Profesor	: Eduardo Engel	Abril 4, 2023
Ayudantes	: Miguel Del Valle y Benjamín Peña	
Curso	: ENECO 630 (Macroeconomía I)	
Semestre	: Otoño 2023	
Guía	: No. 1	
Entrega	: Lunes 10 de abril, antes de las 8am	

1. **Procesos estacionarios con muchos ceros**

   Los ε_t son i.i.d. $N(0,1)$ y $x_t = \varepsilon_t$ y $y_t = (-1)^t \varepsilon_t$. También tenemos una secuencia ξ_t de variables aleatorias independientes que toman los valores 0 y 1 con probabilidad 1/2. Los ξ también son independientes de los ε.

   (a) Muestre que los procesos x e y son débilmente estacionarios (y por lo tanto fuertemente estacionarios ya que son Gausianos).

   (b) Muestre que $z_t \equiv x_t + y_t$ *no* es débilmente estacionario.

   (c) Muestre que $u_t \equiv \xi_t \varepsilon_t$ es débilmente estacionario.

   (d) Las trayectorias de z y u tienen en común que aproximadamente la mitad de los valores son cero, sin embargo, u es stationary y z no. Explique en qué radica la diferencia fundamental entre estos procesos.

   Respuesta:

   (a) Para mostrar que los procesos x e y son débilmente estacionarios se debe mostrar que sus primeros dos momentos son finitos y que no dependen del tiempo. Así, tendremos que para el proceso x:

   $$\begin{aligned} \mathrm{E}x_t &= \mathrm{E}\varepsilon_t = 0 \\ \mathrm{E}x_t^2 &= \mathrm{E}\varepsilon_t^2 = 1 \\ \mathrm{E}x_t x_{t+j} &= \mathrm{E}\varepsilon_t \varepsilon_{t+j} = 0, \quad \forall j = 0,1,2,\ldots \end{aligned}$$

   Por lo que el proceso x es débilmente estacionario. Por otra parte, para el proceso y:

   $$\begin{aligned} \mathrm{E}y_t &= \mathrm{E}(-1)^t \varepsilon_t = (-1)^t \mathrm{E}\varepsilon_t = 0 \\ \mathrm{E}y_t^2 &= \mathrm{E}(-1)^{2t}\varepsilon_t^2 = \mathrm{E}\varepsilon_t^2 = 1 \\ \mathrm{E}y_t y_{t+j} &= \mathrm{E}(-1)^t \varepsilon_t (-1)^{t+j} \varepsilon_{t+j} = (-1)^{2t+j} \mathrm{E}\varepsilon_t \varepsilon_{t+j} = 0, \quad \forall j = 0,1,2,\ldots \end{aligned}$$

   Por lo que el proceso y también es débilmente estacionario.

   (b) Repetimos el proceso anterior, ahora para z.

   $$\begin{aligned} \mathrm{E}z_t &= \mathrm{E}(x_t + y_t) = \mathrm{E}x_t + \mathrm{E}y_t = 0 \\ \mathrm{E}z_t^2 &= \mathrm{E}(x_t + y_t)^2 = \mathrm{E}(x_t^2 + 2x_t y_t + y_t^2) = 1 + 2Ex_t y_t + 1 = 2(1 + (-1)^t) \\ \mathrm{E}z_t z_{t+j} &= \mathrm{E}(x_t + y_t)(x_{t+j} + y_{t+j}) = \mathrm{E}(x_t x_{t+j} + y_{t+j} x_t + x_{t+j} y_t + y_t y_{t+j}) = 0 \end{aligned}$$

Como podemos notar, los momentos son finitos pero se tiene el problema de que la varianza de la variable depende del tiempo, por lo que el proceso no es débilmente estacionario.

(c) Para el caso de u tenemos que:

$$\begin{aligned}
\mathrm{E}u_t &= \mathrm{E}(\xi_t \varepsilon_t) = \mathrm{E}\xi_t \mathrm{E}\varepsilon_t = 0 \\
\mathrm{E}u_t^2 &= \mathrm{E}(\xi_t \varepsilon_t)^2 = \mathrm{E}\xi_t^2 \mathrm{E}\varepsilon_t^2 = \frac{1}{2} \cdot 1 = \frac{1}{2} \\
\mathrm{E}u_t u_{t+j} &= \mathrm{E}(\xi_t \varepsilon_t)(\xi_{t+j}\varepsilon_{t+j}) = \mathrm{E}(\xi_t \xi_{t+j})\mathrm{E}(\varepsilon_t \varepsilon_{t+j}) = 0
\end{aligned}$$

Por lo que vemos que el proceso u es débilmente estacionario.

(d) La principal diferencia en ambas series es que en el proceso z, los valores que se hacen cero tienen una estructura temporal, por lo que sabemos que la serie se hará 0 para todos los valores impares de t. Por otra parte, la serie u también contará con apróximadamente la mitad de valores con ceros, pero la ubicación temporal de estos valores es aleatoria y correspondiente a una variable aleatoria Bernoulli.

Por lo tanto, esta estructura temporal hace que si yo tomo una muestra de la serie u, no puedo hacer inferencia sobre el proceso completo, debido a que las características de la serie dependerán de la muestra que extraiga. Por ejemplo, no es lo mismo sacar una muestra de 7 datos en que existan 4 valores con t par y 3 valores con t impar y viceversa. Este problema no se genera en la serie z, en la cual los momentos están bien definidos a través del tiempo.

2. **Predicciones para un proceso AR(2)**

   Una serie económica trimestral, x_t, sigue un proceso AR(2) estacionario:

   $$x_t = a_1 x_{t-1} + a_2 x_{t-2} + \varepsilon_t.$$

   Donde los ε son un ruido blanco de media nula y varianza σ^2. Las preguntas que siguen son motivadas por el interés de predecir la seria a seis meses plazo, es decir, a dos trimestres.

   (a) Determine $\mathrm{E}[x_{t+1}|x_t, x_{t-1}, ...]$.
   (b) Determine $\mathrm{E}[x_{t+2}|x_t, x_{t-1}, ...]$.
   (c) Denote mediante
   $$e_{2,t} = x_{t+2} - \mathrm{E}[x_{t+2}|x_t, x_{t-1}, ...]$$
   el error de predicción a dos períodos. Muestre que $e_{2,t}$ sigue un proceso MA(q) con q finito. Determine el valor de q.
   (d) Para mostrar que el resultado anterior es general, considere un proceso estacionario x_t que se puede escribir como una media movil infinita:
   $$x_t = \sum_{j=0}^{\infty} b_j \varepsilon_{t-j},$$
   donde $\sum b_j^2 < \infty$. Repita las partes *a)*, *b)* y *c)* concluya que el error de predicción a dos pasos sigue un MA(q) con el mismo valor de q que obtuvo para el AR(2).

   Respuesta:

   a)
   $$\begin{aligned} \mathrm{E}[x_{t+1}|x_t, x_{t-1}, ...] &= \mathrm{E}[a_1 x_t + a_2 x_{t-1} + \varepsilon_{t+1}|x_t, x_{t-1}, ...] \\ &= a_1 x_t + a_2 x_{t-1} \end{aligned}$$

   b)
   $$\begin{aligned} \mathrm{E}[x_{t+2}|x_t, x_{t-1}, ...] &= \mathrm{E}[a_1 x_{t+1} + a_2 x_t + \varepsilon_{t+2}|x_t, x_{t-1}, ...] \\ &= a_1 E(x_{t+1}) + a_2 x_t \\ &= a_1(a_1 x_t + a_2 x_{t-1}) + a_2 x_t \\ &= (a_1^2 + a_2) x_t + a_1 a_2 x_{t-1} \end{aligned}$$

   c)
   $$\begin{aligned} e_{2,t} &= x_{t+2} - \mathrm{E}[x_{t+2}|x_t, x_{t-1}, ...] \\ &= a_1 x_{t+1} + a_2 x_t + \varepsilon_{t+2} - \mathrm{E}[a_1 x_{t+1} + a_2 x_t + \varepsilon_{t+2}|x_t, x_{t-1}, ...] \\ &= a_1 x_{t+1} + a_2 x_t + \varepsilon_{t+2} - (a_1^2 + a_2) x_t - a_1 a_2 x_{t-1} \\ &= a_1(a_1 x_t + a_2 x_{t-1} + \varepsilon_{t+1}) + a_2 x_t + \varepsilon_{t+2} - (a_1^2 + a_2) x_t - a_1 a_2 x_{t-1} \\ &= a_1 \varepsilon_{t+1} + \varepsilon_{t+2} \end{aligned}$$

   Lo que corresponde a un MA(q), con $q = 1$.

d) Partimos por calcular:

$$\begin{aligned}
\mathrm{E}[x_{t+1}|x_t, x_{t-1}, ...] &= \mathrm{E}[\sum_{j=0}^{\infty} b_j \varepsilon_{t+1-j}|x_t, x_{t-1}, ...] \\
&= \sum_{j=1}^{\infty} b_j \varepsilon_{t+1-j}
\end{aligned}$$

$$\begin{aligned}
\mathrm{E}[x_{t+2}|x_t, x_{t-1}, ...] &= \mathrm{E}[\sum_{j=0}^{\infty} b_j \varepsilon_{t+2-j}|x_t, x_{t-1}, ...] \\
&= \sum_{j=2}^{\infty} b_j \varepsilon_{t+2-j}
\end{aligned}$$

Estos dos resultados son dado que sabemos la información hasta t, por lo solo los ruidos blancos después de t tienen esperanza nula.

$$\begin{aligned}
e_{2,t} &= x_{t+2} - \mathrm{E}[x_{t+2}|x_t, x_{t-1}, ...] \\
&= \sum_{j=0}^{\infty} b_j \varepsilon_{t+2-j} - \sum_{j=2}^{\infty} b_j \varepsilon_{t+2-j} \\
&= b_0 \varepsilon_{t+2} + b_1 \varepsilon_{t+1}
\end{aligned}$$

Lo que corresponde a un MA(q) con $q = 1$, al igual que el caso anterior de un AR(2). Esto muestra la generalidad del resultado del error de predicción a dos períodos.

3. **Transmisión de innovaciones correlacionadas**

x_t y y_t siguen procesos estacionarios AR(1) con innovaciones ε_t^x y ε_t^y:

$$\begin{aligned} x_t &= a_x x_{t-1} + \varepsilon_t^x, \qquad (1) \\ y_t &= a_y y_{t-1} + \varepsilon_t^y. \qquad (2) \end{aligned}$$

Las varianzas de ε_t^x y ε_t^y se denotan $\sigma^2_{\varepsilon,x}$ y $\sigma^2_{\varepsilon,y}$, respectivamente, y la correlación entre las innovaciones *contemporaneas* es $\rho > 0$ mientras que las correlaciones de innovaciones no contemporaneas es cero:

$$\rho(\varepsilon_s^x, \varepsilon_t^y) = \begin{cases} \rho & \text{si } s = t, \\ 0 & \text{si } s \neq t. \end{cases}$$

Encuentre la correlación entre x_t y y_t y muestre que es menor o igual que ρ, con igualdad si y solo si $a_x = a_y$. Discuta la intuición del resultado.

Indicación: Recuerde que la correlación entre dos variables aleatorias V y W se define como

$$\rho(V, W) \equiv \frac{\text{Cov}(V, W)}{\sigma(V)\sigma(W)}, \qquad (3)$$

donde $\text{Cov}(V, W)$ denota la covarianza de V y W y $\sigma(V)$ y $\sigma(W)$ las desviaciones estándar correspondientes.

Respuesta: Siguiendo la indicación, calcularemos la Covarianza y las desviaciones estándar de las series x e y. Así, partiendo por las desviaciones estándar tenemos que:

$$Var(x_t) = a_x^2 Var(x_{t-1}) + Var(\varepsilon_t^x)$$

Donde utilizamos el hecho de que $Cov(x_t - 1, \varepsilon_t^x) = 0$. Ahora, como las series son estacionarias sabemos que $Var(x_t) = Var(x_{t-1})$. Por lo tanto:

$$\begin{aligned} Var(x_t) &= \frac{\sigma^2_{\varepsilon,x}}{1 - a_x^2} \\ \sigma(x_t) &= \frac{\sigma_{\varepsilon,x}}{\sqrt{1 - a_x^2}} \end{aligned}$$

Por simetría, la desviación estándar para y vendrá dada por:

$$\sigma(y_t) = \frac{\sigma_{\varepsilon,y}}{\sqrt{1 - a_y^2}}$$

Por la parte de la Covarianza tenemos que:

$$\begin{aligned} Cov(x_t, y_t) &= Cov(a_x x_{t-1} + \varepsilon_t^x, a_y y_{t-1} + \varepsilon_t^y) \\ &= a_x a_y Cov(x_{t-1}, y_{t-1}) + Cov(\varepsilon_t^x, \varepsilon_t^y) + a_x Cov(x_{t-1}, \varepsilon_t^y) + a_y Cov(\varepsilon_t^x, y_{t-1}) \end{aligned}$$

Como no tenemos correlación entre períodos, las últimas dos covarianzas serán iguales a 0. Por su parte, como tanto x como y son estacionarias, tendremos que $Cov(x_t, y_t) = Cov(x_{t-1}, y_{t-1})$. Por lo tanto:

$$Cov(x_t, y_t) = \frac{\rho\sigma_{\varepsilon,x}\sigma_{\varepsilon,y}}{1 - a_x a_y}$$

Así, tendremos que:

$$\begin{aligned}
Corr(x_t, y_t) &= \frac{Cov(x_t, y_t)}{\sigma(x_t)\sigma(y_t)} \\
&= \frac{\rho\sigma_{\varepsilon,x}\sigma_{\varepsilon,y}}{1 - a_x a_y} \frac{\sqrt{1 - a_x^2}\sqrt{1 - a_y^2}}{\sigma_{\varepsilon,x}\sigma_{\varepsilon,y}} \\
&= \rho \left(\frac{\sqrt{(1 - a_x^2)(1 - a_y^2)}}{1 - a_x a_y} \right) \\
&= \rho \left(\frac{\sqrt{1 - a_x^2 - a_y^2 + a_x^2 a_y^2 + (2a_x a_y - 2a_x a_y)}}{1 - a_x a_y} \right) \\
&= \rho \left(\frac{\sqrt{(1 - a_x a_y)^2 - (a_x - a_y)^2}}{(1 - a_x a_y)} \right)
\end{aligned}$$

Donde podemos notar que la correlación será igual o menor a ρ, con igualdad si y solo si $a_x = a_y$.
La intuición de este resultado es más clara al representar los procesos en forma de MA(∞).

$$\begin{aligned}
x_t &= \sum_{j \geq 0} a_x^j \varepsilon_{t-j}^x \\
y_t &= \sum_{j \geq 0} a_y^j \varepsilon_{t-j}^y
\end{aligned}$$

Como podemos notar, la diferencia entre las series en términos de la correlación viene dada por la diferencia en la importancia que asigna cada serie a los shocks pasados, es decir, las diferencias en la persistencia de los shocks explica esta diferencia.
Así, a mayor sea la diferencia entre los parámetros de persistencia de los procesos AR, mayor será la diferencia en las trayectorias de las series y por lo tanto, menor será la correlación entre las mismas.

4. **Representación de Wold**

El proceso x_t satisface

$$x_t = \varepsilon_t - 3\varepsilon_{t-1}$$

donde ε_t es un ruido blanco Gaussiano con media nula y varianza igual a uno que captura los shocks de interés económico que determinan (causan) la serie de interés, x_t.

(a) Muestre que x_t es débilmente estacionario y encuentre la función de autocovarianza y de autocorrelación correspondientes.

(b) ¿Es x_t invertible? Justifique.

(c) Explique por qué la representación de Wold de x_t será de la forma

$$x_t^{\text{Wold}} = v_t - \frac{1}{3}v_{t-1},$$

donde v_t es un ruido blanco Gaussiano con media nula.

(d) Determine la varianza de los v_t. Compare con la varianza de los ε_t. Comente.

(e) En general hemos insistido en que x_t y x_t^{Wold} pueden ser procesos distintos. Explique por qué en este caso particular son iguales.

(f) Usted cuenta con datos $x_t, x_{t-1}, x_{t-2}...$ y desea predecir x_{t+1} usando $\mathrm{E}_t x_{t+1}$. Obtenga una expresión para $\mathrm{E}_t x_{t+1}$ en función de los datos que tiene. Calcule el error cuadrático medio de estas proyecciones y compare con el que obtendría si observara los ε_t. Comente.

(g) Obtenga la función de respuesta al impulso unitario para x_t respecto de ε_t y de x_t respecto de v_t. Compare. Si son iguales, explique por qué. Si son distintas, ¿cuál es útil?

Respuesta:

(a) $\mathrm{E}x_t = \mathrm{E}\varepsilon_t - 3\mathrm{E}\varepsilon_{t-1} = 0$, por lo que no depende de t. Basta mostrar que los segundos momentos no dependen de t.

$$\begin{aligned}\gamma_x(j) &= \mathrm{E}\left(x_{t+j}x_t\right)\\ &= \mathrm{E}\left((\varepsilon_{t+j} - 3\varepsilon_{t+j-1})(\varepsilon_t - 3\varepsilon_{t-1})\right)\\ &= \mathrm{E}\left(\varepsilon_{t+j}\varepsilon_t - 3\varepsilon_{t+j}\varepsilon_{t-1} - 3\varepsilon_{t+j-1}\varepsilon_t + 9\varepsilon_{t+j-1}\varepsilon_{t-1}\right)\end{aligned}$$

Tomando en cuenta que $\mathrm{E}(\varepsilon_t\varepsilon_s) = 0$ para $t \neq s$, tenemos lo siguiente:

$$\gamma_j(x) = \begin{cases} 10, & \text{si } j = 0 \\ -3, & \text{si } j = 1 \\ 0, & \text{si } j > 1 \end{cases}$$

Luego, sólo depende de j, y no de t. Tenemos que x es débilmente estacionario, y además tenemos la función de autocovarianza. Para la función de autocorrelación basta dividir cada elemento por la varianza ($\gamma_0(x)$):

$$\gamma_j(x) = \begin{cases} 1, & \text{si } j = 0 \\ -\frac{3}{10}, & \text{si } j = 1 \\ 0, & \text{si } j > 1 \end{cases}$$

(b) El polinomio MA es $1 - 3z$, que tiene una raíz igual a $1/3$. Dado que la raíz no es mayor a uno en valor absoluto, el proceso no va a ser invertible.

(c) Sabemos que la representación de Wold de un proceso ARMA con un componente MA no invertible se obtiene reemplazando las raíces del polinomio MA menores a uno por su inverso. Al ser un MA(1), esto es lo mismo que reemplazar el coeficiente MA(1) por su inverso en la ecuación, que es lo que se hace aquí. v_t tendrá media nula dado que x_t^{Wold} tiene que tener la misma media que x_t. Los v_t serán Gaussianos, porque al ser el error de la proyección lineal de x_t en base a $x_{t-1}, x_{t-2}....$ (ver el enunciado de la Representación de Wold), son una combinación lineal de variables de un proceso Gaussiano y, por lo tanto, tienen una distribución normal.

(d) Sabemos que la representación de Wold tiene la misma media y la misma función de autocovarianza que el proceso original. En particular, tiene la misma varianza. La varianza de x^{Wold} es la siguiente:

$$\gamma_0(x^{\text{Wold}}) = \frac{10}{9}\sigma_v^2$$

Dado que se tiene que cumplir que $\gamma_0(x^{\text{Wold}}) = \gamma_0(x) = 10$, tenemos que $\sigma_v^2 = 9$. La varianza es diferente, dado que la innovación de la representación de Wold tiene que adaptarse al cambio en el coeficiente para tener primeros y segundos momentos equivalentes al proceso verdadero.

(e) Sabemos que la distribución de procesos Gaussianos está determinada por sus primeros y segundos momentos. Tanto x_t como x_t^{Wold} son Gaussianos, dado que tanto ε como v son Gaussianos. Además, x_t y x_t^{Wold} tienen primeros y segundos momentos equivalentes, por la definición de la representación de Wold. Luego, al ser Gaussianos y tener primeros y segundos momentos equivalentes los procesos son equivalentes. La diferencia con lo visto en clases es que en general se consideran ruidos blancos generalizados, por lo que no se cumple que la distribución completa sea determinada por los primeros y segundos momentos.

(f) Nos tenemos que fijar en la representación de Wold, ya que como x_t no es invertible no podemos expresar ε_t como función de x_t. Entonces:

$$\mathrm{E}_t x_{t+1} = \mathrm{E}_t(v_{t+1} - \frac{1}{3}v_t) = -\frac{1}{3}v_t$$

Por otra parte, dado que la representación de Wold es invertible:

$$v_t = \sum_{j=0}^{\infty} \left(\frac{1}{3}\right)^j x_{t-j}$$

Entonces, podemos expresar $\mathrm{E}_t x_{t+1}$ en términos de los datos que tenemos como sigue:

$$\mathrm{E}_t x_{t+1} = -\frac{1}{3}\sum_{j=0}^{\infty} \left(\frac{1}{3}\right)^j x_{t-j} = -\sum_{j=0}^{\infty} \left(\frac{1}{3}\right)^{j+1} x_{t-j}$$

Para el error cuadrático medio:

$$\begin{aligned}
&\mathrm{E}_t x_{t+1}^{\text{Wold}} = -\frac{1}{3}v_t \\
&x_{t+1}^{\text{Wold}} - \mathrm{E}_t x_{t+1}^{\text{Wold}} = v_{t+1} - \frac{1}{3}v_t + \frac{1}{3}v_t = v_{t+1} \\
&\text{MSE}_1 = \mathrm{E}(v_{t+1}^2) = 9
\end{aligned}$$

Si pudiéramos observar los ε_t:

$$\begin{aligned}
&\mathrm{E}_t x_{t+1} = -3\varepsilon_t \\
&x_{t+1} - \mathrm{E}_t x_{t+1} = \varepsilon_{t+1} - 3\varepsilon_t + 3\varepsilon_t = \varepsilon_{t+1} \\
&\mathrm{MSE}_1 = \mathrm{E}(\varepsilon_{t+1}^2) = 1
\end{aligned}$$

(g) Tenemos:

$$\begin{aligned}
x_t &= \varepsilon_t - 3\varepsilon_{t-1} \\
x_t^{\text{Wold}} &= v_t - \frac{1}{3} v_{t-1}
\end{aligned}$$

Entonces, dado que $IRF_k(x) = \frac{\partial x_{t+k}}{\partial \varepsilon_t}$ y $IRF_k(x^{\text{Wold}}) = \frac{\partial x_{t+k}^{\text{Wold}}}{\partial v_t}$:

$$IRF_k(x) = \begin{cases} 1, & \text{si } k = 0 \\ -3, & \text{si } k = 1 \\ 0, & \text{si } k > 1 \end{cases}$$

$$IRF_k(x^{\text{Wold}}) = \begin{cases} 1, & \text{si } k = 0 \\ -\frac{1}{3}, & \text{si } k = 1 \\ 0, & \text{si } k > 1 \end{cases}$$

Tenemos que las IRF son distintas. La que nos es útil es la IRF de x_t, dado que es ε y no v la variable que captura los shocks que nos interesa analizar.