Tesis de Maestría en Econometría

"Propiedades Finitas de Tests de Quiebres Estructurales Endógenos"

Alumno: Fernando Andrés Delbianco

Director: Martín González Rozada

Universidad Torcuato di Tella

2013

Abstract

En este trabajo se hace un breve repaso de la literatura de quiebres estructurales, y particularmente se describen brevemente cuatro tests que se encuadran dentro de dicha literatura (ya sean como test de raíz unitaria o test de break *per se*): Zivot y Andrews (1992), Bai y Perron (1998), Hansen (2000) y Kim y Perron (2009). Finalmente se hace un sencillo ejercicio de Monte Carlo para ver las propiedades de estos cuatro tests mencionados bajo cinco procesos generadores de los datos (DGP) distintos. Los resultados parecerían afirmar que, dado los DGP seleccionados y confeccionados, el test de Bai y Perron (1998) es el mejor a la hora de capturar un posible quiebre, mientras que el de Kim y Perron (2009) presenta buen desempeño a la hora de captar estacionariedad bajo la presencia de breaks.

***

Contents

1 Introducción

Desde Chow (1960) en la introducción de su trabajo, cuando menciona que después de haber estimado una regresión con p variables, quizás nos interese saber si m observaciones adicionales pertenecen a la misma regresión, se ha desarrollado una extensa literatura con respecto a las incidencias de los quiebres o cambios estructurales, desde sus efectos hasta su detección.

Como menciona Banerjee y Urga (2005) Cualquier descripción del "problema" que enfrentamos con quiebres en la estimación y la inferencia econométrica, puede ser descripta de distintas maneras, como por ejemplo contrastar modelos lineales o no lineales, detectar si los regresores son estacionarios o no, si el quiebre es conocido a priori o no, si el cambio es único o la serie presenta múltiples quiebres, etcétera.

En los 60s y 70s las teorías de "business cycles" fueron pensadas en el contexto de la presencia de tendencias seculares. Entonces, se pensaba que era suficiente descomponer las series en un componente de tendencia y otro que capturara el componente cíclico de las series. Usualmente, la tendencia se pensaba no solo de carácter determinístico, sino también lineal (sobre todo si en economía se trabaja con logaritmos y se piensa que el crecimiento suele presentar un comportamiento exponencial).

Se produjo un viraje importante cuando Nelson y Plosser (1982) señalaron que era necesario pensar la posibilidad de captar un componente estocástico de la tendencia. De esta manera, si las series no eran estacionaras (o de otra manera, eran integradas de algún orden d), probablemente los shocks serían permanentes y se irían acumulando en las series, lo cual es un concepto muy razonable en el ámbito de variables económicas.

Claramente hay una relación muy fuerte en la literatura de "breaks", con la idea de tests de raíces unitarias. Muchos de los desarrollos en quiebres estructurales fueron en un principio concebidos por el temor de no rechazar la hipótesis de raíz unitaria en casos donde era probable que la correcta especificación era tener en cuenta algún cambio permanente sufrido por la serie en algún punto del periodo muestral.

Dentro del área de discusión de series con raíces unitarias, Rappoport y Reichlin (1989) y Perron (1989) realizaron la distinción entre un shock que luego desaparece (abrupta o paulatinamente) y un shock que tiene un efecto permanente. Si se está en presencia de éste último caso, probablemente las series en cuestión deban ser representadas con algún quiebre en sus componentes determinísticos o de tendencia. Si no se incluyen en la especificación del modelo, entonces se está sesgando a los tests de raíces unitarias a no rechazar la hipótesis de raíz unitaria.

Otro aporte importante surgió cuando Zivot and Andrews (1992) advirtió que tal vez no sea aconsejable elegir la especificación del quiebre mediante un análisis previo de la información. De allí que surge toda una literatura de análisis y detección de quiebres estructurales endógenamente, como por ejemplo las especificaciones sugeridas por Bai y Perron (1998) de elegir aquellos quiebres que minimicen el error del modelo, y de esa manera objetivando de alguna manera la elección de los cambios especificados en los componentes determinísticos del modelo. También se propusieron en la agenda de quiebres endógenos, toda una serie de tests para analizar los "breaks" elegidos, i.e. Bai (1999) propuso un ratio para detectar múltiples quiebres conocido como "l breaks vs l+1 breaks", dada la hipótesis nula de la existencia de l cambios estructurales.

Hansen (2000) por otra parte, propuso tener en cuenta la distribución condicional y marginal de los procesos, y de esa manera, distinguir entre simples cambios en los parámetros y la estabilidad estructural de los regresores.

Newbold et al. (1995) señaló el hecho de que si hay un cambio temprano en la serie es probable que un test estándar como Dickey-Fuller no lo capte y rechace erróneamente la hipótesis de raíz unitaria. Harvey et al. (2001), Kim et al. (2000), y Leybourne et al. (1998) hacen un análisis similar, mencionando que si una serie es estacionaria alrededor de una tendencia pero presenta un quiebre (o múltiples), el Dickey-Fuller, nuevamente presenta poca potencia.

Siguiendo esta línea, autores como Lee and Strazicich (2001) o Harveyet al. (2001), aseguran que los procedimientos como el especificado por Zivot and Andrews (1992) a menudo identifican el quiebre de manera incorrecta. Más específicamente, lo ubican un periodo luego del verdadero quiebre. Esto se agrava a medida que el cambio (o break) es considerablemente más grande.

Alternativamente, Andrews (1993) desarrolla ratios basados en Wald, LM y LR tests para cambios estructurales basados en estimadores de GMM.

Asi mismo, Patrick Gagliardini, Fabio Trojani y Giovanni Urga (2005) proponen una nueva clase de estimadores GMM robustos, que incluyen ejercicios de simulación de Monte Carlo para comparar la performance de los tests GMM clásicos y los robustos que proponen.

Dentro de la literatura de quiebres, una cuestión importante a considerar, es si los cambios son aditivos o son realmente innovaciones outliers, lo cual cambia drásticamente el tratamiento de la dinámica y la especificación del modelo a considerar. Perron y Vogelsang (1992) y Vogelsang y Perron (1998) tratan estas cuestiones.

Bai y Perron (2006) aseguran que sus tests para múltiples "breaks" trabajan bien en muestras lo suficientemente grandes, pero sufren sustanciales desviaciones en la potencia de sus tests en muestras pequeñas.

El objetivo del trabajo consiste en el estudio de la incidencia de quiebres estructurales en series de tiempo. Más específicamente, sobre el desempeño de los tests de quiebres en muestras pequeñas, en especial los de carácter endógeno, de la línea que propone Bai Perron (1998, 2003, 2006) y Hansen (2000), o de los tests de raíz unitaria con el quiebre determindado endógenamente, como Zivot y Andrews (1992) y Kim y Perron (2009).

Siguiendo el trabajo de Antoshin, Berg y Souto (2008) y El-Shagi, Giesen (2010) se compararán distintas especificaciones (sobre todo cambiando el tamaño muestral) de los tests. Además, la intención es poder presentar variantes que tengan en cuenta que la muestra es finita, en lugar de utilizar resultados asintóticos para la derivación de los valores críticos y los intervalos de confianza de los tests.

El objetivo es internalizar que la muestra es finita, mediante simulaciones de Montecarlo para los distintos tamaños de muestra. Por otra parte, también se considera conveniente caracterizar los errores, permitiendo que exista autocorrelación. Como resultado, se espera obtener un panorama de la potencia que presentan los tests de cambios estructurales en muestras finitas (específicamente bajo el marco de las series utilizadas para las simulaciones), sin la herramienta de la teoría asintótica.

El trabajo se estructura de la siguiente manera: en la segunda sección se detallan aquellos tests con los que luego en la tercera sección se realizan las simulaciones. La cuarta sección finaliza el trabajo presentando las conclusiones.

2 Tests

Las simulaciones que se realizaran estarán enfocadas en cuatro tests, dos encuadrados en el ámbito de tests de raíz unitaria, como lo es el más antiguo de los tres (Zivot y Andrews, 1992) y el más moderno (Kim y Perron, 2009), y otros dos que son puramente tests de quiebres endógenos (Bai Perron, 1998 y Hansen 2000).

A continuación se pasa a detallar brevemente las estructuras de los mismos.

2.1 Zivot y Andrews (ZA)

El test clásico de Dickey-Fuller (ADF) con sus tres diferentes especificaciones (con tendencia, con drift y sin constante) sin cambios estructurales en las series analizadas, muestra en general que las series son no estacionarias. Sin embargo, este resultado no sorprende cuando se trabaja con series de tiempo relativamente largas.

Los tests de raíz unitaria más usados como el de Dickey-Fuller (1984) [1] o Perron (1989)[2], tienden a no rechazar la hipótesis nula de raíz unitaria en presencia de cambios estructurales y en general entonces tienden a concluir que las series no son estacionarias. Existen entonces los Tests que detectan cuando hubo un quiebre en la serie analizada, como el de Chow, pero esta primera generación requiere información a priori sobre la existencia del break, o usar iteraciones para encontrar dicha información. Un camino diferente se sigue en el trabajo de Zivot y Andrews (1992), el cual halla endógenamente la fecha del cambio estructural.

El test de ZA analiza secuencialmente la posible presencia de cambios estructurales en la serie en cada observación, generando dummies en cada periodo.

[1] Para una explicación detallada del test se puede consultar D. A. Dickey – D. P. Hasza – W. A. Fuller (1984).

[2] Para un análisis más profundo de este test, se puede recurrir al trabajo original de P. Perron (1989).

La *dummy* con mayor nivel de significatividad es tomada como indicadora del periodo en el cual la serie bajo estudio sufre un cambio de régimen. Quitando esta incorporación de dummies secuencial, el test de ZA tiene luego el formato clásico de un test de estacionaridad (un test de raíz unitaria) como el ADF.

ZA no es un test de quiebres estructurales *per se*. Es, como se mencionó, una manera de realizar un test de estacionariedad donde se testea la hipótesis nula de una raíz unitaria contra la hipótesis alternativa de estacionariedad con un punto de quiebre en algún punto desconocido de la serie.

Entonces, este test tiene las características de un test de raíz unitaria con notación similar al de Perron pero dejando que se defina endógenamente el quiebre estructural.

En primer lugar, se procede a repasar la estructura del test original de Perron (1989), para luego describir el cambio en $H0$ y la estimación del quiebre que introducen Zivot y Andrews.

2.1.1 Perron (1989)

En este test, la hipótesis nula es que la serie tiene una raíz unitaria con intercepto, y un cambio exógeno que ocurre en el periodo Tb(donde $1 < Tb < T$). Por otro lado, la hipótesis alternativa es que la serie es estacionaria con una tendencia determinística y un cambio exógeno en Tb. Los tres modelos que se pueden presentar en la hipótesis nula tienen las siguientes formas:

- Modelo (A): $y_t = \mu + dD(T_b)_t + y_{t-1} + e_t$
- Modelo (B): $y_t = \mu_1 + y_{t-1} + (\mu_2 - \mu_1)DU_t + e_t$
- Modelo (C): $y_t = \mu_1 + y_{t-1} + dD(T_b)_t + (\mu_2 - \mu_1)DU_t + e_t$

donde $D(T_b)_t = 1$ si $t = T_b$ y es igual a cero de otra manera; $DU_t = 1$ si $t > T_b$, 0 de otra manera; $A(L)e_t = B(L)v_t$ donde $v_t \sim iid(0, \sigma^2)$, con $A(L), B(L)$ operadores rezago con polinomios de orden p y q respectivamente.

Estos tres modelos, lo que están significando es que el test permite modelar un cambio en el nivel de las series (modelo A), un cambio exógeno en la tasa de crecimiento (modelo B), o ambos cambios (modelo C).

Por el otro lado, las alternativas de estacionareidad, tienen las siguientes estructuras:

- Modelo (A): $y_t = \mu_1 + \beta_t + (\mu_2 - \mu_1)DU_t + e_t$
- Modelo (B): $y_t = \mu + \beta_1 t + (\beta_2 - \beta_1)DT^* + e_t$
- Modelo (C): $y_t = u + \beta_1 t + (\mu_2 - \mu_1)DU + (\beta_2 - \beta_1)DT^* + e_t$

donde $DT_t^* = t - T_b$ si $t > T_b$, o de otra manera.

Siguiendo la estrategia de testeo de Perron (con una estructura ADF), las ecuaciones que se usan para testear raiz unitaria son:

$$y_t = \widehat{\mu}^A + \widehat{\theta}^A DU_t(\widehat{\lambda}) + \widehat{\beta}^A t + +\widehat{d}^A D(T_B)_t + \widehat{\alpha}^A y_{t-1} + \sum_{j=1}^{k} \widehat{c}_j^A \Delta y_{t-j} + \widehat{e}_t \quad (2.1.1)$$

$$y_t = \widehat{\mu}^B + \widehat{\beta}^B t + \widehat{\gamma}^B DT_t^* + \widehat{\alpha}^B y_{t-1} + \sum_{j=1}^{k} \widehat{c}_j^B \Delta y_{t-j} + \widehat{e}_t \quad (2.1.2)$$

$$y_t = \widehat{\mu}^C + \widehat{\theta}^C DU_t(\widehat{\lambda}) + \widehat{\beta}^C t + \widehat{\gamma}^C DT_t^* + \widehat{d}^C D(T_B)_t + \widehat{\alpha}^C y_{t-1} + \sum_{j=1}^{k} \widehat{c}_j^C \Delta y_{t-j} + \widehat{e}_t \quad (2.1.3)$$

Un estadístico que se computa es $t_{\widehat{\alpha}^i}(\lambda), i = A, B, C$. Este estadístico depende de la fracción (donde se encuentra el quiebre) λ, que es el ratio Tb/T. De esta manera, la hipótesis nula se rechaza si:

$$t_{\widehat{\alpha}^i}(\lambda) < \kappa_\alpha(\lambda)$$

donde $\kappa_\alpha(\lambda)$ hace referencia al size de la distribución asintótica de un λ fijo Tb/T; $DU_t(\lambda) = 1$ si $t > T\lambda$, 0 caso contrario; $DT^*(\lambda) = t - T\lambda$ si $t > T\lambda$ y 0 en el caso que no se cumpla la desigualdad mencionada.

Por otro lado, el número de k rezagos es determinado mediante testear la significatividad de $\widehat{c}_j^i$, $(i = A, B, C)$.

2.1.2 Zivot y Andrews (1992)

La hipótesis nula que ZA plantean para los tres modelos es:

$$y_t = \mu + y_{t-1} + e_t \quad (2.1.4)$$

De esta manera se considera que bajo $H0$ la serie es integrada sin cambios estructurales. Luego, la selección de λ es el producto de buscar una dummy[3] que logre que la representación de y_t sea estacionaria. Es decir que la hipótesis alternativa implica estacionariedad con un único quiebre. El objetivo es entonces, estimar el quiebre (la dummy) que más pondere la alternativa de estacionariedad.

λ es elegido de manera de minimizar el estadístico t a una cola, testeando que $\alpha^i = 1$ para $i = A, B, C$, dado que valores pequeños del estadístico denotan el rechazo de la nula. O sea,

$$t_{\widehat{\alpha}^i}[\widehat{\lambda}^i_{\inf}] = \inf_{\lambda \in \Lambda} t_{\alpha^i}, \qquad i = A, B, C, \quad (2.1.5)$$

y donde $\Lambda \in (0, 1)$.

[3] Al igual que en Perron (1989), los breaks se introducen como dummies en cualquiera de los tres modelos pausibles.

Como ahora la nula se encuentra especificada como en 2.1.4, ya no son necesarias las variables dummies en 2.1.1 y 2.1.3. Luego, las ecuaciones que se regresan en ZA para testear raíz unitaria son:

$$y_t = \widehat{\mu}^A + \widehat{\theta}^A DU_t(\widehat{\lambda}) + \widehat{\beta}^A t + \widehat{\alpha}^A y_{t-1} + \sum_{j=1}^{k} \widehat{c}_j^A \Delta y_{t-j} + \widehat{e}_t \tag{2.1.6}$$

$$y_t = \widehat{\mu}^B + \widehat{\beta}^B t + \widehat{\gamma}^B DT_t^*(\widehat{\lambda}) + \widehat{\alpha}^B y_{t-1} + \sum_{j=1}^{k} \widehat{c}_j^B \Delta y_{t-j} + \widehat{e}_t \tag{2.1.7}$$

$$y_t = \widehat{\mu}^C + \widehat{\theta}^C DU_t(\widehat{\lambda}) + \widehat{\beta}^C t + \widehat{\gamma}^C DT_t^*(\widehat{\lambda}) + \widehat{\alpha}^C y_{t-1} + \sum_{j=1}^{k} \widehat{c}_j^C \Delta y_{t-j} + \widehat{e}_t \tag{2.1.8}$$

donde $DU_t(\widehat{\lambda}) = 1$ si $t > T\lambda$, 0 en otro caso; $DT_t^*(\widehat{\lambda}) = t - T\lambda$ si $t > T\lambda$, 0 caso contrario. $\widehat{\lambda}$ es el punto de quiebre estimado por el procedimiento previamente mencionado.

Para cada valor de λ, el número de regresores extra, k, fue determinado usando el mismo procedimiento que Perron[4]. Luego, el estadístico t es computado. El mínimo t obtenido sobre $T-2$ regresiones indicará entonces la fecha de quiebre estimada.

Una vez que la selección de λ deja de ser exógena, endogeneizándose a través de un metodo de estimación, ZA ya no pueden usar los valores críticos computados por Perron. Ahora se rechaza $H0$ si:

$$\inf_{\lambda \in \Lambda} t_{\alpha^i}(\lambda) < \kappa^i_{\text{inf},\alpha}, \qquad i = A, B, C, \tag{2.1.9}$$

con $\kappa^i_{\text{inf},\alpha}$ como el size α crítico (hacia la cola izquierda de la distribución del estadístico) de la distribución asintótica de $\inf_{\lambda \in \Lambda} t_{\alpha^i}(\lambda)$. Aquí, ZA afirman que, por definición, estos valores críticos serán al menos tan grandes como los que se obtienen bajo un λ exógeno. Bajo esta perspectiva entonces, el test de raíz unitaria de Perron estaría sesgado a rechazar la hipótesis nula.

2.2 Bai Perron (BP)

Bai Perron (1998) consideran el siguiente modelo:

[4] Esto significa que la cantidad de extra regresores k, variará (o por lo menos puede variar potencialmente) para cada selección de λ.(Zivot y Andrews (1992), p. 255)

$$\begin{aligned} y_t &= x_t'\beta + z_t'\delta_1 + u_t, t = 1, ..., T_1, \\ y_t &= x_t'\beta + z_t'\delta_2 + u_t, t = 1, ..., T_2, \\ &\vdots \\ y_t &= x_t'\beta + z_t'\delta_{m+1} + u_t, t = 1, ..., T. \end{aligned} \tag{2.2.0}$$

La variable dependiente, observada en el momento t, es y_t.Los vectores de

covariables son $x_t(p \text{ x } 1)$, y $z_t(q \text{ x } 1)$ y sus respectivos vectores de coeficientes son β y $\delta_j(j = 1, ..., m+1).u_t$ hace referencia al error, mientras $(T_1, T_2, ..., T_{m+1})$ son los quiebres, que son tratados como desconocidos.

Si $p = 0$, el modelo pasa a ser un modelo de quiebres estructurales puros. Llamando a la matriz diagonal $\overline{Z} = diag(Z_1, ..., Z_{m+1})$, donde $Z_i = (z_{T_{i-1}} + 1, ...z_{T_i})'$ se puede reescribir el modelo de la siguiente forma:

$$Y = X\beta + \overline{Z}\delta + U, \tag{2.2.1}$$

donde $Y = (y_1, y_2, ..., y_T)', X = (x_1, x_2, ..., x_T)', U = (u_1, u_2, ..., u_T)', \delta = (\delta_1', \delta_2', ..., \delta_T')'$, y $\overline{Z}$ es la matriz diagonal que particiona en $(T_1, ..., T_m)$.

Con un superíndice 0, se indican los datos verdaderos, de manera que el proceso generador viene dado por:

$$Y = X\beta^0 + \overline{Z}^0\delta^0 + U, \tag{2.2.2}$$

La meta del test es estimar los puntos de quiebre, de manera de minimizar la suma de los residuos al cuadrado, S_T:

$$(\widehat{T}_1, ...\widehat{T}_m) = \arg\min S_T(T_1, ..., T_m) \tag{2.2.3}$$

Bai Perron (2003) incorpora los siguientes tests: un test de no quiebres versus un número fijo de quiebres, un tests "doble máximo", y un test de ℓ versus $\ell + 1$ breaks. En las simulaciones que se desarrollarán más adelante se utilizan los estadísticos de los últimos dos tests mencionados, por lo que a continuación se los describirá brevemente.

2.2.1 Tests "doble máximo"

Este test permite la hipótesis nula de la no existencia de quiebre estructural versus un número desconocido de quiebres (con un límite superior, que BP denota m).

El primero de estos tests, $UD\max F_T(M, q)$, es una versión ponderada del test F que utiliza las estimaciones de quiebres obtenidas $(\widehat{\lambda}_j = \widehat{T}_j/T$ con $j = 1, ..., m)$ al minimizar la SSR de los segmentos de largo mínimo h.

El segundo de los tests, $WD\max F_T(M, q)$, se encuentra ponderado de manera de igualar los p-values marginales a través de los distintos valores de m.

2.2.2 Test de ℓ versus $\ell + 1$ breaks

Aquí, el test de ℓ versus $\ell + 1$ quiebres estructurales (Bai, 1999) se denota $\sup F_T(\ell \mid \ell + 1)$. El mismo se aplica a cada segmento que contiene las observaciones $\widehat{T}_{i-1} + 1$ hasta $\widehat{T}_i$ (donde $i = 1, ..., \ell + 1$). Si el estadístico $\sup F_T(1; q)$, es suficientemente grande, se rechaza la hipótesis nula, concluyendo que existe evidencia estadística a favor de un modelo con $\ell + 1$ breaks. Finalmente, la fecha del quiebre que se selecciona es la correspondiente a este valor máximo del test.

2.3 Hansen

Considerando el siguiente modelo,

$$y_{ni} = x'_{ni}\beta_{ni} + e_{ni}; i = 1, 2, ..., n\ ; \sigma = E(e_{ni}^2) < \infty \tag{2.3.1}$$

el trabajo de Hansen (2000) se motiva en las diferencias conceptuales de que un quiebre en β_{ni} es conceptualmente distinto a un quiebre en la distribución de x_{ni}. En otras palabras, testear un quiebre en (2.3.1) es distinto a determinar si x_{ni} es estacionario o no. Si la forma funcional de (2.3.1) se mantiene constante ante un cambio en x_{ni}, entonces entra en juego el concepto de *super exogeneity* de Engle, Hendry y Richard (1983). El problema radica en si β_{ni} no es robusto ante cambios en x_{ni}. Ante este problema, es que Hansen propone su test.

Supone la siguiente forma del cambio estructural en β_{ni} :

$$\beta_{ni} = \begin{cases} \beta\ ,\ i < t_0 \\ \beta + \theta_n,\ i \geqslant t_0 \end{cases} \tag{2.3.2}$$

Con t_0 en el intervalo (t_1, t_2) indicando el momento del quiebre y θ_n la magnitud del mismo. Lo que se testea en este caso es si:
$H_0 : \theta_n = 0$
$H_1 : \theta_n \neq 0$
La forma que se asume que θ_n tiene es:

$$\theta_n = \frac{\delta\sigma}{\sqrt{n}} \tag{2.3.3}$$

con δ fijo cuando $n \to \infty$.

Aquí Hansen menciona que el test standard para contrastar H_0 (no break) versus H_1 es el estadístico de Wald (e.g. Chow (1960)):

$$F_t = \frac{(n-m)\widehat{\sigma}^2 - (n-2m)\widehat{\sigma}_t^2}{\widehat{\sigma}_t^2} \tag{2.3.4}$$

El test de Wald, como se mencionó en la introducción, necesita que se especifique el punto en el que surge el quiebre. Pero Hansen se encuentra interesado en cuando los breaks son desconocidos, por lo que menciona una serie de tests endógenos sugeridos, que luego utilizará para su test. El primero de ellos es el propuesto por Quandt (1960) quien propuso un test basado en un ratio de verosimilitud donde el valor F supremo es tomado sobre $t \in (t_1, t_2)$.Luego se menciona una serie de tests propuestos por Andrews and Ploberger (1994), entre los que se encuentran un test de Wald ponderado exponencialmente, $ExpF_n = \ln \int \exp(F_t/2)dw(t)$, y otro promediado, $AveF_n = \int_t F_t dw(t)$, donde w pondera con $1/(t_2 - t_1)$ a cada entero $t \in (t_1, t_2)$. Finalmente, hace mención a Nyblom (1989), quien considera un cambio estructural aleatorio, donde θ_n y t_0 son variables aleatorias en lugar de parámetros como en los casos anteriores. Nyblom propone entonces un test basado en el multiplcador de Lagrange (LM test). La hipótesis de no quiebre es rechazada para valores suficientemente grandes del estadístico generado por el LM test.

2.3.1 "Fixed Regressor Bootstrap"

Hansen (2000) demuestra en su trabajo que la no estacionariedad o un cambio estructural en la distribución marginal afecta directamente a las distribuciones asintóticas, de "maneras complicadas" (*"in complicated ways").* Es por eso que se propone un regresor basado en la técnica de bootstrap ("Fixed Regressor Bootstrap"), el cual toma su nombre de tratar a los regresores x_{ni} como fijos (incluso si contienen valores rezagados). La forma de este test dependerá de si los errores son homocedásticos o heterocedásticos.

Para el caso homocedástico, suponiendo que $y_{ni}(b)$ es una muestra aleatoria de la distribución normal ($N(0,1)$),

1) regresando $y_{ni}(b)$ sobre x_{ni}, se obtiene la varianza $\widehat{\sigma}^2(b)$

2) regresando $y_{ni}(b)$ sobre x_{ni} y $x_{ni}I(I \leq t)$, se obtiene los residuos $\widehat{\sigma}_t^2(b)$ y entonces:

$$F_t(b) = \frac{(n-m)\widehat{\sigma}^2(b) - (n-2m)\widehat{\sigma}_t^2(b)}{\widehat{\sigma}_t^2(b)} \tag{2.3.5}$$

El estadístico de bootstrap es entonces $Su\,\mathrm{pF}_n(b) = \sup_{t_1 \leq t \leq t_2} F_t(b)$. El p-value del bootstrap es $p_n = 1 - G_n(Sup\ F_n)$, donde $G_n(x) = P(\sup\ F_n(b) \leq x|\Im_n)$. Se rechaza H_0 cuando p_n es pequeño.

Para el caso con heterocedasticidad, se introduce una modificación, de manera que $y_{ni}^h(b) = u_i(b)\widetilde{e}_i$, donde $u_i(b)$ es una muestra iid $N(0,1)$, y $\widetilde{e}_i$ es un término de error (que puede llegar a ser *condicionalmente* heterocedástico). Luego, el estadístico bootstrap se construye igual que antes, reemplazando, $y_{ni}(b)$ por $y_{ni}^h(b)$.

Como $G(\cdot)$ es desconocida, se calcula por simulación, por lo que p_n y p_n^h son los estadísticos bootrsap de interés. Dichos p-values $p_n(J)$ son calculados como el porcentaje de simulaciones en las que el estadístico sup $F_n(j)$ supera el valor muestral del sup F_n.

2.4 Kim - Perron (KP)

Con respecto al test de Zivot y Andrews, Kim y Perron (2009) dicen en su trabajo:

> *"Zivot and Andrews assumed that if a break occurs, it does so only under the alternative hypothesis of stationarity. This is undesirable since a) it imposes an asymmetric treatment when allowing for a break, so that the test may reject when the noise is integrated but the trend is changing; b) if a break is present, this information is not exploited to improve the power of the test."*[5]

Teniendo en cuenta estas dos cuestiones, Kim Perron (2009) proponen un procedimiento de testeo que permite un quiebre estructural tanto bajo la hipótesis nula como la alternativa; y cuando un quiebre estructural se hace presente, la distribución asintótica del test es la misma a la del caso con un quiebre conocido en la serie, y así permitiendo incrementar la potencia mientras se mantiene un *size* correcto del test.

Kim Perron (KP) considera un proceso univariante $\{y_t\}$ generado por alguno de los tres modelos de Outliers Aditivos (*Additive Outliers*, AO), o alguno de los dos modelos de Outliers Innovacionales (*Innovational Outliers*, IO). Para cada modelo, la serie es generada por la suma de una tendencia determinística y un término de error. La tendencia determinística tiene un único quiebre que ocurre en un periodo desconocido en el intercepto, la pendiente, o ambos, dependiendo del modelo.

Los procesos generadores de los datos (DGP) de los modelos AO son:

$$y_t = z(T_1)'_t\phi + u_t = z'_{t,1}\phi_1 + z(T_1)'_{t,2}\phi_2 + u_t \tag{2.4.1}$$

donde $z_{t,1} = (1,t)'$, $\phi_1 = (\mu, \beta)'$,

$$z(T_1)_{t,2} = \begin{cases} DU_t & \text{Modelo A1} \\ B_t & \text{Modelo A2} \\ (DU_t, B_t)' & \text{Modelo A3} \end{cases}, \ \phi_2 = \begin{cases} \mu_b & \text{Modelo A1} \\ \beta_b & \text{Modelo A2} \\ (\mu_b, \beta_b)' & \text{Modelo A3} \end{cases}$$

con $DU_t = B_t = 0$ si $t \leq T_1$, y $DU_t = 1$, $B_t = t - T_1$, si $t > T_1$. Aquí $T_1 = \lambda^C T$, con $0 < \lambda^C < 1$, que denota el verdadero break (y λ^C la verdadera fracción que ese break representa). Notese que DU_t y B_t dependen de T_1 y T pero esta dependencia se encuentra omitida. El error (el ruido de la serie) $\{u_t\}$ es tal que $A(L)u_t = B(L)\varepsilon_t$ donde $\varepsilon_t \backsim i.i.d\ (0, \sigma^2_\varepsilon)$, y $A(L)$ y $B(L)$ son polinomios L de orden $p+1$ y q, respectivamente. Se factorea $A(L)$ como $(1-\alpha L)A^*(L)$ y se asume que $A^*(L)$ y $B(L)$ tienen raíces estricamente fuera del círculo unitario. Las hipótesis nula y alternativa son $H_0 : \alpha = 1$ y $H_1 : |\alpha| < 1$, respectivamente. El modelo ARMA especificado se puede relajar para permitir procesos aún más generales, pero se usa estas especificaciones para facilitar la presentación del test. Los DGP de los modelos de *innovational outlier* (IO) bajo la hipótesis nula están dados por:

[5] Kim, D., Perron, P., "Unit Root Tests Allowing for a Break in the Trend Function at an Unknown Time under Both the Null and Alternative Hypotheses", Journal of Econometrics (2009) 148, 1-13., p. 1.

$$y_t = y_{t-1} + \beta + \Psi^*(L)(d(T_1)'_{t,2}\phi_2 + \varepsilon_t) \tag{2.4.2}$$

donde

$$d(T_1)_{t,2} = \begin{cases} D(T_1)_t & \text{Modelo I1} \\ (D(T_1)_t, DU_t)' & \text{Modelo I3} \end{cases}$$

y $D(T_1)_t = 1$ si $t = T_1 + 0$ y 0 caso contrario. Bajo la hipótesis alternativa, se tiene que:

$$y_t = z'_{t,1}\phi_1 + \Psi(L)(z(T_1)'_{t,2}\phi_2 + \varepsilon_t) \tag{2.4.3}$$

donde

$$z(T_1)_{t,2} = \begin{cases} DU_t & \text{Modelo I1} \\ (DU_t, B_t)' & \text{Modelo I3} \end{cases}, \phi_2 = \begin{cases} \mu_b & \text{Modelo I1} \\ (\mu_b, \beta_b)' & \text{Modelo I3} \end{cases}$$

con $\Psi^*(L)$ y $\Psi(L)$ tales que $\Psi^*(L) = A^*(L)^{-1}B(L)$ y $(1-\alpha L)^{-1}\Psi^*(L) = \Psi(L)$.

Los autores nos hacen notar en este punto que los Modelos A1, A2, A3 (con $\mu_b = c + \beta_b T_1$), I1 y I3 (con $\mu_b = c + \beta_b T_1$) son los mismos que en Perron (1989), excepto que se trata el cambio estructural como desconocido (i.e. la fecha potencial del mismo es desconocida).

Aquí, cabe remarcar que lo que se hace para testear es en primer lugar un test similar al implementado en Perron (1989), pero en lugar de usar la fecha verdadera de quiebre, se utiliza una estimación de la misma. El procedimiento de Perron testea la hipótesis de raíz unitaria sobre la suma de los coeficientes autorregresivos de la regresión sobre la serie a la que se le removió previamente la tendencia (tanto para los modelos AO como IO). El resultado de este test es que $t_\alpha(\lambda^c) \Longrightarrow R(\lambda^c)$. Entonces, utilizando una estimación de λ la condición deseable es que $t_\alpha(\widehat{\lambda}^c) \Longrightarrow R(\lambda^c)$. Si este resultado se cumple, entonces se pueden usar los valores críticos correspondientes al caso donde λ es conocido.

Para estimar el quiebre, Kim y Perron (2009) se concentran sobre el método de minimizar SSR (en particular Perron-Zhu (2005) y Hatanaka-Yamada (1999)). El trabajo de KP entonces luego demuestra que la condición mencionada se cumple bajo ciertos supuestos, dependiendo del caso de DGP que se trate. Al cumplirse esta condición, como se mencionó al comienzo de esta subsección, el *size* mejora al poder trabajar con la distribución como si el quiebre fuera conocido en lugar de desconocido.

3 Simulaciones

Siguiendo los trabajos mencionados, se procede en este apartado a realizar una serie de sencillas simulaciones, en pos de ver algunas de las propiedades finitas de los tests de quiebres utilizados en este trabajo. Los procesos generadores de los datos (a partir de aquí, DGP), fueron a grandes rasgos los siguientes:

- DGP1: una serie con distribución normal con un componente de error ruido blanco.
- DGP2: misma serie que en el DGP anterior, pero con una constante ($c = 10$)
- DGP3: un modelo ARMA. En algunos casos fue puramente AR, en otros ARMA.
- DGP4: un modelo ARMA con un componente exógeno ARMA.
- DGP5: un modelo con autocorrelación. El termino de error sigue un proceso autorregresivo.

Cabe destacar que todos los DGP son procesos estacionarios. Todos los coeficientes autorregresivos fueron fijados en $\rho = 0.5$.

Se simularon luego breaks. Los mismos fueron específicados para suceder en $\lambda = T/2$. Los quiebes se vieron reflejados en algunos casos directamente en la serie, y otras indirectamente a través del término de error (e.g. un cambio de constante de 0 a 5, o un cambio en la varianza, de 1 a 10).

En algunas circunstancias se simularon los tests sobre DGP que contenían dos quiebres estructurales en su serie (en general un quiebre en media y el otro en varianza, nunca los dos quiebres de la misma naturaleza).

En todos los casos, se realizaron 1000 simulaciones, y los tamaños muestrales seleccionados fueron de 30, 50, 100, 200 y 500 onservaciones.

Para las simulaciones se utilizaron diferentes paquetes econométricos, dependiendo del software con que el autor provea sus códigos.

A continuación, se detallan las simulaciones específicamente para cada test con sus correspondientes resultados.

3.1 ZA

Para el caso del test de Zivot y Andrews, se muestran tres tablas (Tablas 1, 2 y 3). La primera, donde el DGP no tiene quiebre, muestra en las últimas dos columnas, los cantidad de casos (de los 1000 simulados) donde el valor crítico de la fecha que el test selecciona como potencial quiebre, supera los umbrales críticos del 1 y 5%.

El test sugiere en estos casos, la mayoría de las veces (y para valores de 200 y 500, todas las veces) un quiebre significativo, a pesar de que las series generadas no contienen un quiebre en su DGP.

La tabla número 2 incorpora un break en el DGP (exactamente en $T/2$). Se incorporan dos tipos de cambios estructurales: un cambio en la varianza del error, de 1 a 10, y un cambio en la media del proceso, de 0 a 5. Se generó un intervalo *ad-hoc* de ± 5 períodos desde la fecha en que es generado el break. Con este intervalo, se generó una indicadora que vale uno cada vez que el break estimado se encuentra dentro de dicho intervalo. El conteo de las veces que esto último sucedió se pueden ver en la columna "$\in$ *Intervalo*". Se observa que al igual que en la Tabla 1, en la mayoría de los casos el valor crítico supero los umbrales del 1 y 5%, pero a medida que la muestra se hace más grande, menor cantidad de veces el break sugerido se encuentra dentro del intervalo (claramente esto debe ocurrir debido a que el intervalo cada vez representa una porción más chica del total de observaciones).

La tabla 3 es similar a la tabla 2, pero con dos quiebres estructurales. El primero es un cambio en media (en $T/3$) y el segundo es un cambio en varianza (2/3 de T). En este caso, la variable indicadora se hace uno cuando el break sugerido por el test se encuentra en el intervalo ± 5 con respecto a alguno de los dos breaks. A pesar de que ahora la variable indicadora es aún más abarcativa con respecto al tamaño de la muestra, no se aprecian aumentos en la cantidad de veces que la variable toma valor uno. Tanto la tabla 2, como la 3, no presentan mayor cantidad de casos donde el estadistico se torna significativo con respecto a la tabla 1, aún cuando esta última tabla ilustra el caso de DGP sin quiebres.

Table 1. ZA simulation with no break

	N	*# Sim*	*1%*	*5%*
DGP1	30	1000	785	912
	50	1000	934	960
	100	1000	953	976
	200	1000	1000	1000
	500	1000	1000	1000
DGP2	30	1000	785	915
	50	1000	928	958
	100	1000	952	978
	200	1000	999	1000
	500	1000	1000	1000
DGP3	30	1000	283	529
	50	1000	597	834
	100	1000	853	940
	200	1000	999	1000
	500	1000	1000	1000
DGP4	30	1000	539	738
	50	1000	882	953
	100	1000	930	963
	200	1000	1000	1000
	500	1000	1000	1000
DGP5	30	1000	740	981
	50	1000	924	955
	100	1000	946	984
	200	1000	1000	1000
	500	1000	1000	1000

Table 2. ZA simulation with break									
			In mean			In variance			
	N	*# Sim*	*1%*	*5%*	ϵ *Intervalo*	*1%*	*5%*	ϵ *Intervalo*
DGP1	30	1000	785	891	424	755	905	456
	50	1000	936	957	93	931	958	290
	100	1000	950	972	165	945	982	173
	200	1000	999	1000	178	999	1000	90
	500	1000	1000	1000	64	1000	1000	30
DGP2	30	1000	769	909	425	764	913	436
	50	1000	943	962	314	927	962	286
	100	1000	944	982	182	957	979	167
	200	1000	1000	1000	50	998	1000	85
	500	1000	1000	1000	22	1000	1000	31
DGP3	30	1000	729	869	437	735	899	454
	50	1000	931	967	297	928	957	288
	100	1000	943	978	147	939	990	156
	200	1000	1000	1000	74	999	1000	82
	500	1000	1000	1000	34	1000	1000	29
DGP4	30	1000	280	584	451	723	873	467
	50	1000	576	817	288	942	966	313
	100	1000	851	915	173	960	985	134
	200	1000	999	997	92	1000	1000	71
	500	1000	1000	1000	33	1000	1000	31
DGP5	30	1000	729	869	437	737	897	465
	50	1000	931	962	314	910	956	322
	100	1000	943	978	147	952	970	141
	200	1000	1000	1000	74	995	1000	69
	500	1000	1000	1000	34	1000	1000	35

Table 3. ZA simulation with two breaks

	N	*# Sim*	*1%*	*5%*	ϵ *Intervalo*
DGP1	30	1000	767	919	438
	50	1000	943	962	314
	100	1000	944	982	182
	200	1000	1000	1000	50
	500	1000	1000	1000	22
DGP2	30	1000	756	903	434
	50	1000	931	967	297
	100	1000	943	978	147
	200	1000	1000	1000	74
	500	1000	1000	1000	34
DGP3	30	1000	741	899	293
	50	1000	934	967	344
	100	1000	943	970	191
	200	1000	1000	1000	195
	500	1000	1000	1000	161
DGP4	30	1000	633	832	286
	50	1000	926	935	273
	100	1000	934	961	136
	200	1000	1000	1000	151
	500	1000	1000	1000	147
DGP5	30	1000	767	919	438
	50	1000	743	895	269
	100	1000	932	976	142
	200	1000	1000	1000	149
	500	1000	1000	1000	110

Se pueden concluir en este ejercicio de Monte Carlo sobre el test de Zivot y Andrews (1992) algunas cuestiones.

En primer lugar, el test no parece reflejar buenos resultados cuando se aplica a una serie que no presenta breaks y es estacionaria, ya que la Tabla 1 nos muestra que en la mayoria de los casos se sugiere un quiebre y se sobrepasan los valores críticos al 1 y 5%. Este resultado es esperable, ya que el DGP en esta tabla no esta asimilado en el test, ni bajo $H0$ (raíz unitaria) ni bajo $H1$ (estacionariedad con un quiebre en la serie).

En segundo lugar, cuando de hecho el DGP usado para las simulaciones se identifica con el que el test plantea como hipótesis alternativa (estacionariedad con cambio estructural) el problema que se puede presentar es que, si bien se rechaza la hipótesis nula, el punto de quiebre sugerido por el test no se encuentra en el intervalo ad hoc sugerido en este trabajo de ± 5. Además, a medida que aumenta el tamaño de la muestra, la variable indicadora de pertenencia al intervalo se hace cada vez menos veces igual a uno. Lo que podría estar sugiriendo que el hecho de que esta variable se haga unas más veces igual a la unidad para las muestras de N=30 o 50 respectivamente, es sencillamente porque el intervalo representa una mayor proporción del total de la serie.

Finalmente, para el tercer caso donde existen dos breaks, la variable indicadora no aumenta considerablemente la cantidad de veces que toma valor uno, a pesar de que ahora el intervalo representa una mayor proporción del período total (porque la indicadora es si el break sugerido toma lugar en los intervalos ± 5 que se forman alrededor de los dos quiebres del DGP subyacente).

Entonces, bajo estos resultados, se puede decir que el test no falla en el sentido de que rechaza en general y correctamente la hipótesis nula, pero el desempeño del test en atinar la estimación del break en algún área cercana al break verdadero no es bueno. Este último resultado no es sorprendente, ya que, como se mencionó ya en la introducción, Lee and Strazicich (2001) o Harvey et al. (2001), aseguran que los tests como el especificado por Zivot and Andrews (1992) a menudo identifican el quiebre de manera incorrecta.

3.2 BP

En el proceso de simulación del test de Bai Perron (1998), se hicieron algunas modificaciones con respecto a la anterior simulación sobre Zivot y Andrews (1992).

Los archivos y códigos utilizados para esta simulación se pueden conseguir en la página de Pierre Perron[6].

En primer lugar se utilizan dos estadistícos y se los compara con sus valores críticos al 5%. Estos dos estadísticos, como se mencionó anteriormente en la sección introductoria del test, son el test de "doble máximo" (*wFtest*), y el test de ℓ versus $\ell + 1$ breaks (*SupFl*). Además se muestran la media de la fecha de los breaks sugeridos, y la cantidad media de cantidad de breaks sugeridos. En la tabla 4 se pueden observar que la cantidad de veces que *SupFl* y *wFtest* fueron menores a 5% fue una cantidad considerable de veces sobre el total de las 1000 simulaciones. A su vez, la media de la cantidad de breaks sugeridos parece tender a cero a medida que aumenta la muestra.

[6] Pierre Perron Homepage: http://people.bu.edu/perron/

Table 4. BP simulation with no break

	N	*# Sim*	*SupFl 5%*	*wFtest 5%*	*Media Break*	*Media # breaks*
DGP1	30	1000	0.876	0.537	1.64	0.118
	50	1000	0.952	0.653	1.619	0.068
	100	1000	0.978	0.811	1.509	0.028
	200	1000	0.993	0.844	2.386	0.026
	500	1000	0.994	0.924	4.617	0.021
DGP2	30	1000	0.814	0.546	0.034	0.002
	50	1000	0.924	0.673	0.076	0.014
	100	1000	0.973	0.788	2.889	0.075
	200	1000	0.984	0.814	4.062	0.07
	500	1000	0.99	0.895	5.363	0.029
DGP3	30	1000	0.862	0.865	1.639	0.105
	50	1000	0.949	0.89	1.46	0.056
	100	1000	0.982	0.755	2.176	0.04
	200	1000	0.992	0.753	1.769	0.02
	500	1000	0.998	0.864	2.326	0.009
DGP4	30	1000	0.848	0.42	2.45	0.17
	50	1000	0.937	0.588	2.541	0.101
	100	1000	0.979	0.788	3.046	0.064
	200	1000	0.987	0.844	3.195	0.032
	500	1000	1	0.896	2.85	0.011
DGP5	30	1000	0.856	0.455	2.428	0.168
	50	1000	0.929	0.569	3.363	0.141
	100	1000	0.978	0.455	3.834	0.076
	200	1000	0.997	0.886	2.939	0.03
	500	1000	0.999	0.915	4.649	0.016

Cuando un quiebre ocurre, se obtiene la tabla número 5. El quiebre especificado aquí es de uno sobre la media del proceso (como en el resto de los casos, de 0 a 5). Aquí las columnas Supfl y *Wftest* muestran la cantidad de veces que los valores críticos superaron el umbral. El test de ℓ versus $\ell+1$ tiende a obtener valores chicos a medida que aumenta la muestra, mientras que el test de "doble máximo" tiende a valores grandes. La media de la fecha del break sugerido se acerca casi con exactitud a la verdadera, y la cantidad de breaks sugeridos también tiende a la verdadera cantidad (o sea, un break).

Dada la posibilidad que el test nos brinda de testear múltiples breaks, se procedió además a generar series con dos cambios estructurales: $\lambda_1 = T/2$ y $\lambda_2 = T/4$.

En la Tabla 6, se puede observar y leer los resultados de manera similar a los que muestra la tabla 5, con la sola diferencia de que se distingue la media del primer break y la media del segundo cambio estructural, ya que en esta tabla el DGP contiene, como se mencionó, dos cambios, uno en T/2 (en media) y otro en T/4 (en varianza).

De nuevo, la media de los breaks tiende a la real, y la cantidad de datos también. Una regularidad para remarcar, es que en general pareciera que se sobre estima el primer cambio estructural y se subestima el segundo (levemente en ambos casos) salvo para el caso de 30 observaciones, donde la media se subestima para ambos breaks.

Table 5. BP simulation with one break

	N	*# Sim*	Supfl *5%*	*Wftest 5%*	*Media Break*	*Media # breaks*
DGP1	30	1000	0.177	0.956	10.768	0.739
	50	1000	0.19	0.997	23.408	0.928
	100	1000	0.05	1	51.62	1.01
	200	1000	0.046	1	101.84	1.008
	500	1000	0.037	1	251.25	1.012
DGP2	30	1000	0.223	0.948	14.425	1.154
	50	1000	0.112	0.975	24.645	1.057
	100	1000	0.043	0.998	49.771	1.014
	200	1000	0.038	1	99.697	1.011
	500	1000	0.018	1	251	1.003
DGP3	30	1000	0.144	0.98	16.003	1.077
	50	1000	0.068	0.994	26.007	1.031
	100	1000	0.049	1	51.004	1.016
	200	1000	0.032	1	100.87	1.006
	500	1000	0.016	1	250.67	1.008
DGP4	30	1000	0.179	0.855	11.38	0.741
	50	1000	0.095	0.937	23.587	0.906
	100	1000	0.065	0.961	52.568	1.013
	200	1000	0.038	0.991	102.93	1.014
	500	1000	0.043	0.997	252.24	1.012
DGP5	30	1000	0.133	0.953	20.201	0.823
	50	1000	0.233	0.959	11.721	0.867
	100	1000	0.06	0.986	49.781	0.97
	200	1000	0.037	0.991	103.1	1.014
	500	1000	0.021	0.998	253.61	1.003

Una vez observadas las características de las tablas generadas por la simulación desarrollada a partir de aplicar el test propuesto por Bai y Perron (1998, 2003) a los DGP generados para el ejercicio, se pueden extraer algunas conclusiones al respecto.

En primera instancia, a diferencia de lo que ocurría en la simulación anterior sobre el test de ZA, el desempeño del test para captar el quiebre estructural es bueno. Salvo en algunas ocasiones donde se subestima o sobreestima (dado que

Table 6. BP simulation with two breaks

	N	*# Sim*	Supfl *5%*	*Wftest 5%*	*Media Break*	*Media break 2*	*Media # breaks*
DGP1	30	1000	0.26	0.984	10.675	14.247	1.805
	50	1000	0.147	0.991	20.901	29.815	1.945
	100	1000	0.055	1	51.012	73.138	2.003
	200	1000	0.033	1	101.42	148.91	2.018
	500	1000	0.036	1	251.96	374.02	2.009
DGP2	30	1000	0.351	0.991	11.001	14.879	1.915
	50	1000	0.124	0.997	19.546	28.994	1.945
	100	1000	0.048	1	52.011	74.112	2.001
	200	1000	0.022	1	102.38	147.92	2.009
	500	1000	0.021	1	252.15	373.88	2.016
DGP3	30	1000	0.191	0.976	10.932	1.74	1.072
	50	1000	0.054	0.989	21.101	35.084	1.791
	100	1000	0.051	1	50.497	74.544	1.955
	200	1000	0.023	1	201.21	149.96	1.993
	500	1000	0.017	1	252.01	371.54	2.006
DGP4	30	1000	0.289	0.985	16.283	8.304	1.415
	50	1000	0.063	0.997	18.578	29.487	1.798
	100	1000	0.03	1	51.873	79.344	1.81
	200	1000	0	1	127.1	143.86	1.922
	500	1000	0.004	1	254.63	370.97	1.984
DGP5	30	1000	0.313	0.991	11.001	14.879	1.915
	50	1000	0.124	0.997	19.546	28.994	1.945
	100	1000	0.048	1	52.011	74.112	2.001
	200	1000	0.022	1	102.38	147.92	2.009
	500	1000	0.021	1	252.15	373.88	2.016

la media de los breaks sugeridos se encuentra muy por debajo o por encima del valor real), se hallaron valores de la media de las 1000 repeticiones con una distancia muy menor al break generado artificialmente para armar los DGP.

A su vez, la media de la cantidad de quiebres sugeridos también tiene un gran desempeño, pues salvo para las muestras muy menores (especialmente de N=30), dicha media de cantidad de quiebres se encuentra muy cercana al valor real. Este resultado se resalta aún más, si pensamos que si un proceso de simulación obtiene una media que se diferencia de la media real en centésimas o milésimas (e.g con valores de 1.008 o 2.003) es casi un resultado perfecto.

En cuanto a los p-values, de los dos tests seleccionados de los que sugieren Bai y Perron, el test de "double maximum" parece tener una eficiencia aceptable cuando no hay breaks en la serie. Especialmente para valores de N a partir de 100, ya que para N=30 o N=50 el size se encuentra entre 0.5 y 0.6 a grandes rasgos. A su vez, dicho test tiende a ser significativo todas las veces cuando la muestra a simular se agranda (para N=200 y 500, y algunas veces para N=100 también). [7]

Por último, el test de ℓ versus $\ell+1$ breaks presenta en todas las simulaciones, tanto para los DGP sin quiebres estructurales, como para los DGP con 1 y 2 cambios en las series, valores bajos del p-value del SupF. Esto termina resultando en que el size al 5% sea mayor a 0.90 aproximadamente en la tabla 4, y menor a 0.10 en general en las tablas 5 y 6.

A modo de conclusión general, se podría decir que el test presenta una performance aceptable a la hora de señalarnos el quiebre potencial de la serie, pero se debe tener cuidado a la hora de mirar los p-values que nos arroja el test luego de correrlo.

[7] Como se mencionó en la introducción, Bai y Perron (2006) aseguran que sus tests para múltiples “breaks” trabajan bien en muestras lo suficientemente grandes, pero sufren sustanciales desviaciones en la potencia de sus tests en muestras pequeñas.

3.3 Hansen

En las tablas 7, 8 y 9 se pueden observar los resultados de las simulaciones con el test de Hansen. En las mismas se puede observar la cantidad de veces (sobre 1000) que los *p-values* hallados fueron, en el caso de la tabla 7 menores a 0.05, y en el caso de las tablas 8 y 9, mayores a 0.05. En otras palabras, se puede observar la cantidad de veces que el test conduce a un resultado erróneo, ya que si el p valor es pequeño, se rechaza $H0$, cuya hipótesis es que no existe break (más específicamente, $\theta_n = 0$). El valor 0.05 es un valor elegido ad-hoc con el sólo propósito de ilustrar la proporción de veces que el valor p aumenta, y no representa selección del valor crítico de rechazo de cada test en particular.

Los códigos de Hansen se encuentran en su página, tanto para R como para Gauss Apthech[8].

Los estadísticos (descriptos brevemente anteriormente) son el supremo (S), el exponencial (E) y el promedio (A), y también se encuentran las columnas para los casos de bootstrap (SB, EB y AB) y para el caso con heterocedasticidad (SH, EH, AH).

Para la tabla 7, donde no ocurrió un cambio estructural en el DGP, los p valores tieden a no ser pequeños en la mayoría de los casos, agrandándose aún mas a medida que aumenta la muestra, y la media del quiebre sugerido se encuentra cercana a un valor cercano (sin tendencia a sobreestimar ni subestimar) a la mitad del período. Esto es buen resultado, considerando que el proceso no contiene break.

En la tabla 8, se pueden observar las mismas estadísticas pero para el caso donde sí ha ocurrido un quiebre, y en este caso en la media. Aquí, la cantidad de veces donde los p valores son grandes, son la gran mayoría (en el orden de más del 80%), salvo para el DGP 4 (el ARMA con otro proceso ARMA como explicativo), donde los p-values hallados son casi todos menores a 0.05.

En la tabla 9, donde se introduce un cambio estructural ubicado en la varianza del error, también la mayoría de los casos son casos con p-values grandes, y en el DGP4 siguen siendo menos casos que en el resto pero ya no tendiendo a cero.

Tanto en la tabla 8, como en la 9, la media de la fecha del break estimado parece contener algo de ruido, pues tras las 1000 simulaciones, se ubica en valores con varios períodos de diferencia con respecto al verdadero valor del quiebre.

A modo de cierre de la simulación del test de Hansen, caben remarcar algunas cuestiones.

El test en general no es aceptable a la hora de trabajar con una serie con break, de los estilos especificados en este trabajo ya que los valores de los p values suelen ser altos para las tablas 8 y 9 (donde los DGP contienen de hecho un cambio en su estructura y los estadísticos deberían mencionar la existencia de

[8] Bruce Hansen Homepage: http://www.ssc.wisc.edu/~bhansen/

Table 7. H simulation with no break

	N	*# Sim*	*S*	*E*	*A*	*SB*	*EB*	*AB*	*SH*	*EH*	*AH*	*Mean Break*
DGP1	30	1000	0.058	0.074	0.062	0.057	0.055	0.052	0.111	0.104	0.103	14.707
	50	1000	0.56	0.068	0.059	0.057	0.056	0.049	0.098	0.096	0.11	25.78
	100	1000	0.053	0.057	0.051	0.049	0.052	0.051	0.072	0.049	0.67	52.161
	200	1000	0.041	0.049	0.049	0.042	0.047	0.049	0.053	0.052	0.057	100.83
	500	1000	0.042	0.046	0.044	0.048	0.043	0.044	0.052	0.043	0.046	256.7
DGP2	30	1000	0.059	0.081	0.066	0.057	0.052	0.054	0.072	0.069	0.062	14.452
	50	1000	0.054	0.063	0.066	0.053	0.056	0.057	0.061	0.057	0.058	24.032
	100	1000	0.044	0.051	0.048	0.053	0.046	0.044	0.055	0.05	0.052	50.781
	200	1000	0.052	0.051	0.048	0.051	0.049	0.045	0.57	0.049	0.053	103.85
	500	1000	0.049	0.052	0.048	0.051	0.049	0.045	0.57	0.049	0.052	248.36
DGP3	30	1000	0.071	0.085	0.061	0.071	0.065	0.053	0.133	0.109	0.11	15.701
	50	1000	0.056	0.061	0.053	0.062	0.051	0.051	0.099	0.089	0.086	25.49
	100	1000	0.046	0.05	0.039	0.053	0.045	0.043	0.088	0.065	0.062	52.158
	200	1000	0.055	0.057	0.038	0.056	0.053	0.04	0.075	0.06	0.051	104.63
	500	1000	0.061	0.065	0.062	0.061	0.065	0.061	0.066	0.064	0.063	254.15
DGP4	30	1000	0.028	0.037	0.028	0.026	0.021	0.024	0.037	0.035	0.041	14.623
	50	1000	0.017	0.017	0.013	0.015	0.012	0.013	0.027	0.22	0.026	23.741
	100	1000	0.021	0.021	0.025	0.021	0.02	0.025	0.03	0.029	0.039	48.361
	200	1000	0.041	0.019	0.017	0.017	0.017	0.018	0.026	0.027	0.029	98.62
	500	1000	0.512	0.481	0.378	0.521	0.489	0.389	0.531	0.489	0.367	255.361
DGP5	30	1000	0.382	0.431	0.345	0.359	0.363	0.313	0.4	0.399	0.349	15.524
	50	1000	0.44	0.464	0.348	0.429	0.42	0.328	0.441	0.422	0.346	26.107
	100	1000	0.488	0.493	0.367	0.494	0.473	0.365	0.501	0.476	0.373	49.385
	200	1000	0.497	0.465	0.351	0.501	0.46	0.351	0.515	0.463	0.349	103.51
	500	1000	0.512	0.481	0.378	0.521	0.489	0.389	0.531	0.489	0.367	255.361

Table 8. H simulation with one break in mean

	N	*# Sim*	*S*	*E*	*A*	*SB*	*EB*	*AB*	*SH*	*EH*	*AH*	*Mean Break*
DGP1	30	1000	0.876	0.871	0.927	0.872	0.898	0.936	0.772	0.796	0.832	20.788
	50	1000	0.855	0.978	0.933	0.873	0.897	0.941	0.831	0.839	0.87	34.907
	100	1000	0.882	0.896	0.934	0.865	0.902	0.941	0.869	0.869	0.892	70.788
	200	1000	0.85	0.89	0.936	0.839	0.892	0.934	0.887	0.903	0.914	142.45
	500	1000	0.886	0.916	0.952	0.879	0.914	0.953	0.935	0.936	0.954	351.62
DGP2	30	1000	0.933	0.912	0.93	0.932	0.936	0.941	0.855	0.966	0.851	17.794
	50	1000	0.939	0.93	0.951	0.938	0.944	0.951	0.872	0.887	0.895	29.864
	100	1000	0.937	0.925	0.948	0.929	0.928	0.949	0.895	0.91	0.92	61.334
	200	1000	0.941	0.942	0.949	0.937	0.944	0.95	0.93	0.931	0.936	123.64
	500	1000	0.946	0.951	0.953	0.938	0.947	0.951	0.95	0.951	0.952	307.82
DGP3	30	1000	0.805	0.799	0.866	0.815	0.843	0.902	0.734	0.75	0.788	21.648
	50	1000	0.855	0.857	0.924	0.847	0.876	0.934	0.811	0.832	0.852	35.873
	100	1000	0.864	0.88	0.933	0.853	0.891	0.934	0.888	0.894	0.903	71.275
	200	1000	0.882	0.91	0.945	0.876	0.91	0.946	0.909	0.92	0.928	142.19
	500	1000	0.865	0.911	0.952	0.858	0.913	0.945	0.941	0.949	0.946	358.62
DGP4	30	1000	0	0	0.003	0.001	0.001	0.002	0.217	0.213	0.01	19.549
	50	1000	0	0	0.001	0	0	0	0.103	0.096	0	29.841
	100	1000	0	0	0.008	0	0	0.003	0.036	0.036	0.004	54.741
	200	1000	0	0	0.006	0	0	0.001	0.004	0.004	0.004	104.72
	500	1000	0	0	0	0	0	0	0	0	0	254.43
DGP5	30	1000	0.851	0.843	0.908	0.846	0.873	0.926	0.756	0.768	0.794	20.983
	50	1000	0.854	0.858	0.919	0.843	0.874	0.825	0.805	0.826	0.855	35.656
	100	1000	0.856	0.879	0.929	0.843	0.884	0.929	0.869	0.882	0.891	71.26
	200	1000	0.865	0.894	0.935	0.857	0.899	0.938	0.909	0.918	0.919	143.28
	500	1000	0.86	0.901	0.944	0.854	0.902	0.941	0.933	0.931	0.943	355.59

Table 9. H simulation with one break in variance

	N	*# Sim*	*S*	*E*	*A*	*SB*	*EB*	*AB*	*SH*	*EH*	*AH*	*Mean Break*
DGP1	30	1000	0.951	0.93	0.941	0.951	0.955	0.944	0.9	0.905	0.893	13.788
	50	1000	0.954	0.923	0.931	0.934	0.935	0.936	0.878	0.885	0.89	23.815
	100	1000	0.958	0.938	0.933	0.85	0.945	0.938	0.922	0.922	0.921	48.358
	200	1000	0.959	0.947	0.944	0.882	0.967	0.951	0.932	0.941	0.938	101.151
	500	1000	0.967	0.958	0.961	0.921	0.989	0.967	0.959	0.949	0.951	258.108
DGP2	30	1000	0.948	0.937	0.951	0.945	0.952	0.962	0.888	0.896	0.889	16.85
	50	1000	0.937	0.926	0.939	0.934	0.37	0.942	0.903	0.905	0.907	28.423
	100	1000	0.947	0.942	0.952	0.935	0.945	0.953	0.91	0.922	0.927	58.022
	200	1000	0.967	0.957	0.968	0.947	0.951	0.964	0.937	0.93	0.941	96.155
	500	1000	0.981	0.985	0.99	0.961	0.973	0.98	0.961	0.959	0.963	246.85
DGP3	30	1000	0.853	0.844	0.912	0.852	0.866	0.916	0.853	0.86	0.87	21.106
	50	1000	0.88	0.875	0.926	0.87	0.884	0.926	0.876	0.877	0.895	35.123
	100	1000	0.888	0.907	0.936	0.882	0.912	0.94	0.906	0.915	0.916	70.954
	200	1000	0.894	0.921	0.95	0.912	0.939	0.961	0.942	0.941	0.936	142.58
	500	1000	0.921	0.948	0.98	0.947	0.954	0.98	0.958	0.982	0.95	289.423
DGP4	30	1000	0.616	0.636	0.874	0.639	0.673	0.863	0.7	0.695	0.714	22.024
	50	1000	0.546	0.586	0.851	0.566	0.606	0.836	0.675	0.672	0.703	36.316
	100	1000	0.467	0.533	0.824	0.478	0.538	0.793	0.67	0.676	0.702	71.645
	200	1000	0.347	0.46	0.775	0.374	0.467	0.713	0.599	0.607	0.662	139.26
	500	1000	0.235	0.38	0.744	0.266	0.383	0.7	0.547	0.566	0.67	339
DGP5	30	1000	0.875	0.867	0.924	0.876	0.889	0.933	0.842	0.859	0.872	20.855
	50	1000	0.859	0.869	0.921	0.852	0.888	0.927	0.855	0.855	0.882	35.715
	100	1000	0.874	0.889	0.935	0.866	0.892	0.935	0.893	0.904	0.909	70.601
	200	1000	0.889	0.901	0.945	0.876	0.905	0.945	0.912	0.905	0.915	115.125
	500	1000	0.912	0.912	0.952	0.89	0.915	0.956	0.934	0.92	0.932	275.89

un quiebre significativo). Pero dichos p valores se mantienen en valores elevados para la tabla 7, donde no se especificó cambios en la estructura del DGP. Esto último puede presentar un problema si se entiende que parece que se puede estar sub estimando la presencia de quiebres.

Por otro lado, la media de los breaks sugeridos no se encuentra tan cercana al valor real como en el caso de la sub sección anterior, donde se testeaba a partir de BP.

Además, no se notan demasiadas diferencias (o por lo menos que aparenten ser grandes) entre los p values generados entre los distintos tests (los originales, los corregidos por heterocedasticidad, o los realizados por bootstrap).

Una mención importante se debe hacer con respecto a los resultados que presentó el test cuando el DGP era un ARMA que contenía un ARMA y el quiebre era en media. Los valores p para ese caso fueron muy pequeños a diferencia de lo que ocurre en el resto de los DGP (los break medios hallados sí son similares al resto de los casos). De hecho, para algunos estadísticos, las 1000 veces simuladas el p value indica la presencia de break.

Esto último es algo importante para remarcar. El test está pensando en algo distinto a los demás tests que aquí se exponen, y es que está estructurado bajo la idea de captar un cambio en la distribución del regresor (que puede afectar β si no se está bajo la presencia de súper exogeneidad). Esto es importante remarcar porque ninguno de los DGP generados aquí presentan cambios estructurales sobre la variable exógena, para poder contrastar los resultados entre los distintos tests (ya que el resto de los tests no enfocan el problema de esa manera). Justamente el DGP que mejor performance presenta cuando existe un break es aquel donde el proceso es un ARMA que contiene un ARMA (a pesar de que el break no se halla en el regresor sino en el error de la serie).

De todas maneras, para poder correr el test, a los DGP que se generaron para el test de Hansen, incluyen siempre un regresor X, que es una variable normal al igual que la serie en estudio, Y.

3.4 KP

Finalmente, se pasaron a realizar las simulaciones con el más moderno (por ser el más recientemente publicado) de los tests en estudio en este trabajo, el test de raíz unitaria propuesto por Kim y Perron (2009). Al igual que en caso de Hansen, se presentan tres tablas, las número 10, 11 y 12, conteniendo las mismas los casos sin quiebre, con quiebre en media y quiebre en varianza del error, respectivamente.

Para estas simulaciones se utilizaron los archivos disponibles en el *home page* de Dupka Kim[9]: *Dbreakestimate.m* y *UR2k.m.* El primero para buscar el quiebre estructural sugerido (para un modelo dinámico) y el segundo para testear raíz unitaria (para un modelo de tipo A2).

Lo que se puede ver en las mencionadas tablas, son la media del quiebre sugerido, y la media del estadístico t calculado. Además, en las tablas 11 y 12, se halla una columna que ilustra, al igual que en el caso de Zivot y Andrews, cuantas veces la variable indicadora de que el break se encuentra en el intervalo ± 5 que contiene al verdadero valor se hizo igual a 1.

Table 10. KP simulation with no break

	N	*# Sim*	*Mean Break*	*Mean t*
DGP1	30	1000	14.941	-2.769
	50	1000	25.234	-3.5012
	100	1000	47.861	-4.6084
	200	1000	99.321	-6.6084
	500	1000	256.981	-10.0449
DGP2	30	1000	12.405	-2.8849
	50	1000	25.445	-3.4936
	100	1000	49.913	-4.6909
	200	1000	101.296	-6.4903
	500	1000	254.98	-10.0653
DGP3	30	1000	14.784	-2.9019
	50	1000	25.739	-3.5838
	100	1000	51.158	-4.906
	200	1000	101.166	-6.7709
	500	1000	249.923	-10.49
DGP4	30	1000	15.244	-2.88
	50	1000	24.85	-3.6207
	100	1000	48.965	-4.7677
	200	1000	101.095	-6.7547
	500	1000	251.395	-10.4727
DGP5	30	1000	14.555	-2.7888
	50	1000	25.183	-3.466
	100	1000	50.111	-5.0118
	200	1000	101.985	-6.8564
	500	1000	251.985	-10.7812

[9]Los archivos se encuentran en *http://people.virginia.edu/~dk4p/*

Table 11. KP simulation with break in mean

	N	*# Sim*	*Mean Break*	*Mean t*	*Interval*
DGP1	30	1000	19.456	-1.98	0.68
	50	1000	28.412	-2.3815	0.75
	100	1000	48.952	-2.8571	0.965
	200	1000	99.815	-3.6633	0.98
	500	1000	249.730	-4.9915	1
DGP2	30	1000	17.2493	-2.23	0.77
	50	1000	27.988	-2.4357	0.798
	100	1000	50.988	-2.9344	0.953
	200	1000	100.304	-3.8207	0.98
	500	1000	250.237	-5.7627	0.98
DGP3	30	1000	14.675	-1.9587	0.173
	50	1000	24.81	-2.006	0.68
	100	1000	49.876	-2.0855	0.708
	200	1000	101.865	-2.3704	0.96
	500	1000	249.885	-3.0737	0.99
DGP4	30	1000	14.808	-1.981	0.921
	50	1000	24.859	-1.9815	0.965
	100	1000	49.884	-2.0232	0.98
	200	1000	99.925	-2.1784	1
	500	1000	249.923	-2.6762	1
DGP5	30	1000	19.956	-2.1451	0.926
	50	1000	24.963	-2.2827	0.942
	100	1000	49.97	-2.356	0.965
	200	1000	99.975	-2.5209	0.98
	500	1000	249.78	-3.0308	1

Table 12. KP simulation with break in variance					
	N	*# Sim*	*Mean Break*	*Mean t*	*Interval*
DGP1	30	1000	20.099	-2.8673	0.501
	50	1000	33.794	-3.4122	0.404
	100	1000	64.024	-4.6132	0.453
	200	1000	114.309	-6.4835	0.634
	500	1000	253.491	-10.0172	0.809
DGP2	30	1000	18.145	-2.2208	0.4015
	50	1000	29.103	-3.1587	0.4241
	100	1000	55.84	-4.6633	0.55
	200	1000	101.15	-6.7001	0.702
	500	1000	248.33	-10.4712	0.849
DGP3	30	1000	14.924	-2.7638	0.396
	50	1000	24.492	-3.5676	0.236
	100	1000	75.288	-4.9616	0.103
	200	1000	98.573	-6.6676	0.12
	500	1000	381.437	-10.4831	0.022
DGP4	30	1000	14.737	-2.7889	0.373
	50	1000	24.584	-3.4102	0.214
	100	1000	47.667	-4.7205	0.097
	200	1000	99.603	-6.702	0.06
	500	1000	242.304	-10.5355	0.025
DGP5	30	1000	21.432	-2.6832	0.351
	50	1000	36.972	-3.5609	0.17
	100	1000	75.259	-4.9008	0.082
	200	1000	153.385	-6.8021	0.035
	500	1000	381.429	-10.7826	0.014

Para redondear la simulación del test de Kim y Perron (2009), se pueden mencionar algunas cuestiones.

En primer lugar, los estadísticos t hallados tienden a aumentar la media a medida que la muestra aumenta su tamaño (en menor escala para el caso de break en media). Esto es una buena cualidad del test, ya que nos estaría indicando que el mismo es capaz de identificar que la serie es estacionaria (i.e. que no tiene raíz unitaria) correctamente. Probablemente requiera de una muestra mayor a 30, para lograr este objetivo al 1% de confianza.

La media del break hallado se encuentra relativamente cerca a la verdadera fecha del quiebre para el caso donde existe un break en la media, pero se encuentra más lejos cuando el cambio estructural ocurre en la varianza. Por este motivo es que la cantidad de veces que la estimación de la fecha de quiebre se halla dentro del intervalo ± 5 es considerablemente mayor para el caso donde el cambio en la serie ocurre en la media. De hecho, en el caso donde el quiebre se da en la varianza, la cantidad de veces que la indicadora toma valor uno se va haciendo cada vez menor a medida que aumenta la muestra. [10]

A grandes rasgos entonces se puede remarcar que el test presenta buenas características en cuanto a rechazar la hipótesis de raíz unitaria, ya que se obtienen en general estadísticos t en aumento a medida que se aumenta la muestra. Con respecto a la fecha del quiebre estructural sugerido, cabe destacar la discrepancia hallada entre los dos tipos de quiebres incluidos en este trabajo: en media y varianza.

[10] Para el caso donde no hubo quiebres en el DGP, la media de los breaks sugeridos se encuentra cercana al valor $T/2$ pero este hecho se debe a que es justo la mitad de la serie y no porque los breaks sugeridos se concentren alli. De hecho la dispersión es mayor que en el caso de la Tabla 11 por ejemplo, ya que si suponemos por un momento que el "verdadero quiebre" se encontraría allí y se fabricara un intervalo para armar la variable indicadora, la misma se hace menor cantidad de veces uno en la Tabla 10 que en la Tabla 11. Este intervalo ficticio no fue incluido en la Tabla del DGP sin quiebre.

4 Conclusiones

En este trabajo se ha realizado un sencillo ejercicio de simulación para ilustrar cuatro importantes tests que permiten la discusión de quiebres estructurales determinados de manera endógena, útiles a la hora de testear la presencia de raíz unitaria en una serie de tiempo.

La cuestión de quiebres estructurales es parte una muy importante a determinar en el estudio de series temporales, y dentro de este tópico, el hecho de endogeneizar la selección de potenciales quiebres estructurales es aún más importante. Bajo esta idea es que se propusieron estudiar y contrastar los tests de raíz unitaria de Zivot y Andrews (1992) y de Kim y Perron (2009), y los tests de quiebres propuestos por Bai y Perron (1998) y Hansen (2000). Todos ellos suman una característica importante y novedosa en el ámbito de breaks como se ha mencionado en la sección segunda del presente trabajo.

Para la tarea de contrastar dichos tests, se realizo un ejercicio de Monte Carlo sobre cinco procesos generadores de los datos (DGP) distintos, de manera de ejecutar las 1000 simulaciones de dicho ejercicio sobre los mismos DGP para cada test y tener un punto de comparación.

Los procesos seleccionados fueron muy sencillos, modelos con variables normales, modelos ARMA estacionarios y procesos con autocorrelación (frecuentes en la literatura de series temporales).

Un punto de comparación es el del correcto rechazo (o no) de la hipótesis nula de raíz unitaria por parte de los tests de ZA y KP. El primero tiene un buen desempeño a partir de N=100, donde rechaza en general unas 900 veces de las 1000 repeticiones (para N=30 en general rechaza alrededor del 70%). En cambio, KP tiene valores críticos del estadístico t relativamente grandes (alrededor de 2) incluso para las muestras más pequeñas, lo que puede estar representando una pequeña ventaja del test KP al tener la posibilidad de tener un quiebre estructural tanto bajo la nula como la alternativa.

Por otro lado, el otro punto de comparación es la correcta selección del quiebre estructural (o al menos estar en un área cercana al mismo). En este ámbito, parece correr con una ventaja por sobre los demás el test de BP, ya que para un break, como para dos breaks, la media de los quiebres sugeridos tiende a estar muy cercana a la verdadera. En el caso de Hansen, la media de los breaks sugeridos se encuentra a mayor distancia que la que se encuentra en BP, con grandes diferencias específicamente para los DGP 3, 4 y 5 (los que tienen los componentes autorregresivos). Pero Hansen presenta el grave problema de que si no acompañamos el test con otros tests, podemos concluir que no existe break en la serie, cuando en realidad no existe quiebre estructural en el regresor, pero sí en la serie.

KP tiene buenos resultados en estimar el break para el caso de break en media, pero la media de los breaks sugeridos se aleja del verdadero valor para el caso de break en varianza (y la distancia no se tiende a reducir a medida que aumenta la muestra, de hecho aumenta). ZA parece tener poco poder a la hora de sugerir el quiebre, ya que para muestras grandes, menos del 10% de las veces el break sugerido se halla en el intervalo ± 5.

Claro que esta comparación es solamente en el ámbito de los DGP seleccionados y varias extensiones posibles se deberían hacer para completar y tener un mejor cuadro de la comparación de los tests. En primer lugar, se podría pensar en incorporar otros tipos de DGP (como series con heterocedasticidad, o para el caso de Hansen en particular, series con quiebres estructurales en los regresores, o la distribución de los mismos mejor dicho)[11].

Otra posible extensión es simular las series con un coeficiente auto regresivo más cercano a la unidad (e.g. 0.8) en lugar de 0.5.

Por otro lado, los quiebres generados en general se ubicaron en la mitad de la serie, o en los tercios (incluso en el 3/4 para BP), pero sería interesante añadir quiebres en las colas (e.g. en las primeras o últimas observaciones) para ver la potencia de los tests cuando el break ocurre muy temprano o muy tarde en el período[12].

Como conclusión final se puede mencionar que lo que este trabajo ilustra es que todos los tests presentan sus ventajas y desventajas a la hora de que el investigador intente determinar endogenámente si existe un cambio estructural significativo en la serie a estudiar. Si se realiza esta estimación mediante una sola metodología se está cayendo en un trabajo con serios riesgos de estar concluyendo erróneamente la presencia o no de un quiebre, así como la estacionariedad o no de la serie. Es por eso, que al no conocerse los verdaderos procesos que llevaron a generar la serie, es recomendable la complementación de todos los tests mencionados. Es tarea del investigador intentar determinar que DGP es mas factible de acuerdo a la serie en cuestión con la que se esté trabajando, y así ponderar de distinta manera (y correctamente) los resultados de los distintos tests[13].

[11] De hecho, Hansen estudia las propiedades de muestra finita del test en su paper con un AR bivariado del siguiente estilo:

$$\begin{aligned} y_i &= \alpha_0 + \alpha_1 y_{i-1} + \alpha_2 y_{i-2} + \alpha_3 y_{i-3} \\ &\quad + \beta_1 x_{i-1} + \beta_2 x_{i-2} + \beta_3 x_{i-3} + e_i \end{aligned}$$

con N=50, $\alpha_1 = 0.5$ y $\beta_1 = 1$.

[12] Kim y Perron mencionan en su trabajo que si se selecciona la fracción verdadera del break en 0.5, en general los resultados no distan de aplicar el cambio estructural en 0.3 o 0.7. Ver Kim (2007) y Kim - Perron (2009), p 14.

[13] Como caso ilustrativo, se puede ver que si uno piensa que el proceso posee un cambio estructural en el regresor, y testea sobre esta hipótesis, correctamente puede determinar que no existe quiebre en la variable independiente. Pero puede incorrectamente concluir que no existe quiebre en toda la serie si no acompaña el test con otro que asuma un DGP de quiebres estructurales diferente.

5 Referencias Bibliográficas Citadas

- Andrews, D. W. K. (1993): “Tests for Parameter Instability and Structural Change with Unknown Change Point”. Econometrica 61.4, pp. 821–856.
- Bai, J. and P. Perron, (1998), “Estimating and Testing Linear Models with Multiple Structural Changes,” Econometrica 66, 47-78.
- ———, (2003a), “Critical values for multiple structural change tests,” Econometrics Journal 6, 72-78.
- ———, (2003b), “Computation and Analysis of Multiple Structural Change Models,” Journal of Applied Econometrics 18, 1-22.
- ———, (2006), “Multiple Structural Change Models: A Simulation Analysis,” in Econometric Theory and Practice: Frontier of Analysis and Applied Research (Essays in Honor of Peter Phillips), ed. by Corbae D., S. Durlauf and B.E. Hansen, Cambridge University Press.
- Bai, J., (1999), “Likelihood ratio tests for multiple structural changes,” Journal of Econometrics 91, 299-323.
- Banerjee and Urga (2005) “Modelling structural breaks, long memory and stock market volatily: an overview”. Journal of Econometrics, p. 1-34.
- Chow, G. C. (1960): “Tests of Equality Between Sets of Coefficients in Two Linear Regressions”. Econometrica 28.3, pp. 591–605.
- Gagliardini, P., Trojani, F., Urga, G., (2004), “Robust GMM tests for structural breaks”. Journal of Econometrics, 10.1016/j.jeconom.2004.09.006.
- Hansen, B. E. (2000): “Testing for structural change in conditional models”, Journal of Econometrics, Elsevier.
- Harvey, D.I., Leybourne, S.J., Newbold, P., (2001). “Innovational outlier unit root tests with an endogenously determined break in level”. Oxford Bulletin of Economics and Statistics 63, 559–575.
- Kim, D., Perron, P., “Unit Root Tests Allowing for a Break in the Trend Function at an Unknown Time under Both the Null and Alternative Hypotheses”, Journal of Econometrics (2009) 148, 1-13.
- Lee, J., Strazicich, M., (2001). “Break point estimation and spurious rejections with endogenous unit root tests”. Oxford Bulletin of Economics and Statistics 62, 535–558.
- Leybourne, S.J., Mills, T.C., Newbold, P., (1998). “Spurious rejections by Dickey–Fuller tests in the presence of a break under the null”. Journal of Econometrics 87, 191–203.
- Nelson, C.R., Plosser, C.I., (1982). “Trends and random walks in macroeconomic time series”. Journal of Monetary Economics 10, 139–162.
- Nunes, L. Kuan, CM ., Newbold, P. , (1995) “Spurious break”, Econometric Theory- Cambridge University Press
- Perron, P., Vogelsang, T.J., (1992). “Testing for a unit root in a time series with a shift in mean, corrections and extensions”. Journal of Business and Economic Statistics 10, 467–470.
- Quandt, R. (1960), "Tests of the hypothesis that a linear regression system obeys two separate regimes", Journal of the American statistical Association, 55, 324-330.

• Rappoport, P., Reichlin, L., (1989). "Segmented trends and nonstationary time series". The Economic Journal 99 (Conference Volume), 168–177.

• Zivot, E. y D. Andrews (1992). "Further Evidence on the Great Crash, the Oil-Price Shock and the Unit-Root Hypothesis", Journal of Business and Economic Statistics, 10(3), 251-270.

5.1 Referencias Bibliográficas extras

• Altissimo, F. and V. Corradi, (2003), "Strong rules for detecting the number of breaks in a time series," Journal of Econometrics 117, 207-244.

• Andrews, D. W. K., I. Lee, and W. Ploberger, (1996), "Optimal change point tests for normal linear regression," Journal of Econometrics 70, 9-38.

• Andrews, D. W. K. (2003): "Tests For Parameter Instability and Structural Change with Unknown Change Point: A Corrrigendum". Econometrica 71.1, pp. 395–397.

• Andrews, D. W. K., I. Lee, and W. Ploberger (1996): "Optimal Change point Tests for Normal Linear Regression". Journal of Econometrics 70.1, pp. 9–38.

• Andrews, D. W. K. and W. Ploberger (1994): "Optimal Tests When a Nuisance Parameter Is Present Only Under the Alternative". Econometrica 62.6, pp. 1383–1414.

• Bai, J. (1997): "Estimating multiple breaks one at a time". Econometric Theory 13.3, pp. 315–352.

• Banerjee, Lumsdaine, Stock (1992), "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence", Journal of Business & Economic Statistics, Vol. 10, No. 3 (Jul., 1992), pp. 271-287

• Brown, R., J. Durbin, and J. Evans (1975): "Techniques for testing the constancy of regression relationships over time". Journal of the Royal Statistical Society B 37, pp. 149–163.

• Campos, J., N. R. Ericsson, and D. F. Hendry, (1996), "Cointegration tests in the presence of structural breaks," Journal of Econometrics 70, 187-220.

• Clements, M.P., Hendry, D.F., (1996). "Intercept corrections and structural change". Journal of Applied Econometrics 11, 475–494.

• Davies, R. B. (1977): "Hypothesis testing when a nuisance parameter is present only under the alternative". Biometrika 64.2, pp. 247–254.

• Diebold, F. X. and C. Chen (1996): "Testing structural stability with endogenous breakpoint: A size comparison of analytic and bootstrap procedures". Journal of Econometrics 70.1, pp. 221–241.

• El-Shagi, M. (2010): "Small Sample Correction for the Alternative CUSUM-Tests." mimeo.

• Garcia, R. and P. Perron, (1996), "An Analysis of the real Interest Rate Under Regime Shifts, "Review of Economics and Statistics 78, 111-125.

• Hansen, B. E. (1990): "Lagrange Multiplier Tests for Parameter Instability in Non-Linear Models." University of Rochester.

- Hansen, B. E. (1992): “Testing for parameter instability in linear models”. Journal of Policy Modeling 14.4, pp. 517–533.
- Hansen, B. E. (1997): “Approximate asymptotic p-values for structural change tests”. Journal of Business and Economic Statistics 15.1, pp. 60–67.
- Hansen, B. E. (2001): “The New Econometrics of Structural Change: Dating Breaks in U.S. Labor Productivity”. The Journal of Economic Perspectives 15.4, pp. 117– 128.
- Hawkins, D. (1987): “A test for a change point in a parametric model based on a maximal Wald-type statistic”. Sankhya: The Indian Journal of Statistics 49.Series A, pp. 368–376.
- Hendry D. (1999), “On detectable and non-detectable structural change”, Structural Change and Economic Dynamic, Elsevier.
- Krishnaiah, P. R. and B. Q. Miao, (1988), “Review about Estimation of Change Points,” in Handbook of Statistics, Vol. 7, ed. By P. R. Krishnaiah and C. R. Rao. New York: Elsevier.
- Lin, C.-F., Terasvirta, T., (1994). “Testing the constancy of regression parameters against continuous structural change”. Journal of Econometrics 62, 211–228.
- Lin, C.-F., Terasvirta, T., (1999). “Testing parameter constancy in linear models against stochastic stationary parameters”. Journal of Econometrics 90, 193–213.
- Liu, J., S. Wu, and J. V. Zidek, (1997), “On Segmented Multivariate Regression,” Statistica Sinica 7, 497-525.
- Montanes, A., Olloqui, I., Calvo, E., (2004). “Selection of the break in Perron-type tests”. Journal of Econometrics, 10.1016 /j.jeconom.2004.09.003.
- Nyblom, J. (1989): “Testing for the constancy of parameters over time”. Journal of the American Statistical Association 84.405, pp. 223–230.
- Perron, P. (2006): “Dealing with Structural Breaks.” Palgrave Handbook of Econometrics.Ed. by T. Mills and K. Patterson. Vol. 1. Palgrave Macmillan, pp. 278–352.
- Pesaram, H. and A. Timmermann, (2000), “Model instability and the choice of observations window,” Mimeo, UCSD and University of Cambridge.
- Stock, J. H. and M. W. Watson (1996): “Evidence on Structural Instability in Macroeconomic Time Series Relations”. Journal of Business & Economic Statistics 14.1, pp. 11–30.
- Vogelsang, T.J., Perron, P., (1998). “Additional tests for unit root allowing the possibility of breaks in the trend function”. International Economic Review 39, 1073–1110.
- Yao, Y.-C. (1988). “Estimating the number of change-points via Schwarz’ criterion.” Statistics and Probability Letters 6, 181-189.
- Zacks, S., (1983), “Survey of Classical and Bayesian Approaches to the Change-Point Problem: Fixed and Sequential Procedures of Testing and Estimation,” in Recent Advances in Statistics, ed. By M. H. Rivzi, J. S. Rustagi, and D. Sigmund. New York: Academic Press, 245-269.
- Zeileis, A. (2005): “A Unified Approach to Structural Change Tests Based on ML Scores, F Statistics, and OLS Residuals”. Econometric Reviews

24.4, pp. 445–466.

- Zeileis, A., C. Kleiber, W. Krämer, and K. Hornik (2003): "Testing and Dating of Structural Changes in Practice". Computational Statistics & Data Analysis 44.1–2, pp. 109–123.
- Zeileis, A., A. Shah, and I. Patnaik (2010): "Testing, monitoring, and dating structural changes in exchange rate regimes". Computational Statistics & Data Analysis 54.6, pp. 1696–1706.