Modelos VAR

Los modelos VAR (Vector Autoregression) son un tipo de modelo estadístico utilizado para analizar la relación entre múltiples variables económicas o financieras a lo largo del tiempo. La teoría matemática detrás de los modelos VAR se basa en la idea de que cada variable en un sistema afecta y es afectada por las otras variables en el mismo sistema.

En términos matemáticos, un modelo VAR puede escribirse como un sistema de ecuaciones de regresión múltiple, donde cada variable depende de su propio valor rezagado y de los valores rezagados de todas las demás variables en el sistema. La forma general de un modelo VAR es:

donde Y_t es un vector de k variables en el tiempo t, c es un vector de constantes, A_i es una matriz de coeficientes de tamaño kxk y u_t es un vector de errores aleatorios. p es el orden del modelo VAR, que especifica cuántos periodos de tiempo se incluyen en el modelo.

Tenemos un conjunto de ecuaciones

$$b_t = c_{21} a_{t-1} + c_{22} b_{t-1} + \epsilon_{b_1 t}$$

Un ejemplo completo en Python utilizando datos reales podría ser el siguiente:

Supongamos que tenemos un conjunto de datos que consiste en la tasa de interés, la inflación y la producción industrial de los Estados Unidos desde 1980 hasta 2020. Queremos estimar un modelo VAR para analizar la relación entre estas tres variables.

Primero, cargamos los datos en Python utilizando la librería pandas:

```
import pandas as pd

data = pd.read_csv('datos_economicos.csv', index_col=0, parse_dates=True)
```

Luego, creamos el modelo VAR utilizando la librería statsmodels:

```
from statsmodels.tsa.vector_ar.var_model import VAR

model = VAR(data)
```

Podemos seleccionar el orden del modelo utilizando el criterio de información de Akaike (AIC):

```
results = model.fit(maxlags=3, ic='aic')
```

Finalmente, podemos utilizar el modelo para hacer predicciones sobre el comportamiento futuro de las variables:

```
forecast = results.forecast(data.values[-3:], steps=3)
```

Para profundizar en el tema de los modelos VAR, se pueden consultar las siguientes referencias:

- Lütkepohl, H. (2005). New introduction to multiple time series analysis. Springer Science & Business Media.
- Enders, W. (2010). Applied econometric time series. John Wiley & Sons.
- Hamilton, J. D. (1994). Time series analysis (Vol. 2). Princeton: Princeton university press.

▼ Herramientas y referencias

Autocorrelation Function (ACF) vs. Partial Autocorrelation Function (PACF) in Time Series Analysis

0. How to use ACF an PACF to identify time series analysis model tutorial
https://www.youtube.com/watch?v=CAT0Y66nPhs

https://youtu.be/5Q5p6eVM7zM

—

▼ Notas Adicionales

▼ Dickey-Fuller

Este Test asume que tenemos un modelo AR(1)

La prueba de Dickey-Fuller es una prueba estadística utilizada para determinar si una serie de tiempo es estacionaria o no. En otras palabras, esta prueba se utiliza para evaluar si la serie de tiempo tiene una tendencia significativa, lo que significa que su media y varianza cambian con el tiempo.

La prueba de Dickey-Fuller se basa en la idea de que si una serie de tiempo es estacionaria, entonces su autocorrelación disminuirá rápidamente a medida que aumente el desfase. Por lo tanto, la prueba compara la tasa de disminución de la autocorrelación de la serie de tiempo con la tasa esperada de disminución si la serie de tiempo es estacionaria.

En Python, se puede realizar una prueba de Dickey-Fuller utilizando la función `adfuller()` de la biblioteca `statsmodels` de la siguiente manera:

```
from statsmodels.tsa.stattools import adfuller

# Crear una serie de tiempo
serie_tiempo = [0, 1, 2, 3, 4, 5, 6, 7, 8, 9]

# Realizar la prueba de Dickey-Fuller
resultado = adfuller(serie_tiempo)

# Imprimir los resultados de la prueba
print('Estadística de prueba: ', resultado[0])
print('Valor p: ', resultado[1])
print('Valores críticos: ', resultado[4])
```

En este ejemplo, creamos una serie de tiempo simple y luego realizamos la prueba de Dickey-Fuller utilizando la función `adfuller()`. Los resultados de la prueba se imprimen en la consola y consisten en la estadística de prueba, el valor p y los valores críticos. Estos resultados se pueden utilizar para determinar si la serie de tiempo es estacionaria o no.

Referencias:

- Documentación de `statsmodels.tsa.stattools.adfuller()`: **https://www.statsmodels.org/stable/generated/statsmodels.tsa.stattools.adfuller.html**
- Dickey, D. A., & Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. Journal of the American statistical association, 74(366a), 427-431.

Ejemplo 2:

```
import pandas as pd
import numpy as np
from statsmodels.tsa.stattools import adfuller
import matplotlib.pyplot as plt

# Importar los datos de la serie de tiempo
data = pd.read_csv('https://raw.githubusercontent.com/jbrownlee/Datasets/master/airline-passengers.csv')
data.columns = ['Month', 'Passengers']
data['Month'] = pd.to_datetime(data['Month'])
data.set_index('Month', inplace=True)

# Visualizar la serie de tiempo
plt.plot(data)
plt.title('Serie de tiempo de pasajeros de aerolíneas')
plt.xlabel('Año')
plt.ylabel('Número de pasajeros')
plt.show()

# Realizar la prueba de Dickey-Fuller
result = adfuller(data['Passengers'])

# Imprimir los resultados de la prueba
print('Estadística de prueba: ', result[0])
print('Valor p: ', result[1])
print('Valores críticos: ', result[4])
```

En este ejemplo, importamos los datos de una serie de tiempo de pasajeros de aerolíneas y visualizamos la serie de tiempo. Luego, realizamos la prueba de Dickey-Fuller utilizando la función `adfuller()` y imprimimos los resultados de la prueba.

Los resultados de la prueba de Dickey-Fuller son los siguientes:

```
cssCopy code
Estadística de prueba: 0.8153688792060433
Valor p: 0.991880243437641
Valores críticos: {'1%': -3.4816817173418295, '5%': -2.8840418343195267, '10%': -2.578770059171598}
```

Podemos interpretar estos resultados de la siguiente manera:

- La estadística de prueba es 0.815, que es mayor que los valores críticos en los niveles de significancia del 1%, 5% y 10%. Esto sugiere que no podemos rechazar la hipótesis nula de que la serie de tiempo no es estacionaria.

- El valor p es 0.992, lo que indica que no hay suficiente evidencia para rechazar la hipótesis nula de que la serie de tiempo no es estacionaria.
- Los valores críticos son -3.482 (1%), -2.884 (5%) y -2.579 (10%). La estadística de prueba es mayor que estos valores, lo que sugiere que no podemos rechazar la hipótesis nula.

En conclusión, podemos concluir que la serie de tiempo de pasajeros de aerolíneas no es estacionaria según los resultados de la prueba de Dickey-Fuller.

> Conclusion
>
> Si se acepta la hipótesis nula (H0) en la prueba de Dickey-Fuller, significa que no hay suficiente evidencia estadística para rechazar la hipótesis de que la serie de tiempo tiene una raíz unitaria, lo que implica que la serie de tiempo no es estacionaria.
>
> En otras palabras, si se acepta la H0 en la prueba de Dickey-Fuller, significa que la serie de tiempo tiene una tendencia significativa y no es estacionaria. Esto puede ser indicativo de que la serie de tiempo tiene una tendencia al alza o a la baja a lo largo del tiempo, lo que significa que la media y la varianza de la serie de tiempo cambian con el tiempo.
>
> Si se acepta la H0 en la prueba de Dickey-Fuller, puede ser necesario realizar transformaciones en la serie de tiempo, como diferenciación, para convertir la serie de tiempo en una serie de tiempo estacionaria antes de realizar un análisis adicional o modelar la serie de tiempo.

▼ Raices unitarias

Las raíces unitarias son raíces de la ecuación característica de un modelo de series de tiempo que tienen un valor de 1. En otras palabras, una raíz unitaria es una solución de la ecuación característica que hace que el polinomio de la ecuación sea igual a cero cuando se le resta 1 a una de las raíces.

En el contexto de las series de tiempo, las raíces unitarias son importantes porque están estrechamente relacionadas con la estacionariedad de la serie de tiempo. Si una serie de tiempo tiene una o más raíces unitarias, se dice que la serie de tiempo tiene una "raíz unitaria" y que es "no estacionaria".

En particular, si una serie de tiempo tiene una raíz unitaria, entonces su media y su varianza no son constantes en el tiempo y la serie de tiempo es "integrada de orden 1" o I(1). Esto significa que la serie de tiempo puede tener una tendencia o una deriva en el tiempo y es necesario aplicar técnicas de preprocesamiento, como la diferenciación, para convertir la serie de tiempo en una serie estacionaria antes de modelarla.

La presencia de raíces unitarias en una serie de tiempo puede afectar significativamente la precisión de los modelos de predicción y, por lo tanto, es importante identificarlas y tratarlas adecuadamente en el análisis de series de tiempo.

Las raíces unitarias están relacionadas con la presencia de una tendencia o deriva en la serie de tiempo. Las ecuaciones que se utilizan para identificar las raíces unitarias en la serie de tiempo son las siguientes:

1. Modelo AR(1) con una raíz unitaria:
    - $y_t = c + \phi * y_{t-1} + \epsilon_t$
      donde c es una constante, phi es el coeficiente autoregresivo y ϵ_t es el término de error.
2. Modelo AR(2) con dos raíces unitarias:
    - $y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \epsilon_t$
      donde c es una constante, ϕ_1 y ϕ_2 son los coeficientes autoregresivos y ϵ_t es el término de error.
3. Modelo de caminata aleatoria:
    - $y_t = y_{t-1} + \epsilon_t$
      donde y_t es el valor de la serie de tiempo en el tiempo t, y_{t-1} es el valor en el tiempo t-1 y ϵ_t es el término de error.

En estos modelos, la presencia de una raíz unitaria indica que la serie de tiempo no es estacionaria y que es necesario aplicar técnicas de preprocesamiento, como la diferenciación, para convertirla en una serie estacionaria antes de modelarla.

Es importante tener en cuenta que hay otras ecuaciones que se utilizan en el análisis de series de tiempo para identificar diferentes tipos de raíces y patrones de estacionariedad, pero las ecuaciones mencionadas anteriormente son las más comunes para identificar raíces unitarias.

En un modelo ARIMA (Autoregressive Integrated Moving Average), las raíces se encuentran en la parte característica del modelo, que es la parte autoregresiva (AR). Un modelo AR(p) tiene una ecuación característica polinómica de orden p, que se expresa como una combinación lineal de los coeficientes autoregresivos. Las raíces de la ecuación característica son las soluciones de la ecuación polinómica.

Si todas las raíces de la ecuación característica tienen un valor absoluto menor que 1, entonces la serie de tiempo es estacionaria. Si alguna de las raíces tiene un valor absoluto igual o mayor que 1, entonces la serie de tiempo no es estacionaria. En este último caso, la serie de tiempo tiene una tendencia o una dependencia de orden superior, lo que implica que su media y/o varianza cambian con el tiempo.

En resumen, las raíces en las series de tiempo son importantes porque proporcionan información sobre la estacionariedad de la serie de tiempo y permiten determinar el tipo de modelo adecuado para modelar la serie de tiempo.

El resultado de ADF_Statistics en la biblioteca de statsmodels de Python es una medida de la estacionariedad de una serie de tiempo, y su interpretación depende del valor p y de los valores críticos de la distribución de la prueba de Dickey-Fuller. Si se rechaza la hipótesis nula de que la serie de tiempo tiene una raíz unitaria, entonces se puede concluir que la serie de tiempo es estacionaria.

- Si el valor p es menor que el nivel de significancia elegido (usualmente 0.05), entonces se puede rechazar la hipótesis nula de que la serie de tiempo tiene una raíz unitaria y se concluye que la serie de tiempo es estacionaria.
- Si el valor p es mayor que el nivel de significancia elegido, entonces no se puede rechazar la hipótesis nula de que la serie de tiempo tiene una raíz unitaria y se concluye que la serie de tiempo es no estacionaria.
- El valor crítico de la distribución de la prueba de Dickey-Fuller también se puede comparar con el valor de ADF_Statistics. Si el valor de ADF_Statistics es menor que el valor crítico, entonces se puede

rechazar la hipótesis nula de que la serie de tiempo tiene una raíz unitaria y se concluye que la serie de tiempo es estacionaria.