

Universidad Torcuato di Tella

Maestría en Econometría

Propiedades de las pruebas de raíz unitaria  
en muestras pequeñas

Carlos González, 14L774

Junio 2019



## Contents

1. Resumen.....	3
2. Introducción .....	4
3. Revisión de la literatura.....	5
4. Series estacionarias vs no estacionarias .....	9
4.1 Series estacionarias en tendencia .....	11
4.2 Series estacionarias en diferencia.....	11
4.2.1 Procesos autoregresivos de orden $p$ , AR( $p$ ).....	11
4.3 Identificación incorrecta en el tipo de estacionariedad .....	13
5. Test de raíz unitaria .....	13
5.1 Raíz unitaria .....	13
5.1.1 Dickey-Fuller (DF) unit root test .....	14
5.1.2 Augmented Dickey-Fuller (ADF) unit root test.....	16
5.1.3 Phillips-Perron (PP) unit root test .....	18
5.1.4 Kwiatkowski–Phillips–Schmidt–Shin (KPSS) stationarity test .....	20
5.2 Una crítica a las pruebas de raíz unitaria.....	23
6. Poder estadístico y tamaño de una prueba.....	25
6.1 Prueba de hipótesis.....	25
6.2 Poder estadístico de una prueba de hipótesis .....	27
6.3 Tamaño empírico de una prueba .....	28
7. Proceso generador de datos (DGP) .....	28
8. Resultados .....	30
9. Conclusiones .....	44
10. Bibliografía .....	47

# 1. Resumen

El objetivo del presente trabajo es analizar las propiedades en muestras pequeñas de las pruebas de raíz unitaria (y de estacionariedad) estándares o más conocidas, a saber: Dickey-Fuller (1981), Phillips-Perron (1988) y Kwiatkowski-Phillips-Schmidt-Shin (KPSS, 1992). Considerando dos especificaciones de modelos, un autoregresivo de primer orden,  $AR(p)$ , y un modelo autoregresivo de promedio móvil,  $ARMA(p,q)$ , se realiza un experimento Monte Carlo de 1000 simulaciones para calcular de forma empírica el tamaño y la potencia estadística de cada prueba, cuando varía el valor del componente autoregresivo y la dimensión temporal de la serie. Los resultados de las simulaciones muestran que, en muestras muy pequeñas, ninguna de las pruebas de raíz unitaria tiene un desempeño aceptable, especialmente cuando el valor del componente autoregresivo es cercano a uno. No obstante lo anterior, en términos relativos la prueba Phillips-Perron (1988) mostró el mejor desempeño en muestras pequeñas. La prueba Dickey-Fuller (1981), por su parte, resultó ser la prueba más sensible al tamaño de la muestra. El desempeño de esta prueba fue el que más mejoró conforme aumentó el tamaño de la muestra. Asintóticamente, las pruebas Dickey-Fuller (1981) y Phillips-Perron (1988) son prácticamente equivalentes. Finalmente, la prueba KPSS (1992) mostró un tamaño empírico de la prueba invariante ante cambios en el tamaño de la muestra. Resultados relativamente similares se encontraron para el poder estadístico de la prueba. En términos globales, la prueba KPSS (1992) mostró los resultados más estables para todos los tamaños de muestra.

## 2. Introducción

La estacionariedad es un supuesto esencial en la teoría econométrica. Si una serie de tiempo no es estacionaria, los estimadores no serán consistentes. Aplicar Mínimos Cuadrados Ordinarios (OLS) a variables no estacionarias puede llevar a conclusiones erróneas sobre los parámetros y, por lo tanto, a la verdadera relación entre las variables. En la literatura econométrica, este fenómeno se le denomina regresiones espurias, mencionado inicialmente por Granger y Newbold en 1974. Una forma práctica para detectar relaciones espurias es un alto  $R^2$ , coeficientes estadísticamente muy significativos y un estadístico de Durbin-Watson muy bajo. Un ejemplo de lo anterior es cuando dos variables comparten una tendencia creciente en el tiempo. Una regresión simple aplicando OLS es muy probable que capture la tendencia determinística de cada variable y se la atribuya a la variable independiente. Muchas series económicas y financieras muestran un comportamiento con tendencia o de no estacionariedad en la media. Algunos ejemplos son los precios de los activos, tipos de cambio y casi todas las series macroeconómicas.

Una tarea importante de la econometría es determinar el orden de integración de una serie como requisito previo para la modelación univariada o multivariada. La econometría ofrece una amplia gama de pruebas para determinar si una serie es estacionaria. La idoneidad de una prueba en particular depende en buena medida de la causa que genera la falta de estacionariedad en el proceso generador de datos (DGP); sin embargo, en la práctica generalmente se observa la aplicación de estas pruebas de forma indistinta, sin considerar que la potencia estadística puede ser afectada por la decisión que se toma sobre la forma funcional del DGP y/o el tamaño de la muestra. Estas pruebas en su mayoría son asintóticas, por lo que tienden a funcionar bien en muestras grandes pero no necesariamente en muestras chicas.

El objetivo de este trabajo es analizar las propiedades en muestras pequeñas de las pruebas de raíz unitaria estándares o más conocidas. Estudios previos han abordado este tipo de análisis para casos específicos sobre la falta de estacionariedad. Por ejemplo, la presencia de un quiebre estructural o cuando el DGP es un AR(1) con  $\rho = 1$  (*pure random walk*). Entre estos estudios se pueden mencionar: Arltová y Fedorová (2016),

Hacker y Hatemi-J (2010), Schwert (1989), Rehman y Zaman (2009), Cavalieri (2003), Leybourne y Newbold (2000), Kapetanios (2002), IMF (2008), entre otros.

Las pruebas de raíz unitaria a analizar son las siguientes: (i) Dickey-Fuller (DF): es prueba de raíz unitaria cuando el DGP es un AR(1). (ii) Augmented Dickey-Fuller (ADF): extensión de la prueba Dickey-Fuller para procesos AR de mayor orden. (iii) Phillips-Perron (PP): es una prueba ADF que permite mayor orden de autocorrelación. (iv) Kwiatkowski-Phillips-Schmidt-Shin (KPSS): prueba de estacionariedad alrededor de una tendencia determinística.

Considerando diferentes formas funcionales (DGP) y a partir de un experimento Monte Carlo de 1000 simulaciones, se calculan de forma empírica el tamaño y la potencia estadística de cada prueba cuando varía algún parámetro del DGP y la dimensión temporal de la serie cambia ( $T = 30, 50, \dots, 500$ ). El trabajo empieza con una revisión de la literatura empírica del tema (Sección 3). Posteriormente, se realiza una revisión teórica sobre procesos no estacionarios y las pruebas de raíz unitaria (Sección 4 y 5). Un breve repaso sobre prueba de hipótesis, poder estadístico y tamaño empírico de una prueba se desarrolla en Sección 6. Las distintas formas funcionales del DGP y los resultados de las simulaciones se discuten en Sección 7 y 8, respectivamente. Finalmente, el documento termina con algunas conclusiones en Sección 9.

### 3. Revisión de la literatura

Estudios empíricos se han enfocado en mostrar que las pruebas de raíz unitaria estándares fallan en rechazar la hipótesis nula de raíz unitaria en diversas series económicas. Por ejemplo, Nelson y Plosser (1982) estudiaron las propiedades dinámicas de varias series macroeconómicas y financieras de EE.UU. Los autores fueron pioneros en mostrar que la prueba de Dickey-Fuller (1979) tiende a no rechazar la hipótesis nula de raíz unitaria en trece de catorce series económicas y financieras agregadas de EE.UU. Como conclusión general, encuentran que *shocks* de corto plazo tienen un impacto de largo plazo en la mayoría de las series analizadas.

A partir de esta publicación, mucha de la investigación empírica sobre el tema se enfocó en tratar de entender el por qué de estos resultados. El principal objetivo fue comprender el impacto de largo plazo que tiene un *shock* en series económicas con un

componente fuerte de tendencia. Por ejemplo, Perron (1989) mostró que si la Gran Depresión en 1929 y la crisis del petróleo en 1973 son tratados como eventos que provocaron un quiebre estructural en las series económicas y financieras analizadas por Nelson y Plosser (1982), muchos de los resultados cambiaban significativamente. En ese sentido, Perron (1989) argumentó que los resultados de Nelson y Plosser (1982) estaban sesgados a no rechazar la hipótesis nula de raíz unitaria en las series por no tomar en cuenta cambios estructurales en las series analizadas. Estos resultados no cambiaron mucho al aplicar versiones aumentadas de la prueba de Said y Dickey (1984) y permitir en la estructura del modelo autocorrelación en los errores.

Mucha de la literatura sobre el desarrollo de pruebas de raíz unitaria se basa en la comparación con los resultados obtenidos por Nelson y Plosser (1982). A manera de ejemplo, para Nelson y Plosser (1982), Murray y Nelson (2002), Kilian y Ohanian (2002), la serie macroeconómica GNP anual de EE.UU es estacionaria en diferencia. Sin embargo, Perron (1989), Zivot y Andrews (1992), Diebold y Senhadji (1996), concluyeron que la serie es estacionaria en tendencia.

Rehman y Zaman (2009) argumentan que muchas de las discrepancias entre los estudios empíricos sobre los resultados de Nelson y Plosser (1982), se debe a las decisiones sobre las especificaciones en el modelo como paso previo a implementar las pruebas de raíz unitaria (por ejemplo, si incluir o no una tendencia determinística). Los autores argumentan que si bien en experimentos Monte Carlo estas especificaciones se pueden controlar, en casos de datos reales no siempre se conocen las verdaderas especificaciones del proceso generador de datos (DGP). Usando data de *real GNP* de EE.UU., Rehman y Zaman (2009) aplican la prueba de Ng-Perron (2001) y encuentran que la hipótesis nula de raíz unitaria no se rechaza cuando incluyen una tendencia lineal; sin embargo, al incluir dicha especificación en la prueba, la hipótesis nula se rechaza al 10% de nivel de significancia. Estos resultados son hasta cierto punto esperables, pues la fortaleza de la prueba de Ng-Perron (2001) está en la selección de la cantidad de rezagos del modelo. Similares discrepancias encuentran los autores cuando hacen algunos ejercicios de simulación. Lo anterior también aplica para especificaciones erróneas sobre la longitud o cantidad de los rezagos, quiebres estructurales y distribución de los errores (homocedásticos vs heterocedásticos). En otras palabras, los autores argumentan que partir de especificaciones equivocadas sobre la verdadera forma funcional del DGP puede llevar a conclusiones erróneas.

Una de las pruebas más utilizadas en la literatura de raíz unitaria es el test de Dickey-Fuller (1979); sin embargo, como argumentan Hacker y Hatemi-J (2010), este tipo de pruebas se desarrollan sobre la base de tres tipos de modelos. Un modelo general que incluye constante y una tendencia lineal determinística, y dos modelos restringidos, sin constante y otro sin tendencia lineal determinística. En la práctica no es fácil determinar *a priori* qué modelo es el más adecuado utilizar para la prueba. En ese sentido, algunos autores como Enders (2004) han propuesto implementar una estrategia de testeo secuencial, empezando por el modelo sin restricciones y avanzar, en caso de ser necesario, hacia los modelos más restringidos. Sin embargo, como mencionan Hacker y Hatemi (2010), este tipo de estrategias reducen el nivel de significancia de la prueba al desarrollar múltiples testeos a la vez (*mass significance problem*). Los autores proponen una alternativa, que si bien se basa siempre en una estrategia de testeo secuencial, desecha *a priori* algunas especificaciones de modelos fruto del conocimiento previo que se puede tener sobre el comportamiento de las series a analizar. Hacker y Hatemi (2010) basan su estrategia en Elder y Kennedy (2004). Para ello desarrollan un experimento Monte Carlo con 5000 simulaciones, partiendo del siguiente modelo general sin restricciones,  $\Delta y_t = a + by_{t-1} + ct + \varepsilon_t$ , donde  $\varepsilon_t \sim N(0,1)$  y son independiente e idénticamente distribuido (iid). Modificando los valores de los parámetros ( $a = 0, 0.25, 1$ ,  $b = 0$  y  $c = 0, 0.4$ ), los autores aplican la prueba de Dickey-Fuller (1979). Los resultados de las simulaciones muestran que, en términos generales, la estrategia de Enders (2004) reduce el poder estadístico y el nivel empírico de la prueba debido al problema de *mass significance*. En cambio, al desechar de forma *a priori* algunas formas funcionales en los modelos, mejora tanto el poder estadístico de la prueba, así como el tamaño empírico. Más concretamente, Hacker y Hatemi (2010) encontraron que bajo la estrategia Enders (2004), cuando el modelo de partida no tiene intercepto y tendencia lineal determinística, la frecuencia de concluir que la serie tiene una raíz unitaria es bastante bajo; es decir, el tamaño empírico de la prueba es más grande relativo al tamaño nominal. Lo mismo sucede –aunque en menor frecuencia– cuando el modelo de partida no tiene tendencia lineal determinística pero sí un *drift* débil. Cuando la serie tiene una raíz unitaria y un componente de tendencia lineal determinística, la prueba Dickey-Fuller (1979) falla con mucha frecuencia en rechazar la hipótesis nula de raíz unitaria, haciendo caer el tamaño empírico de la prueba. Lo anterior es esperable, pues como argumentan los autores, la prueba Dickey-

Fuller (1979) no está diseñada para detectar una raíz unitaria cuando ésta existe en combinación con una tendencia lineal determinística.

La estrategia propuesta por Elder y Kennedy (2004) aplicada por Hacker y Hatemi (2010) consiste en correr dos pruebas de Dickey-Fuller (1979) de forma secuencial. Una primera prueba verifica si la serie tiene una raíz unitaria. Si la primera prueba no rechaza la hipótesis nula de raíz unitaria, el segundo test verifica si la serie tiene un *drift*; en cambio, si la primera prueba rechaza la hipótesis nula de raíz unitaria, el segundo test verifica si la serie tiene un componente de tendencia lineal determinística. Partiendo siempre del modelo general sin restricciones, Hacker y Hatemi (2010) modifican el valor de los parámetros ( $a = 0, 0.25, 1, 2$   $b = -0.05, -0.1, -0.5, 0$  y  $c = 0, 0.1, 0.2, 0.4$ ). Los resultados muestran que, cuando el modelo de partida solo tiene raíz unitaria, el tamaño empírico del primer test es bastante cercano a su valor nominal. El segundo test en cambio tiene un tamaño empírico bastante más alto. Cuando el verdadero proceso generador de datos tiene raíz unitaria con *drift*, el tamaño empírico de la prueba es bastante cercano a su valor nominal. El poder del test parece estar correlacionado positivamente con la magnitud del *drift*, es decir, si el *drift* es alto, el poder estadístico es cercano a 100%, cayendo significativamente cuando el *drift* es bajo. Cuando el verdadero DGP tiene un valor bajo para el parámetro del autorezago  $y_{t-1}$ , la primera prueba muestra un poder estadístico bastante bajo. La segunda prueba muestra un tamaño empírico bastante más alto que el nominal si la serie también posee un componente de constante, pero muestra un alto poder estadístico si tiene constante y tendencia lineal. Finalmente, la prueba parece mostrar un alto poder estadístico cuando el verdadero DGP tiene un valor intermedio del autorezago, constante y tendencia lineal determinística (excepto cuando el valor de la tendencia es baja).

Otro estudio que utiliza simulaciones Monte Carlo es Arltová y Fedorová (2016). Los autores parten de un proceso autoregresivo de primer orden  $AR(1)$  sin constante,  $y_t = \phi y_{t-1} + \varepsilon_t$  donde  $\varepsilon_t \sim N(0,1)$ . A partir de este modelo, van modificando los valores de  $\phi_i, i = 0.01, 0.02, \dots, 0.99$  y la dimensión temporal de la serie  $T = 25, 50$  y  $100$ . Con base a ejercicio de 3000 simulaciones, aplican las pruebas de Dickey-Fuller (1979), Dickey-Fuller (1981), Ng-Perron (2001), Phillips-Perron (1988), ADF-GLS (1996) y KPSS (1992). Los resultados de las simulaciones se miden en términos del poder estadístico de la prueba; es decir, que tan probable es el test de rechazar la hipótesis nula



cuando es falsa, y en términos del nivel empírico del test, es decir, que tantas veces se equivoca la prueba en rechazar la hipótesis nula cuando es falsa (error tipo I). Los resultados muestran que, como conclusión general, no existe un test ideal que tenga el mejor performance para todos los tamaños de muestra y los valores de  $\phi$ . En muestras pequeñas ( $T = 25$ ), las pruebas de Dickey-Fuller (1981) y Phillips-Perron (1988) parecen tener un mejor desempeño; sin embargo, cuando el valor de  $\phi$  es pequeño, los autores encontraron que el test de KPSS tiene un mejor desempeño. Para muestras de tamaño medio ( $T = 25$  y  $100$ ), los autores encontraron que la prueba Dickey-Fuller (1981) tiene un buen desempeño. Estos resultados también se confirman cuando la muestra es grande ( $T = 500$ ). Las pruebas de Dickey-Fuller (1981), Phillips-Perron (1988) y Ng-Perron (2001) muestran un buen performance, incluso para valores pequeños de  $\phi$ . Cuando la dimensión temporal y el valor de  $\phi$  son grandes, la prueba de Ng-Perron (2001) y ADF-GLS (1996) tienen el mejor desempeño. Otra conclusión general es que la prueba de Ng-Perron (2001) funciona bien cuando, independientemente del tamaño de muestra, el valor de  $\phi$  es pequeño. Finalmente, y al igual que mucha de la literatura empírica sobre las pruebas de raíz unitaria, los autores recomiendan complementar el uso de este tipo de pruebas con el test de KPSS.

## 4. Series estacionarias vs no estacionarias

Una serie de tiempo  $\{y_t\}_{t=1}^T$  se define como fuertemente (o completamente) estacionaria si la distribución conjunta es invariante en el tiempo. Lo anterior implica que todos los momentos de la distribución como la media, la varianza, etc., no dependen del tiempo, ni las correlaciones cambien con el tiempo (por ejemplo, si existe autocorrelación de primer orden, ésta no cambia con el tiempo).

En la práctica es difícil probar si una serie temporal es fuertemente estacionaria, principalmente por problemas de disponibilidad de datos. La mayoría de las series económicas son muestras no necesariamente muy grandes en la dimensión temporal. Por lo anterior, un concepto más útil y usado en la práctica es el de estacionariedad débil o en covarianza. La estacionariedad débil solo requiere que los primeros dos momentos de la distribución (media y varianza) y la covarianza sean independientes del tiempo.

Al modelar series de tiempo es útil pensarlas como una realización de un proceso estocástico  $(x_1, x_2, \dots, x_n)$ . En general, un proceso estocástico se puede describir como una distribución de probabilidad  $n$ -dimensional  $p(x_1, x_2, \dots, x_n)$  y, por lo tanto, la relación entre una realización y el proceso estocástico detrás se puede asemejar a la relación entre una población y una muestra. Los primeros dos momentos de esta distribución vienen dados por:

- $\mathbb{E}(x_1), \mathbb{E}(x_2), \dots, \mathbb{E}(x_n)$
- $\mathcal{V}(x_1), \mathcal{V}(x_2), \dots, \mathcal{V}(x_n)$
- $Cov(x_i, x_j), i < j$

Como se mencionó anteriormente, un proceso estocástico se dice que es estacionario en sentido estricto si las propiedades de la distribución de probabilidad no se ve afectada por la dimensión tiempo:

$$p(x_1, x_2, \dots, x_n) = p(x_{1+l}, x_{2+l}, \dots, x_{n+l}) \quad (1)$$

Por otra parte, un proceso se dice que es débilmente estacionario si tiene media y varianza constante. La covarianza del proceso es independiente del tiempo y solo es función del nivel de rezago ( $k$ ):a

$$\mathbb{E}(x_1) = \mathbb{E}(x_2) = \dots = \mathbb{E}(x_t) = \mu \quad (2)$$

$$\mathcal{V}(x_1) = \mathcal{V}(x_2) = \dots = \mathcal{V}(x_t) = \sigma^2 \quad (3)$$

$$Cov(x_t, x_{t-k}) = Cov(x_{t+l}, x_{t-k+l}) = \gamma_k \quad (4)$$

El ejemplo clásico de una serie de tiempo no estacionaria es el del paseo aleatorio o *random walk*, el cual se define como:

$$y_t = y_{t-1} + \varepsilon_t \quad (5)$$

Donde  $\varepsilon_t \sim iid(0, \sigma^2)$  y donde para todo  $n$ , se cumple que:

i.  $E_t y_{t+n} = y_t, E_{t+m} y_{t+m+n} = y_{t+m} \neq y_t = E_t y_{t+n},$

ii.  $E_t \sigma_{t+n}^2 = n\sigma^2$

La primera condición implica que la serie de tiempo tiene una serie indeterminada, y la segunda condición implica que la varianza de  $y_{t+n}$  condicional en la información disponible hasta tiempo  $t$  es dependiente del tiempo. Ambas condiciones violan las condiciones de serie temporal estacionaria.

## 4.1 Series estacionarias en tendencia

Las series de tiempo estacionarias en tendencia son bastante común en las series macroeconómicas. Una serie estacionaria en tendencia se caracteriza por tener una media que crece alrededor de una tendencia fija o determinística. Visualmente, este tipo de series crecen de forma relativamente estable y alrededor de la pendiente de la curva sin desviarse o tener mayores cambios bruscos. En estos casos, extraer la tendencia de la serie (*detrending*) las vuelve estacionarias. Por ejemplo:

$$y_t = \alpha + \mu t + \varepsilon_t \quad (6)$$

Donde  $\varepsilon_t \sim N(0, \sigma^2)$ . Extraer la tendencia de la serie consiste en sustraer  $\mu t$  de cada observación. En la práctica, lo difícil es identificar la tendencia de la serie. Como se puede observar, la media de esta serie tiende a variar con el tiempo:

$$\mathbb{E}(y_t) = \alpha + \mu t \quad (7)$$

Sin embargo, la varianza es constante:

$$\mathcal{V}(y_t) = \mathbb{E}((\alpha + \mu t + \varepsilon_t - (\alpha + \mu t))^2) = \sigma^2 \quad (8)$$

## 4.2 Series estacionarias en diferencia

### 4.2.1 Procesos autoregresivos de orden $p$ , AR( $p$ )

Un proceso AR( $p$ ) tiene raíz unitaria si el polinomio en  $L$  ( $1 - \phi_1 L - \dots - \phi_p L^p$ ) tiene una raíz igual a uno. Como se mencionó anteriormente, el ejemplo clásico de un proceso autoregresivo con raíz unitaria es el camino aleatorio o *random walk*:

$$y_t = y_{t-1} + \varepsilon_t \quad (9)$$

Donde  $\varepsilon_t \sim N(0, \sigma^2)$ , son independientes e idénticamente distribuidos. El *random shock* representan las innovaciones de la serie. Para calcular la media y la varianza del proceso es necesario ir un periodo para atrás:

$$y_{t-1} = y_{t-2} + \varepsilon_{t-1} \quad (10)$$

Sustituyendo (10) en (9) se obtiene lo siguiente:

$$y_t = y_{t-2} + \varepsilon_{t-1} + \varepsilon_t \quad (11)$$

Repitiendo este proceso indefinidamente se obtiene la siguiente expresión:

$$y_t = y_0 + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{t-1} + \varepsilon_t \quad (12)$$

Asumiendo que  $y_0$  es fijo, se puede calcular la media y la varianza de este proceso:

$$\mathbb{E}(y_t) = \mathbb{E}(y_0 + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{t-1} + \varepsilon_t) = y_0 \quad (13)$$

Y, por lo tanto, conociendo  $y_0$ , la varianza condicional es:

$$\mathcal{V}(y_t) = \mathcal{V}(y_0 + \varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{t-1} + \varepsilon_t) \quad (14)$$

$$\mathcal{V}(y_t) = \mathcal{V}(y_0) + \mathcal{V}(\varepsilon_1) + \mathcal{V}(\varepsilon_2) + \dots + \mathcal{V}(\varepsilon_{t-1}) + \mathcal{V}(\varepsilon_t) = t\sigma^2$$

Como se puede observar, la varianza de un proceso con raíz unitaria depende del tiempo  $y$ , por lo tanto, es infinito. Otra forma de verlo es que  $y_{t+n} = y_{t-1} + \sum_{i=0}^n \varepsilon_{t+i}$  y por lo tanto los efectos de  $\varepsilon_t$  nunca mueren, pero si se toma la primera diferencia de  $y_t$  se obtiene que  $\Delta y_t = \varepsilon_t$ , donde  $\Delta y_t = y_t - y_{t-1}$ . En este caso, la primera diferencia hace estacionaria la serie.

Una serie no estacionaria se dice de orden uno,  $I(1)$ , si la primera diferencia de la serie  $\Delta y_t = y_t - y_{t-1}$  es estacionaria,  $I(0)$ . Mas generalmente, una serie es integrada de orden  $d$ ,  $I(d)$ , si es necesario diferenciar  $d$  veces la serie para alcanzar estacionariedad  $I(0)$ . Una serie es  $I(1)$  si contiene una raíz unitaria.

Como afirma Cochrane (2018), en la práctica no es recomendable diferenciar más de una vez las series económicas, ya que en ocasiones se pierde mucha información sobre la variabilidad de la serie original o puede introducir estructuras de correlación en los errores que no necesariamente están presentes en la serie original.

## 4.3 Identificación incorrecta en el tipo de estacionariedad

En la práctica no es fácil determinar si una serie de tiempo es estacionaria en diferencia o estacionaria en tendencia, y hay casos que incluso es una combinación de ambas; por ejemplo, una serie puede tener una tendencia determinística y una raíz unitaria. El no identificar adecuadamente entre una serie estacionaria en diferencia o en tendencia puede llevar a conclusiones erróneas sobre el verdadero proceso generador de datos. Por ejemplo, extraer la tendencia a una serie que es estacionaria en diferencias no elimina el efecto persistente de los errores, como se explicó en Sección 3.2. En este caso, lo que pasa es que simplemente los errores son transformados al extraer la media de los mismos.

Por el contrario, si se toma la primera diferencia a una serie que es estacionaria en tendencia generalmente se introduce un estructura de promedio móvil en los errores. Por ejemplo, al tomar la primera diferencias en tiempo  $t$  y en  $t + 1$  de  $y_t = \alpha + \mu t + \varepsilon_t$  introduce una estructura de promedios móviles en los errores de la serie:  $\Delta y_t = t + (\varepsilon_t - \varepsilon_{t-1})$  y  $\Delta y_{t+1} = t + (\varepsilon_{t-1} - \varepsilon_t)$ .

## 5. Test de raíz unitaria

### 5.1 Raíz unitaria

El término "raíz unitaria" utilizado en las pruebas estadísticas se desprende del proceso *random walk*. Un proceso de camino aleatorio con la misma especificación de (9) puede reescribirse de la siguiente forma:

$$(1 - L)y_t = \varepsilon_t \quad (15)$$

Donde  $L$  representa el operador de rezago. Un proceso autoregresivo  $u_t$  siempre satisface la siguiente ecuación:

$$(1 - \rho(L))u_t = \varepsilon_t \quad (16)$$

Donde  $\rho(L)$  es el polinomio del operador rezado  $L$  (sin constante) y  $\varepsilon_t$  es ruido blanco. El proceso (16) se dice que es estacionario si y solo si las raíces del polinomio  $(1 -$

$\rho(L) = 0$  cae fuera del círculo unitario (son mayores que uno en valor absoluto). Una raíz igual a uno es denominada como raíz unitaria.

En las siguientes subsecciones se realiza una revisión teórica de las pruebas principales y estándares de raíz unitaria.

### 5.1.1 Dickey-Fuller (DF) unit root test

Esta prueba fue desarrollado por Dickey y Fuller en 1979. Es una de las pruebas más sencillas y más utilizadas para testear la presencia de una raíz unitaria. Partiendo de un proceso autorregresivo de orden uno  $AR(1)$ :

$$Y_t = \rho Y_{t-1} + e_t, \quad t = 1, 2, \dots, \quad (17)$$

Donde  $Y_0 = 0$ ,  $\rho$  es un número real, y  $\varepsilon_t$  es ruido blanco con varianza unitaria. La serie  $Y_t$  converge a una serie estacionaria cuando  $t \rightarrow \infty$  si  $|\rho| < 1$ . Cuando  $|\rho| = 1$  el modelo tiene una raíz unitaria y su varianza es  $t\sigma^2$  (*random walk*). Por último, si  $|\rho| > 1$  la serie no es estacionaria y su varianza crece de forma exponencial en la medida que  $t$  aumenta.

Dada la muestra  $Y_1, Y_2, \dots, Y_n$  de tamaño  $n$ , el estimador de máxima verosimilitud de  $\rho$  viene dado por la siguiente expresión:

$$\hat{\rho} = \left( \sum_{t=1}^n Y_{t-1}^2 \right)^{-1} \sum_{t=1}^n Y_t Y_{t-1} \quad (18)$$

Rubin (1950) demostró que  $\hat{\rho}$  es un estimador consistente de  $\rho$ . Por su parte, White (1958) demostró que distribución asintótica de  $n^{\frac{1}{2}}(\hat{\rho} - \rho)$  cuando  $|\rho| < 1$  es normal. Cuando  $|\rho| > 1$ , la distribución asintótica de  $|\rho|^n(\rho^2 - 1)^{-1}(\hat{\rho} - \rho)$  es Cauchy. Por último, cuando  $\rho = 1$  la distribución asintótica de  $n(\hat{\rho} - 1)$  es el ratio de dos integrales definidos como un proceso de Wiener.

En su paper seminal de 1979, Dickey y Fuller derivaron la distribución asintótica de  $\hat{\rho}$  y de  $\hat{\tau}$  para el caso cuando  $|\rho| = 1$ , donde  $\hat{\tau}$  es el estadístico de prueba y está definido como:

$$\hat{t} = (\hat{\rho} - 1)S_e^{-1} \left( \sum_{t=2}^n Y_{t-1}^2 \right)^{1/2} \quad (19)$$

Donde:

$$S_e^2 = (n - 2)^{-1} \sum_{t=2}^n (Y_t - \hat{\rho}Y_{t-1})^2 \quad (20)$$

Existen tres versiones de la prueba Dickey-Fuller (1979). Sustrayendo  $y_{t-1}$  de ambos lados de la ecuación (17) se obtiene la siguiente expresión:

$$\Delta Y_t = (\rho - 1)Y_{t-1} + e_t = \delta Y_{t-1} + e_t \quad (21)$$

Donde  $\Delta$  es el operador de primeras diferencias. Las tres versiones de este test vienen dadas por las siguientes expresiones:

**Modelo A.** *Test de raíz unitaria:*

$$\Delta Y_t = \delta Y_{t-1} + e_t \quad (22)$$

**Modelo B.** *Test de raíz unitaria con drift:*

$$\Delta Y_t = \alpha_0 + \delta Y_{t-1} + e_t \quad (23)$$

**Modelo C.** *Test de raíz unitaria con drift y tendencia determinística:*

$$\Delta Y_t = \alpha_0 + \alpha_1 t + \delta Y_{t-1} + e_t \quad (24)$$

En todos los casos, la forma de probar la hipótesis nula que la serie tiene una raíz unitaria es testear que  $\delta = 0$ . La hipótesis alternativa es que  $\delta < 0$ . Es una prueba a una cola, y dado que el test se aplica a los residuos de (17) no es posible usar los valores críticos de una distribución  $t$  estándar. Dickey-Fuller (1979) calcularon los estadísticos de la prueba que se conocen como los (tau)  $\tau$ -estadísticos. Se dice que test de Dickey-Fuller tiene bajo poder estadístico para distinguir entre un proceso con raíz unitaria ( $\delta = 0$ ) y un proceso muy cercano a tener raíz unitaria ( $\delta$  muy cercano a cero).

Si el proceso es estacionario, la serie debería mostrar una tendencia que tienda a una media constante (tendencia determinística). Por lo tanto, valores grandes de la serie

deberían ser seguidos por valores chicos (o cambios negativos) y viceversa, valores chicos en la serie deberían ser seguidos por valores grandes (o cambios positivos). Lo anterior significa que valores en nivel de la serie en tiempo  $t$  deberían ser un buen predictor de valores de la serie en tiempo  $t + 1$ . De forma contraria, si la serie no es estacionaria, cambios positivos o negativos deberían ocurrir de forma indistinta sin depender de los valores presentes de la serie en nivel. Por último, si la serie tiene autocorrelación en los errores, los resultados de la prueba estarán sesgados.

Utilizando los datos provenientes de Friedman y Schwartz (1963), Dickey-Fuller (1979) aplicaron la prueba a la serie económica de la velocidad del dinero. Las observaciones son anuales y abarcan el periodo de 1869 a 1960. Previamente, Gould y Nelson (1974) habían concluido que la serie del logaritmo de la velocidad del dinero era consistente con un modelo  $AR(1)$ :  $X_t = X_{t-1} + e_t$ . Dickey y Fuller (1979) aplicaron los modelos A y B de la prueba y concluyeron que al 10% de nivel de significancia, no era posible rechazar la hipótesis nula de que la serie tiene raíz unitaria.

Como segundo ejemplo, los autores aplicaron la prueba a la serie trimestral del logaritmo del índice de producción elaborado por la Reserva Federal de Estados Unidos. La muestra abarcaba el periodo de 1950 a 1977. Los resultados mostraron que, a un 5% de nivel de significancia, la prueba rechazó la hipótesis nula de que la serie tiene una raíz unitaria.

### 5.1.2 Augmented Dickey-Fuller (ADF) unit root test

La prueba Dickey-Fuller (1981) es una versión aumentada de la prueba de raíz unitaria de Dickey y Fuller (1979). Por lo anterior, los procedimientos de ambas pruebas son bastante parecidos. Se dice que es una versión aumentada de la prueba Dickey y Fuller (1979) en el sentido que aplica para procesos auto regresivos de mayor orden y modelos más complicados de series de tiempo, como por ejemplo, series que tienen estructuras de promedios móviles en los errores.

La prueba Dickey y Fuller (1981) parte del siguiente modelo:

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (25)$$



Donde  $\alpha$  es la constante del modelo,  $\beta$  es el coeficiente de la tendencia que depende del tiempo, y  $\rho$  es el orden de rezago del proceso autoregresivo. Al igual que en el caso de la prueba de Dickey-Fuller (1979), la hipótesis nula es la presencia de una raíz unitaria ( $\gamma = 1$ ). La hipótesis alternativa depende de la versión de la prueba que se aplique pudiendo ser estacionaria en diferencia o en tendencia ( $\gamma < 0$ ). Existen tres versiones del modelo:

**Modelo A.** *Random walk* ( $\alpha = 0$  y  $\beta = 0$ ):

$$\Delta y_t = \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (26)$$

**Modelo B.** *Random walk con drift* ( $\beta = 0$ ):

$$\Delta y_t = \alpha + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (27)$$

**Modelo C.** *Random walk con con drift y tendencia determinística:*

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (28)$$

Como se mencionó anteriormente, la prueba Dickey-Fuller (1981) permite ejecutar la prueba Dickey-Fuller (1979) de raíz unitaria para procesos de mayor orden de correlación, incluidos en el parámetro  $\rho$ . La longitud de  $\rho$  o el máximo orden de correlación en la prueba se define basándose en los estadísticos de criterios de información como Akaike, BIC o Hannan-Quinn. En términos generales, la prueba Dickey-Fuller (1981) aplica una corrección paramétrica a la estructura de autocorrelación. El estadístico de esta prueba se define como:

$$DF_t = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad (29)$$

Si el valor del estadístico de prueba es menor que valor crítico de la prueba Dickey-Fuller (1981), la hipótesis nula de que  $\gamma = 1$  se rechaza y, por lo tanto, se dice que la serie es estacionaria. La intuición detrás de la prueba Dickey-Fuller (1981) es la siguiente: si la serie está caracterizada por un proceso de raíz unitaria, entonces los valores rezagados de la serie en nivel ( $y_{t-1}$ ) no proveen información relevante en la predicción en los cambios de la variable dependiente ( $y_t$ ), adicional al que provee la variable rezagada en primera diferencia ( $\Delta y_{t-k}$ ).

Al igual que en Dickey y Fuller (1979), Dickey y Fuller (1981) aplican la prueba a la serie del logaritmo del índice de producción elaborado por la Reserva Federal de Estados Unidos. Los datos son trimestrales y abarcan el periodo 1950-1977. Los autores asumen que la serie está representada por la siguiente expresión:

$$Y_t = \alpha + \beta t + \gamma Y_{t-1} + \delta_1 \Delta Y_{t-1} + \varepsilon_t \quad (30)$$

Donde  $\varepsilon_t$  es una variable aleatoria, independiente e idénticamente distribuido (iid) con media cero y varianza  $\sigma^2$ .

La hipótesis nula de que  $\alpha = \beta = 0$  y  $\gamma = 1$  (modelo A, la serie es un proceso de *random walk*) es rechazada al 2.5% nivel de significancia. De igual forma, los autores aplican la prueba con base a la hipótesis nula de que  $\beta = 0$  y  $\gamma = 1$  (modelo B, *Random walk con drift*). A un nivel de significancia de 5%, los resultados llevan a no rechazar la hipótesis nula; por lo tanto, la serie es un proceso autoregresivo de primer orden con raíz unitaria y un posible *drift*. Sin embargo, estos mismos resultados pueden ser rechazados a un nivel de significancia de 10% bajo las pruebas Dickey y Fuller (1979) y Dickey y Fuller (1981).

### 5.1.3 Phillips-Perron (PP) unit root test

La prueba de raíz unitaria de Phillips-Perron (1988) es una extensión de la prueba Dickey-Fuller (1979). Basándose en Phillips (1987), los autores aplican esta prueba para los casos cuando la especificación del modelo incorpora: (i) un *drift*, o (ii) un *drift* con una tendencia lineal. En términos generales, la diferencia de esta prueba es que aplica un enfoque no paramétrico en la estimación de algunos parámetros que no son de interés (*nuisance parameters*) dentro del desarrollo de la prueba. En ese sentido, una ventaja de la prueba Phillips-Perron (1988) es que no asume una forma funcional de los errores de la serie al ser una prueba no paramétrica, permitiendo incluir una gama más amplia de modelos en los cuales se desconoce la verdadera estructura de los errores. Por ejemplo, modelos ARIMA con problemas de heterocedastidad en los errores. Este método parece tener ventajas cuando hay un componente de promedio móvil en la serie (Phillips y Perron, 1988). No obstante lo anterior, una desventaja es que la prueba es asintótica; es decir, tiende a funcionar bien en muestras grandes pero no necesariamente en muestras chicas.

La prueba parte del siguiente modelo:

$$Y_t = \alpha Y_{t-1} + u_t, \quad t = 1, 2, \dots, \quad (31)$$

La hipótesis nula la prueba es que  $\alpha = 1$ . Las condiciones iniciales del modelo son establecidas en  $t = 0$ .  $Y_0$  puede ser una variable aleatoria o puede ser un valor fijo e independiente del tamaño de muestra ( $T$ ). El término de error o innovaciones cumple con las siguientes condiciones:

- $E(u_t) = 0 \forall t$
- $\sup_t E|u_t|^{\beta+\varepsilon} < \infty$  para  $\beta > 2$  y  $\varepsilon > 0$
- Cuando  $T \rightarrow \infty$ ,  $\sigma^2 = \lim E(T^{-1}S_T^2)$  existe y  $\sigma^2 > 0$  donde  $S_t = u_1 + \dots + u_t$

La segunda y tercera condición permite controlar la presencia de heterocedasticidad y dependencia temporal o autocorrelación en el proceso. Los autores consideraron dos variantes del modelo general:

**Modelo A.** Serie con *drift*:

$$Y_t = \hat{\mu} + \hat{\alpha}Y_{t-1} + \hat{u}_t, \quad t = 1, 2, \dots, \quad (32)$$

**Modelo B.** Serie con tendencia determinística y *drift*:

$$Y_t = \tilde{\mu} + \tilde{\beta} \left( t - \frac{1}{2}T \right) + \tilde{\alpha}Y_{t-1} + \tilde{u}_t, \quad t = 1, 2, \dots, \quad (33)$$

Donde  $(\hat{\mu}, \hat{\alpha})$  y  $(\tilde{\mu}, \tilde{\beta}, \tilde{\alpha})$  son coeficientes estimados por el método de mínimos cuadrados ordinarios (MCO). Los estadísticos  $t$  están definidos como:

$$t_{\hat{\alpha}} = \frac{(\hat{\alpha} - \alpha) \{ (\sum y_{t-1} - \bar{y}_{-1})^2 \}^{\frac{1}{2}}}{\hat{S}} \quad (34)$$

$$t_{\hat{\mu}} = \frac{(\hat{\mu} - \mu) \{ (\sum y_{t-1} - \bar{y}_{-1})^2 / \sum y_{t-1}^2 \}^{\frac{1}{2}}}{\hat{S}} \quad (35)$$

$$t_{\tilde{\mu}} = \frac{(\tilde{\mu} - \mu)}{(\tilde{S}^2 c_1)^{1/2}}$$

(36)

$$t_{\tilde{\beta}} = \frac{(\tilde{\beta} - \beta)}{(\tilde{S}^2 c_2)^{1/2}} \quad (37)$$

$$t_{\tilde{\alpha}} = \frac{(\tilde{\alpha} - \alpha)}{(\tilde{S}^2 c_3)^{1/2}} \quad (38)$$

Donde  $\hat{S}$  y  $\tilde{S}$  son los desvíos estándares de (32) y (33),  $c_i$ ,  $i = 1, 2, 3$  es el elemento  $i$  de la diagonal principal de la matriz  $(X'X)^{-1}$  de los coeficientes estimados, y  $\bar{y}_{-1} = T^{-1} \sum y_{t-1}$ . Siguiendo el procedimiento de Dickey y Fuller (1979), los autores estiman la distribución asintótica de los coeficientes de las regresiones (32) y (33) y todos los estadísticos  $t$  especificados en (34), (35), (36), (37) y (38), bajo el supuesto de que los datos fueron generados por (31).

Como se mencionó anteriormente, y al igual que la prueba Dickey-Fuller (1981), esta prueba incorpora en su estructura la posibilidad de un mayor orden de autocorrelación, haciendo que el valor de la variable rezagada ( $y_{t-1}$ ) sea endógena y, por lo tanto, invalidando los estadísticos  $t$  de la prueba Dickey-Fuller (1979). La prueba de Phillips-Perron (1988) realiza una corrección no paramétrica en los estadístico  $t$  de la prueba Dickey-Fuller (1979). Se dice que esta prueba es robusta ante la no especificación en la estructura de autocorrelación y heterocedasticidad de los errores de la prueba. Sin embargo, Davidson y Mackinson (2004) encontraron que esta prueba tiene un peor desempeño que Dickey-Fuller (1981) en muestras pequeñas. Por su parte, Pesarán (2015) mostró que asintóticamente, las pruebas Phillips-Perron (1988) y Dickey-Fuller (1981) son equivalentes.

#### 5.1.4 Kwiatkowski–Phillips–Schmidt–Shin (KPSS)

##### stationarity test

Esta prueba fue desarrollada por Denis Kwiatkowski, Peter Phillips, Peter Schmidt y Yongcheol Shin en 1992, y es conocida como la prueba KPSS. A diferencia de las pruebas anteriores, la hipótesis nula de esta prueba es que la serie es estacionaria alrededor de una tendencia determinística (*trend-stationary*). La hipótesis alternativa de la prueba es que la serie tiene una raíz unitaria.

Los autores de la prueba parten de que la serie puede ser expresada como la suma de tres componentes, a saber: un componente de tendencia determinístico, un *random walk* y un error estacionario:

$$y_t = \xi_t + r_t + \varepsilon_t \quad (39)$$

donde  $r_t = r_{t-1} + u_t$  es el *random walk* y  $u_t \sim iid(0, \sigma_u^2)$ . El valor inicial de  $r_0$  se considera fijo y sirve como la constante en el modelo.

La prueba KPSS es un test de *Lagange Multiplier* (LM) sobre la hipótesis de que el *random walk* tiene varianza cero; es decir,  $\sigma_u^2 = 0$ . Dado que  $\varepsilon_t$  se asume que es estacionario, bajo la hipótesis nula,  $y_t$  sería estacionaria en tendencia.

Bajo el supuesto de la hipótesis nula (*random walk* es normal con cero varianza) y que los errores  $\varepsilon_t$  son *white noise*, los estadísticos de prueba LM (a un lado) para la hipótesis de estacionariedad en tendencia, son los mismos que en Nabeya y Tanaka (1988) conocidos como *locally best invariant* (LBI). Sin embargo, argumentan los autores, el supuesto de que los errores  $\varepsilon_t$  son *white noise* puede ser poco creíble en algunas ocasiones. En ese sentido, proceden basándose en Phillips (1987) y Phillips y Perron (1988) y derivan la distribución asintótica de los estadísticos de prueba bajo condiciones generales de que el error es estacionario. Los estadísticos de prueba no son estándares e involucran movimientos Brownianos de mayor orden.

Siendo  $e_t$ ,  $t = 1, 2, \dots, T$  los residuos de una regresión de  $y$  sobre un intercepto y un coeficiente de tendencia, la suma parcial de residuos se define como:

$$S_t = \sum_{i=1}^t e_i, \quad t = 1, 2, \dots, T \quad (40)$$

El estadístico de prueba LM y LBI se define como:

$$LM = \sum_{t=1}^T \frac{S_t^2}{\hat{\sigma}_\varepsilon^2} \quad (41)$$

Donde  $\hat{\sigma}_\varepsilon^2$  es una estimación de la varianza de los errores de la regresión anterior (39).

Un modelo alternativo incluido por los autores es cuando  $\xi = 0$ . En este caso, la hipótesis nula de la prueba es que la serie es estacionaria alrededor de un nivel ( $r_0$ ). Los

estadísticos de prueba son los mismos LM y LBI que en el caso anterior. Nyblon y Makelainen (1983) proveen los estadísticos de prueba LBI.

Finalmente, los autores aplican el test de estacionariedad a la base de datos de series macroeconómicas de Nelson y Plosser (1982). Los autores encuentran que no pueden rechazar la hipótesis nula de estacionariedad en tendencia (con ocho rezagos en la especificación de la varianza de largo plazo,  $l = 8$ , y con un nivel de significancia de la prueba de 5%) para cinco series: producción industrial, precios del consumidor, salarios reales, velocidad del dinero y stock prices.

Combinando estos resultados con los resultados de la prueba Dickey-Fuller (1981), los autores concluyen que la tasa de desempleo es estacionaria. Por el contrario, dado que la prueba rechaza la hipótesis de nula de estacionariedad en tendencia, consideran que cuatro series tiene raíz unitaria: índice de precios al consumidor, salarios reales, velocidad del dinero y stock prices. Los autores concluyen que tres de las series probablemente tienen raíz unitaria (real GNP, nominal GNP y tasa de interés); sin embargo, aclaran que la evidencia en contra de la hipótesis nula de estacionariedad en tendencia no es concluyente dado que es solo marginalmente significativa. Finalmente, los autores consideran que para seis series (real per cápita GNP, empleo, tasa de desempleo, GNP deflactor, salarios y masa monetaria) la información que proveen los datos no es lo suficientemente informativa para que la prueba pueda distinguir entre las hipótesis de estacionariedad en tendencia y raíz unitaria. Por último, para la serie de producción industrial, la prueba rechaza ambas hipótesis y, por lo tanto, mencionan que no es claro qué se debe concluir.

Como se puede observar, una dificultad práctica de a prueba KPSS (1992) viene dada por la naturaleza del diseño de la misma. La ausencia de una raíz unitaria no significa necesariamente que la serie sea estacionaria, sino, estacionaria en tendencia. En ese sentido, es posible que con esta prueba se esté ante una serie que no sea estacionaria, no tenga una raíz unitaria pero que sí sea estacionaria en tendencia. En ambos procesos (de raíz unitaria y estacionariedad en tendencia) la media de la serie puede tender a crecer o decrecer en el tiempo; sin embargo, las series estacionarias en tendencia son *mean-reverting* en presencia de un *shock* (Bohn, 2005).

Esta prueba se considera un buen complemento a los tests usuales de raíz unitaria. Por ejemplo, si Dickey-Fuller (1981) falla en rechazar la hipótesis nula de raíz unitaria, y a su vez la prueba KPSS (1992) rechaza estacionariedad, entonces se considera que existe más evidencia a favor de la existencia de una raíz unitaria en la serie.

Por último, la prueba KPSS (1992) también se utiliza para testear integración fraccionada (Lee y Schmidt, 1996). Si las pruebas convencionales fallan en rechazar la hipótesis nula de raíz unitaria, y el KPSS también rechaza estacionariedad, entonces se puede considerar como evidencia a favor de la existencia de integración fraccionada para series  $(1 - L)^d y_t = \varepsilon_t$ , donde  $d$  es un número real<sup>1</sup> (Lee y Schmidt, 1996).

## 5.2 Una crítica a las pruebas de raíz unitaria

Cochrane (1991) argumenta que –dado que cualquier series de tiempo con raíz unitaria se puede descomponer en un proceso estacionario y un *random walk*, y que éste último componente usualmente tienen una varianza pequeña– las pruebas de raíz unitaria y/o de estacionariedad en tendencia, suelen tener un poder estadístico bajo en muestras pequeñas. Adicionalmente, Cochrane (1991) argumenta que en muestras muy chicas, existen procesos con raíz unitaria que tienen una función de verosimilitud y de autocorrelación muy parecida (o difícil de distinguir) a la de un proceso estacionario, y que –por lo tanto– en estos casos, la inferencia que se realice es muy parecida entre ambos procesos.

Cochrane (1991) parte de un proceso que puede ser estacionario en diferencias:

$$(1 - L)y_t = \mu + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} = \mu + a(L)\varepsilon_t \quad (42)$$

O estacionario en tendencia:

---

<sup>1</sup> El modelo más simple de una serie que es integrada fraccionalmente es  $(1 - L)^d y_t = \varepsilon_t$ , donde si  $d = 0$  la serie es simplemente ruido blanco, y si  $d = 1$  es un *random walk*. Más generalmente, un modelo autoregresivo integrado de promedios móviles (ARIMA) se puede escribir de la forma  $(A(L)y_t) = \varepsilon_t$ , donde  $A(L)$  es el polinomio del operador rezago. Por su parte, un modelo ARFIMA, donde FI se refiere a que la serie es integrada fraccionalmente, se puede escribir como  $(A(L)y_t)^d = \varepsilon_t$ , donde  $d$  no es necesariamente un número entero. Si  $0 < |d| < 1$  las autocorrelaciones se desvanecen de forma más lenta que en un modelo ARIMA. Si  $0.5 < |d| < 1$  la serie no es estacionaria ya que la suma de las autocorrelaciones no están acotadas.

$$y_t = \mu t + \sum_{j=0}^{\infty} a_j \varepsilon_{t-j} = \mu t + a(L)\varepsilon_t \quad (43)$$

Donde  $E(\varepsilon_t | y_{t-1}, y_{t-2}, \dots) = 0$ ,  $var(\varepsilon_t) = \sigma_\varepsilon^2$ ,  $a_0 = a(0) = 1$ ,  $a(L)$  es invertible y  $\sum_{j=0}^{\infty} a_j^2 < \infty$ . La ecuación (43) es un caso especial de la ecuación (42), donde el polinomio  $a(L)$  se puede expresar como  $(1 - L)b(L)$ . Beveridge y Nelson (1981) proponen una descomposición de la ecuación (42) en dos componentes: un proceso estacionario y un *random walk* con un *drift*:

$$y_t = z_t + c_t \quad (44)$$

Donde:

$$z_t = \mu + z_{t-1} + a(1)\varepsilon_t, \quad a(1) = \sum_{j=0}^{\infty} a_j \quad (45)$$

$$c_t = a^*(L)\varepsilon_t, \quad a_k^* = - \sum_{j=k+1}^{\infty} a_j \quad (46)$$

Una representación alternativa de la ecuación (42) viene dada por:

$$(1 - L)y_t = \mu + a(L)\varepsilon_t = \mu + (a(1) + (1 - L)a^*(L))\varepsilon_t \quad (47)$$

Donde  $a(L)$  y  $a^*(L)$  satisfacen las condiciones de las ecuaciones (42) y (46), respectivamente. Cochrane (1991) argumenta que las representaciones (42), (45) y (47) son equivalentes. Por lo tanto, una representación de un proceso estacionario en diferencias, de un proceso con raíz unitaria y de un *random walk* son equivalentes. En otras palabras, esto significa que, dado que las ecuaciones (45), (46) y (47) provienen de la representación en ecuación (42), todo proceso con dicha representación tiene el mismo proceso de descomposición. La representación de un proceso estacionario en tendencia (43) también puede tener la misma descomposición pero con el componente de *random walk* ( $z_t$ ) igual a cero.

Generalizando esta idea, Cochrane (1991) de hecho argumenta que siempre existe una representación de un proceso estacionario en diferencias cuyos estimadores y/o estadísticos de prueba son parecidos a los estimadores y/o estadísticos de prueba



derivados de un proceso estacionario. Esto lo demuestra con un proceso estacionario en diferencias cuya función de verosimilitud y autocorrelación son parecidas a las de una serie estacionaria en tendencia.

Schwert (1987), Lo y MacKindlay (1989) y Blough (1988) habían demostrado previamente que las pruebas de raíz unitaria o de estacionariedad en tendencia pueden tener un bajo poder estadístico, especialmente en muestras pequeñas, y cuando la raíz es menor pero muy cercano uno. La importancia de esto radica en que la teoría asintótica sobre la distribución de algunos estimadores o estadísticos de prueba es muy sensible a la existencia de una raíz unitaria en la serie. De hecho, Cochrane (1991) argumenta que es discontinua en la medida que la raíz se aproxima a uno o cuando la componente de *random walk* se aproxima a cero. Por lo tanto, en muchos casos aun cuando sea posible distinguir entre procesos estacionarios y con raíz unitaria (o procesos cercanos a la raíz unitaria), es difícil determinar cuál es la mejor aproximación a la distribución de los estimadores y estadísticos de prueba en muestras chicas.

## 6. Poder estadístico y tamaño de una prueba

### 6.1 Prueba de hipótesis

Una prueba de hipótesis se define como una regla de decisión o un procedimiento que permite decidir entre dos opciones rivales. Más formalmente, suponiendo que un espacio de parámetros  $\Theta$  se puede separar en dos conjuntos disjuntos  $\Theta_0$  y  $\Theta_1$ , que  $\underline{x} = \{x_1, \dots, x_n\}$  es una muestra de tamaño  $n$  de variables aleatorias, y que  $x \in \mathcal{X}$ , donde  $\mathcal{X}$  es el espacio muestral de todas las posibles muestras de  $x$ . Además, suponiendo que  $\underline{x}$  tiene una distribución  $f(\underline{x})$  que pertenece a la familia  $\mathcal{F} = \{f(x, \theta): \theta \in \Theta\}$ , donde  $\theta$  es el parámetro o vector de parámetros de interés. Se desea conocer si  $\theta$  está contenido en  $\Theta_0$  o en  $\Theta_1$ , es decir, se quiere probar las siguientes hipótesis:  $H_0: \theta \in \Theta_0$  vs  $H_1: \theta \in \Theta_1$ .  $H_0$  se conoce como la hipótesis nula de la prueba y  $H_1$  es la hipótesis alternativa. Una regla de decisión para elegir entre éstas opciones es rechazar la hipótesis nula si  $\underline{x} \in \mathcal{R}$  y no rechazar la hipótesis nula si  $\underline{x} \in \mathcal{R}^c$ , donde  $\mathcal{R}$  es una región definida de rechazo en todo el espacio muestral  $\mathcal{X}$ , y  $\mathcal{R}^c$  es el complemento o región de no rechazo. En terminos binarios, esta regla de decisión se puede definir como:

$$\phi(\underline{x}) = \begin{cases} 1, & \underline{x} \in \mathcal{R} \\ 0, & \underline{x} \in \mathcal{R}^c \end{cases} \quad (48)$$

Es decir, si la regla de decisión es igual a uno,  $\phi(\underline{x}) = 1$ , se rechaza la hipótesis nula y, por lo tanto, la decisión es  $\theta \in \Theta_0$ . De forma contraria, cuando la regla de decisión es igual a cero,  $\phi(\underline{x}) = 0$ , la hipótesis nula no se rechaza, y la decisión es  $\theta \in \Theta_1$ . La función  $\phi(\cdot)$  se conoce como función de prueba, y depende de los datos a través del estadístico de prueba,  $T = T(\underline{x})$ .

Las funciones de prueba no son infalibles, por lo que es importante calcular la probabilidad de cometer errores. Por ejemplo, si  $\theta \in \Theta_0$  pero el resultado de la regla de decisión es  $\phi(\underline{x}) = 1$ , entonces se comete lo que se conoce como error tipo I, es decir, rechazar la hipótesis nula cuando es cierta. La probabilidad de cometer error tipo I está definida como:

$$P_\theta[\phi(\underline{x}) = 1] = E_\theta[\phi(\underline{x})] \quad (\theta \in \Theta_0) \quad (49)$$

Por el contrario, si  $\theta \in \Theta_1$  y el resultado de la regla de decisión es  $\phi(\underline{x}) = 0$ , entonces se comete lo que se conoce como error tipo II, es decir, no rechazar la hipótesis nula cuando es falsa. La probabilidad de cometer este error es la siguiente:

$$P_\theta[\phi(\underline{x}) = 0] = 1 - E_\theta[\phi(\underline{x})] \quad (\theta \in \Theta_1) \quad (50)$$

Lo ideal es buscar una función de prueba que minimice la probabilidad de cometer ambos errores; sin embargo, existe un *trad-off* entre ambos. Por ejemplo, definiendo dos funciones de prueba,  $\phi_1$  y  $\phi_2$ :

$$\begin{aligned} \mathcal{R}_1 &= \{\underline{x}: \phi_1(\underline{x}) = 1\} \\ \mathcal{R}_2 &= \{\underline{x}: \phi_2(\underline{x}) = 1\} \end{aligned} \quad (51)$$

Donde  $\mathcal{R}_1 \subset \mathcal{R}_2$ , lo cual implica que:

$$E_\theta[\phi_1(\underline{x})] \leq E_\theta[\phi_2(\underline{x})] \quad (52)$$

Y, por lo tanto, esto también implica que:

$$1 - E_{\theta}[\phi_1(\underline{x})] \geq 1 - E_{\theta}[\phi_2(\underline{x})] \quad (53)$$

Por lo tanto, al intentar disminuir la probabilidad de un error cuando  $\theta \in \Theta_0$  se corre el riesgo de aumentar el otro error cuando  $\theta \in \Theta_1$ . El enfoque clásico para especificar una función de prueba,  $\phi(\cdot)$ , consiste en fijar la probabilidad de cometer error tipo I ( $\alpha > 0$ ) que cumpla:

$$E_{\theta}[\phi_1(\underline{x})] \leq \alpha \quad (54)$$

Una vez definido el nivel del test ( $\alpha$ ), se busca la función de prueba que maximice  $E_{\theta}[\phi_1(\underline{x})]$  para  $\theta \in \Theta_1$ . Definiendo una función de potencia (considerando a  $\phi$  fijo) como la probabilidad bajo  $\theta$  cuando la regla de decisión rechaza la hipótesis nula:  $\pi_{\phi}(\theta) = P_{\theta}[\phi(\underline{x}) = 1]$ , donde:

$$\pi_{\phi}(\theta) = \begin{cases} \text{Probabilidad de error tipo I, si } \theta \in \Theta_0 \\ 1 - \text{prob. error tipo II, si } \theta \in \Theta_1 \end{cases} \quad (55)$$

Entonces, se busca que  $\pi_{\phi}(\theta_0)$  sea bajo, porque está relacionado con el error tipo I, y que  $\pi_{\phi}(\theta_1)$  dado que está relacionado con la potencia de la prueba, aunque ambos están relacionados inversamente. Habiendo fijado el nivel de la prueba ( $\alpha$ ) previamente, las reglas de decisión que sirven son aquellas que cumplen la condición de  $\pi_{\phi}(\theta_0) \leq \alpha$  eligiendo la de mayor  $\pi_{\phi}(\theta_1)$ . Si bien en la práctica no se conoce el valor verdadero de  $\alpha$ , por el Lema de Neyman-Pearson se garantiza que el nivel de una prueba es igual a  $\text{Sup}_{\theta \in \Theta_0} \pi_{\phi} \leq \alpha$ . Un test de nivel  $\alpha$  de  $H_0: \theta \in \Theta_0$  vs  $H_1: \theta \in \Theta_1$  se dice más potente para detectar a  $\theta_1 \in \Theta$  si para cualquier otro test  $\phi'$  también de nivel  $\alpha$  se garantiza que  $\pi_{\phi}(\theta_1) \geq \pi_{\phi'}(\theta_1)$ .

## 6.2 Poder estadístico de una prueba de hipótesis

El poder de un test se define como la probabilidad de rechazar una hipótesis nula que es falsa:

$$\text{Poder del test} = \text{Pr}(\text{Rechazar } H_0 \mid H_0 \text{ es falsa}) \quad (56)$$

También se puede definir el poder de un test en términos del error tipo II. Se comete error tipo II (denominado por la letra griega beta,  $\beta$ ) cuando no se rechaza una hipótesis

nula que es falsa. En ese sentido, el poder de un test también se define como uno menos la probabilidad de cometer dicho error:

$$\text{Poder del test} = 1 - \beta \quad (57)$$

Un problema con las pruebas de raíz unitaria es que generalmente tienen un poder estadístico bajo. El bajo poder estadístico lleva a que las pruebas fallen en rechazar la hipótesis nula más de lo habitual, concluyendo erróneamente que las series tienen una raíz unitaria.

Para realizar la estimación empírica del poder estadístico de las pruebas de raíz unitaria, se calcula la tasa de rechazo cuando la hipótesis nula es falsa. La proporción de veces que se rechaza la hipótesis nula es, precisamente, el poder empírico del test. Es decir, es la proporción de veces en que la prueba rechaza cuando efectivamente tiene que rechazar.

### 6.3 Tamaño empírico de una prueba

El tamaño de un test estadístico –o también llamado el nivel de significancia de la prueba– es la probabilidad de rechazar una hipótesis nula cuando es verdadera; es decir, es la probabilidad de cometer error tipo I. Generalmente, el tamaño del test es denotado por la letra griega alfa ( $\alpha$ ):

$$\text{Tamaño del test} = \Pr(\text{Rechazar } H_0 \mid H_0 \text{ es verdadero}) \quad (58)$$

Un test tiene propiedades pobres en muestras pequeñas si el valor reportado del *p-valor* es una estimación pobre del verdadero tamaño del test. Comúnmente, el principal problema es que el *p-valor* reportado es mucho menor que el verdadero tamaño del test, por lo que se tiende a rechazar la hipótesis nula más de lo que realmente se debería (Cameron y Trivedi, 2009).

## 7. Proceso generador de datos (DGP)

Schwert (1987) afirma que existen muchas razones para pensar que la mayoría de las series macroeconómicas (o al menos una buena cantidad) si bien pueden seguir un proceso autoregresivo de orden  $p$ ,  $AR(p)$ , generalmente contienen un componente de

promedio ponderado. Schwert (1987) argumenta que, como bien mostró Box y Jenkins (1976), la suma de dos variables aleatorias no correlacionadas (una de las cuales sigue un proceso  $ARIMA(p, d, q)$  y la otra no está serialmente correlacionada), siguen de forma conjunta un proceso  $ARIMA(p, d, Q)$  donde  $Q = \max \{(p + d), q\}$ . Por lo tanto, si una variable económica al momento de medirla tiene un error aleatorio y sigue un proceso puramente  $AR(p)$ , el resultado final contendrá un componente de promedio ponderado. Por ejemplo, si una serie de tiempo sigue un proceso *random walk*, al cual se le agrega un error independiente y no está serialmente correlacionado, el resultado será una serie que sigue un proceso  $ARIMA(0,1,1)$ . Series agregadas no estacionarias también lleva a procesos  $ARIMA(0,1,1)$ .

Sin embargo, como concluye el propio Schwert (1987), si bien existen muchas razones técnicas para pensar que las series económicas no sigue un proceso puramente  $AR(p)$ , el argumento más sólido a favor de utilizar procesos ARIMA es que éstos modelos generalmente ajustan bien a los datos. Análisis empíricos han utilizado modelos ARIMA para modelar series económicas como la tasa de inflación, el índice de precios al consumidor, la tasa de desempleo, entre otras.

En razón de lo anterior, en el presente trabajo se decidió trabajar con dos modelos generales. Un primer modelo es un proceso autoregresivo,  $AR(p)$  de la forma siguiente:

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad t = 1, 2, \dots, N \quad (59)$$

Donde  $\varepsilon_t \sim N(0,1)$ , e  $y_t \sim N(0,1)$ . En el software R,  $y_t$  se construye como la suma acumulada de una variable aleatoria que proviene de una distribución normal con media cero y varianza igual a uno. El comando en R es "cumsum(rnorm(y))". El proceso  $AR(1)$  se corre con el comando "arima.sim" el cual genera simulaciones a partir de un proceso ARMA previamente definido.

En este caso en concreto, se decidió realizar un experimento Monte Carlo de 1 000 simulaciones, con muestras de tamaño 30, 50, 100, 250 y 500. El valor del componente autoregresivo ( $\rho$ ) varía entre 0.05 y 0.95. Mas formalmente,  $\rho \in \{0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00\}$ . El objetivo es evaluar el comportamiento de las pruebas de raíz unitaria cuando el valor del componente autoregresivo se aproxima a los extremos de uno y cero, y varía el tamaño de la muestra. En primera instancia, se esperaría que el poder estadístico de

las pruebas de raíz unitaria tenga una relación inversa con el valor del componente autoregresivo y positiva con el tamaño de la muestra. Esto significa que el poder estadístico debería tender a disminuir cuando  $p$  se acerque a uno y a mejorar cuando el tamaño de muestra aumenta.

El segundo modelo es un modelo autoregresivo de promedio móvil,  $ARMA(p,q)$ , de primer orden tanto para el componente autoregresivo ( $p$ ), como para el componente de promedio móvil ( $q$ ). La especificación del modelo es la siguiente:

$$y_t = \rho_1 y_{t-1} + \varepsilon_t + \theta_1 \varepsilon_{t-1}, \quad t = 1, 2, \dots, N \quad (60)$$

Donde  $\varepsilon_t \sim N(0,1)$ , e  $y_t \sim N(0,1)$ . Al igual que en el caso del modelo  $AR(p)$ ,  $y_t$  se construye como la suma acumulada de una variable aleatoria que proviene de una distribución normal con media cero y varianza igual a uno. El componente de promedio móvil también varía entre 0.05 y 1. Mas formalmente,  $q \in \{0.05, 0.10, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, 0.60, 0.65, 0.70, 0.75, 0.80, 0.85, 0.90, 0.95, 1.00\}$ .

## 8. Resultados

Dado que se parte de una especificación sin raíz unitaria en ambos modelos cuando el componente autoregresivo es menor que uno ( $p < 1.0$ ), la cantidad de veces en que las pruebas Dickey-Fuller (1981) y Phillips y Perron (1988) rechazan la hipótesis nula de raíz unitaria es el poder estadístico de la prueba calculado de forma empírica. En otras palabras, es la cantidad de veces en que las pruebas son capaces de tomar una decisión correcta, rechazando una hipótesis nula cuando ésta es falsa. En la práctica, de las 1 000 simulaciones se cuentan los  $p$ -valores que arrojan la pruebas menores a un nivel de significancia de 1 y 5%. La proporción respecto del total de simulaciones es, precisamente, el poder estadístico de la prueba.

De forma contraria, el tamaño empírico de la prueba se calcula cuando el valor del componente autoregresivo en ambos modelos es igual a uno (la serie tiene raíz unitaria), y las pruebas Dickey-Fuller (1981) y Phillips y Perron (1988) rechazan la hipótesis nula de raíz unitaria. Es decir, cuando las pruebas se equivocan al rechazar una hipótesis nula que es cierta. Al igual que en el caso del poder estadístico, en la práctica se cuentan los

$p$ -valores menores a un nivel de significancia de 1 y 5%. La proporción respecto del total de simulaciones es, precisamente, el tamaño empírico de la prueba.

Para la prueba KPSS (1992) la lógica funciona a la inversa. La hipótesis nula de esta prueba es que la serie es estacionaria; por lo tanto, cuando el componente autoregresivo es menor a uno, la hipótesis nula es cierta. La cantidad de veces en que la prueba rechaza dicha hipótesis es el tamaño empírico de la prueba. Nuevamente, se cuentan la cantidad de  $p$ -valores menores a un nivel de significancia de 1 y 5%. De forma complementaria, cuando el valor del componente autoregresivo es igual a uno, la serie es no estacionaria y, por lo tanto, la hipótesis nula no es cierta. La cantidad de veces en que la prueba rechaza a un nivel de 1 y 5% dicha hipótesis, viene a ser cálculo empírico del poder estadístico de la prueba.

En total se corrieron 10 especificaciones de modelos, al considerar las combinaciones entre ambos modelos y los diferentes tamaños de muestras. En todos los casos, los componentes autoregresivos y de promedio móvil varían entre 0.05 y 1. Las diferentes especificaciones de los modelos se detallan a continuación:

**Tabla 1. Especificación de los modelos y tamaño de muestra**

No.	Modelo	Especificación
1	Modelo 1a	AR(1), n=30
2	Modelo 1b	AR(1), n=50
3	Modelo 1c	AR(1), n=100
4	Modelo 1d	AR(1), n=250
5	Modelo 1e	AR(1), n=500
6	Modelo 2a	ARMA(1,1), n=30
7	Modelo 2b	ARMA(1,1), n=50
8	Modelo 2c	ARMA(1,1), n=100
9	Modelo 2d	ARMA(1,1), n=250
10	Modelo 2e	ARMA(1,1), n=500

**Modelo 1a y 2a: AR( $p$ ) y ARMA( $p,q$ ),  $n = 30$**

Los resultados de las simulaciones reflejan que, con un tamaño de muestra igual a treinta, la prueba que tiene un mejor desempeño es Phillips y Perron (1988). Para el modelo AR( $p$ ), el poder estadístico de dicha prueba es aceptable (mayor o igual a 80%) para valores del componente autoregresivo menores a 0.20 y 0.35, cuando el nivel de significancia de la prueba es 1 y 5%, respectivamente (ver Tabla 2). Después de dichos valores y en la medida que  $p$  se acerca a uno, el poder estadístico de la prueba cae drásticamente. Cabe destacar que el poder de esta prueba a un nivel de significancia de 5% es de 0.99 para valores pequeños del componente autoregresivo. La prueba Dickey-

Fuller (1981) tiene un poder estadístico bastante bajo (no mayor de 5%) para todos los valores del componente autoregresivo (ver gráfico 1, panel A y B). Por su parte, el poder estadístico de la prueba KPPS (1992) es relativamente alto (58 y 75%) cuando el nivel de de significancia de la prueba se fija en 1 y 5%, respectivamente.

El tamaño empírico de la prueba es bastante cercano al nominal (1 y 5%) en el caso de las pruebas Phillips y Perron (1988) y Dickey-Fuller (1981). En el caso de la prueba KPSS (1992), el tamaño de la prueba es aceptable hasta valores de 0.15 en el componente autoregresivo. Después de este valor, el tamaño de la prueba crece aceleradamente en la medida que  $p$  se acerca a uno. Con un valor de  $\rho$  igual a 0.95 y un nivel de significancia de 5%, el tamaño empírico de la prueba alcanza un valor de 67% (ver gráfico 1, panel B).

**Tabla 2. AR( $p$ ) y ARMA( $p,q$ ),  $n=30$**

$p$ y $q$	AR( $p$ )						ARMA( $p,q$ )					
	alpha = 0.01			alpha = 0.05			alpha = 0.01			alpha = 0.05		
	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**
0.05	0.05	0.88	0.01	0.19	0.99	0.04	0.04	0.80	0.01	0.17	0.98	0.05
0.10	0.04	0.79	0.01	0.18	0.97	0.06	0.05	0.68	0.01	0.17	0.94	0.07
0.15	0.04	0.72	0.02	0.18	0.96	0.06	0.04	0.47	0.01	0.15	0.85	0.08
0.20	0.04	0.67	0.01	0.14	0.94	0.09	0.04	0.31	0.01	0.13	0.71	0.09
0.25	0.04	0.58	0.02	0.14	0.91	0.09	0.04	0.17	0.03	0.15	0.57	0.10
0.30	0.04	0.48	0.02	0.16	0.85	0.10	0.04	0.07	0.03	0.14	0.40	0.13
0.35	0.04	0.39	0.02	0.14	0.78	0.11	0.03	0.04	0.04	0.14	0.28	0.15
0.40	0.04	0.32	0.02	0.14	0.69	0.12	0.04	0.01	0.04	0.12	0.16	0.16
0.45	0.03	0.24	0.04	0.11	0.60	0.17	0.02	0.01	0.05	0.12	0.10	0.20
0.50	0.02	0.16	0.05	0.10	0.53	0.17	0.03	0.00	0.07	0.12	0.07	0.22
0.55	0.03	0.12	0.05	0.10	0.42	0.19	0.03	0.00	0.08	0.13	0.03	0.24
0.60	0.02	0.09	0.08	0.09	0.31	0.23	0.04	0.00	0.11	0.11	0.02	0.28
0.65	0.01	0.06	0.09	0.07	0.25	0.29	0.03	0.00	0.14	0.12	0.01	0.35
0.70	0.02	0.04	0.12	0.08	0.19	0.30	0.03	0.00	0.18	0.12	0.01	0.38
0.75	0.01	0.02	0.19	0.07	0.14	0.40	0.03	0.00	0.22	0.11	0.01	0.44
0.80	0.01	0.02	0.21	0.06	0.13	0.45	0.03	0.00	0.26	0.11	0.00	0.47
0.85	0.01	0.01	0.30	0.06	0.08	0.52	0.03	0.00	0.34	0.12	0.00	0.55
0.90	0.02	0.01	0.37	0.06	0.06	0.59	0.04	0.00	0.41	0.12	0.00	0.60
0.95	0.01	0.01	0.46	0.05	0.05	0.67	0.04	0.00	0.50	0.11	0.00	0.70
1.00	0.01	0.01	0.58	0.04	0.05	0.75	0.01	0.01	0.59	0.05	0.04	0.75

\* Se calcula el poder estadístico de la prueba cuando el valor de componente autoregresivo está entre 0.05 y 0.95. El tamaño empírico de la prueba se calcula cuando  $\rho$  es igual a uno.

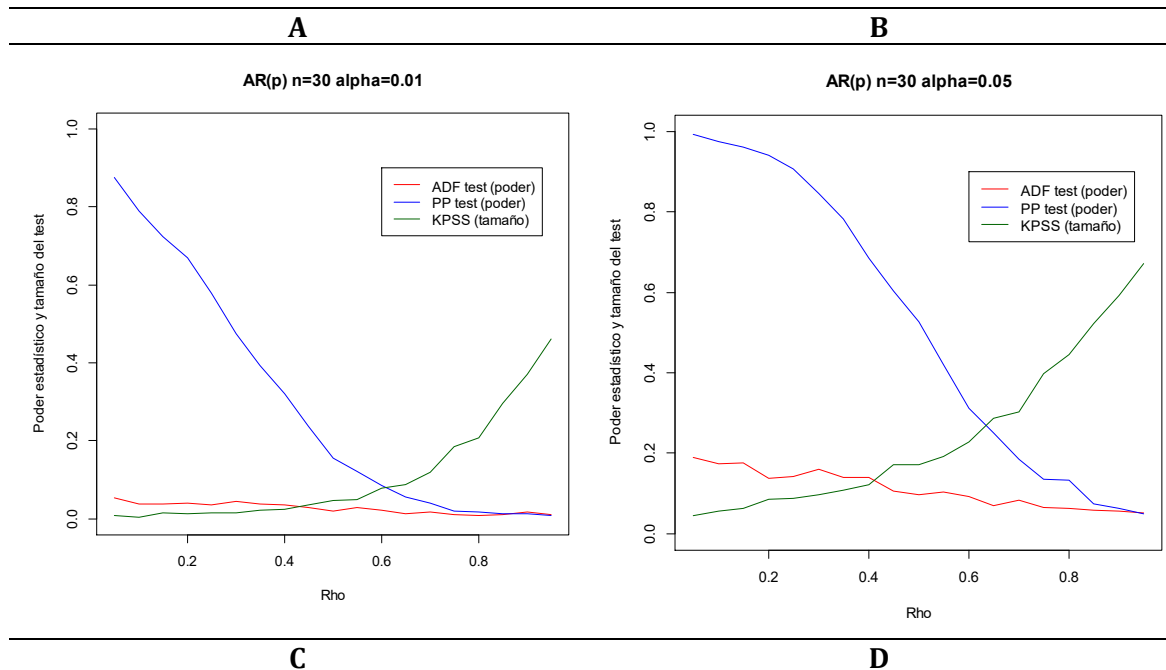
\*\* Se calcula el tamaño empírico de la prueba cuando el valor de componente autoregresivo está entre 0.05 y 0.95. El poder estadístico de la prueba se calcula cuando  $\rho$  es igual a uno.

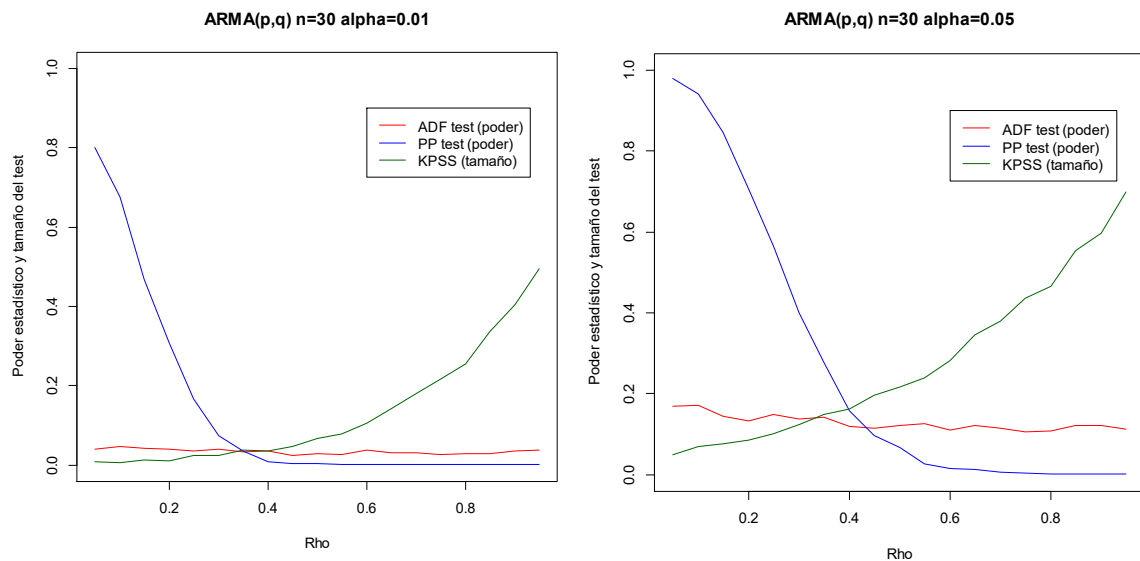
Fuente: Elaboración propia



Los resultados del modelo **ARMA** ( $p,q$ ) son bastante parecidos al  $AR(p)$  en términos del desempeño general de las pruebas (ver gráfico 1, panel C y D). El mejor desempeño nuevamente lo refleja la prueba Phillips y Perron (1988). Para valores bajos del componente autoregresivo, el poder estadístico es bastante aceptable (~80%) pero decrece rápidamente en la medida que éste crece y se acerca a uno. La prueba Dickey-Fuller (1981) nuevamente refleja un desempeño bastante pobre para todos los valores de  $\rho$ . El tamaño empírico de la prueba es bastante cercano al valor nominal en ambas pruebas. Por su parte, la prueba KPSS (1992) refleja un tamaño empírico cercano al nominal cuando el componente autoregresivo es menor o igual a 0.45 con un nivel de significancia de 1%. El poder estadístico de esta prueba es relativamente bueno (entre 60 y 75%), especialmente a un nivel de significancia de 5% (ver Tabla 2).

**Gráfico 1.  $AR(p)$  y  $ARMA(p,q)$ ,  $n=30$**





Fuente: Elaboración propia

### Modelo 1b y 2b: $AR(p)$ y $ARMA(p,q)$ , $n = 50$

En la segunda especificación de los modelos la muestra aumentó a 50. Como era de esperarse, el desempeño de todas las pruebas mejoraron. Nuevamente, el mejor desempeño lo refleja la prueba Phillips y Perron (1988). Como puede observarse en el gráfico 2 (panel A y B), para el modelo  $AR(p)$  esta prueba refleja un poder estadístico bastante aceptable cuando el valor del componente autoregresivo es menor o igual a 0.45 y 0.60, con un nivel de significancia de 1 y 5%, respectivamente. La prueba Dickey-Fuller (1981) mejoró, pero aún muestra un poder estadístico bastante bajo para valores de  $\rho$  cercanos a uno. El tamaño empírico de ambas pruebas es bastante cercano al valor nominal. Por su parte, la prueba KPSS (1992) muestra un poder estadístico relativamente alto, especialmente a un 5% de nivel de significancia (ver tabla 2). Nuevamente, el tamaño de esta prueba es aceptable para valores pequeños del componente autoregresivo, pero se aleja rápidamente del valor nominal en la medida  $\rho$  crece y se acerca a uno.

Tabla 3.  $AR(p)$  y  $ARMA(p,q)$ ,  $n=50$

$p$ y $q$	$AR(p)$						$ARMA(p,q)$					
	$\alpha = 0.01$			$\alpha = 0.05$			$\alpha = 0.01$			$\alpha = 0.05$		
	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**
0.05	0.25	1.00	0.01	0.56	1.00	0.05	0.20	1.00	0.01	0.52	1.00	0.06
0.10	0.22	1.00	0.01	0.52	1.00	0.05	0.19	1.00	0.02	0.49	1.00	0.07
0.15	0.22	1.00	0.02	0.52	1.00	0.07	0.14	0.97	0.02	0.45	1.00	0.08
0.20	0.16	0.99	0.02	0.47	1.00	0.08	0.16	0.88	0.02	0.43	1.00	0.10

0.25	0.17	0.99	0.02	0.46	1.00	0.09	0.13	0.75	0.03	0.42	0.98	0.12
0.30	0.15	0.96	0.02	0.42	1.00	0.10	0.12	0.52	0.04	0.38	0.92	0.14
0.35	0.14	0.92	0.03	0.38	1.00	0.12	0.10	0.30	0.05	0.34	0.80	0.16
0.40	0.11	0.87	0.04	0.36	0.99	0.14	0.12	0.14	0.06	0.36	0.67	0.19
0.45	0.11	0.77	0.05	0.35	0.98	0.16	0.10	0.07	0.09	0.34	0.52	0.22
0.50	0.10	0.63	0.06	0.30	0.94	0.19	0.09	0.03	0.09	0.29	0.32	0.23
0.55	0.08	0.52	0.08	0.27	0.87	0.23	0.10	0.01	0.11	0.30	0.20	0.27
0.60	0.07	0.39	0.11	0.22	0.79	0.25	0.08	0.01	0.14	0.28	0.12	0.31
0.65	0.06	0.27	0.16	0.20	0.64	0.32	0.08	0.00	0.15	0.25	0.07	0.34
0.70	0.04	0.16	0.17	0.17	0.51	0.33	0.08	0.00	0.19	0.23	0.04	0.40
0.75	0.04	0.09	0.22	0.16	0.36	0.42	0.07	0.00	0.25	0.22	0.02	0.47
0.80	0.03	0.05	0.27	0.11	0.25	0.50	0.05	0.00	0.33	0.17	0.01	0.55
0.85	0.01	0.05	0.38	0.09	0.16	0.59	0.04	0.00	0.41	0.17	0.01	0.64
0.90	0.01	0.02	0.52	0.06	0.10	0.71	0.03	0.00	0.53	0.13	0.00	0.72
0.95	0.01	0.01	0.62	0.05	0.05	0.80	0.03	0.00	0.66	0.12	0.01	0.82
1.00	0.01	0.01	0.74	0.04	0.04	0.86	0.02	0.01	0.73	0.06	0.05	0.87

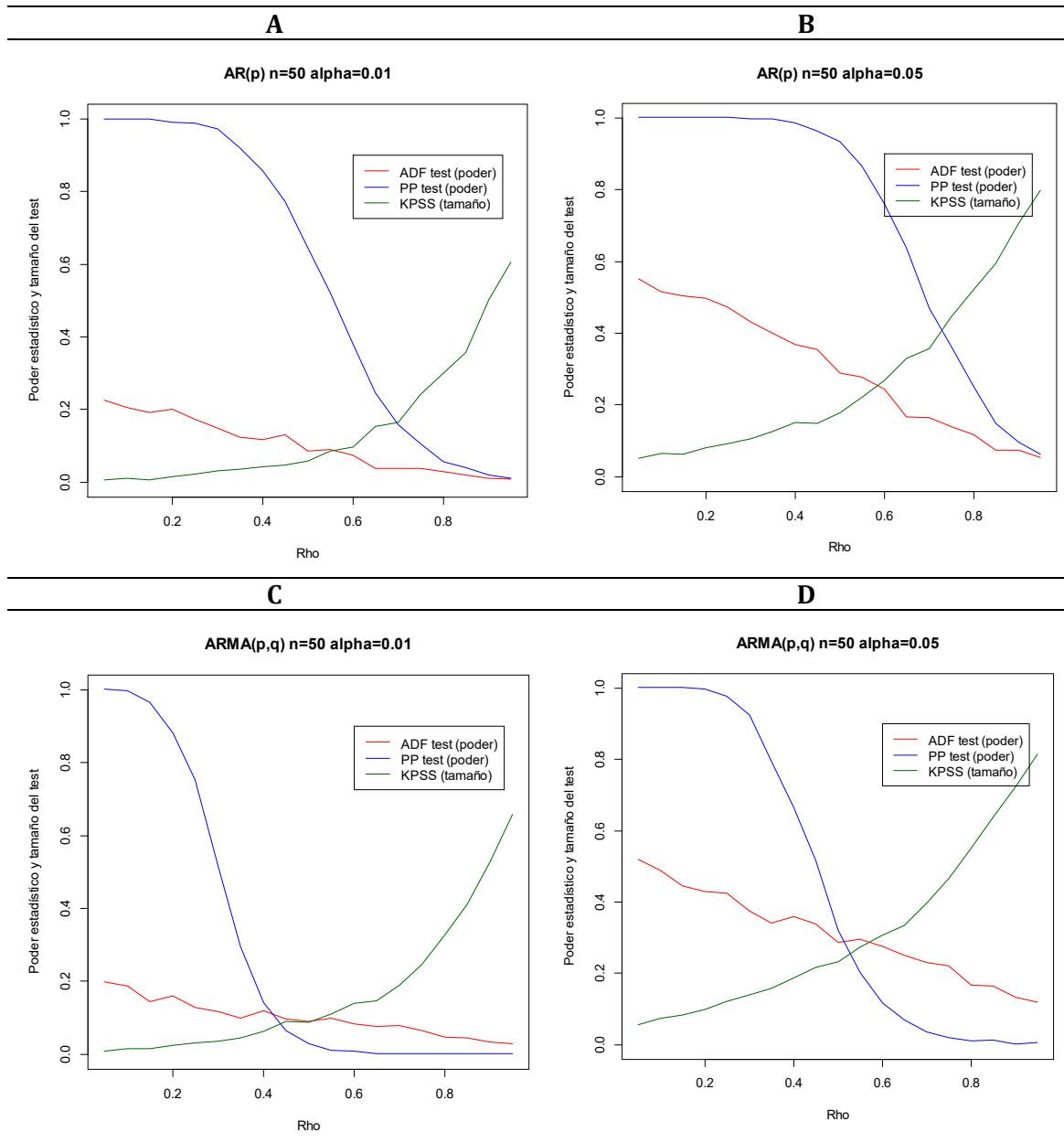
\* Se calcula el poder estadístico de la prueba cuando el valor de componente autoregresivo está entre 0.05 y 0.95. El tamaño empírico de la prueba se calcula cuando  $\rho$  es igual a uno.

\*\* Se calcula el tamaño empírico de la prueba cuando el valor de componente autoregresivo está entre 0.05 y 0.95. El poder estadístico de la prueba se calcula cuando  $\rho$  es igual a uno.

Fuente: Elaboración propia

Los resultados de las simulaciones del modelo **ARMA** ( $p,q$ ) reflejan que el mejor desempeño siempre viene dado por la prueba Phillips y Perron (1988): muestra un alto poder estadístico para valores bajos del componente autoregresivo y el tamaño empírico de la prueba es bastante cercano al nominal. La prueba Dickey-Fuller (1981) –si bien refleja un bajo poder estadístico– fue la que más mejoró su desempeño al aumentar la muestra de 30 a 50. Esto podría estar sugiriendo que dicha prueba es la más sensible al tamaño de muestra. El tamaño empírico de esta prueba se alejó ligeramente del valor nominal. Por su parte, la prueba KPSS (1992) mejoró en cuanto al poder estadístico pero aún refleja tamaños empíricos que rápidamente empeoran en la medida que  $\rho$  crece y se acerca a uno.

**Gráfico 2. AR(p) y ARMA(p,q), n=50**



Fuente: Elaboración propia

**Modelo 1c y 2c: AR(p) y ARMA(p,q), n = 100**

En general, el desempeño de todas las pruebas mejora al aumentar la muestra a 100. En el caso del modelo **AR(p)**, la prueba que nuevamente presentó el mejor desempeño fue Phillips y Perron (1988). El poder estadístico de la prueba es bastante alto cuando el valor del componente autoregresivo es menor o igual que 0.70 y 0.75, a un nivel de significancia de 1 y 5%, respectivamente.

Cabe destacar que una vez más la prueba que más mejoró su desempeño fue Dickey-Fuller (1981). Si bien el poder estadístico de la prueba aún no alcanza valores

aceptables para un nivel de significancia de 1%, cuando el valor de  $\rho$  es menor que 0.5, el poder estadístico de la prueba a un nivel de significancia de 5% es bastante alto y empieza a asemejarse a los niveles mostrados por la prueba Phillips y Perron (1988) (ver gráfico 3, panel B). El tamaño empírico de la prueba es ligeramente inferior al valor nominal en el caso de Dickey-Fuller (1981), y ligeramente superior en el caso de Phillips y Perron (1988) (ver tabla 4). No se observa mayor diferencia en el comportamiento del tamaño empírico de la prueba KPSS (1992) al aumentar la muestra de 50 a 100. Para valores muy pequeños del componente autoregresivo, el tamaño empírico es satisfactorio pero rápidamente crece en la medida que se acerca a uno. Esta prueba mejoró bastante en términos del poder estadístico.

**Tabla 4. AR( $p$ ) y ARMA( $p,q$ ),  $n=100$**

$p$ y $q$	AR( $p$ )						ARMA( $p,q$ )					
	alpha = 0.01			alpha = 0.05			alpha = 0.01			alpha = 0.05		
	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**
0.05	0.75	1.00	0.01	0.95	1.00	0.05	0.75	1.00	0.01	0.95	1.00	0.07
0.10	0.70	1.00	0.02	0.93	1.00	0.07	0.65	1.00	0.01	0.93	1.00	0.06
0.15	0.68	1.00	0.02	0.93	1.00	0.08	0.66	1.00	0.02	0.89	1.00	0.07
0.20	0.67	1.00	0.02	0.91	1.00	0.09	0.63	1.00	0.02	0.89	1.00	0.09
0.25	0.61	1.00	0.02	0.90	1.00	0.09	0.57	1.00	0.02	0.87	1.00	0.09
0.30	0.63	1.00	0.01	0.89	1.00	0.08	0.53	1.00	0.02	0.82	1.00	0.10
0.35	0.56	1.00	0.01	0.87	1.00	0.09	0.46	1.00	0.04	0.79	1.00	0.12
0.40	0.51	1.00	0.03	0.83	1.00	0.10	0.39	0.99	0.04	0.76	1.00	0.15
0.45	0.47	1.00	0.03	0.80	1.00	0.11	0.38	0.97	0.04	0.73	1.00	0.15
0.50	0.42	1.00	0.04	0.76	1.00	0.14	0.31	0.87	0.04	0.64	0.99	0.16
0.55	0.37	1.00	0.05	0.71	1.00	0.14	0.26	0.77	0.07	0.58	0.99	0.20
0.60	0.30	0.99	0.08	0.66	1.00	0.20	0.23	0.59	0.08	0.55	0.95	0.21
0.65	0.25	0.95	0.10	0.58	1.00	0.23	0.13	0.35	0.11	0.40	0.84	0.27
0.70	0.19	0.83	0.13	0.49	0.98	0.29	0.09	0.24	0.12	0.34	0.71	0.29
0.75	0.16	0.64	0.16	0.41	0.92	0.34	0.06	0.11	0.18	0.24	0.52	0.35
0.80	0.08	0.40	0.21	0.30	0.75	0.41	0.05	0.06	0.23	0.18	0.34	0.43
0.85	0.05	0.18	0.30	0.20	0.52	0.51	0.02	0.01	0.32	0.10	0.13	0.53
0.90	0.03	0.07	0.41	0.12	0.26	0.66	0.01	0.01	0.40	0.07	0.06	0.65
0.95	0.03	0.03	0.62	0.08	0.12	0.80	0.01	0.00	0.59	0.04	0.03	0.80
1.00	0.01	0.02	0.84	0.04	0.07	0.93	0.01	0.02	0.80	0.05	0.08	0.91

\* Se calcula el poder estadístico de la prueba cuando el valor de componente autoregresivo está entre 0.05 y 0.95. El tamaño empírico de la prueba se calcula cuando  $\rho$  es igual a uno.

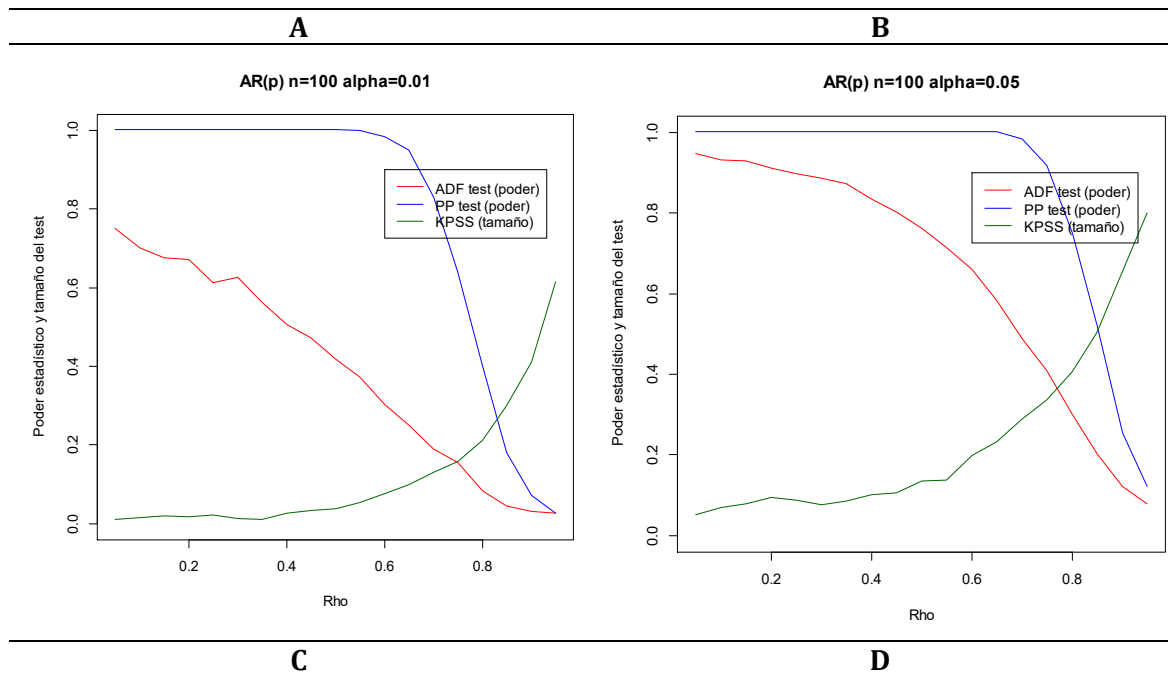
\*\* Se calcula el tamaño empírico de la prueba cuando el valor de componente autoregresivo está entre 0.05 y 0.95. El poder estadístico de la prueba se calcula cuando  $\rho$  es igual a uno.

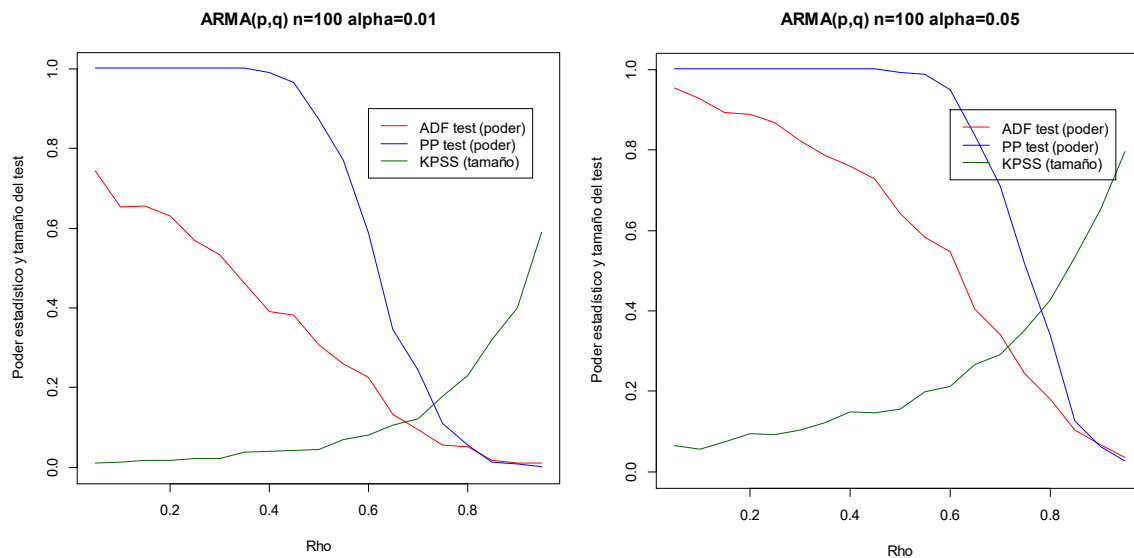
Fuente: Elaboración propia

Los resultados de la especificación ARMA( $p,q$ ) son muy parecidos a los del modelo AR( $p$ ) y muestran una mejora sustancial con respecto a la misma especificación pero

con un tamaño de muestra de 50. El mejor desempeño lo muestra la prueba Phillips y Perron (1988). El poder estadístico de esta prueba es aceptable para valores iguales o menores que 0.65 (a un nivel de significancia de 5%) y de 0.50 (a un nivel de significancia de 1%). Nuevamente, la prueba que más mejoró su desempeño fue Dickey-Fuller (1981), pues el poder estadístico de la prueba es aceptable para valores del componente autoregresivo iguales o menores a 0.30, a un nivel de confianza de 5%. El tamaño empírico de esta prueba es igual al valor nominal, y ligeramente superior en el caso de la prueba Phillips y Perron (1988). La prueba KPSS(1992) no mejoró en términos del tamaño empírico, pero si aumentó el poder estadístico a niveles cercanos al 90% (ver tabla 4).

**Gráfico 3. AR(p) y ARMA(p,q), n=100**





Fuente: Elaboración propia

### Modelo 1d y 2d: $AR(p)$ y $ARMA(p,q)$ , $n = 250$

Los resultados de las simulaciones refleja que todas las pruebas mejoraron sustancialmente su desempeño al aumentar la muestra a 250 (ver gráfico 4). El comportamiento de las pruebas Dickey-Fuller (1981) y Phillips y Perron (1988) se asemejan cada vez más. En el caso del modelo  $AR(p)$ , la prueba Phillips y Perron (1988) muestra un alto poder estadístico cuando el componente autoregresivo es menor o igual a 0.90 y 0.85, a un nivel de significancia de 1 y 5%, respectivamente. Esto es importante pues significa que esta prueba empieza a poder discriminar adecuadamente entre una serie estacionaria y no estacionaria (o procesos cercanos a la raíz unitaria) con un tamaño de muestra de 250. Una vez más, los resultados muestran que la prueba Dickey-Fuller (1981) fue la que más mejoró su desempeño en términos del poder estadístico. Su comportamiento se asemeja mucho a la prueba Phillips y Perron (1988), pues para valores de  $\rho$  de 0.75 y 0.80 (al 1 y 5% de nivel de significancia), el poder estadístico es alto. El tamaño empírico de la prueba es igual al valor nominal en el caso de Dickey-Fuller (1981), y ligeramente superior en el caso de Phillips y Perron (1988) (ver tabla 5). La prueba KPSS (1992) mostró mejoras en el cálculo empírico del tamaño de la prueba cuando el nivel de significancia es de 1%, pero los resultados fueron invariantes cuando el nivel de significancia se fija en 5%. El poder estadístico de esta prueba mejoró a niveles cercanos al 98%.

**Tabla 5. AR( $p$ ) y ARMA( $p,q$ ),  $n=250$**

$p$ y $q$	AR( $p$ )						ARMA( $p,q$ )					
	alpha = 0.01			alpha = 0.05			alpha = 0.01			alpha = 0.05		
	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**
0.05	1.00	1.00	0.01	1.00	1.00	0.05	1.00	1.00	0.02	1.00	1.00	0.06
0.10	1.00	1.00	0.02	1.00	1.00	0.06	1.00	1.00	0.01	1.00	1.00	0.06
0.15	1.00	1.00	0.01	1.00	1.00	0.06	1.00	1.00	0.01	1.00	1.00	0.07
0.20	1.00	1.00	0.01	1.00	1.00	0.06	1.00	1.00	0.02	1.00	1.00	0.09
0.25	1.00	1.00	0.02	1.00	1.00	0.07	1.00	1.00	0.01	1.00	1.00	0.07
0.30	1.00	1.00	0.01	1.00	1.00	0.08	1.00	1.00	0.02	1.00	1.00	0.10
0.35	1.00	1.00	0.03	1.00	1.00	0.10	1.00	1.00	0.02	1.00	1.00	0.10
0.40	1.00	1.00	0.03	1.00	1.00	0.09	0.99	1.00	0.03	1.00	1.00	0.11
0.45	1.00	1.00	0.03	1.00	1.00	0.12	0.98	1.00	0.03	1.00	1.00	0.11
0.50	0.99	1.00	0.03	1.00	1.00	0.12	0.99	1.00	0.03	1.00	1.00	0.12
0.55	0.98	1.00	0.03	1.00	1.00	0.12	0.97	1.00	0.05	1.00	1.00	0.16
0.60	0.98	1.00	0.05	1.00	1.00	0.15	0.93	1.00	0.06	1.00	1.00	0.16
0.65	0.96	1.00	0.08	1.00	1.00	0.18	0.89	1.00	0.06	0.99	1.00	0.19
0.70	0.90	1.00	0.10	0.99	1.00	0.23	0.82	1.00	0.08	0.98	1.00	0.21
0.75	0.82	1.00	0.13	0.98	1.00	0.26	0.68	0.99	0.14	0.92	1.00	0.31
0.80	0.72	1.00	0.17	0.94	1.00	0.33	0.49	0.92	0.19	0.85	1.00	0.37
0.85	0.43	0.97	0.23	0.78	1.00	0.44	0.29	0.66	0.23	0.66	0.96	0.42
0.90	0.21	0.61	0.40	0.51	0.90	0.62	0.12	0.23	0.40	0.37	0.66	0.63
0.95	0.04	0.12	0.65	0.20	0.36	0.84	0.02	0.03	0.64	0.11	0.18	0.83
1.00	0.01	0.02	0.95	0.05	0.07	0.98	0.01	0.02	0.94	0.05	0.07	0.98

\* Se calcula el poder estadístico de la prueba cuando el valor de componente autoregresivo está entre 0.05 y 0.95. El tamaño empírico de la prueba se calcula cuando  $\rho$  es igual a uno.

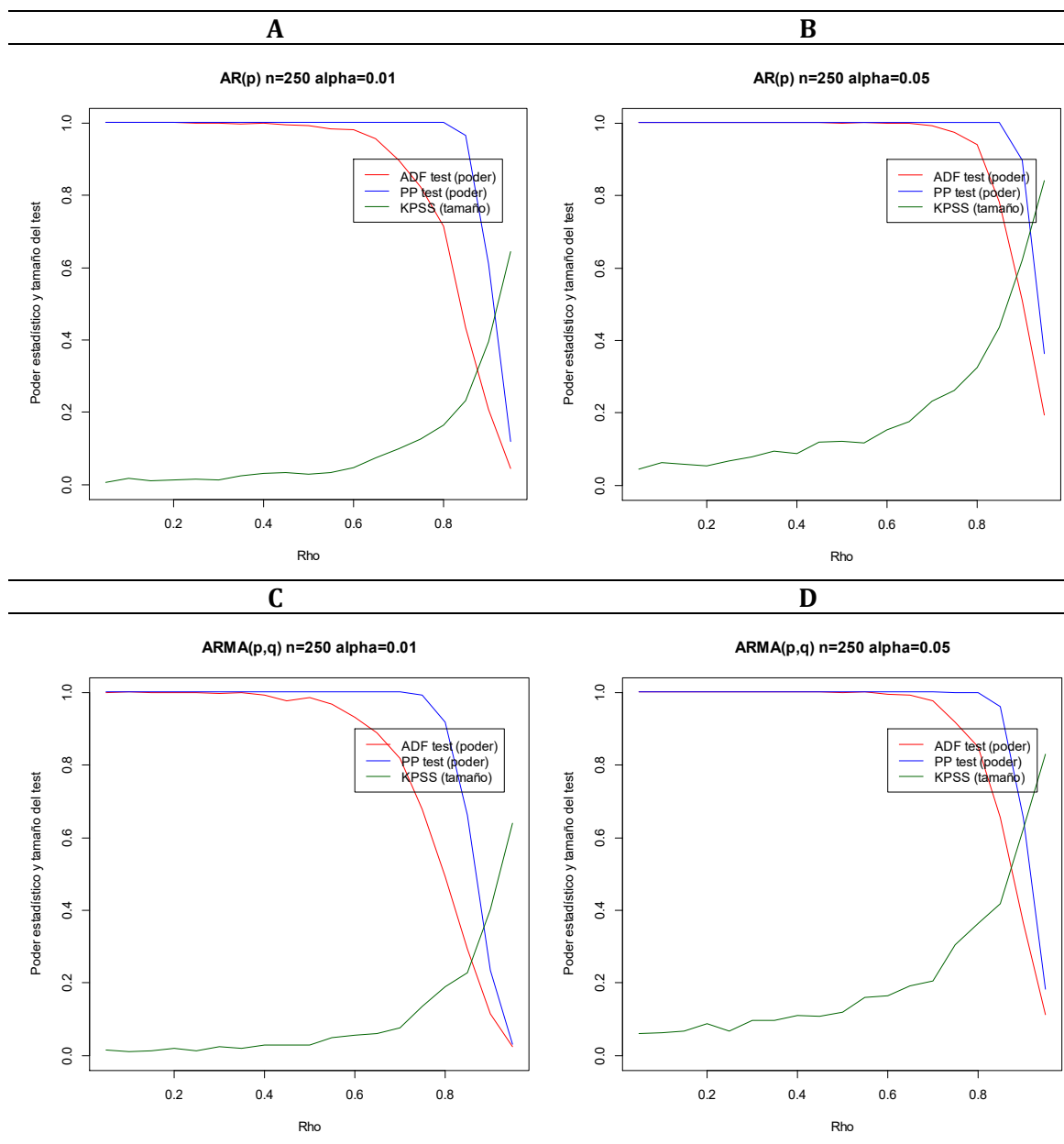
\*\* Se calcula el tamaño empírico de la prueba cuando el valor de componente autoregresivo está entre 0.05 y 0.95. El poder estadístico de la prueba se calcula cuando  $\rho$  es igual a uno.

Fuente: Elaboración propia

En general, las pruebas mostraban un mejor desempeño bajo el modelo AR( $p$ ); sin embargo, con una muestra de 250, el comportamiento entre ambos modelos, AR( $p$ ) y ARMA( $p,q$ ), son cada vez más parecidos (ver gráfico 4). El poder estadístico de la prueba Phillips y Perron (1988) es bastante aceptable en casi toda la grilla de valores del componente autoregresivo, y la prueba Dickey-Fuller (1981) mejoró considerablemente su desempeño acercándose cada vez más a los niveles de la prueba de Phillips y Perron (1988). Al igual que en el modelo AR( $p$ ), el tamaño empírico de la prueba es ligeramente inferior al valor nominal en el caso de Dickey-Fuller (1981), y ligeramente superior en el caso de Phillips y Perron (1988) (ver tabla 5). La prueba KPSS (1992) no mostró mejoras en el tamaño empírico de la prueba. Cuando el valor del componente autoregresivo crece y se acerca a uno, el tamaño empírico de aleja considerablemente del valor nominal. El poder estadístico de esta prueba mejoró, acercándose a valores de 98%.



Gráfico 4.  $AR(p)$  y  $ARMA(p,q)$ ,  $n=250$



Fuente: Elaboración propia

### Modelo 1e y 2e: $AR(p)$ y $ARMA(p,q)$ , $n = 500$

Finalmente, cuando el tamaño de muestra aumenta a 500, el desempeño de todas las pruebas son prácticamente iguales (ver gráfico 5). Bajo el modelo  $AR(p)$ , el poder estadístico de las pruebas Phillips y Perron (1988) y Dickey-Fuller (1981) es muy alto para valores del componente autoregresivo de entre 0.90 y 0.95. En otras palabras, con un tamaño de muestra de 500 ambas pruebas tienen un alto poder para discriminar entre series estacionarias y no estacionarias (o procesos cercanos a la raíz unitaria).

Adicionalmente, el tamaño empírico de ambas pruebas es muy cercano al valor nominal

(ver tabla 6). Por su parte, la prueba KPSS (1992) se mantuvo en prácticamente el mismo nivel de poder estadístico, entre 96 y 98% (ver gráfico 5, panel A y B), pero no mejoró en cuanto al cálculo del tamaño empírico de la prueba (ver tabla 6).

**Tabla 6. AR( $p$ ) y ARMA( $p,q$ ),  $n=500$**

$p$ y $q$	AR( $p$ )						ARMA( $p,q$ )					
	alpha = 0.01			alpha = 0.05			alpha = 0.01			alpha = 0.05		
	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**	ADF*	PP*	KPSS**
0.05	1.00	1.00	0.01	1.00	1.00	0.05	1.00	1.00	0.01	1.00	1.00	0.05
0.10	1.00	1.00	0.01	1.00	1.00	0.05	1.00	1.00	0.01	1.00	1.00	0.06
0.15	1.00	1.00	0.01	1.00	1.00	0.06	1.00	1.00	0.02	1.00	1.00	0.08
0.20	1.00	1.00	0.02	1.00	1.00	0.07	1.00	1.00	0.01	1.00	1.00	0.06
0.25	1.00	1.00	0.02	1.00	1.00	0.06	1.00	1.00	0.01	1.00	1.00	0.06
0.30	1.00	1.00	0.02	1.00	1.00	0.07	1.00	1.00	0.02	1.00	1.00	0.07
0.35	1.00	1.00	0.02	1.00	1.00	0.07	1.00	1.00	0.02	1.00	1.00	0.10
0.40	1.00	1.00	0.02	1.00	1.00	0.08	1.00	1.00	0.03	1.00	1.00	0.08
0.45	1.00	1.00	0.02	1.00	1.00	0.08	1.00	1.00	0.02	1.00	1.00	0.08
0.50	1.00	1.00	0.03	1.00	1.00	0.10	1.00	1.00	0.03	1.00	1.00	0.09
0.55	1.00	1.00	0.03	1.00	1.00	0.10	1.00	1.00	0.04	1.00	1.00	0.10
0.60	1.00	1.00	0.04	1.00	1.00	0.12	1.00	1.00	0.05	1.00	1.00	0.13
0.65	1.00	1.00	0.04	1.00	1.00	0.15	1.00	1.00	0.04	1.00	1.00	0.15
0.70	1.00	1.00	0.06	1.00	1.00	0.15	1.00	1.00	0.05	1.00	1.00	0.14
0.75	1.00	1.00	0.06	1.00	1.00	0.17	1.00	1.00	0.09	1.00	1.00	0.22
0.80	1.00	1.00	0.12	1.00	1.00	0.26	1.00	1.00	0.10	1.00	1.00	0.24
0.85	0.99	1.00	0.15	1.00	1.00	0.33	0.99	1.00	0.18	1.00	1.00	0.35
0.90	0.85	1.00	0.27	0.99	1.00	0.47	0.91	0.98	0.28	0.99	1.00	0.48
0.95	0.30	0.60	0.51	0.67	0.90	0.73	0.38	0.31	0.51	0.73	0.73	0.73
1.00	0.01	0.01	0.96	0.04	0.04	0.99	0.02	0.01	0.97	0.04	0.05	0.99

\* Se calcula el poder estadístico de la prueba cuando el valor de componente autoregresivo está entre 0.05 y 0.95. El tamaño empírico de la prueba se calcula cuando  $\rho$  es igual a uno.

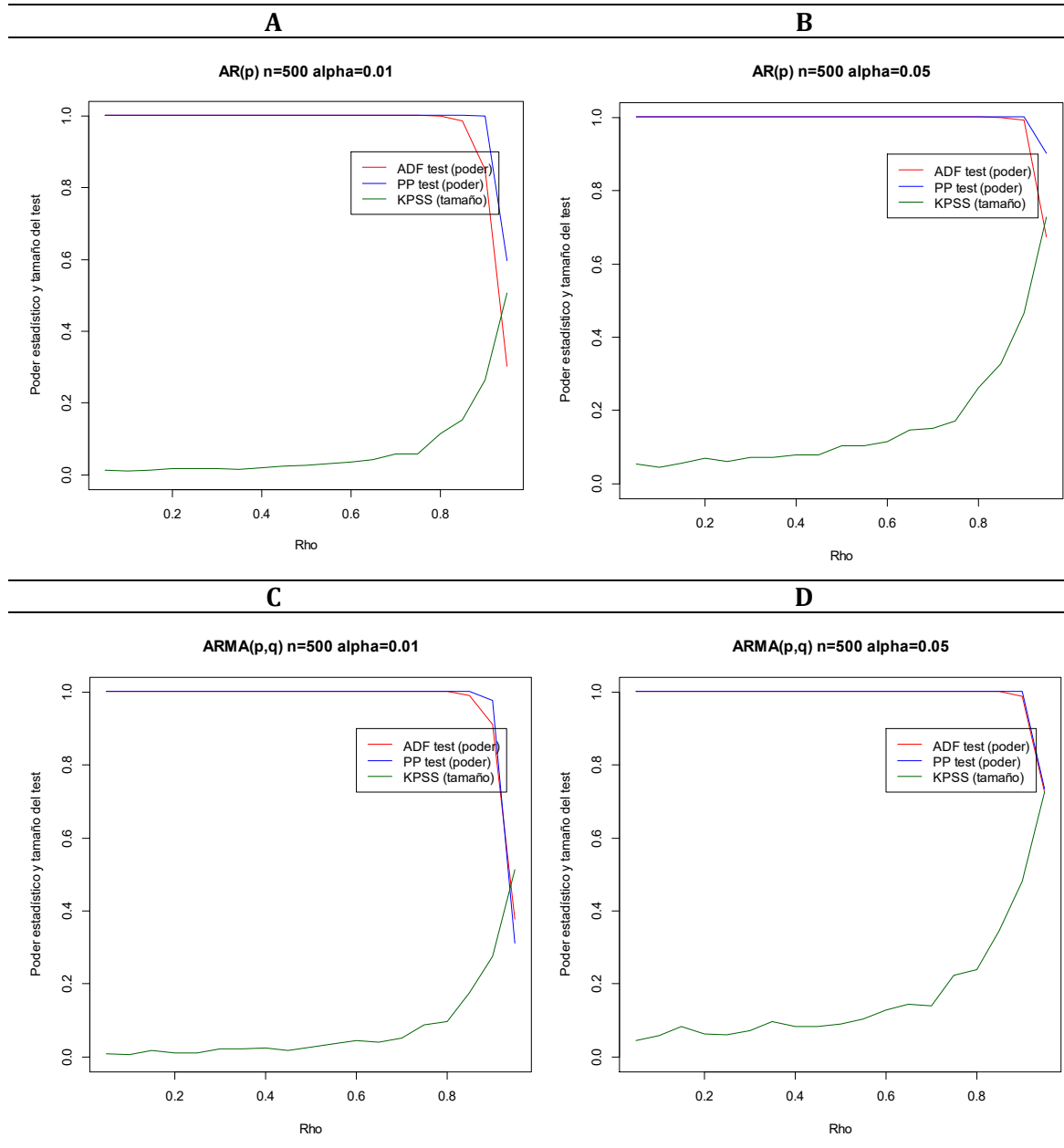
\*\* Se calcula el tamaño empírico de la prueba cuando el valor de componente autoregresivo está entre 0.05 y 0.95. El poder estadístico de la prueba se calcula cuando  $\rho$  es igual a uno.

Fuente: Elaboración propia

Los resultados del modelo ARMA( $p,q$ ) son prácticamente iguales a los del modelo AR( $p$ ), pero mejoraron con respecto a la misma especificación con un tamaño de muestra de 250. En general, las pruebas Phillips y Perron (1988) y Dickey-Fuller (1981) muestran un alto poder estadístico para prácticamente toda la grilla del componente autoregresivo (excepto cuando  $\rho$  adopta el valor de 0.95), lo cual, al igual que en el modelo anterior, significa que estas pruebas tienen un alto poder para discriminar entre series estacionarias y no estacionarias (o procesos cercanos a la raíz unitaria) cuando la muestra es de tamaño 500. El tamaño empírico de la prueba es igual al nominal para la prueba Phillips y Perron (1988) y ligeramente diferente en el caso de la prueba Dickey-

Fuller (1981). Los resultados de la prueba KPSS (1992) son muy parecidos a los anteriores, tanto en términos del poder estadístico como en términos del tamaño empírico de la prueba. Esto sugiere que esta prueba es poco sensible a cambios en el tamaño de la muestra.

**Gráfico 5. AR(p) y ARMA(p,q), n=500**



Fuente: Elaboración propia

## 9. Conclusiones

El objetivo de este trabajo fue analizar las propiedades en muestras pequeñas de las pruebas de raíz unitaria estándares o más conocidas, a saber: Dickey-Fuller (1981), Phillips-Perron (1988) y Kwiatkowski-Phillips-Schmidt-Shin (KPSS, 1992).

Considerando dos especificaciones de modelos, un autoregresivo de primer orden,  $AR(p)$ , y un modelo autoregresivo de promedio móvil,  $ARMA(p,q)$ , se realizó un experimento Monte Carlo de 1000 simulaciones para calcular de forma empírica el tamaño y la potencia estadística de cada prueba, cuando varía el valor del componente autoregresivo y la dimensión temporal de la serie.

En términos generales, los resultados de las simulaciones mostraron cuatro grandes conclusiones. Una primera conclusión es que en muestras muy pequeñas, ninguna de las pruebas de raíz unitaria mostró un desempeño aceptable, especialmente cuando el valor del componente autoregresivo es cercano a uno. El poder estadístico de las pruebas fue bajo y el tamaño empírico se alejó del valor nominal. Esta primera conclusión confirma los resultados de Schwert (1987), Lo y MacKindlay (1989) y Blough (1988), quienes habían encontrado previamente que las pruebas de raíz unitaria tienen poder estadístico bajo, especialmente en muestras pequeñas y cuando la raíz es menor pero muy cercano uno.

Una segunda conclusión que se puede extraer es que, en términos comparativos, la prueba Phillips-Perron (1988) mostró el mejor desempeño en muestras muy pequeñas. Como se mencionó anteriormente, una ventaja de la prueba de Phillips-Perron (1988) es que no asume una forma funcional de los errores de la serie, pues es una prueba no paramétrica. En ese sentido, esta prueba se aplica a una mayor cantidad de casos en los cuales se desconoce la verdadera estructura de los errores. Una supuesta desventaja es que la prueba es asintótica; es decir, tiende a funcionar bien en muestras grandes pero no necesariamente en muestras chicas. Esta conclusión se confirma de alguna forma en el presente trabajo, pues si bien comparativamente fue la prueba con mejor desempeño en muestras pequeñas, su desempeño no fue aceptable para todos los valores del componente autoregresivo. En muestras grandes, esta prueba también mostró el mejor desempeño. El poder estadístico fue alto en casi toda grilla del valor de  $\rho$ , y el tamaño empírico es muy cercano al nominal.

Una tercera conclusión que mostraron las simulaciones es que la prueba Dickey-Fuller (1981) es la más sensible al tamaño de la muestra. Esto significa que, comparativamente, esta prueba fue la que más mejoró su desempeño conforme aumentó el tamaño de la muestra. En muestras muy pequeñas, el poder estadístico de esta prueba era extremadamente bajo, aunque el tamaño empírico de la prueba era muy cercano al nominal. Cuando la muestra aumentó a 500, la prueba mejoró sustancialmente su desempeño. Los resultados fueron muy semejantes a los de la prueba Phillips-Perron (1988), reflejando un alto poder estadístico para casi todos los valores del componente autoregresivo y el tamaño empírico fue muy cercano a su valor nominal. Lo anterior confirma los resultados de Pesarán (2015), quien encontró que, asintóticamente, las pruebas Phillips-Perron (1988) y Dickey-Fuller (1981) son equivalentes. Lo anterior también es importante de destacar pues con un tamaño de muestra de 500, ambas pruebas tienen un alto poder para discriminar entre series estacionarias y no estacionarias (o procesos cercanos a la raíz unitaria).

Finalmente, una cuarta conclusión que se puede extraer de los resultados es que la prueba KPSS (1992) mostró un tamaño empírico de la prueba prácticamente invariante ante cambios en el tamaño de la muestra. Para valores muy pequeños del componente autoregresivo, el tamaño empírico de la prueba es muy cercano al nominal pero se aleja rápidamente cuando éste crece y se acerca a uno. Este comportamiento se repitió para todos los tamaños de muestra. Por su parte, el poder estadístico de la prueba en muestras pequeñas fue relativamente alto, y mejoró al aumentar el tamaño de la muestra. En términos globales, se puede decir que la prueba KPSS (1992) mostró los resultados más estables para todos los tamaños de muestra. Como se mencionó anteriormente, una desventaja práctica de la prueba KPSS (1992) está en la naturaleza misma del diseño de la prueba. La ausencia de una raíz unitaria no significa necesariamente que la serie sea estacionaria, sino, estacionaria en tendencia. En ese sentido, es posible que con esta prueba se esté ante una serie que no sea estacionaria, no tenga una raíz unitaria pero que sí sea estacionaria en tendencia. Por lo anterior, la prueba KPSS (1992) se considera un buen complemento a las pruebas estándares de raíz unitaria. Por ejemplo, si la prueba Dickey-Fuller (1981) falla en rechazar la hipótesis nula de raíz unitaria, y a su vez la prueba KPSS (1992) rechaza estacionariedad, entonces se podría considerar que existe suficiente evidencia de que la serie tiene una raíz unitaria.

El consenso general es que las pruebas de raíz unitaria son capaces de distinguir entre series con raíz unitaria y series muy estacionarias; sin embargo, no siempre son capaces de distinguir entre diferentes formas de no estacionariedad (por ejemplo, estacionariedad en tendencia o estacionariedad en diferencias) o cuando el componente autoregresivo en un modelo  $AR(p)$  está muy cercano a la unidad. No poder identificar adecuadamente la fuente de no estacionariedad en las series puede llevar a conclusiones erróneas. Si bien hay pruebas de raíz unitaria que en determinadas condiciones funcionan mejor que otras, como afirma Stock (1994) no existe una prueba general que sea dominante en todos los sentidos. Lo anterior recalca la importancia de analizar las series con base al conocimiento que se pueda tener sobre ellas, tratando de identificar previamente cuál es el verdadero proceso generador detrás de los datos antes de aplicar cualquiera de las pruebas estándar de raíz unitaria.

## 10. Bibliografía

1. Arltová, Markéta y Fedorová, Darina (2016). *Selection of unit root test on the basis of length of the time series and value of AR(1) parameter*. Statistika, 2016, 96 (3).
2. Beveridge, Stephen y Charles R. Nelson, (1981). *A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'*. Journal of Monetary Economics 7, 151-174.
3. Blough, Stephen R. (1988). *On the impossibility of-testing for unit roots and cointegration in finite samples*. Johns Hopkins University working paper in Economics no. 211.
4. Box, George E.P. y Jenkins, Gwilym M. (1976). *Time series analysis: Forecasting and control*. Rev. edition. Holdem-Day, San Francisco, CA.
5. Cameron, A. Colin y Pravin K. Trivedi. (2009). *Microeconometrics Using Stata*. EE.UU.: Stata Press.
6. Cochrane, John H. (1991). *A critique of the application of unit root tests*. Journal of Economic Dynamics and Control 15 (1991) 275-284.
7. Cochrane, John H. (2018). *A Brief Parable of Over-Differencing*.
8. Davidson, Russell y MacKinnon, James G. (2004). *Econometric Theory and Methods*. Oxford University Press, 2004
9. Dickey, David A. y Fuller, Wayne A (1979). *Distribution of the estimators for autoregressive time series with a unit root*. Journal of the American Statistical Association, Vol. 74. No. 366 (Jun., 1979), pp. 427- 431.
10. Dickey, David A. y Fuller, Wayne A (1981). *Likelihood ratio statistics for autoregressive time series with a unit root*. Econometrica, Vol. 49, No. 4 (July 1981).
11. Diebold, F. y A. Senhadji (1996). *The Uncertain Root in Real GNP: Comment*. *American Economic Review*, 86, 1291-1298.
12. Dufour, J. M., y King, M. L. (1991). *Optimal Invariant Tests for the autocorrelation Coefficient in Linear Regressions with Stationary or Nonstationary AR(1) Errors*," *Journal of Econometrics*, 47, 115-143.
13. Elder J. and Kennedy P. E. (2001). *Testing for Unit Roots: What Should Students Be Taught?*. *Journal of Economic Education*, 32(2): 137-46.
14. Enders, W. (2004). *Applied Econometric Time Series*. Second Edition. John Wiley & Sons: United States.
15. Friedman, Milton, and Schwartz, A.J. (1963). *A Monetary History of the United States 1867-1960*, Princeton, N.J.: Princeton University Press.

16. Gould, John P., and Nelson, Charles R. (1974). *The Stochastic Structure of the Velocity of Money*. *American Economic Review*, 64, 405-417.
17. Granger, C. W. J. y Newbold, P. (1974). Spurious regressions in econometrics. *Journal of Econometrics* 2 (1974) 111-120.
18. Hacker, R.Scott y Hatemi-J, Abdunnasser (2010). *The properties of procedures dealing with uncertainty about the intercept and deterministic trend in unit root testing*. Paper No. 212, Centre of Excellence for Science and Innovation Studies (CESIS), The Royal Institute of Technology
19. Kilian, L. y L. Ohanian (2002). *Unit Roots, Trend Breaks and Transitory Dynamics: A Macroeconomic Perspective*. *Macroeconomic Dynamics*, 6, 614-631.
20. Lee, D. y P. Schmidt (1996). *On the power of the KPSS test of stationarity against fractional-integrated alternatives*. *Journal of Econometrics*, 73, 1996, 285-302.
21. Lo, Andrew y Craig MacKinlay (1989). *The size and power of the variance ratio test in finite samples: A Monte Carlo investigation*. *Journal of Econometrics* 40, 203-238.
22. Murray, C., y C. Nelson (2000). *The uncertain trend in U.S. GDP*. *Journal of Monetary Economics* 46, 79-95.
23. Nelson, C. R. & Plosser, C. I. (1982). *Trends and random walks in macroeconomic time series: Some evidence and implications*. *J. Monet. Econ.* 10, 139-62.
24. Ng, S. y P. Perron, 2001, Lag length selection and the construction of unit root tests with good size and power. *Econometrica* 69, 1519-1554.
25. Pesaran, M. H. (2015). *Time Series and Panel Data Econometrics*. Oxford University Press.
26. Phillips, Peter C. B. (1987). *Time series regression with a unit root*. *Econometrica* 55, 277-301.
27. Phillips, Peter C. B. and Perron, Pierre (1988). *Testing for unit root in time series regression*. *Biometrika*, Vol. 75, No. 2 (Jun., 1988), pp. 335-346.
28. Rehman, Atiq-ur y Zaman, Asad (2009). *Impact of model specification decisions on unit root tests*. Munich Personal RePec Archive (MPRA), paper No. 19963.
29. Rubin, Herman (1950). *Consistency of Maximum-Likelihood Estimates in the Explosive Case*, in *Statistical Inference in Dynamic Economic Models*, ed. T.C. Koopmans, New York: John Wiley & Sons.
30. Said, S. E. & Dickey, D. A. (1984). *Testing for unit roots in autoregressive-moving average of unknown order*. *Biometrika* 71, 599-607.
31. Schwert, G. William (1987). *Effects of model specification on tests for unit roots in macroeconomic data*. *Journal of Monetary Economics* 20, 73.



32. Schwert, G. William (1989). *Tests for unit roots: a Monte Carlo investigation*. Journal of Business and Economic Statistics, April 1989, Vol. 7, No. 2.
33. White, John S. (1958). *The Limiting Distribution of the Serial Correlation Coefficient in the Explosive Case*. Annals of Mathematical Statistics, 29, 1188-1197.