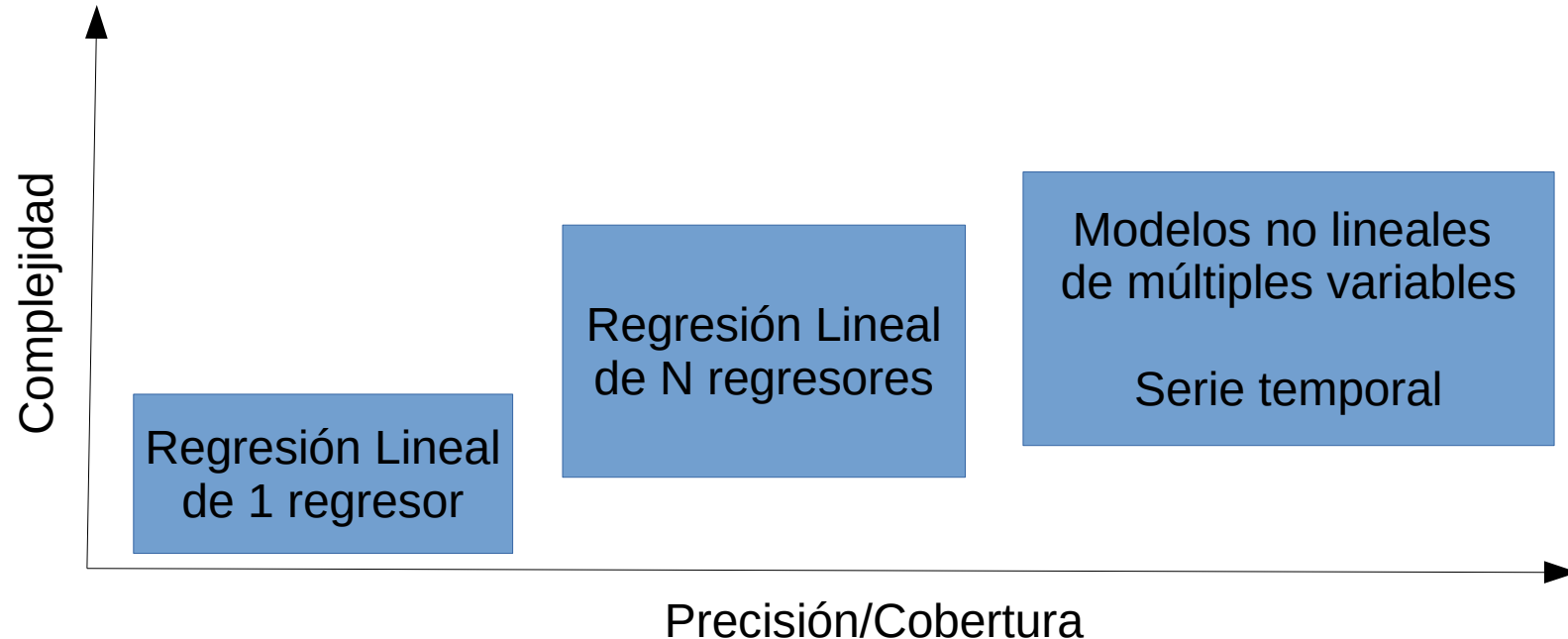
Fecha de cierre de evaluaciones online y entrega de T3: **Domingo 27 de Octubre**

Resumen de contenidos

Proceso continuo de modelamiento

En general seguimos interesados primeramente en la capacidad predictiva de los modelos que se construyen en base a los datos.

En el contexto del ejercicio de las notas podemos nos interés predecir el aumento en las notas dependiendo del cambio en el tamaño de las clases. **En el contexto de series temporales nos interesa predecir el comportamiento futuro de los observables relevantes**



Contenidos de Series Temporales

Explicaciones generales

Definición y ejemplos

Diagnóstico del problema usando caso de estudio

Selección de caso práctico, relevante y autocontenido

Presentación de Notebook con caso de estudio (**Guía para T3**)

Exploración de los datos

 Limpieza, formato, ingeniería de datos

 Estacionariedad

Autocorrelacion

 Modelo de autorregresión de primer orden

Detalles técnicos

Limitaciones del modelo de Regresión Lineal

Modelos autorregresivos (AR), ARIMA y autorregresión vectorial (VAR)

Análisis estadístico de los resultados

Contenidos de la Clase

Explicaciones generales **Definición y ejemplos**

Diagnóstico del problema usando caso de estudio

Selección de caso práctico, relevante y autocontenido

Presentación de Notebook con caso de estudio (Guía para T3)

Exploración de los datos

 Limpieza, formato, ingeniería de datos

 Estacionariedad

Autocorrelacion

Modelo de autorregresión de primer orden

Desarrollo de contenidos

Definición de Serie Temporal

Matemáticamente, una **serie temporal** es un conjunto de observaciones $\{x_t\}$ indexadas por un parámetro t , donde **t representa el tiempo (normalmente, un número entero)**

$$\{x_t\}_{t=1}^n = \{x_1, x_2, x_3, \dots, x_n\}$$

En términos prácticos, considerando la disponibilidad de datos tabulados, una **serie temporal** puede estar representada por una tabla con **N filas y al menos dos columnas**, la primera con la estampa de tiempo (Time) asociado a la observación y la segunda columna contiene el valor que representa la medición de alguna cantidad (X,Y,Z,etc)

En el estudio del clima: Fecha, Temperatura, Precipitaciones, etc

En el mercado de valores: Fecha y Hora, Precio, Compra, Venta, etc

En políticas publicas: Mes, Inflación, Desempleo, Producto Interno Bruto, Notas

Mientras mas entradas tiene la serie temporal más conocimiento tenemos sobre el proceso que se esta estudiando. **Una serie temporal bien comportada tiene un intervalo de tiempo constante y datos existentes para todos los valores intermedios**

Predicciones económicas para el futuro

¿cuál es la mejor predicción sobre el valor de una variable en una fecha futura?

Utilización del modelo fuera del rango de ajuste, **probablemente sustentado por efectos estacionarios** asociados a las series de tiempo

Efectos causales dinámicos de Y a partir de una variación de X

¿cuál es el efecto sobre las muertes en accidentes de tráfico de una ley que obliga a que los pasajeros utilicen cinturones de seguridad, tanto en un momento inicial como posteriormente cuando los conductores se acomodan a la ley?

Esta pregunta se ve bastante profunda o poco clara aun. Es planteada en S&W, dado las restricciones de tiempo quizás solo veremos parte de este tema

Características cualitativas de una serie temporal

Para identificar un conjunto de datos en el contexto de regresión lineal, buscamos una tendencia mas menos constante de los datos. En el contexto de series temporales podemos encontrar tendencias pero también otros elementos relevantes

Estacionariedad, Tendencias, Puntos de Ruptura

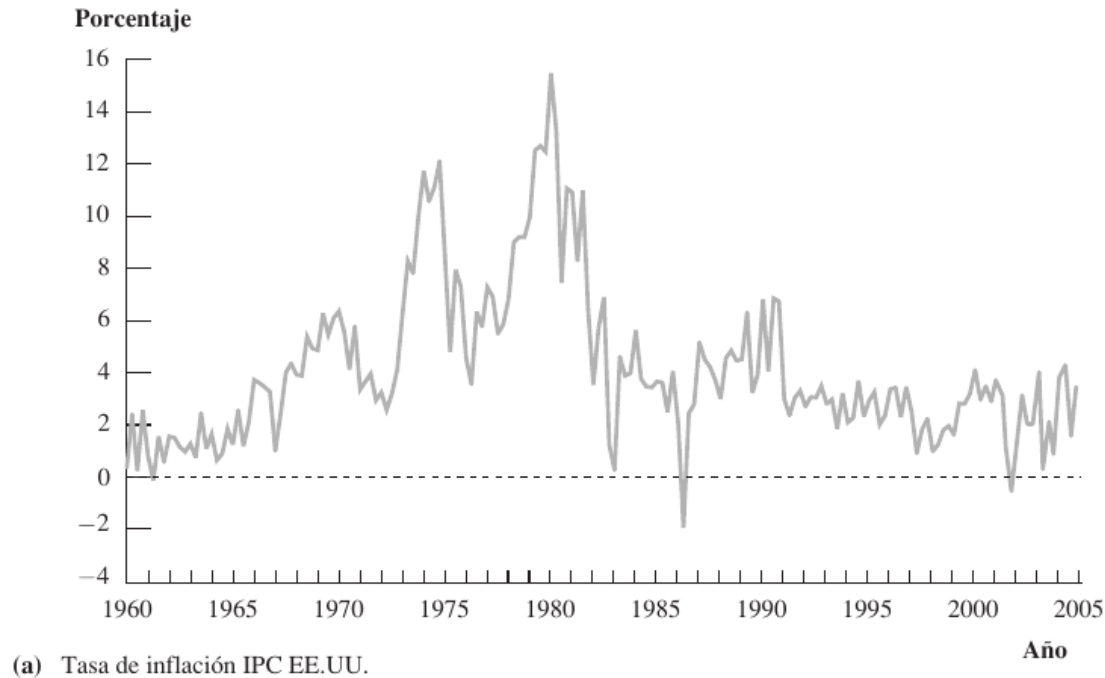
El supuesto de que el futuro será como el pasado resulta muy importante en la regresión de series temporales, lo suficiente como para que se le dote de su propio nombre, «estacionariedad». Las variables de series temporales pueden no ser estacionarias en varios sentidos, pero dos de ellos son especialmente relevantes para el análisis de regresión de series temporales con datos económicos: (1) las series pueden tener movimientos persistentes a largo plazo, es decir, la serie puede presentar tendencias; y (2) la regresión poblacional puede ser inestable en el tiempo, es decir, la regresión poblacional puede tener puntos de ruptura («breaks»). Estas desviaciones de la estacionariedad ponen en peligro las predicciones e inferencias basadas en la regresión de series temporales. Afortunadamente, existen procedimientos estadísticos para detectar las tendencias y los puntos de ruptura y, una vez detectados, para ajustar la especificación del modelo. Estos procedimientos se presentan en las Secciones 14.6 y 14.7.

Caso de estudio para Serie Temporal

Selección caso de estudio: Tasa de Inflación

En el capítulo 14 de S&W podemos encontrar el siguiente gráfico en la discusión inicial sobre series temporales

FIGURA 14.1 Inflación y desempleo en Estados Unidos, 1960-2004



Parece una serie temporal interesante. Presenta

Rasgos estacionarios
(repetición de patrones)

Tendencias

Quiebres

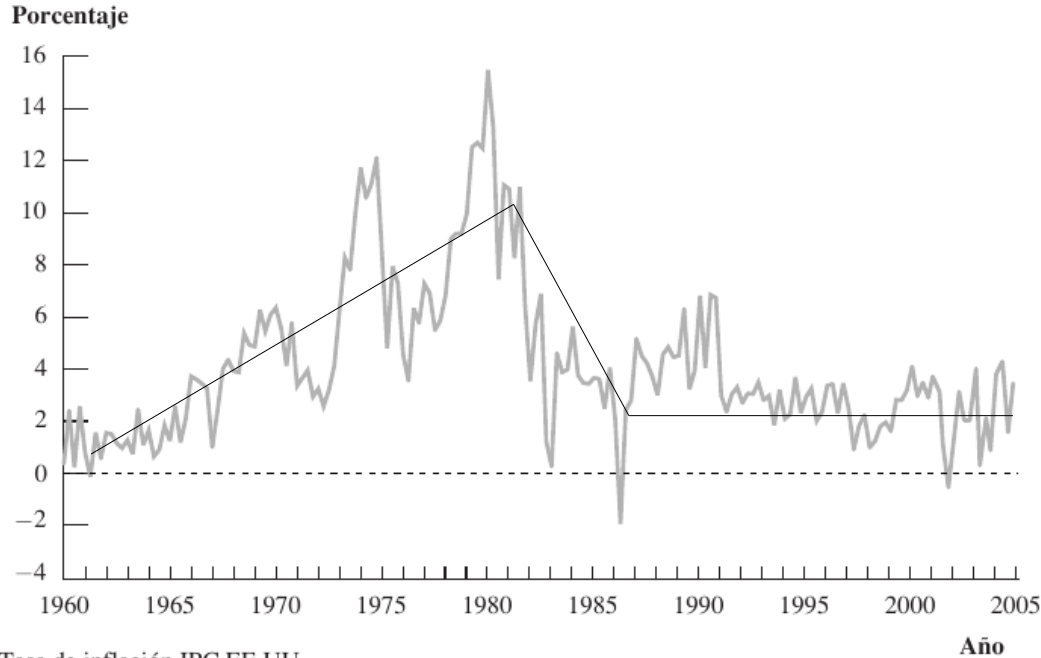
¿Tenemos el CSV?

Si, lo tenemos

Notar que los valores de tiempo son densos, es decir tenemos valores para todo t

Caso de estudio: Tasa de Inflación

FIGURA 14.1 Inflación y desempleo en Estados Unidos, 1960-2004



(a) Tasa de inflación IPC EE.UU.

Parece una serie temporal interesante. Presenta

Rasgos estacionarios
(repetición de patrones)

Tendencias

Quiebres

¿Tenemos el CSV?

Si, lo tenemos

Notar que los eventos en tiempos pasados pueden influenciar valores medidos en tiempos futuros. Los eventos en una serie temporal no satisfacen el requerimiento de i.i.d

Variación del precio de los productos con el tiempo

La variación de los precios puede modelarse considerando la dinámica de la oferta y la demanda

Variables: Precio de un producto, Oferta del producto, Demanda del producto.

La **ley de oferta y demanda** identifica una relación aproximada entre estas variables

Cuando la demanda de un producto aumenta se espera un aumento en el precio del producto

Cuando aumenta el precio de un producto aumenta la oferta de este

Al aumentar la oferta disminuye el precio del producto

Dado las fuerzas de la oferta y la demanda los precios tienden a oscilar en torno a un valor estable donde la oferta iguala a la demanda

Tasa de inflación: variación en términos porcentuales de los precios

El problema empírico estudiado en este capítulo es la **predicción de la tasa de inflación**, es decir, el aumento general en términos porcentuales de los precios.

$$IPC = \left(\frac{\text{Costo de la canasta en el período actual}}{\text{Costo de la canasta en el período base}} \right) \times 100$$

*Para Chile el valor base es 2018

Donde el periodo base se define un periodo en particular arbitraria en el pasado con respecto al cual se desea comparar

Este ejemplo nos permite revisar conceptos básicos del análisis de series temporales que son análogos al tratamiento estándar, incluyendo selección de los datos, descripción del problema, exploración de datos, limpieza y transformación de datos

Modelos de regresión de series temporales en los que los regresores son los valores pasados de la variable dependiente; estos **modelos «autorregresivos»** utilizan la **historia de la inflación para predecir su futuro**

TABLA 14.1

La inflación en los Estados Unidos en el 2004 y el primer trimestre de 2005

Trimestre	IPC EE.UU.	Tasa de Inflación a una tasa anualizada (Inf_t)	Primer retardo (Inf_{t-1})	Variación de la inflación (ΔInf_t)
2004:I	186,57	3,8	0,9	2,9
2004:II	188,60	4,4	3,8	0,6
2004:III	189,37	1,6	4,4	-2,8
2004:IV	191,03	3,5	1,6	1,9
2005:I	192,17	2,4	3,5	-1,1

La tasa de inflación anualizada es la variación porcentual en el IPC en el trimestre actual respecto al trimestre anterior, multiplicada por cuatro. El primer retardo de la inflación es su valor en el trimestre anterior, y la variación de la inflación es la tasa de inflación actual menos su primer retardo. Todas las cifras se han redondeado al decimal más cercano..

us_macro_quarterly
_imputed.csv

	freq	GDPC1	JAPAN_IP	PCECTPI	CPIAUCSL
194	7/1/2003	13372.357	99.142200	87.769	184.43
195	10/1/2003	13528.710	102.001053	88.124	184.80
196	1/1/2004	13606.509	103.271654	88.797	186.57
197	4/1/2004	13706.247	104.789317	89.421	188.60
198	7/1/2004	13830.828	105.636384	89.942	189.37
199	10/1/2004	13950.376	103.730482	90.652	191.03
200	1/1/2005	14099.081	105.777562	91.122	192.17

Aproximadamente la
misma tabla

¿Podemos derivar las
otras columnas?

Acrónimos de las columnas del dataset

En las paginas del Federal Reserve Bank of Sant Louis (FRED) podemos encontrar múltiples series de tiempo de tipo económico

<https://fred.stlouisfed.org/>

CPIAUCSL = Consumer Price Index for All Urban Consumers

<https://fred.stlouisfed.org/series/CPIAUCSL>

PCECTPI = Chain-type Price Index

<https://fred.stlouisfed.org/series/PCECTPI>

JAPAN_IP = Production, Sales, Work Started and Orders: Production Volume: Economic Activity: Industry (Except Construction) for Japan

<https://fred.stlouisfed.org/series/JPNPROINDQISMEI>

GDPC1 = Real Gross Domestic Product

<https://fred.stlouisfed.org/series/GDPC1>

Download, graph, and track [825,000 US and international time series](#) from [114 sources](#).

Reproducción del gráfico usando datos disponibles

	freq	GDPC1	JAPAN_IP	PCECTPI	CPIAUCSL
194	7/1/2003	13372.357	99.142200	87.769	184.43
195	10/1/2003	13528.710	102.001053	88.124	184.80
196	1/1/2004	13606.509	103.271654	88.797	186.57
197	4/1/2004	13706.247	104.789317	89.421	188.60
198	7/1/2004	13830.828	105.636384	89.942	189.37
199	10/1/2004	13950.376	103.730482	90.652	191.03
200	1/1/2005	14099.081	105.777562	91.122	192.17
201	4/1/2005	14172.695	105.989329	91.728	193.67
202	7/1/2005	14291.757	105.636384	92.734	196.60



	freq	GDPC1	JAPAN_IP	PCECTPI	CPIAUCSL	Inf_T	Inf_T-1	Delta_Inf_T
194	7/1/2003	13372.357	99.142200	87.769	184.43	2.971541	-0.654415	3.625956
195	10/1/2003	13528.710	102.001053	88.124	184.80	0.802472	2.971541	-2.169068
196	1/1/2004	13606.509	103.271654	88.797	186.57	3.831169	0.802472	3.028696
197	4/1/2004	13706.247	104.789317	89.421	188.60	4.352254	3.831169	0.521085
198	7/1/2004	13830.828	105.636384	89.942	189.37	1.633086	4.352254	-2.719168
199	10/1/2004	13950.376	103.730482	90.652	191.03	3.506363	1.633086	1.873277
200	1/1/2005	14099.081	105.777562	91.122	192.17	2.387060	3.506363	-1.119304
201	4/1/2005	14172.695	105.989329	91.728	193.67	3.122236	2.387060	0.735176
202	7/1/2005	14291.757	105.636384	92.734	196.60	6.051531	3.122236	2.929295

Trimestre	IPC EE.UU.	Tasa de Inflación a una tasa anualizada (Inf_t)	Primer retardo (Inf_{t-1})	Variación de la inflación (ΔInf_t)
2004:I	186,57	3,8	0,9	2,9
2004:II	188,60	4,4	3,8	0,6
2004:III	189,37	1,6	4,4	-2,8
2004:IV	191,03	3,5	1,6	1,9
2005:I	192,17	2,4	3,5	-1,1

Limpieza de datos

Dataframe
con variables
auxiliares

	freq	GDPC1	JAPAN_IP	PCECTPI	CPIAUCSL	Inf_T	Inf_T-1	Delta_Inf_T	Delta_Inf_T-1
0	1/1/1955	2683.766	NaN	15.755	26.79	NaN	NaN	NaN	NaN
1	4/1/1955	2727.452	NaN	15.771	26.76	-0.447928	NaN	NaN	NaN
2	7/1/1955	2764.128	NaN	15.834	26.78	0.298954	-0.447928	0.746882	NaN
3	10/1/1955	2780.762	NaN	15.878	26.86	1.194922	0.298954	0.895968	0.746882
4	1/1/1956	2770.032	NaN	15.943	26.86	0.000000	1.194922	-1.194922	0.895968
...
247	10/1/2016	16851.420	99.125073	111.583	242.18	2.710455	1.837544	0.872911	-0.853227
248	1/1/2017	16903.240	99.291726	112.198	243.95	2.923445	2.710455	0.212990	0.872911
249	4/1/2017	17031.085	101.324890	112.273	244.01	0.098381	2.923445	-2.825065	0.212990
250	7/1/2017	17163.894	101.724856	112.699	245.30	2.114667	0.098381	2.016287	-2.825065
251	10/1/2017	17271.702	103.491376	113.458	247.30	3.261313	2.114667	1.146645	2.016287

252 rows × 9 columns

Observamos la presencia de NaN en las variables originales del dataset como JAPAN_IP, además vemos que las nuevas columnas generan valores NaN dado su naturaleza recursiva. Restringiendo el set de datos para que comience el primer trimestre de 1960 evadimos todos los valores nulos

Limpieza de datos

```
data_inflacion.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 252 entries, 0 to 251  
Data columns (total 9 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   freq            252 non-null   object  
1   GDPC1           252 non-null   float64  
2   JAPAN_IP        232 non-null   float64  
3   PCECTPI         252 non-null   float64  
4   CPIAUCSL        252 non-null   float64  
5   Inf_T           251 non-null   float64  
6   Inf_T-1         250 non-null   float64  
7   Delta_Inf_T     250 non-null   float64  
8   Delta_Inf_T-1   249 non-null   float64  
dtypes: float64(8), object(1)  
memory usage: 17.8+ KB
```



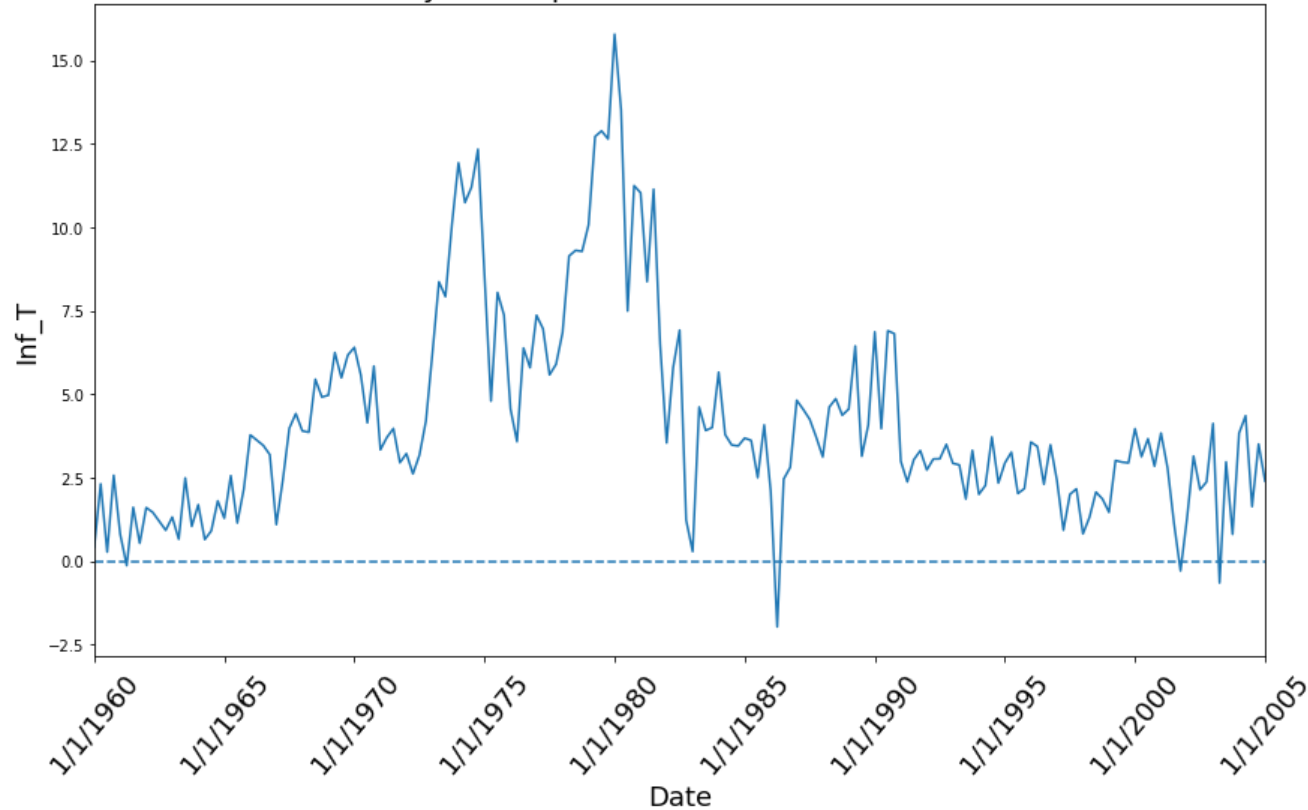
```
df_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 181 entries, 20 to 200  
Data columns (total 9 columns):  
#   Column          Non-Null Count  Dtype  
---  ---  
0   freq            181 non-null   object  
1   GDPC1           181 non-null   float64  
2   JAPAN_IP        181 non-null   float64  
3   PCECTPI         181 non-null   float64  
4   CPIAUCSL        181 non-null   float64  
5   Inf_T           181 non-null   float64  
6   Inf_T-1         181 non-null   float64  
7   Delta_Inf_T     181 non-null   float64  
8   Delta_Inf_T-1   181 non-null   float64  
dtypes: float64(8), object(1)  
memory usage: 12.9+ KB
```

Trimestres entre 1955-2017

Trimestres entre 1960-2005

Inflación y desempleo en Estados Unidos, 1960-2004



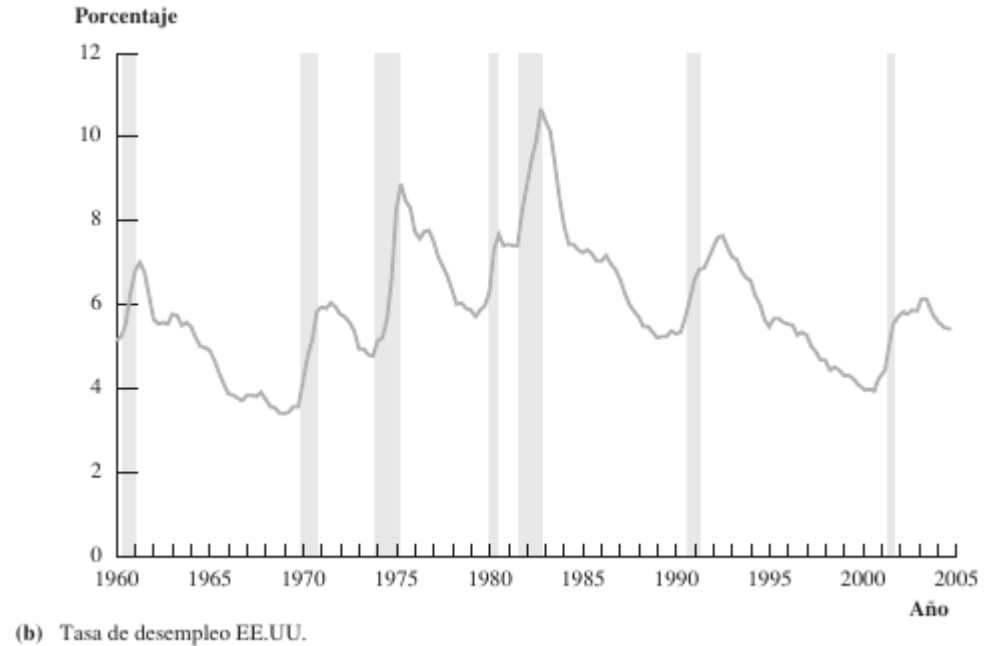
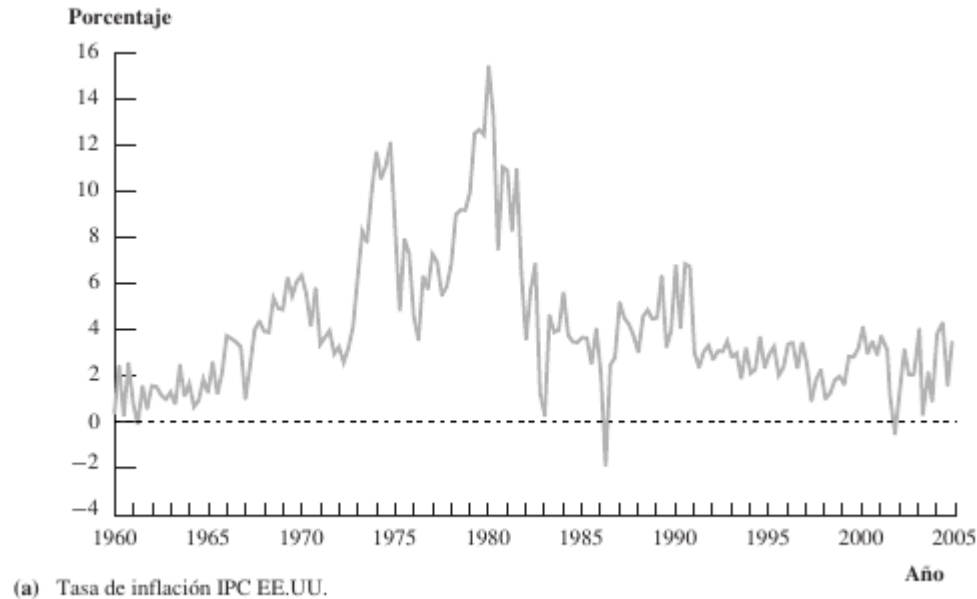
Estos datos son altamente consistentes con los datos presentados en el capítulo 14 del libro S&W sobre series temporales, por lo cual podemos estar tranquilos para seguir con este caso de estudio. Inf_T es la tasa de variación porcentual del IPC anualizado.

$$\text{Inf}_t = 4 \times 100 \times \frac{\text{IPC}_t - \text{IPC}_{t-1}}{\text{IPC}_{t-1}}$$

El factor 4 viene de anualizar la tasa, es decir la variación de un trimestre se multiplica por 4 para obtener la de un año.

¿Por que podría ser útil predecir Inf_T ?

Hasta el momento hemos visto que la métrica Inf_T nos entrega datos en formato de serie temporal. Lo cual permite estudiar el tema con un ejemplo específico. Además debemos recordar que estos datos pueden ser útiles para entender otras variables económicas como la tasa de desempleo



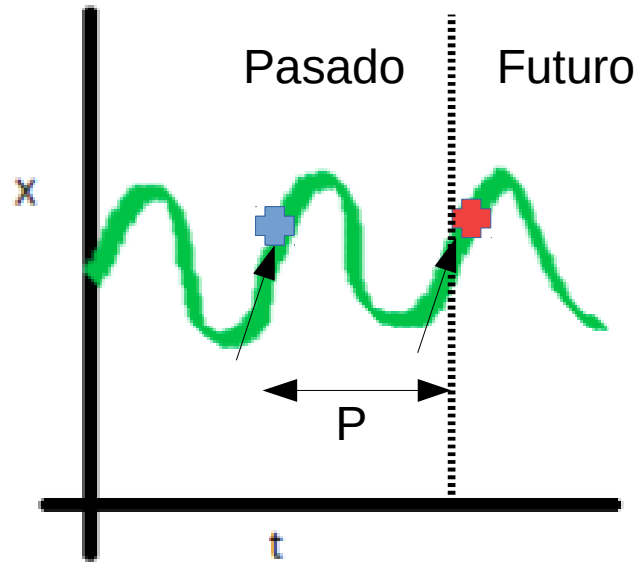
Parece argumentable relacionar periodos de alto desempleo con un aumento de la inflación medida como Inf_T

Exploración de estadísticos y modelamiento

Estacionariedad y modelos de autorregresión

Considerando las características de estacionariedad es razonable asumir que un valor futuro de la serie temporal dependerá de los valores pasados

Serie temporal = Estacionariedad = Autoregresión



■ Evento conocido análogo al evento que mediremos

↗ Datos anteriores a la medición de un evento dado

■ Evento que queremos predecir

“Si los valores vienen subiendo con cierta pendiente entonces el siguiente valor debería ser mas alto”

$$x(t + P) = x(t)$$

Definición de elementos clave

La variable t es utilizada para identificar el valor de la variable temporal para un evento dado. En general puede estar en unidades de días, meses, **trimestres**, etc. Sea la variable Y el parámetro a estudiar (IPC por ejemplo), tenemos las siguientes definiciones

Valor de Y en el tiempo t $\longrightarrow Y_t$

Valor de Y en el tiempo $t-1$ $\longrightarrow Y_{t-1}$

Valor de Y en el tiempo $t-j$ $\longrightarrow Y_{t-j}$

Variación de Y entre t y $t-1$ $\longrightarrow \Delta Y_t = Y_t - Y_{t-1}$

Ultimo valor de t $\longrightarrow T$

$$Y_t = \text{IPC}_t$$

$$\text{Inf}_t \approx 4 \times 100 \times \Delta \ln Y_t$$

En particular, cuando tratamos valores que pueden crecer exponencialmente, como los índices económicos es conveniente utilizar logaritmos. En estos casos podemos notar que la diferencia de logaritmos es similar a la diferencia porcentual entre los valores

$$\Delta \ln(Y_t) = \ln(Y_t) - \ln(Y_{t-1}) = \ln(Y_{t-1} + \Delta Y_t) - \ln(Y_{t-1}) \approx \frac{\Delta Y_t}{Y_{t-1}}$$

Autoregresión y Autocorrelación

Dado el argumento anterior podemos suponer que el valor de la variable Y_t puede ser influido o estar determinado por los valores de Y_{t-1} , Y_{t-2} , ..., Y_{t-j} .

En términos más cuantitativos podemos evaluar la correlación entre valores de Y medidos en diferentes tiempos, para así entender el *tamaño* de la dependencia. En los casos donde la correlación es calculada sobre una misma variable en diferentes puntos en el tiempo se habla de autocorrelación o correlación serial

$$\widehat{\text{cov}}(Y_t, Y_{t-j}) = \frac{1}{T} \sum_{t=j+1}^T (Y_t - \bar{Y}_{j+1,T})(Y_{t-j} - \bar{Y}_{1,T-j})$$

$$\hat{\rho}_j = \frac{\widehat{\text{cov}}(Y_t, Y_{t-j})}{\widehat{\text{var}}(Y_t)},$$

Dado un valor de retardo fijo, la covarianza es calculada sobre todas las posibilidades, donde la suma recorre los valores entre $j+1$ y T para permitir el primer par de forma consistente. El promedio es calculado considerando una separación entre los valores que juegan el rol de valor futuro y aquellos que juegan el rol de valor pasado

Procedimiento para calcular autocorrelación

	freq	GDPC1	JAPAN_IP	PCECTPI	CPIAUCSL	Inf_T	Inf_T-1	Delta_Inf_T	Delta_Inf_T-1
0	1/1/1960	3123.162	12.184435	17.424	29.40	0.408580	2.466598	-2.058018	0.400482
1	4/1/1960	3111.310	12.676183	17.516	29.57	2.312925	0.408580	1.904345	-2.058018
2	7/1/1960	3119.057	13.222570	17.583	29.59	0.270544	2.312925	-2.042381	1.904345
3	10/1/1960	3081.300	13.850916	17.661	29.78	2.568435	0.270544	2.297891	-2.042381
4	1/1/1961	3102.251	14.615858	17.694	29.84	0.805910	2.568435	-1.762525	2.297891
...
176	1/1/2004	13606.509	103.271654	88.797	186.57	3.831169	0.802472	3.028696	-2.169068
177	4/1/2004	13706.247	104.789317	89.421	188.60	4.352254	3.831169	0.521085	3.028696
178	7/1/2004	13830.828	105.636384	89.942	189.37	1.633086	4.352254	-2.719168	0.521085
179	10/1/2004	13950.376	103.730482	90.652	191.03	3.506363	1.633086	1.873277	-2.719168
180	1/1/2005	14099.081	105.777562	91.122	192.17	2.387060	3.506363	-1.119304	1.873277

181 rows × 9 columns

Valores
futuros

Valores
pasados

Luego de crear las variables correspondientes a los valores futuros y pasados la autocorrelación es simplemente el valor de la correlación entre las columnas correspondientes

Autoregresión y Autocorrelación

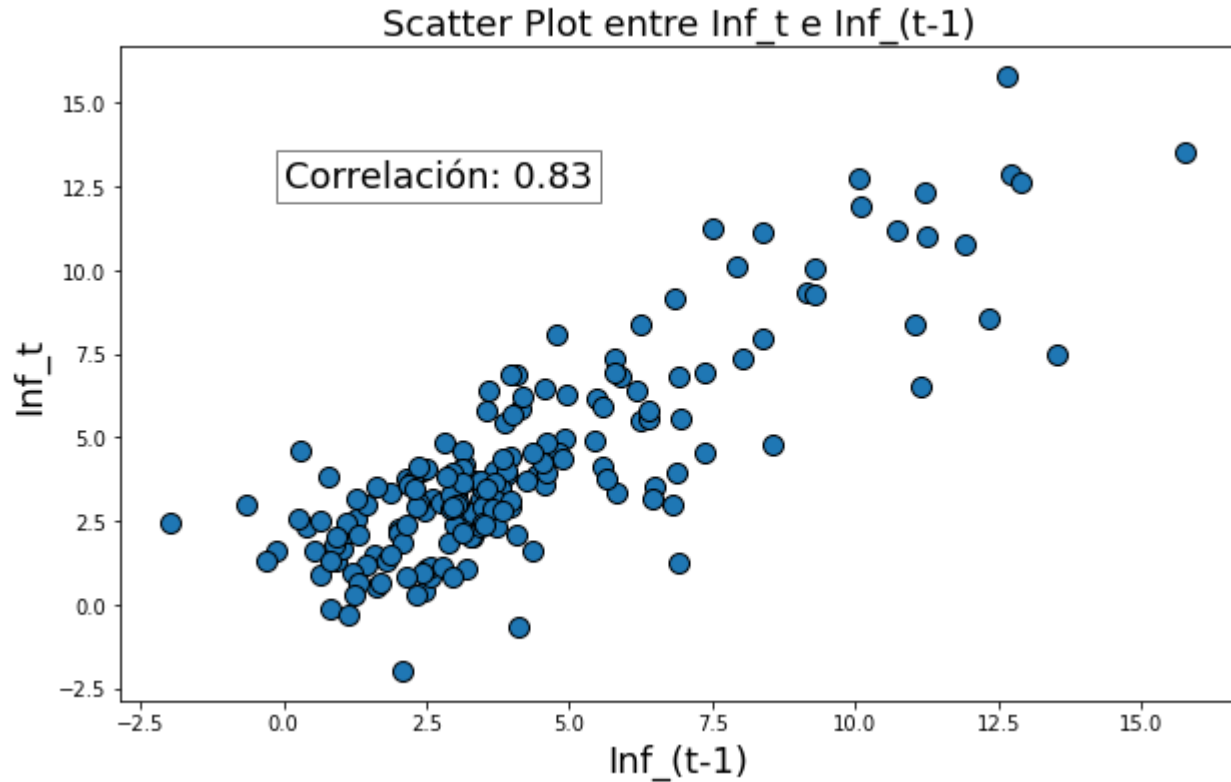
TABLA 14.2 Primeras cuatro autocorrelaciones muestrales de la tasa de inflación en EE.UU. y su variación, 1960:I-2004:IV

Retardo	Autocorrelación de:	
	Tasa de inflación (Inf_t)	Variación de la tasa de inflación (ΔInf_t)
1	0,84	-0,26
2	0,76	-0,25
3	0,76	-0,29
4	0,67	-0,06

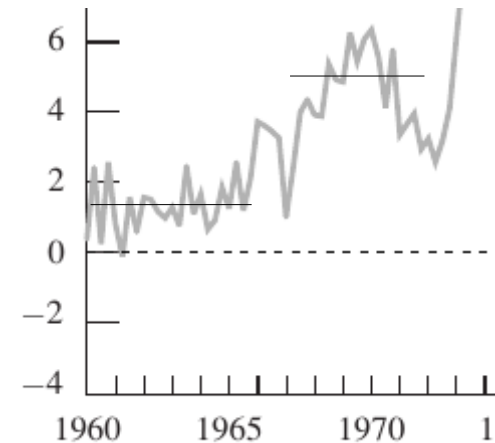
Podemos notar que la autocorrelación de la tasa de inflación es bastante alta incluso hasta un retardo de 4 unidades temporales (4 trimestres).

Por otro lado podemos notar que la variación de la tasa de inflación presenta una autocorrelación negativa hasta el cuarto nivel de retardo

Scatter plot entre Inf_t e Inf_t-1

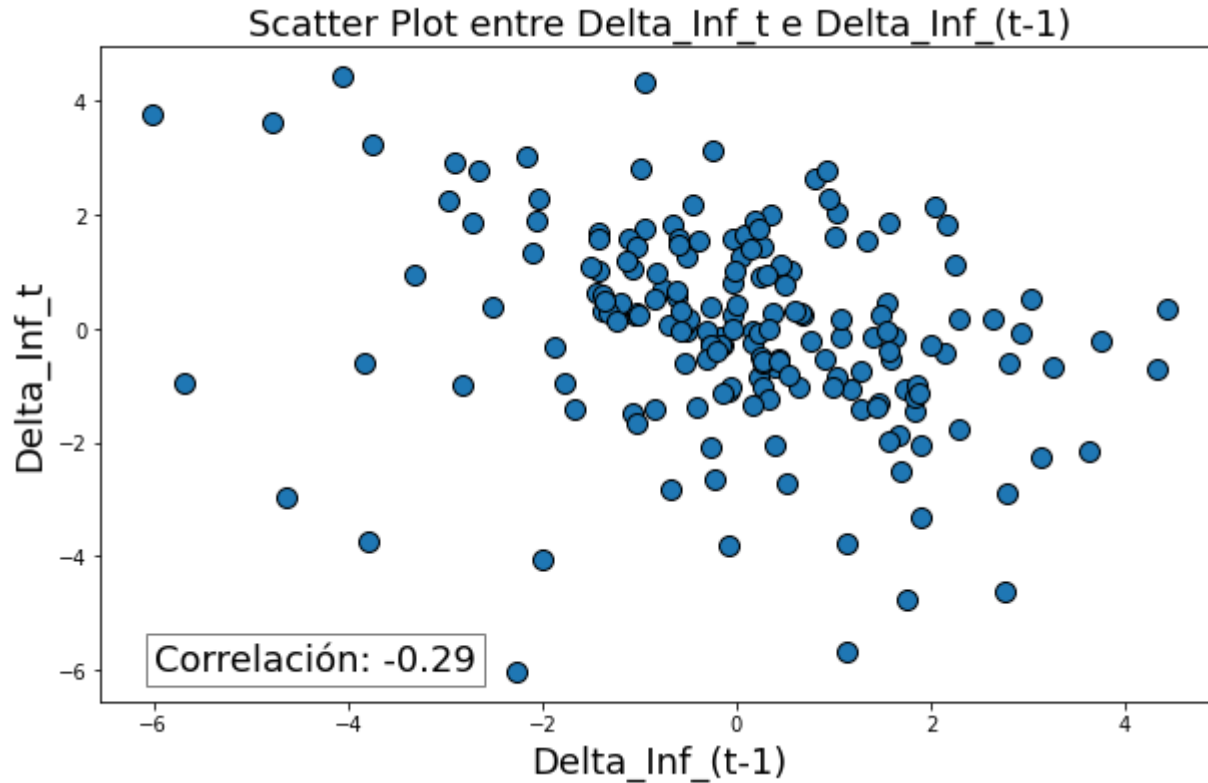


Podemos notar que además de estar altamente correlacionados los valores tienden a ser muy similares, es decir, el término del intercepto de la recta asociada es cercano a cero

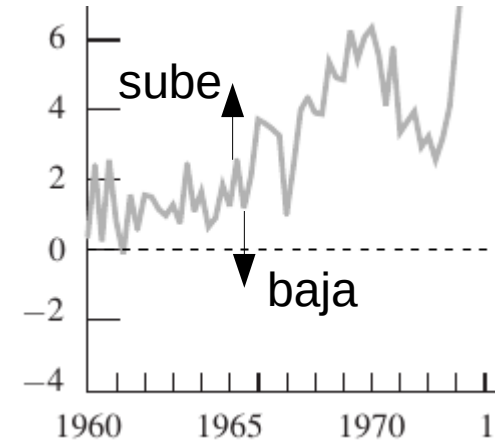


Podemos concluir que en promedio la tendencia del valor de la inflación medido en t es muy similar a su valor medido en $t-1$, excepto por los saltos abruptos que son menos en general

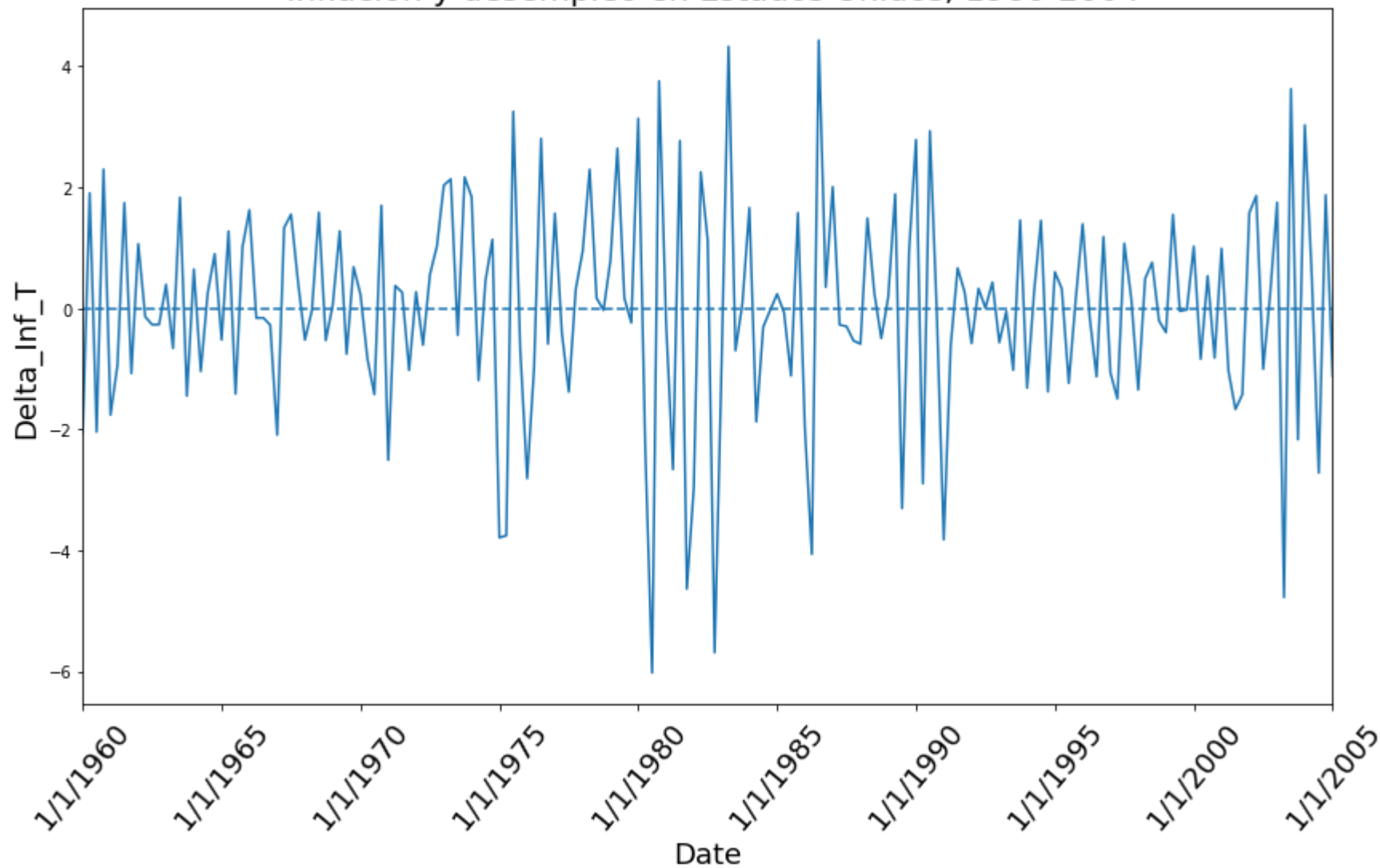
Scatter plot entre Delta_Inf_t e Delta_Inf_t-1



La correlación negativa en la variación de la inflación tiende a cambiar de dirección en tiempos secuenciales



Inflación y desempleo en Estados Unidos, 1960-2004



Modelo autorregresivo de primer orden o AR(1)

El modelo mas simple para intentar predecir el valor futuro de una variable conociendo sus valores pasados es considerar un modelo autorregresivo usando una regresión lineal simple con un retardo de una unidad

$$\hat{Y}_{t,i} = \beta_0 + \beta_1 Y_{t-1,i} + e_i$$

Por ejemplo, podemos considerar como variable de interés para modelar, la tasa de variación de la inflación, es decir nos interesa predecir el cambio en la inflación entre dos trimestres consecutivos, lo cual nos indicaría si los precios subirán o bajaran en el siguiente periodo

Para encontrar el valor de los coeficientes de la regresión lineal usamos nuevamente MCO

$$\widehat{\Delta Inf_t} = 0,017 - 0,238 \Delta Inf_{t-1},$$

(0,126) (0,096)

Donde podemos observar los valores de la dispersión muestral de cada coeficiente calculados siguiendo las expresiones obtenidas en el capítulo de regresión lineal simple

Evaluación de AR(1) y modelos de series temporales en general

En general, los coeficientes del modelo AR(1) son encontrados utilizando MCO sobre un conjunto de datos históricos y la predicción del modelo es evaluada en el momento inmediatamente siguiente

	freq	GDPC1	JAPAN_IP	PCECTPI	CPIAUCSL	Inf_T	Inf_T-1	Delta_Inf_T
194	7/1/2003	13372.357	99.142200	87.769	184.43	2.971541	-0.654415	3.625956
195	10/1/2003	13528.710	102.001053	88.124	184.80	0.802472	2.971541	-2.169068
196	1/1/2004	13606.509	103.271654	88.797	186.57	3.831169	0.802472	3.028696
197	4/1/2004	13706.247	104.789317	89.421	188.60	4.352254	3.831169	0.521085
198	7/1/2004	13830.828	105.636384	89.942	189.37	1.633086	4.352254	-2.719168
199	10/1/2004	13950.376	103.730482	90.652	191.03	3.506363	1.633086	1.873277
200	1/1/2005	14099.081	105.777562	91.122	192.17	2.387060	3.506363	-1.119304
201	4/1/2005	14172.695	105.989329	91.728	193.67	3.122236	2.387060	0.735176
202	7/1/2005	14291.757	105.636384	92.734	196.60	6.051531	3.122236	2.929295

→ Datos para el ajuste MCO

→ Datos para evaluación de la predicción

$$\text{Error en la Prediccion} = Y_{T+1} - \hat{Y}_{T+1|T}$$

Inflación y desempleo en Estados Unidos, 1960-2004

