

Raíz Unitaria, Mínimos Cuadrados Generalizados y Modelos con Cambio Estructural

Gabriel Rodríguez

Universidad de Ottawa
Banco Central de Reserva del Perú

XXIV Encuentro de Economistas del Banco Central de Reserva del Perú
Diciembre 13-15, 2006

1. Introducción

- La potencia de los tests de raiz unitaria
- Estado actual de la literatura: Stock (1994), Elliott, Rothenberg y Stock (1996), Phillips y Xiao (1999), Maddala y Kim (2000), Haldrup y Jansson (2006)
- Referencias para la presente investigación: Hansen (1995), Elliott y Jansson (2003), Perron and Rodríguez (2003)
- Objetivo de la presente investigación
- Principales Resultados

2. El Proceso Generador de Datos (DGP)

- Tenemos el siguiente proceso generador de datos:

$$y_t = d_{yt} + u_{yt}, \quad (1)$$

$$x_t = d_{xt} + u_{xt}, \quad (2)$$

$$A(L) \begin{pmatrix} [1 - \rho L]u_{y,t} \\ u_{x,t} \end{pmatrix} \equiv A(L) u_t(\rho) = e_t \quad (3)$$

$$\rho = 1 + cT^{-1} \quad (4)$$

- La expresión (4) corresponde al marco teórico denominado local a la unidad (“local to unity”) propuesto por Phillips (1987), Chan y Wei (1987).
- $A(L)$ is una matriz polinomial de orden k and L es el operador de retardos.
- x_t , es un vector $m \times 1$ conteniendo un número arbitrario de variables estacionarias (co-variables) que tienen información adicional respecto de la variable y_t .
- Las expresiones d_{yt} y d_{xt} denotan los componentes determinísticos de y_t y x_t .

- La expresión

$$\Omega = A(1)^{-1}\Sigma A(1)^{-1'} = \begin{bmatrix} \omega_{yy} & \omega_{yx} \\ \omega'_{yx} & \Omega_{xx} \end{bmatrix} \quad (5)$$

es la densidad espectral a la frecuencia cero (con una escala de 2π) de $u_t(\rho)$

- Es posible definir $R^2 = \omega_{yy}^{-1}\omega_{yx}\Omega_{xx}^{-1}\omega'_{yx}$ como una medida de la correlación de largo plazo entre los choques a x_t y las cuasi-diferencias de y_t a la frecuencia cero.
- El valor R^2 representa la contribución de las co-variables (x_t) a la explicación de y_t . Este valor se encuentra en el rango de 0 a 1.
- Cuando $R^2 = 0$, las co-variables (x_t) no contienen información sobre el comportamiento de la variable potencialmente integrada y_t . Este es el tradicional caso donde no existen co-variables (x_t) en el test de raíz unitaria.
- Cuando $R^2 = 1$, esto significa que existe perfecta correlación entre las co-variables (x_t) y las cuasi-diferencias de y_t . En este caso, las sumas parciales de x_t cointegran con y_t .
- En la mayoría de casos, R^2 se encuentra en el intervalo (0,1) y la presente investigación se interesa en el análisis de la performance del test de raíz unitaria cuando R^2 aumenta.

Supuestos

- $|A(z)| = 0$ tiene sus raíces fuera del círculo unitario
- $E_{t-1}(e_t) = 0$, $E_{t-1}(e_t e_t') = \Sigma$ y $\sup_t E \|e_t\|^{2+\eta} < \infty$ (a.s.) para algún $\eta > 0$, donde Σ es una matriz definida positiva y $E_{t-1}(\cdot)$ denota esperanza condicional con respecto a $\{e_{t-1}, e_{t-2}, \dots\}$
- u_0, u_1, \dots, u_{t-k} , son $O_p(1)$

Modelos

- Utilizamos los mismos dos modelos trabajados en Perron y Rodríguez (2003)
- El primer modelo incluye un intercepto, una tendencia lineal y un quiebre en la pendiente de la función de tendencia. Esto significa que los componentes determinísticos de la variable y_t son

$$d_{yt} = \beta_1 + \beta_3 t + \beta_5 (t - T_B) \mathbf{1}(t > T_B),$$

donde $\mathbf{1}(\cdot)$ denota la función indicadora y T_B es la fecha (desconocida) del punto de ruptura.

- El segundo modelo incluye un intercepto, una tendencia lineal, un quiebre en el intercepto y en la pendiente. Esto significa que los componentes determinísticos de la variable y_t son:

$$d_{yt} = \beta_1 + \beta_3 t + \beta_5 (t - T_B) \mathbf{1}(t > T_B) + \beta_6 \mathbf{1}(t > T_B).$$

- Los componentes determinísticos de las co-variables (x_t) incluyen un intercepto y una tendencia lineal en los dos casos. Esto es, $d_{xt} = \beta_2 + \beta_4 t$.
- Es fácil observar que el primer caso puede obtenerse a partir del segundo caso imponiendo $\beta_6 = 0$.

- Para resumir hasta aquí:
 - Tenemos un modelo VAR que incluye las co-variables x_t y las cuasi-diferencias de y_t . Nos interesa verificar si y_t tiene una raíz unitaria bajo la hipótesis nula contra la hipótesis alternativa que y_t tiene una raíz menor que la unidad.
 - Utilizando el contexto univariado, King (1980, 1988) y Dufour y King (1991) han mostrado que la hipótesis de raíz unitaria puede ser examinada utilizando el marco teórico univariado propuesto por Neyman-Pearson.
 - Elliott, Rothenberg y Stock (1996) han desarrollado una familia de estadísticos que son óptimos en un punto (de la hipótesis alternativa) utilizando datos en cuasi-diferencias y en modelos sin cambio estructural.
 - Perron y Rodríguez (2003) han extendido el estadístico óptimo en un punto a modelos con cambio estructural y asumiendo que la fecha del quiebre es desconocida.

Un estadístico óptimo en un punto

- Como primera etapa, derivamos la distribución de un test óptimo (denotado por P) asumiendo que todos los parámetros de “*nuisance*” son conocidos excepto el punto de ruptura.
- Tenemos los siguientes supuestos: i) $A(L) = I$, en consecuencia $\Omega = \Sigma$; ii) e_t es distribuido según una distribución normal con $u_{y_0} = 0$; iii) $T_B = T\lambda$ para algún $\lambda \in (0, 1)$.
- Basado en los supuestos arriba mencionados, el objetivo es testear la hipótesis nula denotada por $c = 0$ contra la hipótesis alternativa local que $c = \bar{c} < 0$ donde $c = T(\rho - 1)$ y $\bar{c} = T(\bar{\rho} - 1)$.
- En el contexto univariado cuando el punto de ruptura es desconocido, Perron y Rodríguez (2003) han mostrado cual es la formula a seguir. En nuestro caso, el contexto es multivariado. Por eso adecuamos la sugerencia de Perron y Rodríguez (2003) a nuestro objetivo. De acuerdo con ello, la formula del test óptimo es la siguiente

$$P^i(1, \bar{\rho}) = \inf_{\lambda \in [0,1]} \sum_{t=1}^T \hat{u}_t^i(\bar{\rho}, \lambda)' \Sigma^{-1} \hat{u}_t^i(\bar{\rho}, \lambda) - \inf_{\lambda \in [0,1]} \sum_{t=1}^T \hat{u}_t^i(1, \lambda)' \Sigma^{-1} \hat{u}_t^i(1, \lambda) - \bar{c}. \quad (6)$$

donde, en el presente caso, $\hat{u}_t^i(r) = z_t(r) - d_t(r)' \hat{\beta}^i$ y $r = \bar{\rho}, 1$.

- Cuando $t > 1$, tenemos:

$$\begin{aligned} z_t(r) &= [(1-rL)y_t, x_t']' \\ d_t(r)' &= \begin{bmatrix} 1-r & 0 & (1-rL)t & 0 & (1-rL)(t-T_B)\mathbf{1}(\cdot) & (1-r)\mathbf{1}(\cdot) \\ 0 & I_m & 0 & I_m t & 0 & 0 \end{bmatrix} \end{aligned}$$

- Cuando $t = 1$, tenemos:

$$\begin{aligned} z_1(r) &= [y_1, x_1']' \\ d_1(r)' &= \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & I_m & 0 & I_m & 0 & 0 \end{bmatrix}. \end{aligned}$$

- Asimismo tenemos que

$$\hat{\beta}^i = [S_i(\sum_{t=1}^T d_t(r) \Sigma^{-1} d_t(r)') S_i]^{-1} [S_i \sum_{t=1}^T d_t(r) \Sigma^{-1} z_t(r)],$$

donde D^- es la matriz inversa de Moore Penrose D . El próximo teorema establece la distribución asintótica del estadístico óptimo.

Theorem 1 Sean $\{y_t\}$ y $\{x_t\}$ generados por las expresiones (1)-(4), con $A(L) = I$, $e_t \sim i.i.d. N(0, \Sigma)$, y los supuestos (i)-(iii) se mantienen. Para testear la hipótesis nula de $\rho = 1$ ($c = 0$) contra la hipótesis alternativa de $\rho = \bar{\rho} = 1 + \bar{c}/T$ con $c = \bar{c}$ fijo cuando $T \rightarrow \infty$, el test P tiene la siguiente distribución asintótica para los casos $i = 1$ and 2 :

$$P^i(1, \bar{\rho}) \Rightarrow \Lambda_1(c, \bar{c}, R^2) + \Lambda_2^i(c, \bar{c}, \lambda, R^2) - \bar{c} \quad (7)$$

donde $\Lambda_1(c, \bar{c}, R^2) = (\bar{c}^2 - 2c\bar{c})(1 + \frac{R^2}{1-R^2}) \int W_{1c}^2 - 2\bar{c} \int W_{1c} dW_1 + 2\bar{c} \frac{R}{\sqrt{1-R^2}} \int W_{1c} dW_2$, W_1 y W_2 son movimientos Brownianos univariados e independientes. W_{1c} es un proceso denominado Ornstein-Uhlenbeck definido por $W_{1c}(s) = c \int_0^s e^{c(s-h)} W_1(h) dh + W_1(s)$.

- La potencia asintótica depende de \bar{c} , el cual corresponde a un punto particular de la hipótesis alternativa. Esto significa que las funciones de potencia asintótica son tangentes a la potencia máxima (“envelope”) en el punto $c = \bar{c}$.
- La distribución asintótica no es estándar.
- La distribución depende del parámetro R^2 . Cuando $R^2 = 0$, esto significa que no existe co-variables (x_t) que sean correlacionadas con las cuasi-diferencias de y_t . En este caso re-encontramos la misma distribución asintótica derivada por Perron y Rodríguez (2003). Cuando $R^2 > 0$, la distribución asintótica es una función de R^2 , indicando que información extra o adicional en las co-variables (x_t) pueden hacer una diferencia en la performance del test.
- Utilizando el Theorema 1, $T = 1,000$, y 10,000 replicaciones, las funciones de potencia asintótica y los “envelopes” de potencia para diferentes valores de R^2 son calculados. Los resultados de la simulación para $R^2 = 0.0, 0.3, 0.5, 0.7, 0.9$ son mostrados en la Figura 1.

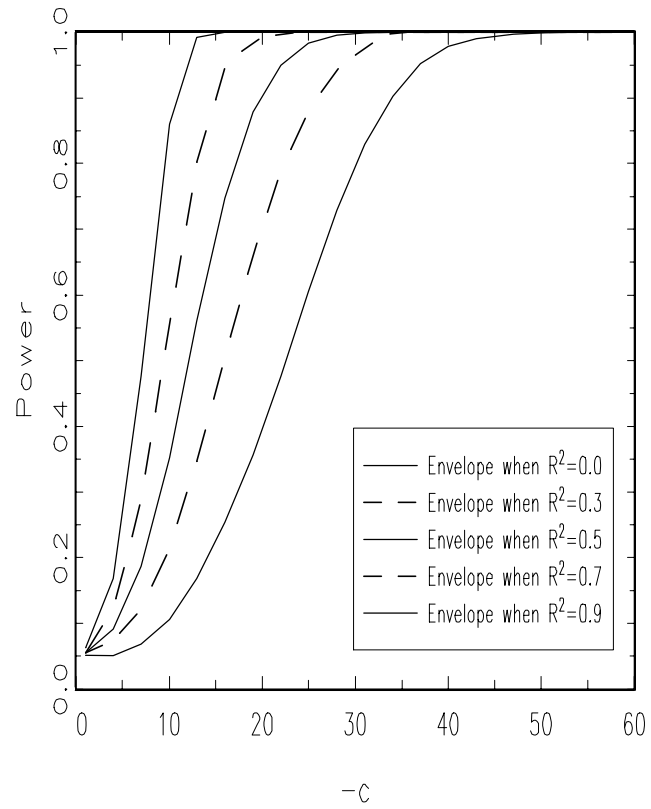


Figure 1. Power Envelopes for $R^2 = 0.0, 0.3, 0.5, 0.7, 0.9$.

- Se observa que las potencias máximas (“*envelope*”) alcanzan sus límites inferiores cuando $R^2 = 0$. Cuando $R^2 > 0$, la potencia lograda aumenta de manera considerable.
- Ejemplos:
 - Cuando $c = -10$, la potencia máxima para $R^2 = 0.0, 0.3, 0.5, 0.7, 0.9$ son 10.61%, 21.27%, 35.14%, 55.33%, and 85.95% respectivamente.
 - La potencia se incrementa en 100% desde $R^2 = 0.0$ hasta $R^2 = 0.3$; en 65.2% para $R^2 = 0.3$ hasta 0.5; en 57.5% para $R^2 = 0.5$ hasta 0.7; y en 55.3% para $R^2 = 0.7$ hasta 0.9.
 - La evidencia sugiere que tomando en cuenta la información contenida en las co-variables (x_t) se puede incrementar la potencia de manera significativa. Este resultado es consistente con lo hallado por Hansen (1995), y Elliott y Jansson (2003) en modelos sin cambio estructural.

“A Feasible Test”

- Hasta ahora hemos asumido que todos los parámetros de “*nuisance*” son conocidos excepto por el punto de ruptura. Asumimos también que los errores son distribuidos según una distribución normal. Además, en la práctica, R^2 , que mide la correlación entre x_t y la variable a ser testeada, no es conocido.
- Es necesario desarrollar un test que estime todos los parámetros de “*nuisance*”. Aquí, derivamos dicho test adoptando el enfoque propuesto por Elliott y Jansson (2003) para las primeras tres etapas. Las otras etapas son modificadas con la intención de estimar el punto de ruptura. La construcción del test es la siguiente:

- Estimar R^2 . Esto es logrado estimando un modelo VAR:

$$A(L)z_t(1, \lambda) = d_t + e_t,$$

donde λ es el punto de quiebre, d_t son los componentes determinísticos específicos al modelo estudiado. Utilizar los residuos obtenidos del modelo VAR para construir

$$\begin{aligned}\hat{\Sigma} &= T^{-1} \sum_{t=k+1}^T \hat{e}_t(\lambda)\hat{e}_t(\lambda)', \\ \hat{\Omega} &= \hat{A}(1, \lambda)^{-1}\hat{\Sigma}\hat{A}(1, \lambda)^{-1'}.\end{aligned}$$

Asi obtenemos $\hat{R}^2 = \hat{\omega}_{yx} \hat{\Omega}_{xx}^{-1} \hat{\omega}'_{yx} / \hat{\omega}_{yy}$.

- Estimación de los parametros de “*nuisance*” para la cuasi-diferenciación y eliminación de la tendencia de los datos bajo la hipótesis nula y alternativa, respectivamente. Esto es, construimos $\tilde{u}_t^i(r, \lambda) = z_t(r, \lambda) - d_t(r, \lambda)' \tilde{\beta}^i(r, \lambda)$ donde $r = 1, \rho, y$

$$\tilde{\beta}^i(r, \lambda) = \left[\sum_{t=1}^T d_t(r, \lambda) \hat{\Omega}^{-1} d_t(r, \lambda)' \right]^{-1} \left[\sum_{t=1}^T d_t(r, \lambda) \hat{\Omega}^{-1} z_t(r, \lambda) \right].$$

- Estimar un modelo VAR:

$$\tilde{A}(L) \tilde{u}_t^i(r, \lambda) = \tilde{e}_t^i(r, \lambda),$$

utilizando $\tilde{u}_t^i(r, \lambda)$ obtenidos de la etapa precedente. A partir de esto, se construye la matriz de variance-covarianzas

$$\tilde{\Sigma}^i(r, \lambda) = T^{-1} \sum_{t=k+1}^T \tilde{e}_t^i(r, \lambda) \tilde{e}_t^i(r, \lambda)'$$

- Seleccionar el punto de quiebre y construcción del test. El punto de quiebre es seleccionado a partir de la selección de las matrices de varianza-covarianza $\tilde{\Sigma}(1)$ y $\tilde{\Sigma}(\bar{\rho})$ minimizando $\sum_{t=k+1}^T \hat{e}_{y,t}^i(r, \lambda) \hat{e}_{y,t}^i(r, \lambda)'$ para $r = 1, \bar{\rho}$, respectivamente. La construcción del test estadístico es la siguiente:

$$\tilde{P}^i(1, \bar{\rho}) = T \{ \text{tr}[\tilde{\Sigma}(1)^{-1} \tilde{\Sigma}(\bar{\rho})] - [m + \bar{\rho}] \}. \quad (8a)$$

- Este test tiene potencia asintótica que es tangente al “envelope” de potencia en el punto \bar{c} y bajo los supuestos especificados.
- El siguiente Theorema establece la distribución limite del estadístico:

Theorem 2 *Sea $\{y_t\}$ y $\{x_t\}$ dos series de tiempo generadas según las ecuaciones (1)-(4), los supuestos (i)-(iii) se mantienen, asuma que los componentes determinísticos son especificados correctamente para $i = 1, 2$. Para testear la hipótesis nula de $\rho = 1$ ($c = 0$) contra la hipótesis alternativa de $\rho = \bar{\rho} = 1 + \bar{c}/T$ con $c = \bar{c}$ fijo cuando $T \rightarrow \infty$, la distribución limite del test que es definida en (8a) tiene la siguiente distribución asintótica:*

$$\tilde{P}^i(1, \bar{\rho}) \Rightarrow \Lambda_1(c, \bar{c}, R^2) + \Lambda_2^i(c, \bar{c}, \lambda, R^2) - \bar{c}. \quad (9)$$

- Como ya se mencionó anteriormente, la distribución limite del test (y los resultados asintóticos) dependen de la selección de \bar{c} para eliminar la tendencia en los componentes determinísticos.
- Hay también dependencia de R^2 .

- Idealmente se debería calcular la máxima potencia (“*power envelope*”) correspondientes para cada valor de R^2 . The selección of \bar{c} para la cuasi-diferenciación también podría depender del valor de R^2 . Sin embargo, en la practica, podemos usar $\bar{c} = -22.5$, el cual es el mismo valor seleccionado por Perron y Rodríguez (2003) cuando $R^2 = 0$. Visto el hecho que la potencia del test aumenta cuando R^2 aumenta, la elección de \bar{c} se vuelve menos importante (ver Elliott y Jansson, 2003).
- Los valores criticos han sido creados tabulando la distribution asintótica utilizando 1,000 “*steps*” para aproximar el proceso de Wiener en el intervalo $[0, 1]$ como las sumas parciales de variables aleatorias *i.i.d.* $N(0, 1)$.
- Los valores criticos finitos son también calculados para ambos casos analizados. Utilizamos $T = 100$ y 200 con datos generados como una marcha aleatoria con condición inicial igual a cero y errores *i.i.d.* $N(0, 1)$.
- El número de retardos es igual a cero en la estimación VAR de la primera etapa.

Evaluación del Test

- Figure 2 muestra la potencia asintótica del test propuesto para los dos casos y para $R^2 = 0.0, 0.3, 0.5, 0.7, 0.9$.
- El “*feasible test*” tiene potencia que es tangente a la máxima potencia en un punto y nunca cae.
- Importantes ganancias en potencia son obtenidos debido a la inclusión de co-variables (x_t). Considere la ganancia en la potencia asintótica cuando se tiene $R^2 = 0.5$ y la hipótesis local alternative es tal que $c = -10$. En este caso, la potencia asintótica aumenta en 217% (potencia es 34.52% cuando $R^2 = 0.5$, y 10.86% when $R^2 = 0.0$). Utilizando la hipótesis alternativa de $c = -4$, las ganancias de potencia son menores pero son apreciables (87% aproximadamente).

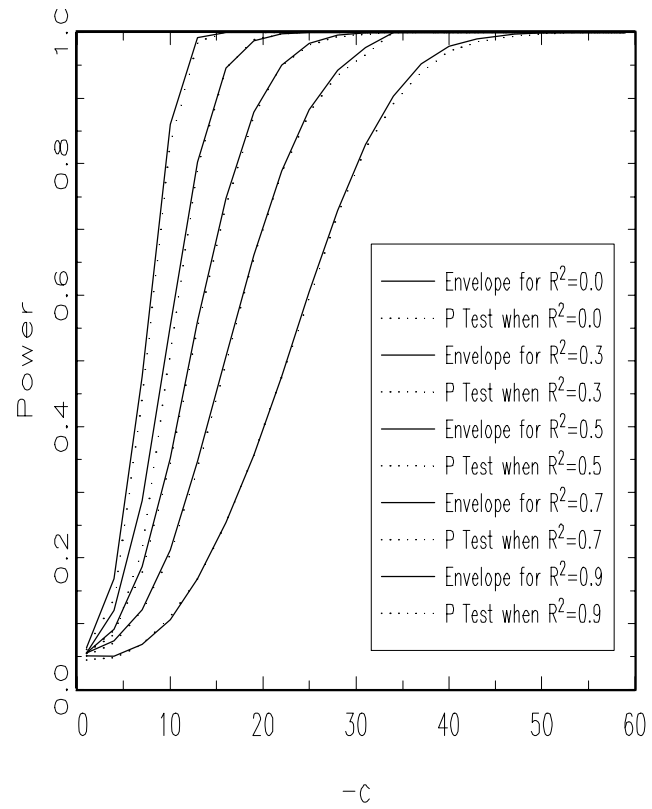


Figure 2. Power Envelopes and Asymptotic Power Functions of the Feasible Point Optimal Test

- La performance del test ha sido también evaluada utilizando un análisis de Monte Carlo en muestras pequeñas. Para simular el tamaño y la potencia del test, $\{y_t\}$ y $\{x_t\}$ son generados utilizando (1)-(4) con errores normales y matriz de varianza-covarianza conocida. Las varianzas toman el valor de la unidad y las autocovarianzas son 0.3, 0.5, 0.7, 0.9 respectivamente. Utilizando 10,000 replications, $T = 100, 200$, valores criticos en muestras finitas son calculados (ver Table 1).
- Dos factores influncian la performance del test en muestras finitas:
 - El número de observaciones. La potencia aumenta cuando T aumenta. Por ejemplo, cuando $\rho = 0.90$, $R^2 = 0.81$, el poder se incrementa en 15% cuando T aumenta de 100 a 200 para el Caso 1 y 26% para el Case 2.
 - El otro factor es el incremento de la correlation entre la co-variables (x_t) y la variable y_t . A medida que R (o R^2) es cercano a la unidad, la potencia se incrementa de manera significativa. Por ejemplo, la potencia aumenta en 19% cuando R aumenta de 0.3 a 0.5 cuando $\rho = 0.90$.

- Para el Caso 1, cuando $T = 100$ y $\rho = 0.96$, el test tiene una potencia de 7.36% sin la utilización de co-variables (x_t). La potencia aumenta a 11.57% cuando $R^2 = 0.49$. Esto es una ganancia aproximada de 57.2%. Para el Caso 2 existe una ganancia de 52.4%.
- Cuando $T = 200$, $\rho = 0.96$, la potencia es de 11.87% sin la presencia de co-variables (x_t). Dicha potencia aumenta a 29.19% cuando $R^2 = 0.49$, esto es, una ganancia aproximada de 146% . Para el Caso 2, dicha ganancia asciende a 120%.

Table 2. Finite Sample Size and Power; Case 1
($\bar{c} = -22.5$ when constructing the test)

	$T = 100$					$T = 200$				
$R =$	0.0	0.30	0.50	0.70	0.90	0.0	0.30	0.50	0.70	0.90
$R^2 =$	0.0	0.09	0.25	0.49	0.81	0.0	0.09	0.25	0.49	0.81
ρ										
1.00	0.0501	0.0501	0.0501	0.0501	0.0501	0.0501	0.0501	0.0500	0.0500	0.0500
0.98	0.0565	0.0567	0.0610	0.0755	0.1033	0.0733	0.0752	0.0873	0.1206	0.2121
0.96	0.0736	0.0750	0.0795	0.1157	0.2082	0.1187	0.1378	0.1679	0.2919	0.6508
0.94	0.0896	0.0967	0.1127	0.1838	0.3956	0.2029	0.252	0.3214	0.5796	0.9661
0.92	0.1344	0.1462	0.1634	0.2894	0.6482	0.3455	0.3555	0.5306	0.8325	0.9997
0.90	0.1839	0.1948	0.2312	0.4284	0.8633	0.5218	0.5409	0.7441	0.9525	1.0000
0.88	0.2390	0.2592	0.3252	0.5825	0.9685	0.6796	0.7223	0.8821	0.9902	1.0000
0.86	0.3156	0.3777	0.4226	0.7150	0.9942	0.8030	0.8511	0.9498	0.9986	1.0000

Number of lags in VAR is zero; 10,000 replications.

Table 3. Finite Sample Size and Power; Case 2
($\bar{c} = -22.5$ when constructing the test)

	$T = 100$					$T = 200$				
$R =$	0.0	0.30	0.50	0.70	0.90	0.0	0.30	0.50	0.70	0.90
$R^2 =$	0.0	0.09	0.25	0.49	0.81	0.0	0.09	0.25	0.49	0.81
ρ										
1.00	0.0500	0.0501	0.0501	0.0500	0.0501	0.0501	0.0501	0.0500	0.0501	0.0501
0.98	0.0585	0.0595	0.0615	0.0775	0.0952	0.0628	0.0674	0.0832	0.1078	0.1882
0.96	0.0765	0.0775	0.0798	0.1166	0.1886	0.1239	0.1323	0.1706	0.2726	0.5875
0.94	0.0965	0.0978	0.1136	0.1788	0.3513	0.2269	0.2588	0.3209	0.5329	0.8996
0.92	0.1427	0.1440	0.1613	0.2738	0.5768	0.3646	0.4100	0.5417	0.7947	0.9726
0.90	0.1921	0.1954	0.2262	0.4048	0.7867	0.5269	0.6557	0.7408	0.9251	0.9914
0.88	0.2529	0.2616	0.3216	0.5549	0.9114	0.6597	0.7198	0.8669	0.9735	0.9961
0.86	0.3277	0.3519	0.4168	0.6798	0.9550	0.7750	0.8580	0.9473	0.9892	0.9992

Number of lags in VAR is zero; 10,000 replications.

Conclusiones

- A partir de investigaciones previas (Hansen 1995; Elliott y Jansson 2003; Perron y Rodríguez 2003), nosotros introducimos un arbitrario número de co-variables (x_t) en modelos de cambio estructural con punto de quiebre desconocido.
- La distribución asintótica derivada tiene potencia que es tangente a la máxima potencia (“*power envelope*”) en un punto de la hipótesis alternativa.
- El estadístico propuesto es una función de \bar{c} y R^2 (la correlación de las co-variables (x_t) y las cuasi-diferencias de y_t).
- Las simulaciones en muestras grandes (asintótica) muestran ganancias importantes en potencia cuando $R^2 > 0$.
- En muestras finitas, las simulaciones Monte Carlo han mostrado que las ganancias en potencia son sustanciales sin deterioración del nivel del test.
- También se ha derivado la distribución asintótica de un test que permite la existencia de correlación. Las simulaciones muestran que este test alcanza una potencia similar a la máxima potencia (“*power envelope*”) cuando $T \rightarrow \infty$.