Econometría II: Procesos integrados y Raíz Unitaria

Profesor: Mijail Yapor

Facultad de Ciencias Económicas y de Administración - UDELAR

2016

Índice

Comentarios iniciales (1)

1. Como hemos visto en el curso, un proceso puede ser no estacionario en media, en varianza, en las autocorrelaciones o en otra característica de la distribución de las variables.
2. Los procesos no estacionarios más importantes son los procesos integrados.

Definición (Proceso Integrado)

Se dice que un proceso es integrado de orden d, I(d), si es necesario realizar d diferencias para obtener la transformación estacionaria I(0).

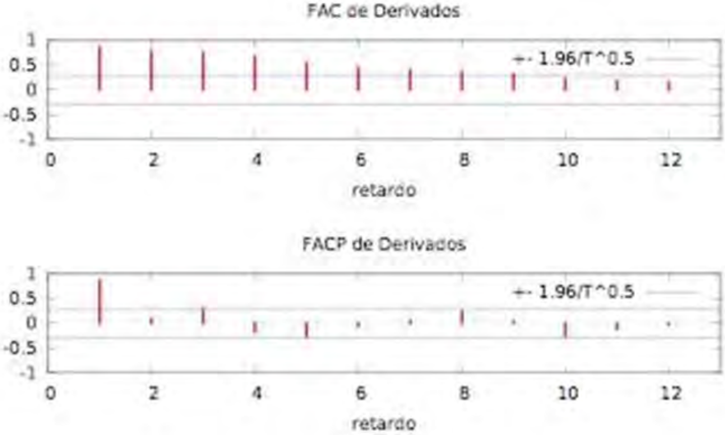
Comentarios iniciales (2)

3. La mayoría de las series reales no son estacionarias, pero sus diferencias relativas, o sus diferencias luego de aplicar logaritmos, sí lo son. Estos son procesos integrados de orden 1.

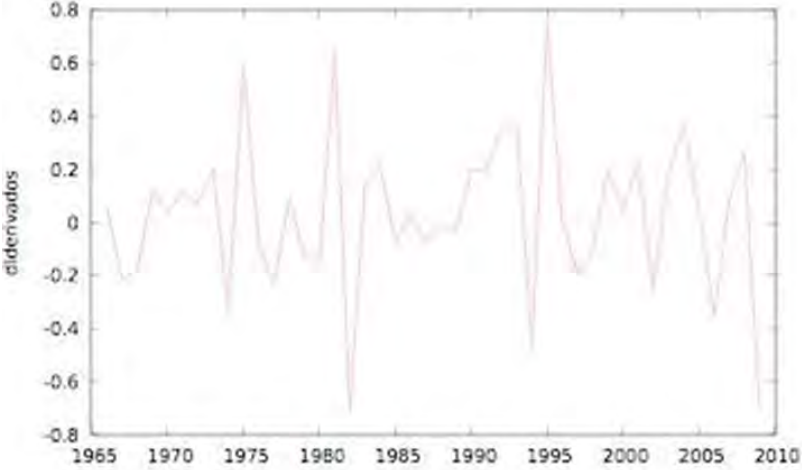
Ejemplos:

1) Exportaciones anuales de derivados del petróleo. Periodo: 1965-2009. Fuente: INE.
2) IPC de Montevideo. Frecuencia mensual. Período: Julio 1937 - Diciembre 2010. Fuente: INE.

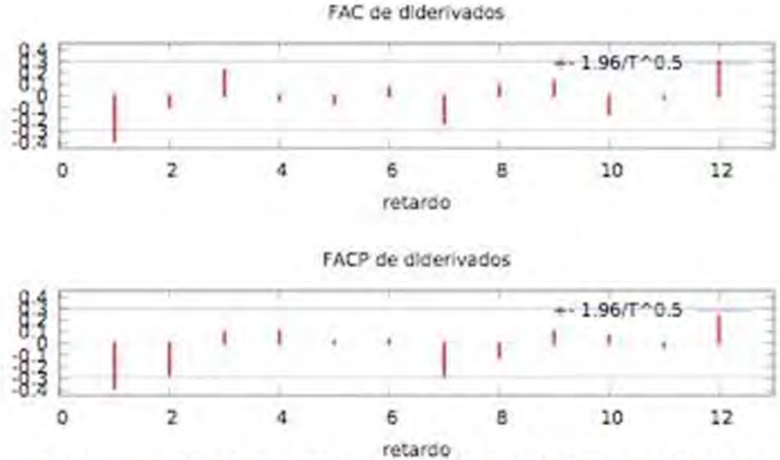
EXPORTACIONES ANUALES DE DERIVADOS DE PETROLEO. 1965-2009. FUENTE: INE



CORRELOGRAMA DE LA SERIE EN NIVELES EXPORTACIONES ANUALES DE DERIVADOS DE PETROLEO

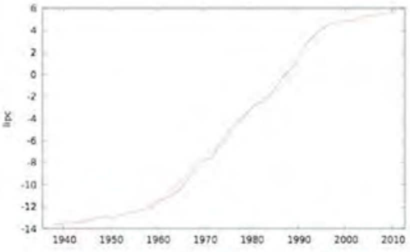


DIFERENCIA DEL LOGARITMO DE LAS EXPORTACIONES ANUALES DE DERIVADOS DE PETROLEO

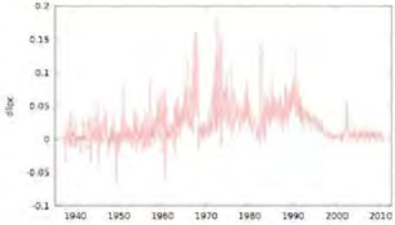


CORRELOGRAMA DE LA DIFERENCIA DEL LOGARITMO DE LAS EXPORTACIONES ANUALES DE DERIVADOS DE PETROLEO

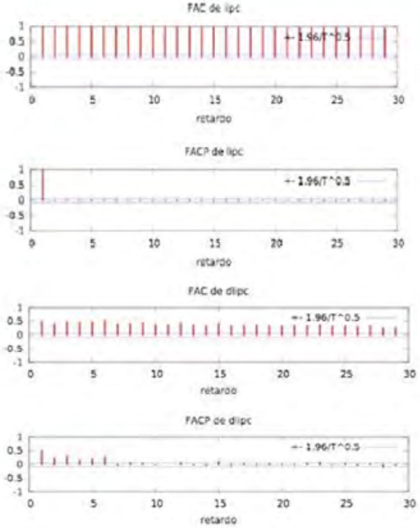
En ciertos casos no basta con una diferencia...



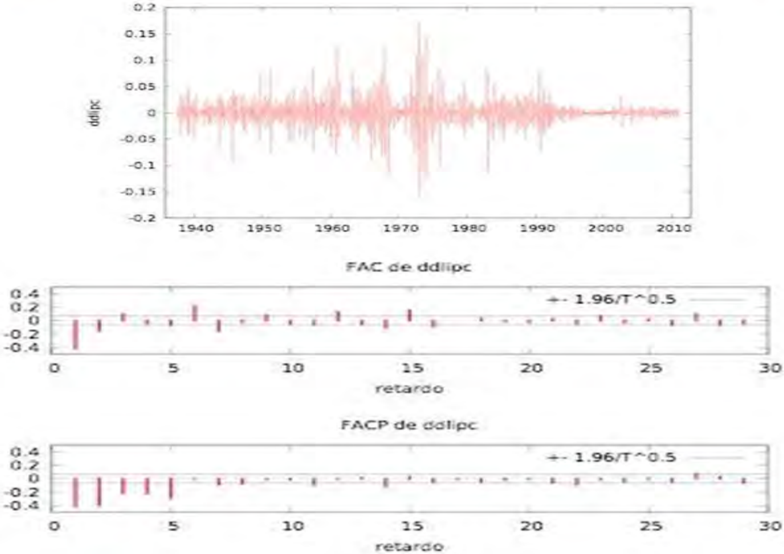
IPC mensual (julio 1937–diciembre 2010) en logs. Fuente: INE.



Diferencia del log del IPC mensual



Segunda diferencia del log del IPC mensual y su correlograma



El Paseo Aleatorio

- Consideremos un modelo del tipo AR(1):

$$y_t = A(t) + \rho y_{t-1} + \varepsilon_t \,, \;\; \text{con} \;\; t = 1, 2, ..., T$$

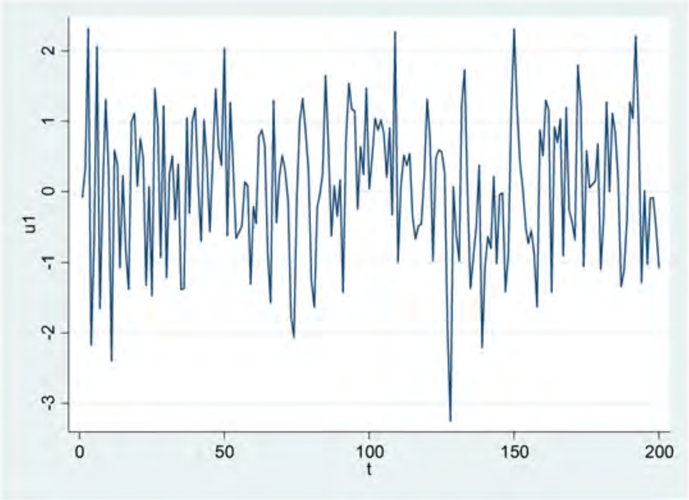
- Donde $\varepsilon_t \sim iid(0, \sigma_\varepsilon^2)$ es un proceso ruido blanco con media 0 y varianza constante.
- Adicionalmente trabajaremos con el caso en que $A(t) = \alpha$ (constante)
- Si $|\rho| < 1$, el proceso es estacionario; mientras que si $\rho = 1$, el proceso autorregresivo es no estacionario.

- Un proceso AR(1) con $\rho = 1$ se llama paseo aleatorio, y a la constante c se la denomina deriva del proceso.
- Una característica importante que diferencia los procesos estacionarios de los no estacionarios es el papel de las constantes en el modelo.
- En un proceso estacionario, la constante no es tan importante, ya que la forma del proceso y sus propiedades básicas son las mismas si tiene media cero o media diferente de cero.
- Sin embargo, en un proceso no estacionario, las constantes, de existir, representan una propiedad permanente del proceso.

Procesos estacionarios	Paseo Aleatorio (RANDOM WALK)
1. $\rho = 0{,}1$	6. $c = 0,\ \rho = 1$
2. $\rho = 0{,}25$	7. $c = 0{,}2,\ \rho = 1$
3. $\rho = 0{,}50$	8. $c = 1,\ \rho = 1$
4. $\rho = 0{,}75$	
5. $\rho = 0{,}95$	

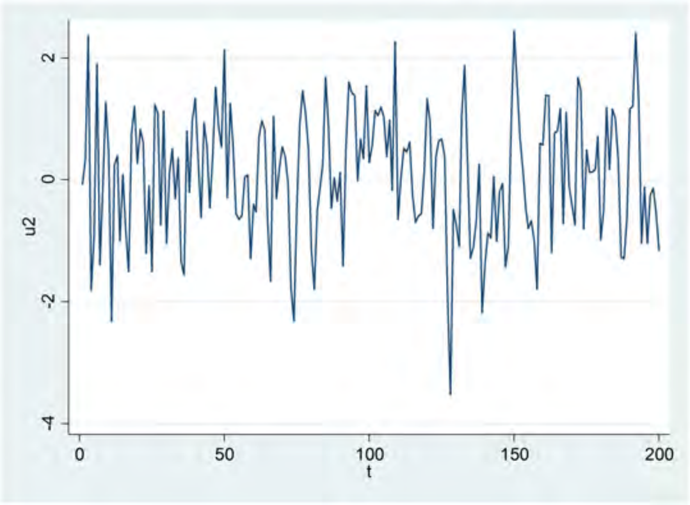
Ejemplo 1: $\rho = 0{,}1$

$$y_t = 0{,}1 \times y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0,1)$$



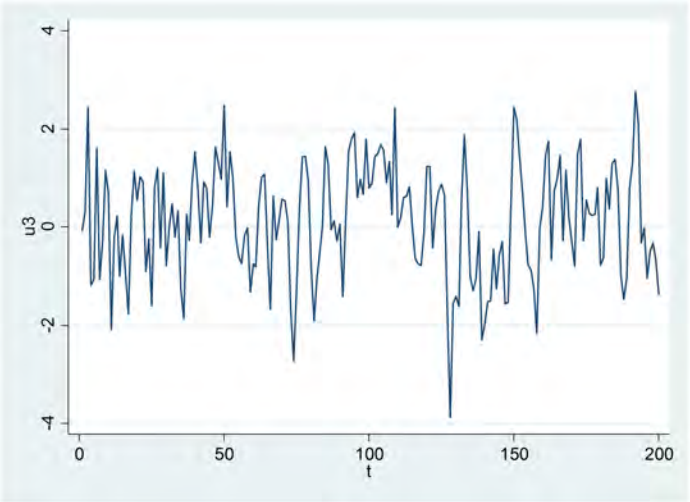
Ejemplo 2: $\rho = 0{,}25$

$$y_t = 0{,}25 \times y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0,1)$$



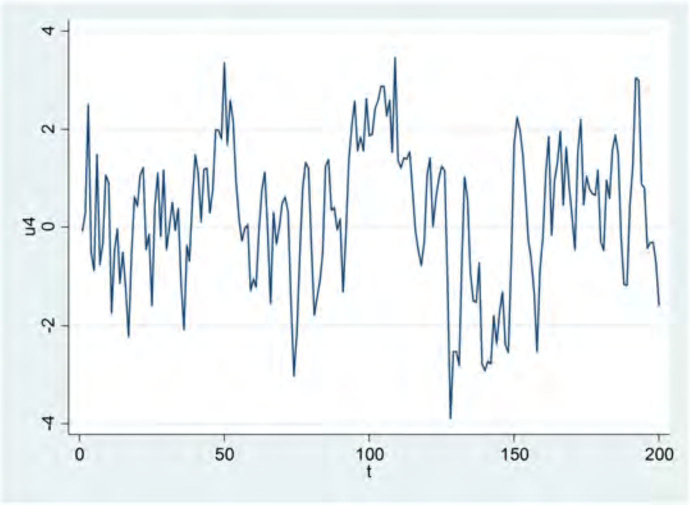
Ejemplo 3: $\rho = 0{,}50$

$$y_t = 0{,}50 \times y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0,1)$$



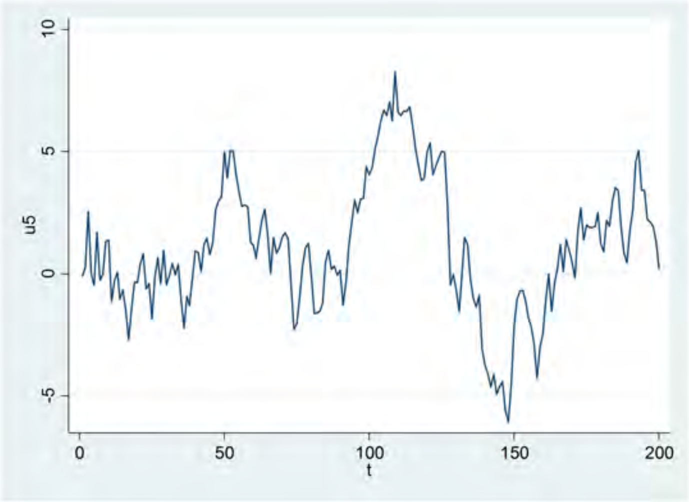
Ejemplo 4: $\rho = 0{,}75$

$$y_t = 0{,}75 \times y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0,1)$$



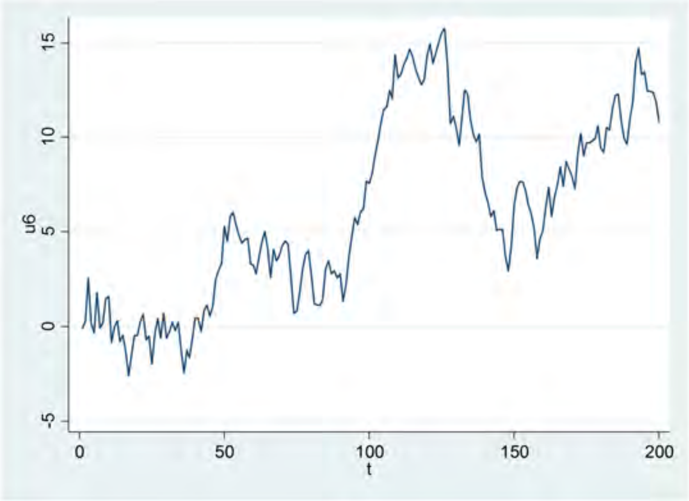
Ejemplo 5: $\rho = 0{,}95$

$$y_t = 0{,}95 \times y_{t-1} + \varepsilon_t, \;\; \varepsilon_t \sim iid(0,1)$$



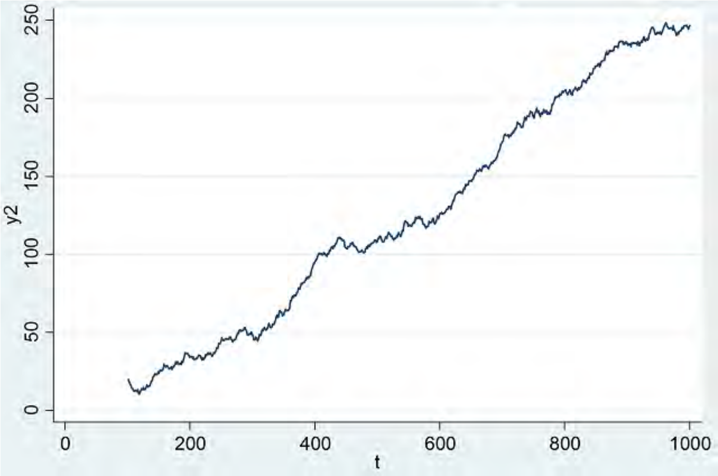
Ejemplo 6: $\rho = 1$ y $c = 0$

$$y_t = y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0,1)$$



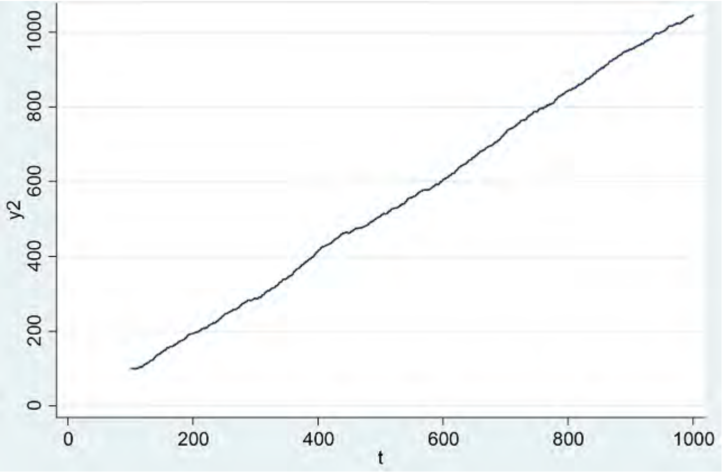
Ejemplo 7: $\rho = 1$ y $c = 0{,}2$

$$y_t \quad = \quad 0{,}2 + y_{t-1} + \varepsilon_t, \;\; \varepsilon_t \sim iid(0,1)$$



Ejemplo 8: $\rho = 1$ y $c = 1$

$$y_t = 1 + y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim iid(0,1)$$



- En los casos 1 a 6, los procesos autorregresivos son estacionarios con media cero, en donde se observa con claridad que si bien todas las series deambulan en torno a su valor medio, a medida que el coeficiente ρ aumenta los períodos de sostenimiento por encima o debajo del valor central son mayores (se incremente la "memoria"del proceso)
- En el caso del ejemplo 6, el proceso no tiene deriva y la serie deambula alrededor del valor medio (en este caso permanece durante largos periodos por encima y por debajo del valor central de la serie).
- En los modelos 7 y 8, con deriva positiva, el proceso muestra una tendencia aproximadamente lineal (con pendientes 0.2 y 1 respectivamente). En otros términos, la serie presenta un crecimiento sistemático que viene marcado por la constante del proceso (c): el nivel de la serie hoy, es la del período anterior (parcial en el 7 e íntegra en el caso 8) más la constante c, lo que supone un crecimiento lineal determinista en este último caso

Momentos del Paseo Aleatorio

Recordando que el proceso era:

$$y_t = A(t) + \rho y_{t-1} + \varepsilon_t\,, \;\; \texttt{con} \;\; t = 1, 2, ..., T$$

Y sustituyendo recursivamente, se tiene:

$$y_t = ct + \varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + ... + \varepsilon_1$$

Tomando esperanzas:

$$E[y_t] = ct$$

Por tanto, si c es diferente de 0, la media del proceso aumenta o disminuye linealmente con el tiempo, dependiendo del signo de c. En el caso $c = 0$, la media es constante y nula.

En el caso de la varianza se tiene que:

$$Var(y_t) = E[\varepsilon_t + \varepsilon_{t-1} + \varepsilon_{t-2} + ... + \varepsilon_1]^2 = \sigma_\varepsilon^2 t.$$

Este resultado implica que la varianza aumenta linealmente con el tiempo y tiende a más infinito con t. Esta propiedad indica que, al aumentar el tiempo, aumenta también la incertidumbre sobre la situación del proceso.

Para el cálculo de las autocovarianzas entre t y k períodos, sabemos que:

$$y_{t+k} = c(t+k) + \varepsilon_{t+k} + \varepsilon_{t+k-1} + \varepsilon_{t+k-2} + ... + \varepsilon_1$$

Por lo tanto:

$$Cov(t, t+k) \;=\; E\Big[(y_t - ct)(y_{t+k} - c(t+k))\Big] = \sigma_\varepsilon^2 t$$

Esta covarianza aumenta linealmente con el tiempo, pero además depende de los instantes en que se calcule:

$$Cov(t, t-k) \;=\; \sigma_\varepsilon^2 (t-k) \neq Cov(t, t+k)$$

La función de autocorrelación queda definida como:

$$\rho(t, t+k) \;=\; \frac{Cov(t, t+k)}{\sqrt{Var(y_t)Var(y_{t+k})}} = \frac{t}{\sqrt{t(t+k)}} = \left(1 + \frac{k}{t}\right)^{-1/2}$$

Por lo tanto, para valores de t elevados los coeficientes de autocorrelación serán próximos a 1 y los mismos decrecerán en forma aproximadamente lineal con k.

Contrastes de Raíz Unitaria - Comentarios iniciales (1)

- La mayoría de las series económicas y financieras se caracteriza por tener un comportamiento tendencial o no estacionario en media.
- Ejemplos claros de esto son los precios de activos financieros, los tipos de cambio y los niveles de los agregados macroeconómicos, como ser el PBI, VAB, etc.
- En este sentido, la econometría debe encargarse de determinar la manera más adecuada para representar la tendencia presente en los datos. En particular, previamente al ajuste de un modelo ARMA, los datos deben transformarse en estacionarios y para ello debe removerse la tendencia.

Comentarios iniciales (2)

- Los dos procedimientos más comunes para remover comportamientos no estacionarios en media consisten en: diferenciar los datos o ajustar una regresión, en general, lineal del tiempo.
- La diferenciación es adecuada cuando el proceso generador de datos es I(1), mientras que es el ajuste de una regresión del tiempo es apropiado cuando los datos son estacionarios en torno a una tendencia.
- En este sentido, los contrastes de raíces unitarias pueden ser utilizados para determinar la transformación más adecuada.
- Otra aplicación importante de los contrastes de raíces unitarias es que pueden usarse para determinar relaciones de equilibrio de largo plazo entre variables económicas no estacionarias.

Contraste Dickey-Fuller básico

El objetivo es determinar si una serie presenta una raíz unitaria, suponiendo que la serie presenta un comportamiento autoregresivo de primer orden.

Retomando la versión sencilla del modelo, el paseo aleatorio sin deriva (Random Walk Without Drift):

$$y_t \quad = \quad \rho y_{t-1} + \varepsilon_t, \;\; \varepsilon_t \sim iid(0, \sigma_\varepsilon^2), \;\; t = 1, 2, ..., T$$

Las hipótesis nula y alternativa serán:

- Ho: $\rho = 1$ ($\rho(L) = 0$ tiene una raíz unitaria)
- H1: $|\rho| < 1$ ($\rho(L) = 0$ tiene raíz fuera del círculo unitario)

Cuando hacemos este contraste en realidad estamos contrastando si y_t es un proceso $I(0)$ frente a que es $I(d)$ (generalmente $d = 1$). El contraste es a una sola cola, pues si $\rho > 1$ el proceso será explosivo, algo muy difícil de encontrar en los procesos económicos.

El estadístico de prueba es:

$$t_{\rho=1} = \frac{\hat{\rho} - 1}{\sqrt{Var(\hat{\rho})}}$$

Bajo la hipótesis alternativa (y_t estacionaria), se tiene que:

$$\sqrt{T}(\hat{\rho} - \rho) \to^d N(0, 1 - \rho^2)$$
$$\hat{\rho} \to^d N(\rho, \frac{1}{T}(1 - \rho^2))$$

Dado este resultado, ahora bajo H_0 cierta, el resultado anterior conduce a

$$\hat{\rho} \to^d N(1, 0)$$

lo cual no tiene sentido alguno.

Reformulación del contraste de Dickey-Fuller

Retomando el modelo básico:

$$y_t = \rho y_{t-1} + \varepsilon_t$$

Restando y_{t-1} en ambos lados se tiene:

$$\begin{aligned} \Delta y_t &= \gamma_a y_{t-1} + \varepsilon_t \\ \gamma_a &= \rho - 1 \end{aligned}$$

y las hipótesis del contraste son:

- Ho: $\gamma_a = 0 \Rightarrow y_t$ es $I(1)$
- H1: $\gamma_a < 0 \Rightarrow y_t$ es $I(0)$

Observación: siendo la diferencia de y_t la variable independiente, el contraste sigue siendo para saber si se tiene raíz unitaria.

El estadístico de prueba ahora es:

$$t = \frac{\widehat{\gamma}_a}{\sqrt{\widehat{Var}(\hat{\gamma}_a)}}$$

Nuevamente es posible afirmar que el estadístico es válido, sin embargo su distribución asintótica no converge a una distribución normal estándar. Para realizar la prueba se utiliza la distribución estimada realizada por Dickey y Fuller (1979).

De esta manera, el contraste básico de raíz unitaria se realiza de la siguiente forma:

- Obtener el estadístico $t = \frac{\widehat{\gamma}_a}{\sqrt{\widehat{Var}(\widehat{\gamma}_a)}}$ en el modelo $\Delta y_t = \gamma_a y_{t-1} + \varepsilon_t$
- Utilizar la distribución Dickey-Fuller para obtener los valores críticos correspondientes a la prueba:
  - Ho: $\gamma_a = 0$
  - H1: $\gamma_a < 0$
- Para muestras de tamaño grande (más de 300) los valores críticos son:

Nivel de significación	**1 %**	**2.5 %**	**5 %**	**10 %**
Valor Crítico Dickey-Fuller	-2.58	-2.23	-1.95	-1.62

Generalización I: componentes determinísticos

- Cuando contrastamos RU, es crucial especificar las hipótesis nula y alternativas apropiadamente para caracterizar las propiedades de la tendencia de los datos.
- Si los datos observados tienen media diferente de cero o una tendencia creciente o decreciente, las hipótesis nula y alternativa deberían reflejar estos aspectos. Para ello, en la regresión auxiliar del contraste, se pueden agregar términos determinísticos (constante y/o tendencia lineal) que a su vez, influirán en la distribución del estadístico.
- Las propiedades de la tendencia de los datos bajo la hipótesis alternativa debe determinar la forma del contraste de regresión usado.
- Generalizaremos incluyendo constante (deriva) y/o tendencia (determinística) en la ecuación.

Los distintos casos entonces son:

Paseo Aleatorio sin deriva (Modelo a)

$$\Delta y_t = \gamma_a y_{t-1} + \varepsilon_t$$

Paseo Aleatorio con deriva (Modelo b)

$$\Delta y_t = \alpha_b + \gamma_b y_{t-1} + \varepsilon_t$$

Paseo Aleatorio con deriva y tendencia determinística (Modelo c)

$$\Delta y_t = \alpha_c + \beta_c t + \gamma_c y_{t-1} + \varepsilon_t$$

- Los valores críticos a utilizar en la prueba de raíz unitaria dependen del modelo a utilizar.
- Se complejiza el problema porque surge el tema de la decisión respecto a qué modelo es el válido. Dos "bibliotecas" al respecto:
  - el investigador decide qué modelo en base a su conocimiento del proceso generador de los datos y la literatura relacionada,
  - es necesario realizar una secuencia de contrastes para decidir cuál es el modelo más apropiado (veremos el procedimiento de Dolado et al. (1990))

Paseo Aleatorio con deriva

$$\Delta y_t = \alpha_b + \gamma_b y_{t-1} + \varepsilon_t$$

- Para muestras de tamaño grande (más de 500) los valores críticos son:

Nivel de significación	**1 %**	**2.5 %**	**5 %**	**10 %**
Valor Crítico Dickey-Fuller	-3.43	-3.12	-2.86	-2.57

- La inclusión de una constante en la regresión mueve hacia la izquierda la distribución de los estadísticos.
- Esta formulación es apropiada para series económicas y financieras sin tendencia como las tasas de interés y los tipos de cambio.

Paseo Aleatorio con deriva y tendencia determinística

$$\Delta y_t = \alpha_c + \beta_c t + \gamma_c y_{t-1} + \varepsilon_t$$

- Para muestras de tamaño grande (más de 500) los valores críticos son:

Nivel de significación	**1 %**	**2.5 %**	**5 %**	**10 %**
Valor Crítico Dickey-Fuller	-3.96	-3.66	-3.41	-3.12

- La inclusión de una constante y un término de tendencia en la regresión mueve hacia la izquierda la distribución de los estadísticos.
- Esta formulación es apropiada para series con tendencia como los precios de activos o los niveles de los agregados macroeconómicos como el PIB real.

Generalización II: rezagos de la variable dependiente

- El contraste descrito hasta aquí, es válido si la serie temporal está bien caracterizada por un AR(1) con errores ruido blanco.
- Muchas series económicas y financieras tienen una estructura dinámica más complicada de la que puede ser capturada por un modelo AR(1).
- *Said & Dickey*(1984, 1985) aumentan el contraste de raíces unitarias básico para poderlo aplicar a series que pueden ser ajustadas por un ARMA(p,q) con órdenes desconocidos, método que se llama Contraste de Dickey-Fuller Aumentado.
- En todos los casos la hipótesis nula del problema es que la serie y_t es I(1) contra la alternativa de que es I(0), asumiendo que la dinámica de los datos se puede representar por un proceso del tipo ARMA.

- Contraste de Dickey-Fuller Aumentado (ADF): se incluyen rezagos de la variable dependiente en la especificación.

$$
\begin{aligned}
ModeloA : \Delta y_t &= \gamma_a y_{t-1} + \sum_{j=1}^{p} \theta_j \Delta y_{t-j} + \upsilon_t \\
ModeloB : \Delta y_t &= \alpha_b + \gamma_b y_{t-1} + \sum_{j=1}^{p} \theta_j \Delta y_{t-j} + \nu_t \\
ModeloC : \Delta y_t &= \alpha_c + \beta_c t + \gamma_c y_{t-1} + \sum_{j=1}^{p} \theta_j \Delta y_{t-j} + \nu_t
\end{aligned}
$$

En este caso los valores críticos no cambian respecto al modelo de referencia en muestras grandes, pero si (aunque de forma suave) en muestras pequeñas porque se pierden observaciones al incluir mayores rezagos de la variable dependiente. De todas formas los valores críticos presentados por Dickey y Fuller son adecuados y por tanto los estadísticos a utilizar en la prueba ADF son los mismos que los presentados para la prueba DF.

- A partir de los modelos anteriores, una primer cuestión a tratar es elegir el orden p del polinomio autorregresivo para Δy_t. El problema radica en que, de no incluir estos rezagos en el modelo y que Δy_t sea efectivamente un proceso autorregresivo, obtendríamos estimaciones inconsistentes para γ debido a la correlación serial que quedará capturada en el término de error.
- Por su parte, si p es demasiado pequeńo seguirá habiendo correlación serial en los residuos y los resultados estarían sesgados.
- Si p es demasiado grande, se reduce la potencia del contraste.
- En la práctica, existen diversas fórmulas para determinar dicho valor, sobre las que no nos detendremos en el curso. Una muy habitual es usar los criterios de información ya conocidos para comparar entre modelos. En general, el orden inicial estará dado en el ejercicio.

Estrategia de Dolado et. al.(1990) para el contraste ADF

1. Se estima en primer lugar un Modelo C, es decir el menos restringido que incluye constante y tendencia. Si se rechaza $H0)\gamma = 0$, culmina el proceso pues se concluye que no hay raíz unitaria.
2. Si no se rechaza la $H0)$ anterior (se admite una raíz unitaria), se pasa a probar la significación del componente tendencial β_c. Dado que en este punto se está bajo la hipótesis de que $H0)$ es cierta, se usa el estadístico τ_{β_T}.
3. Si el término β resulta ser significativo, se prueba de nuevo la presencia de una raíz unitaria vía $H0)\gamma = 0$, pero utilizando la tabla de una normal N(0,1). Sea cual fuera el resultado, finalizamos la prueba admitiendo o rechazando la presencia de una raíz unitaria.

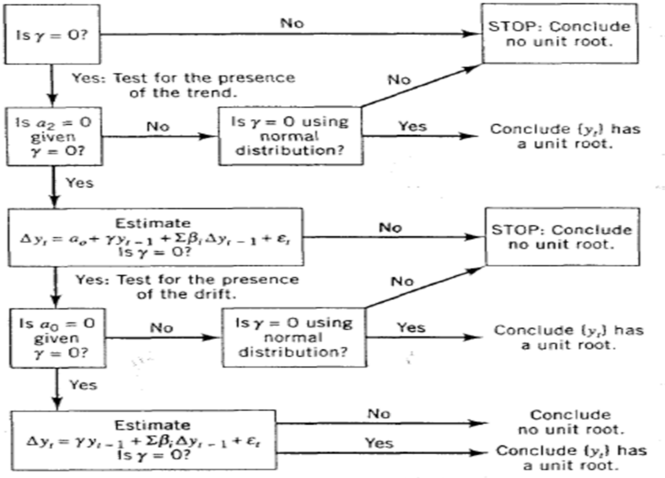
4. Si el termino β no resulta ser significativo (tendencia no significativa), se replantea el modelo inicialmente estimado, pasándose a un Modelo B (sólo con constante). Se realiza la prueba $H0)\gamma = 0$. Si se rechaza $H0$) no hay raíz unitaria y se culmina el proceso. Si no se rechaza $H0$), se prueba si es necesario poner la constante α a través del estadístico $\tau_{\alpha\mu}$. Si el término constante resulta significativo, se usa una tabla de la N(0,1) para probar la existencia de una raíz unitaria, haciendo el test sobre γ, concluyendo aquí si existe o no dicha raíz.
5. Si la constante no resulta significativa, se utiliza el Modelo A probándose la existencia de una raíz unitaria. Aquí no se usa la N(0,1) sino el test DF conocido.

Modelo	Hipótesis nula	Est.	Valores críticos T=50 95%	Valores críticos T=50 99%	Valores críticos T=100 95%	Valores críticos T=100 99%	Valores críticos asintóticos 95%	Valores críticos asintóticos 99%
$\Delta y_t = \gamma_a y_{t-1} + \varepsilon_t$	$\gamma_a = 0$	τ	-1,95	-2.62	-1,95	-2.6	-1,95	-2.58
$\Delta y_t = \alpha_b + \gamma_b y_{t-1} + \varepsilon_t$	$\gamma_b = 0$	τ_μ	-2,93	-3,58	-2,89	-3,51	-2,86	-3,43
	$\alpha_b = \gamma_b = 0$	ϕ_1	4,86	7,06	4,71	6,7	4,59	6,43
	$\alpha_b = 0$ dado $\gamma_b = 0$	$\tau_{\alpha\mu}$	2,56	3,28	2,54	3,22	2,52	3,18
$\Delta y_t = \alpha_c + \beta_c t + \gamma_c y_{t-1} + \varepsilon_t$	$\gamma_c = 0$	τ_τ	-3,5	-4,15	-3,45	-4,04	-3,41	-3,96
	$\alpha_c = \beta_c = \gamma_c = 0$	ϕ_2	5,13	7,02	4,88	6,5	4,68	6,09
	$\beta_c = \gamma_c = 0$	ϕ_3	6,73	9,31	6,49	8,73	6,25	8,27
	$\beta_c = 0$ dado $\gamma_c = 0$	$\tau_{\beta\tau}$	2,81	3,60	2,79	3,53	2,78	3,46
	$\alpha_c = 0$ dado $\gamma_c = 0$	$\tau_{\alpha\tau}$	3,14	3,87	3,11	3,78	3,08	3,71

Esquema Dolado, Jenkinson y Sosvilla-Rivero (1990) (extraído de Enders (1995))

Figure 4.7 A procedure to test for unit roots.

Estimate $\Delta y_t = a_o + \gamma y_{t-1} + a_2 t + \Sigma \beta_i \Delta y_{t-i} + \varepsilon_t$



Bibliografía:

- Enders, W (1995): Applied Econometrics Time Series.
- Stock, J. H. y M. M. Watson (2011): Introduction to Econometrics.
- Wooldridge, J.M (2014): Introducción a la econometría: un enfoque moderno.

Econometría II: Procesos integrados y Raíz Unitaria

Profesor: Mijail Yapor

Facultad de Ciencias Económicas y de Administración - UDELAR

2016