

Taller I
Pronósticos en Sección Cruzada – Tendencia, Estacionalidad y Ciclos.
2023 - I

Instrucciones para la entrega:

1. El taller debe ser entregado en grupos de máximo 4 personas y mínimo de 3 personas. No se aceptarán trabajos que no sean realizados en grupo, con el número indicado de personas.
2. Los talleres deben ser entregados de **manera ordenada** en un solo archivo PDF. La solución de cada punto debe estar organizada numeral por numeral de menor a mayor. Si lo hacen así, nombren el archivo de PDF -Solución Taller-

También pueden entregar el taller punto por punto en un archivo PDF distinto. La solución de cada punto debe estar organizada numeral por numeral de menor a mayor. Debe haber tantos PDF como puntos del taller y el nombre de cada archivo debe ser -Solucion X- donde X es el número del punto (p.ej. Solución 1, es la solución al punto 1 del taller).

3. Cada pregunta empírica debe estar acompañada por el M-File o Do-File y la base de datos relacionada. En caso que sea necesario revisar la programación de las respuestas dadas en el PDF, **los códigos deben correr** y así se corroborará la validez de sus respuestas.
4. Los M-FILE o Do-File deben tener las secciones y comentarios respectivos donde se describe el paso a paso de lo que realizan. En caso de usar archivos de excel para realizar gráficas o estimaciones secundarias descríbanlo en el M-FILE, Do-File y en el archivo de PDF donde responden formalmente el taller. Todos los archivos deben ser adjuntados.
5. Todos los PDF y M-File - Do-File enviados deben estar marcados al inicio con el nombre y el código de cada uno de los integrantes del grupo.
6. Todos los archivos usados deben ser enviados al correo electrónico szapata@uniandes.edu.co en un archivo comprimido cuyo nombre será el primer apellido de cada uno de los integrantes del grupo a más tardar el **24 de febrero de 2023**.
7. **No seguir las instrucciones y/o no entregar la solución del taller de manera ordenada y comprensible causará que el taller sea calificado sobre 3.**

1. Pronósticos con variables exógenas y simulaciones

Ustedes son unos reconocidos agentes de opinión que están muy preocupados por el efecto que puede tener en la producción de petróleo el hecho de parar la exploración para encontrar este insumo en el país. Dado que su conocimiento en el tema no es tan profundo inician la tarea de entenderlo mejor, para que su opinión sea lo más informada posible.

Para empezar su análisis deciden visitar la página de la oficina de administración de información de energía (EIA) y elaboran una base de datos anual desde el año 1998, que contiene información para 28 países con la producción de petróleo promedio diaria en miles de barriles, las reservas probadas anuales en miles de millones de barriles, el PIB anual en dólares (PPP) y el precio promedio del

petróleo tipo brent el cual proviene de las bases de datos de commodities del FMI – (esta información se encuentra en el archivo de Excel datos).

Al realizar una revisión de literatura se percatan que teóricamente la producción de petróleo es una función de la cantidad de pozos explorados, la probabilidad de que estos pozos tengan petróleo y su cantidad de reservas existentes. Para iniciar con la estimación y entender mejor cómo podría afectar la no exploración de petróleo la producción de un país, ustedes deciden usar un modelo de proyección lineal, de la siguiente forma:

$$Y_{it} = \alpha_0 + \sum_{i=1}^I \beta_i x_{it} + \varepsilon_{it}$$

Donde Y_{it} es el logaritmo de la producción para cada país i en cada punto del tiempo t , x_{it} representa cada una de las variables exógenas de la muestra para cada país i , salvo que se defina lo contrario (en dónde inequívocamente estarán las reservas), α_0 es la constante de la estimación, β_i el parámetro que define cómo cada variable x_{it} afecta la proyección de la producción de petróleo y ε_{it} es el error de pronóstico de su estimación.

Para seleccionar cuál es el mejor modelo de proyección lineal siguen los siguientes pasos, entre los cuales se encuentra recordar algunos aspectos teóricos básicos:

1.1. Análisis Preliminares (10 Puntos)

1. ¿Cuál es el pronóstico óptimo para la producción de petróleo en un año dado? Discutan los supuestos necesarios para que el pronóstico propuesto sea, de hecho, óptimo¹.
2. Conviertan la producción, las reservas, el PIB y el precio del petróleo a términos logarítmicos. De igual forma, conviertan la variable países en una variable categórica.
3. Usando las observaciones del logaritmo de la producción, el logaritmo de las reservas, el logaritmo del PIB y el logaritmo del precio del petróleo, haga un correlograma de estas variables²:
 - a. ¿De acuerdo con el correlograma cuál es la variable que ustedes esperan afecte en mayor medida el logaritmo de la producción de petróleo?
 - b. ¿A partir del correlograma cómo esperan que sean las relaciones (correlaciones parciales) del logaritmo de las reservas, el PIB y los precios del petróleo con la producción de petróleo?
 - c. ¿A partir del correlograma consideran que la producción del petróleo ha sido importante para el PIB de estos países?

¹ Para este punto discutan los supuestos definidos cuando se realiza la estimación con datos de sección cruzada (i.e. un punto en el tiempo con muchos datos i).

² Utilizando el comando `removevars` debe eliminar todas las variables categóricas y las fechas, una vez hayan hecho eso pueden usar el comando `corrplot()` para estimar el correlograma.

4. Haga dos diagramas de dispersión. En el primero, usarán en el eje X el logaritmo de las reservas y en el eje Y el logaritmo de la producción; en el segundo, usarán en el eje X el logaritmo del PIB y en el eje Y el logaritmo de la producción:
 - a. ¿Se observa evidencia de alguna relación no lineal (i.e. cuadrática, cubica, etc) entre alguna de las variables X y la variable Y, si es así para cuál o cuáles?
5. Grafiquen un boxplot utilizando como variable cualitativa País y cuantitativa el logaritmo de la producción. Reporten la gráfica y expliquen si es relevante incluir los la variable cualitativa en una regresión que explique el comportamiento del logaritmo de la producción.
6. Estimen los siguientes modelos por MCO:

$$Y_{it} = \alpha_0 + \beta_1 R_{it-1} + \beta_2 R_{it-1}^2 + \beta_3 PIB_{it} + \beta_4 OIL_{it} + \varepsilon_{it} \quad [1]$$

Y_{it} : Es el logaritmo de la producción para cada país i, en cada periodo t.

R_{it-1} : Es el logaritmo de las reservas de petróleo para cada país i, en cada periodo t-1.

PIB_{it} : Es el logaritmo del PIB para cada país i, en cada periodo t.

OIL_{it} : Es el logaritmo del precio del petróleo, en cada periodo t.

$$Y_{it} = \alpha_0 + \beta_1 R_{it-1} + \beta_2 R_{it-1}^2 + \beta_3 PIB_{it} + \beta_4 OIL_{it} + \sum_{j=1}^{27} \gamma_j g_j + \varepsilon_{it} \quad [2]$$

Donde:

Y_{it} : Es el logaritmo de la producción para cada país i, en cada periodo t.

R_{it-1} : Es el logaritmo de las reservas de petróleo para cada país i, en cada periodo t-1.

PIB_{it} : Es el logaritmo del PIB para cada país i, en cada periodo t.

OIL_{it} : Es el logaritmo del precio del petróleo, en cada periodo t.

g_i : Es la variable categórica de cada país, que incluye 27 países para eliminar cualquier posible problema de multicolinealidad con la constante.

Una vez los hayan estimado, reporten el R^2 , el R^2 ajustado y el error cuadrático medio e interprétenlos desde una perspectiva de pronósticos. ¿Cuál modelo es mejor, el modelo 1 o el modelo 2? ¿Incluir la variable categórica de país mejoraría la proyección o lo empeoraría? ¿Dada la interpretación del R^2 de los modelos (1) y (2) que ustedes seleccionaron, creen que la selección de variables a partir de correlaciones para hacer pronósticos es suficiente?

Adicionalmente, comenten si los parámetros estimados tienen los signos esperados. ¿Si alguno no tiene el signo esperado, esa variable no es informativa para hacer proyecciones?

7. A partir de las estimaciones del numeral 6, estimen con el modelo [2] la **producción** de petróleo dentro de muestra de todos los países en el año 2021. Con la gráfica de su predilección reporten las estimaciones y comenten los resultados.

1.2. Estimación de Intervalos y Densidades (30 Puntos)

Supongamos que después de la inspección previa ustedes deciden que el mejor modelo es [2] y con este evaluarán el efecto de la disminución de las reservas sobre la producción de **Colombia**. Para esto deciden usar los siguientes valores de las variables exógenas:

$$PIB_{it} = LN(710,2)$$

$$OIL_{it} = LN(70,8)$$

Que son los valores de la muestra del PIB en Colombia y el precio del petróleo en el año 2021. Con respecto a las reservas (R_{it-1}) ustedes no saben cuánto podrían caer las reservas si disminuye la exploración de petróleo. Para informar este valor ustedes lo hacen a partir de los hallazgos descritos en el documento de James Cust y Torfinn Handing (Institutions and the Location of Oil Exploration) publicado en 2019, en donde los autores estiman que países con instituciones más proclives a explorar tienen en promedio 55% más pozos exploratorios³. Dado lo anterior, y para tener una orden de magnitud, deciden utilizar como medida de reservas:

$$R_{it} = LN(0,92)$$

Que es el 45% del total de las reservas observadas en Colombia en 2021.

Con esta información ustedes realizan la proyección del total de **producción** para hacerse una idea de cómo podría caer la producción ante una disminución de las reservas de 55%. Para esto elaboran los siguientes ejercicios:

1. Suponen que $\varepsilon_{it} \sim N(0, \sigma^2)$. Sin hacer simulaciones, calculen un intervalo de confianza de 95% de confianza para su pronóstico. Reporten pronóstico puntual e intervalo inferior y superior de pronóstico.
2. Siguen suponiendo que $\varepsilon_{it} \sim N(0, \sigma^2)$. Ahora construyan un intervalo de confianza de 95% de confianza para su pronóstico usando $R = 50$, $R = 500$, $R = 10,000$. Comparen los intervalos con los del intervalo teórico asumido en 1.
3. Grafiquen las simulaciones del punto 2, incluyendo punto, intervalo y densidad (usen el comando subplot de MATLAB para incluir varios gráficos en uno).
4. Ahora supongan que no conoce la distribución de ε_{it} , construyan un intervalo de confianza de 95% para su pronóstico asumiendo $R=10.000$.
5. Grafiquen las simulaciones del punto 4, incluyendo punto, intervalo y densidad.
6. Ahora, a la estimación realizada en el punto 4 agréguele incertidumbre en los parámetros⁴. Construyan un intervalo de 95% de confianza para su pronóstico asumiendo $R=10.000$ tanto para las repeticiones de ε como para las simulaciones de los parámetros.

³ De acuerdo con los autores este porcentaje corresponde al número de pozos adicionales que tienen los países más proclives a explorar, controlando por calidad institucional y diferencias culturales alrededor del mundo.

⁴ Asuma normalidad de los parámetros.

7. Grafiquen las simulaciones del punto 6, incluyendo punto, intervalo y densidad.
8. Finalmente, suponen que $\varepsilon_{it} \sim N(0, \sigma_i^5)$ -errores heteroscedasticos-, pero asuma **total certidumbre** en los parámetros. Construyan un intervalo de 95% de confianza para su pronóstico asumiendo $R=10.000$ tanto para las simulaciones de épsilon como para las simulaciones de los parámetros.
9. Grafiquen las simulaciones del punto 8, incluyendo punto, intervalo y densidad. ¿Se generó algún cambio con respecto a las estimaciones previas?
10. Teniendo en cuenta sus respuestas de todo este ejercicio, analicen sus resultados, concluyan y discutan qué tipo de información adicional necesitarían para estimar los efectos de la eliminación de la exploración de petróleo sobre la producción de petróleo en Colombia.

2. Tendencia, Estacionalidad y Ciclos

2.1. Análisis de propiedades modelos AR, MA y ARMA (30 Puntos)

1. Se tiene una ecuación en diferencias de primer orden definida por la siguiente ecuación:

$$y_t = \theta y_{t-1} + w_t$$

Donde supondremos que w_t es una variable conocida, y por el momento no importan los supuestos que hagamos sobre su distribución.

Resuelvan:

- 1.a. Utilizando sustitución recursiva muestre que y_t es una función del valor y_{-1} , y la historia de w entre el periodo cero y el periodo t .

- 1.b. calculen $\left(\frac{\partial y_t}{\partial w_0}\right)$.

- 1.c. Repita el punto 1.a. pero suponiendo que el punto de partida no es cero sino el periodo t y se repite j pasos adelante. Dado esto, muestre que y_{t+j} es una función del valor y_{t-1} , y la historia de w entre el periodo t y el periodo $t+j$. Calculen $\left(\frac{\partial y_{t+j}}{\partial w_t}\right)$ y reporten si existe alguna diferencia entre el resultado del punto 1b y el resultado obtenido acá.

- 1.d. Suponga que usted tiene un modelo de demanda de dinero, donde se relaciona el logaritmo de las tenencias reales de dinero (m_t) con el logaritmo del agregado total del ingreso real (I_t), el logaritmo de las tasas de interés en las cuentas bancarias (r_{bt}) y el logaritmo de las tasas de interés de los préstamos bancarios (r_{ct}).

$$m_t = 0,50 + 0,65m_{t-1} + 0,14I_t - 0,023r_{bt} - 0,023r_{bt} - 0,045r_{ct}$$

A partir de la anterior ecuación estimen qué sucedería con la demanda de dinero dos meses adelante si el ingreso real de la economía se incrementa en 1% en el periodo t [i.e $\left(\frac{\partial m_{t+2}}{\partial I_t}\right)$].

⁵ Asuma normalidad de los parámetros.

2. Para entender un proceso de media móvil, supongan que ustedes ganan 1 USD si al lanzar una moneda cae cara y pierden 1 USD si cae sello. Supongamos que en la variable ε_t se reportan los resultados de las apuestas de cada lanzamiento en cada punto del tiempo t (i.e. para cada apuesta en el punto del tiempo t , ε_t es +1 USD ó -1 USD). De esta forma, para cada lanzamiento de moneda en t , su ganancia media de los últimos cinco lanzamientos es:

$$w_t = 1/5\varepsilon_t + 1/5\varepsilon_{t-1} + 1/5\varepsilon_{t-2} + 1/5\varepsilon_{t-3} + 1/5\varepsilon_{t-4}$$

Dado w_t y suponiendo que ε_t es ruido blanco:

- Encuentre el valor esperado de sus ganancias. Encuentra el valor esperado dado que $\varepsilon_{t-3} = \varepsilon_{t-2} = \varepsilon_{t-4} = 1$.
 - Encuentre la varianza de w_t . Encuentre la varianza de w_t dado que $\varepsilon_{t-3} = \varepsilon_{t-2} = \varepsilon_{t-4} = 1$.
 - Encuentre $\text{cov}(w_t, w_{t-1})$, $\text{cov}(w_t, w_{t-2})$, $\text{cov}(w_t, w_{t-5})$.
3. Consideren el proceso $x_t = \phi x_{t-1} + \varepsilon_t$, donde ε_t es ruido blanco y $|\phi| < 1$. A partir de esta información, demuestren formalmente y expliquen la importancia de tener un número de datos que tienda a infinito. Comprueben las implicaciones de esto en términos de su media no condicional, su varianza no condicional y su auto covarianza no condicional.
4. Se tiene un proceso MA(q) definido por:

$$y_t = \alpha_0 + \varepsilon_t + \gamma_1\varepsilon_{t-1} + \gamma_2\varepsilon_{t-2} + \gamma_3\varepsilon_{t-3} + \dots + \gamma_q\varepsilon_{t-q}$$

Donde ε_t es ruido blanco e independiente. Estimen la media no condicional, la varianza no condicional y las auto covarianzas desde cero hasta el rezago q . Si el proceso es un MA(2) muestren cómo serían las auto covarianzas desde el rezago cero hasta el rezago q .

2.2. Estimación AR, MA y ARMA (30 Puntos)

1. En el archivo de excel denominado IPC se encuentra la serie mensual del IPC de Colombia desde enero de 2000⁶. Por el momento, suponga que esta variable sigue un proceso AR(2): $y_t = c + \vartheta_1 y_{t-1} + \vartheta_2 y_{t-2} + \varepsilon_t$ donde ε_t es un proceso ruido blanco.

1.1. Para la serie del IPC(SA) realicen los siguientes procedimientos:

- Conviertan la serie del IPC(SA) a logaritmo natural, estime su primera diferencia y describa sus principales características (asegúrense de multiplicar por 100 este valor para interpretar todos los resultados como variación porcentual mensual)⁷.

⁶ La serie IPC proviene del Banco de la República, la serie IPC (SA) es el IPC ajustado por estacionalidad, este último es una estimación realizada con X-13ARIMA-SEATS.

⁷ Recuerde que ante variaciones muy pequeñas de una serie la diferencia del logaritmo natural es aproximadamente igual a la variación porcentual.

- b) Estimen y grafiquen las 20 primeras autocorrelaciones de la primera diferencia logarítmica del IPC e Interpreten los resultados.
 - c) Estimen y grafiquen las 20 primeras autocorrelaciones parciales de la primera diferencia logarítmica del IPC e interpreten los resultados.
 - d) Estimen los parámetros c y θ por MCO, tomando como variable endógena la primera diferencia del logaritmo natural del IPC.
 - e) Realicen el pronóstico condicional 12 periodos delante de la variación mensual del IPC con el modelo AR(2) obtenido por medio de MCO. Transformen la serie para obtener el nivel proyectado en del IPC en diciembre de 2023 y estimen la inflación anual de diciembre de 2023 vs. diciembre de 2022 según este modelo. ¿Creen que la inflación al finalizar el año será la que les está arrojando este modelo?⁸ Comenten sus resultados.
2. Ahora, usando la serie de la primera diferencia del logaritmo del IPC(SA), suponga que esta serie sigue un proceso ARMA:

$$y_t = \frac{\Theta(L)}{\Phi(L)} \varepsilon_t$$

- a) Estime el componente cíclico y justifique su estimación con estadísticas y criterios de información del modelo estimado.
- b) Construya y grafique la serie de residuales. Justifique que se trata de ruido blanco.
- c) Documente pruebas de normalidad y aquellas relevantes para justificar los supuestos de su modelo.
- d) Construya una gráfica en la que el eje x serán todos los meses del año 2021 y 2022, en donde se presente la primera diferencia del logaritmo natural del IPC(SA) por cada mes observado y su pronóstico desde enero de 2023 hasta diciembre de 2023⁹.
- e) Con sus estimaciones construya el pronóstico de la variación anual del IPC para el año 2023¹⁰. ¿Cuál es su pronóstico del IPC para el año 2023? ¿El DANE acaba de publicar el IPC de enero de 2023, sus resultados son aceptables o distan mucho de lo observado? ¿En este momento, para el año 2023 el Ministerio de Hacienda proyecta una inflación al cerrar el año de 7,2% y el Banco de la República 8,7%, al comparar estos valores con su proyección de la variación anual consideran que estas instituciones están siendo muy pesimistas, muy optimistas o realistas? ¿comente sobre los límites superior e inferior de la inflación anual para todos los meses del año 2023 estimados por ustedes?

Nota: El grupo que tenga el menor error de pronóstico en la inflación anual de febrero que será reportada por el DANE el próximo mes de marzo de 2023, obtendrá un bono en el primer parcial. Por esto, les pedimos el favor de reportar la proyección de sus inflaciones anuales de cada mes de 2023 a 3 decimales en una tabla al final de la solución del taller.

⁸ Para esto pueden usar como ayuda el archivo de Excel que se llama archivo auxiliar – solo debe incluir sus pronósticos mensuales en la tabla indicada. No olviden incluirlo en los archivos que nos enviarían para la calificación del taller (acá no deben incluir límites).

⁹ En este punto pueden usar los comandos de Matlab o el programa de su preferencia para realizar los pronósticos.

¹⁰ Para esto use como ayuda el archivo de Excel que se llama archivo auxiliar – solo debe incluir sus pronósticos mensuales en la tabla indicada. No olviden incluirlo en los archivos que nos enviarían para la calificación del taller.