

COYUNTURA Y PREDICCIÓN

Curso 2022-2023

Tema 3 – Tarea 3

A resolver en la EB virtual del lunes 5 de diciembre

Instrucciones:

Los datos del ejercicio se incluyen en el archivo Excel “Tema3_Tarea3.xlsx”, disponible en:
https://github.com/otoperalias/Coyuntura/blob/main/clases/datos/Tema3_Tarea3.xlsx

Parte 1.

-Modelizar la serie “pibv” con el modelo SARIMA. Para analizar cómo de exacta es la capacidad de predecir de la serie, dividir esta en dos partes, una “training dataset” consistente en las primeras 85 observaciones y una “test dataset” con el resto.

-Predecir hacia el futuro 4 periodos (desde el tercer trimestre de 2019 hasta el segundo de 2020, ambos inclusive).

Parte 2.

-Modelizar la serie “pibv” con el modelo SARIMAX. Como variable exógena, de entre los indicadores disponibles en la hoja “otros”, usar aquel que más se correlacione con la serie “pibv”. Las correlaciones calcularlas usando la variación interanual. Para transformar la frecuencia de los datos, podéis usar la función **resample** (ver pág. siguiente).

Para analizar cómo de exacta es la capacidad de predecir de la serie, hacer lo mismo que lo indicado anteriormente en la parte 1.

-Predecir hacia el futuro 1 periodo (es decir, predecir el valor para el tercer trimestre de 2019).

I. Completar la siguiente tabla:

	SARIMA	SARIMAX
Especificación del modelo usado (p,d,q)(P,D,Q)+exog <small>* Para SARIMAX, indicar, además de los parámetros del modelo, la variable exógena usada.</small>		
AIC de dicho modelo		
Número de observaciones usadas para calcular el modelo		
mape entre la serie predicha y la real en el periodo definido por la test dataset		
coef. Correlación entre la serie predicha y la real en el periodo definido por la test dataset		
Predicción correspondiente al tercer trimestre de 2019		

II. Además, debéis preparar los siguientes gráficos:

-Tanto para la parte 1 como la parte 2: Gráficos mostrando la comparación de la predicción del modelo con la serie real durante el periodo definido por la “test dataset”.

-Solo para la parte 1: Un gráfico con la serie real y la predicción hacia el futuro 4 periodos.

IMP: replicar el tipo de gráficos vistos en clase (siguiendo el mismo estilo) y mostrando intervalos de confianza cuando sea oportuno.

Ayuda:

- (1) Para crear la tabla con los datos de los indicadores que se pueden usar como variables exógenas en el modelo SARIMAX, primero hay que importar los datos y luego transformarlos en una frecuencia igual a nuestra serie de pibv.

Notar que debéis transformar la columna de periodo en un formato que Python lo interprete como “DateTime” (o sea, fecha).

Una vez que tenéis el index de la tabla de otros indicadores como DateTime, debéis especificar que se tratan de datos mensuales:

```
otros.index=otros.index.to_period("M")
```

Entonces, para transformar los datos mensuales en trimestrales, usamos el método resample:

```
otrosq=otros.resample("Q", label="right").sum()
```

Y, finalmente, para obtener el mismo formato temporal que la tabla del pib:

```
otrosq.index=otrosq.index.to_timestamp()
```

- (2) Cuando se calculan los coeficientes de correlación, recordar que las series no pueden tener valores nulos (si los tuviesen, hay que usar la función **.dropna()**). Además, tener en cuenta que las series han de tener la misma dimensión o longitud. Si, por ejemplo, la serie del indicador *matric* tiene una observación más que la serie del PIB, tendrías que escribir:

```
np.corrcoef(vpib.dropna(), otros["matric"][:-1].dropna())[0,1]
```

(observar que se añade [0,1] para extraer el elemento [0,1] de la matriz de correlación-autocorrelación. Si no ponemos eso, Python nos da dicha matriz entera)