

GRUPO**ANTÓN MARÍA PAZ****FIGLIOLIA JULIETA ALDANA****PONCIO FEDERICO****SABATER ANNA****WANG JIA QI**

NOTA: Por limitación de espacio las salidas de STATA comentadas se encuentran adjuntas en el archivo logfile y dofile.

PUNTO 1**Ejercicio 1a**

Se construye la variable t , que indica el período mensual en cada observación desde el año 1980 hasta 1988. El intervalo de tiempo de estudio está expresado como 1980m1-1988m12.

Ejercicio 1b

Se estima por MCO el modelo del logaritmo de $uclms$ descompuesto por una tendencia determinística lineal llamada t , las 12 variables dicotómicas indicativas del mes y se excluye la constante para evitar la multicolinealidad.

Se realiza el test de Dickey-Fuller con la activación de la opción *trend* para evaluar la presencia de tendencia determinística o estacionalidad en la serie. La hipótesis nula testea si hay raíz unitaria, es decir, si la serie es no estacionaria.

El test reporta un p-valor de 0.0525 y se concluye que, solo en el caso de un nivel de significancia del 10%, se rechaza la H_0 y la serie del logaritmo de desempleo no presenta raíz unitaria y es estacionaria. Sin embargo, a un nivel de significancia del 5% no se puede rechazar la hipótesis nula y existe evidencia estadística a favor de la presencia de tendencia estocástica en la serie. Además, el p-valor reportado del coeficiente *trend*, que representa la estacionalidad, fue de 0.023 y por lo tanto es estadísticamente significativa la presencia de tendencia determinística en la serie a un nivel del 5%.

En conclusión, la serie de logaritmo de desempleo presenta evidencia empírica a favor de comportamientos tendenciales estocásticos y determinísticos a un nivel de significancia del 5%.

Ejercicio 1c

Se grafica la serie original junto con la predicción de la estimación solo con tendencia lineal llamada $lnuc_trend$ y los residuos de la regresión de las dummies mensuales denominados $lnuc_sa$.

Se observa que la serie del logaritmo de los pedidos de desempleo presenta oscilaciones con relación a la tendencia lineal decreciente a lo largo del tiempo estudiado. Por otro lado, la serie tiene un comportamiento similar al de los residuos de las dummies, pero más errático. Esto podría ser porque los residuos capturan los efectos de estacionalidad mientras que la serie original todavía presenta tanto su tendencia como su ruido blanco. Por eso, el ajuste no es perfecto. De todas formas, se ve una fuerte correspondencia entre la serie y el residuo de las dummies, lo cual exhibe el componente estacional en los pedidos de desempleo solicitados en Indiana entre enero 1980 y diciembre 1988.

Para ayudar a la comparación entre ambas series, se creó el gráfico “series_acercadas” en donde se le restó a la serie original la estimación de la tendencia lineal. Se confirman las observaciones originales que esta nueva serie sigue en general el comportamiento de los residuos de las dummies, pero de forma más errática.

Ejercicio 1d

Se vuelve a estimar el modelo completo, con tendencia, estacionalidad y la variable dicotómica ez , que toma valor 1 en los meses en los que existieron incentivos fiscales para la formación de empresas en Indiana.

De acuerdo a la teoría de Wooldridge, hay que incluir la tendencia lineal en la regresión para destendenciar los datos, así como considerar las dummies estacionales y obtener los residuos para desestacionalizar la serie.

Ya que, si las variables son no estacionarias y tienen tendencia, entonces la predicción puede estar sesgada, ser ineficiente y no será válida la inferencia estadística basada en MCO.

En efecto, **una vez descompuesta la tendencia lineal decreciente y la estacionalidad**, se observa que la variable ez es estadísticamente significativa y su coeficiente es de -0.508. Hay evidencia empírica a favor de que la presencia de incentivos fiscales para la formación de empresas en Indiana reduce el pedido de seguros de desempleo en un $100*[1-\exp(-0.508)]=39.83\%$.

Ejercicio 1e

El supuesto que se debe hacer es la no presencia de otros factores que estén relacionados con los incentivos fiscales (ez) y causen también una reducción en el logaritmo de los pedidos de desempleos. Ya que, si hay otras variables no

observadas que están correlacionadas con la variable dependiente, *luclms*, y el término del error, entonces la estimación por MCO estará sesgada e inconsistente.

PUNTO 2

Preparación de la base

Ejercicio 2a, 2b y 2c

Desarrollados en el do-file.

Análisis de estacionariedad

Ejercicio 2d

Se grafica cada una de las series y se observa que:

- *Priv*: la variable exhibe un fuerte comportamiento tendencial creciente desde enero 2012 hasta agosto 2017. Por lo tanto, la serie no presenta estacionariedad ya que su media no es cero y es ascendente con el correr del tiempo. Además de la presencia de tendencia, se observan unas caídas bruscas en los valores de la media y un cierto patrón de regularidad entre los meses del 2015-2016.
- *Pub*: la variable también muestra un componente tendencial creciente a lo largo del período estudiado. A su vez, se observa la presencia de breaks o quiebre estructural. Uno menos pronunciado al inicio de la serie y otro más significativo al comienzo del año 2015. En consecuencia, la serie tampoco es estacionaria.
- *Dpriv*: la variable diferenciada del sector público exhibe ahora un comportamiento errático y sin componentes claros. A simple vista, se comporta como un ruido blanco. Luego se realizará un test estadístico Dickey-Fuller para verificarlo.
- *Dpub*: la variable diferenciada del sector privado también exhibe ahora un comportamiento errático pero éste es menos que en la variable *Dpriv*. A simple vista, se comporta como un ruido blanco y luego se realizará un test estadístico Dickey-Fuller para comprobarlo.

Ejercicio 2e

La primera diferencia de las series es su variación entre los períodos $t-1$ y t ; es decir, $\Delta Y_t = Y_t - Y_{t-1}$. Para las variables *priv* y *pub* se interpreta como la diferencia entre un mes y el anterior del número de trabajadores del sector privado y público, respectivamente.

Ejercicio 2f

Para cada una de las variables en niveles, se realiza un test Dickey-Fuller aumentado para una raíz unitaria autorregresiva, ya que se realiza un contraste de que la serie sea estacionaria alrededor de una tendencia temporal determinística, es decir, una tendencia que sea una función determinística del tiempo. Por eso, se agrega la opción “*trend*”.

Bajo la hipótesis nula $\delta = 0$ y Y_t es un proceso AR(p) estacionario. Bajo la hipótesis alternativa, $\delta < 0$, por lo que la serie no tiene raíz unitaria pero tiene una tendencia temporal determinística. Se concluye que:

- Tanto para las variables en niveles, *priv* y *pub*, no se rechaza la hipótesis nula de raíz de unitaria, entonces se verifica que ambas tienen tendencia estocástica pero no tendencia determinística dado que *trend* es no significativa al 10%.

Por lo tanto, no se puede hacer estimaciones utilizando las variables en niveles ya que se podría caer en regresiones espurias. A su vez, cabe mencionar que, aunque no se rechace la hipótesis nula de presencia de raíz unitaria, no significa que la serie tenga una raíz unitaria, y aun podría ser razonable aproximar la verdadera raíz autorregresiva como si fuera igual a 1 y por lo tanto utilizar las diferencias de la serie en lugar de sus niveles

Ejercicio 2g

También, para cada una de las variables diferenciadas se realiza un test de raíz unitaria y se concluye que:

- *Dpriv*: Se rechaza la hipótesis nula de raíz unitaria, entonces la variable es una serie estacionaria y tampoco tiene tendencia determinística dado que *trend* no es significativa a un nivel del 5%. Estos resultados estacionarios acerca de la variable son los deseados para poder hacer inferencia válida.
- *Dpub*: Se rechaza la hipótesis nula de raíz unitaria, entonces *dpriv* es una serie estacionaria y no tiene tendencia determinística dado que *trend* es no significativa al 10%. Estos resultados estacionarios también son los deseados.

Modelos univariados: Metodología Box-Jenkins

Ejercicio 2h

Se obtienen los correlogramas hasta 12 lags de las variables. De acuerdo a la teoría, si la serie exhibe un comportamiento autorregresivo, su función AC va a decaer en forma exponencial y esto daría la pauta de modelar un AR. Justamente se observan estas características en los gráficos. Los dos primeros rezagos mensuales de la variable son estadísticamente distintos de cero dado que sobresalen de las bandas de nulidad. En otras palabras, la variable diferenciada del sector privado se acuerda muy bien del mes anterior pero rápidamente se empieza a olvidar del mes más lejano. Por otro

lado, la función parcial de autocorrelación (PAC) expone que sólo el primer rezago mensual es estadísticamente significativo, lo cual sugiere que el orden del modelo autorregresivo sea de uno.

A su vez, no se observa que el comportamiento de las barras de las funciones AC y PAC sean compatibles con un modelo de medias móviles (MA).

Ejercicio 2i

Como se menciona en el inciso anterior, el correlograma muestra que puede ser un modelo ARMA(2,0). Este es el que decidimos estimar en base a nuestra intuición. Nos encontramos con que los residuos son ruido blanco y que el primer rezago de la variable es significativo al 1% pero el segundo no lo es.

Ejercicio 2j

Se estiman todos los procesos ARMA posibles desde ARMA(0,0) hasta ARMA(2,2). Recuerde acortar la muestra de cada corrida para asegurarse que todos los modelos tengan la misma cantidad de observaciones. Usando criterios de información seleccione el mejor modelo dentro de todos los estimados.

El mejor modelo, el más parsimonioso, va a ser aquel que tenga el menor criterio de información. Por lo tanto, al observar la tabla con todos los modelos posibles, vemos que el mejor modelo estimado es el ARMA(1,0). Con un criterio de información AIC de 517,43 este modelo es superior al resto. Cabe mencionar que el modelo intuido en los incisos anteriores tiene un criterio de información muy cercano al del ARMA(1,0) con solo 0,55 centésimas de diferencia.

		Modelos							
		ARMA(1,0)	ARMA(0,1)	ARMA(1,1)	ARMA(1,2)	ARMA(2,1)	ARMA(2,2)	ARMA(2,0)	ARMA(0,2)
AR	L.1	0.487***		0.6405**	0.4828	-0.0768	1.4723***	0.4146***	0.4146***
	L.2					0.3964	-0.6968**	0.1462	0.1462
	L.3								
MA	L.1		0.3405**	-0.20	-0.0494	0.5005	-1.1036**		0.4466**
	L.2				0.1314		0.4883*		0.3125*
	L.3								
Criterios de información									
AIC		517.4299	523.779	518.4378	519.7369	519.1548	520.7073	517.9808	518.1342
BIC		524.044	530.393	527.2566	530.7604	530.1783	533.9354	526.7996	526.9529
Residuos son Ruido Blanco?		SI	SI	SI	SI	SI	SI	SI	SI
* significativo al 0.05									
** significativo al 0.01									
*** significativo al 0.001									

Ejercicio 2k

El mejor modelo identificado en el punto anterior es ARMA(1,0). Las evidencias a favor fueron recolectadas con el comando *varsoc*, la tabla armada en Excel y el reporte de los criterios de información AIC y BIC tras correr el loop *forvalues*. Se observa que ARMA(1,0) es el modelo que cumple con una significatividad al nivel del 1% y el menor criterio de información, es decir, es la estimación más parsimoniosa de todas las propuestas. Este modelo difiere con el ARMA(2,0) estimado en el punto 2i.

La ecuación del proceso generador de datos estimado es la siguiente:

$$dpriv_t = \phi * dpriv_{t-1} + \epsilon_t$$

$$\widehat{dpriv}_t = 0.487 * dpriv_{t-1}$$

Se evalúan los residuos con un test White Noise. No se rechaza la hipótesis nula y se comprueba que los residuos son ruido blanco. Además, el gráfico de los residuos exhibe un comportamiento errático típico de un ruido blanco. Hay evidencia a favor de que el proceso generador de los datos es estacionario ya que se rechaza la hipótesis nula de raíz unitaria del test Dickey-Fuller y se comprueba la estacionariedad de la primer diferencia mensual en la cantidad de trabajadores del sector privado.

Ejercicio 2l

Se grafica el cambio mensual en el empleo privado junto con las predicciones del modelo estimado en el punto 2k. La predicción de ARMA(1,0) no ajusta perfectamente con el comportamiento la variable *dpriv*, de hecho, falla en la precisión del tiempo en que ocurren los cambios en el movimiento de la serie y no ajusta correctamente la magnitud de los picos pronunciados. Una recomendación sería mejorar la estimación y considerar los quiebres estructurales con dummies temporales para capturar esos momentos del tiempo y conocer con exactitud las fechas en que ocurrieron. De todas maneras, luego de haber observado el resto de los gráficos, el modelo ARMA(1,0) es el que mejor predicción hizo entre todos los propuestos.