Integración Fraccional y Outliers Aditivos

Gabriel Rodríguez

Banco Central de Reserva del Perú

Diciembre 2007

Contenido

- Introducción
- Procesos Integrados Fraccionales
- Detección de Outliers
- Experimento Monte Carlo
- Resultados
- Conclusiones

1 Introducción

- Procesos TS, DS y Fraccionales [$I(0)$, $I(1)$, $I(d)$].

- Efectos de Outliers sobre modelos ARMA (Cheng y Liu, 1993).

- Efectos de Outliers sobre tests de raíz unitaria (Franses y Haldrup, 1994; Vogelsang, 1999; Perron y Rodríguez, 2003).

- Métodos de estimación: MCO (Geweke, Porter-Hudak, GPH, 1983), MV (Sowell, 1992).

2 Modelos de Memoria Larga (Integrados Fraccionales)

- Características:
  - Correlograma muy persistente
  - Función de autocorrelación no sumable
  - Infinito espectro a la frecuencia cero

- Modelo fraccional integrado ruido blanco ($y_t = Y_t - \mu$)

$$(1 - L)^d y_t = \epsilon_t$$

El modelo es estacionario y tiene una representación promedios móviles infinita si $|d| < 0.5$.

- Modelo fraccional integrado autoregresivo promedios móviles ARFIMA(p,d,q):

$$(1 - L)^d[y_t - \phi_1 y_{t-1} - .... - \phi_p y_{t-p}] = \epsilon_t + \theta_1 \epsilon_{t-1} + ... + \theta_q \epsilon_{t-q}.$$

- Asuma $(1 - L)^d y_t = u_t$ donde u_t es un proceso linear estacionario con función de densidad espectral dada por $f_u(\lambda)$. $f_u(\lambda)$ es finita y continua en el intervalo $[-\pi, \pi]$. El método de GPH (1983) consiste en estimar el parámetro d de la siguiente regresión

$$\log[f_y(\omega_j)] = \log f_u(0) - d \log[4 \sin^2(\omega_j/2)] + \log f_u(\omega_j)/f_u(0).$$

- Utilizando ordenadas espectrales $\omega_1, \omega_2, ..., \omega_m$, el parámetro d puede ser estimado utilizando el periodograma de y_t, es decir $I_y(\omega_j)$:

$$\log I_y(\omega_j) = a - d \log[4 \sin^2(\omega_j/2) + v_j$$

donde $j = 1, 2, ..., n$ y $v_j = \log f_u(\omega_j)/f_u(0)$ y $v_j \sim i.i.d.\ (0, \pi^2/6)$.

- Asumiendo que $n = g(T)$, donde $g(T)$ es tal que $lim_{T\to\infty} g(T) = \infty$, $lim_{T\to\infty} g(T)/T = 0$ y $lim_{T\to\infty}\ log(T)^2/g(T) = 0$, tenemos que el estimador MCO de d

$$\frac{\widehat{d} - d}{var(\widehat{d})^{1/2}} \Rightarrow N(0, 1).$$

- En las estimaciones: $n = g(T) = T^{1/2}$

3 Detección de Outliers

- Detección de outliers en modelos ARMA (Cheng y Liu, 1993).

- Detección de outliers y raíz unitaria (Franses y Haldrup, 1994; Vogelsang, 1999; Perron y Rodríguez, 2003).

- Procedimiento de Vogelsang (1999): estadístico τ.

- Procedimiento de Perron y Rodríguez (2003) I: estadístico τ_c.

- Procedimiento de Perron y Rodríguez (2003) II: estadístico τ_d.

4 Experimento Monte Carlo

- Diseño del experimento similar a Vogelsang (1999) y Perron y Rodríguez (2003).

- Proceso generador de datos

$$
\begin{aligned}
y_t &= n_t + \sum_{i=1}^{m} \delta_i D(T_{ao,i})_t + u_t, \\
(1-L)^d u_t &= v_t,
\end{aligned}
$$

  - $v_t \sim i.i.d.\ N(0,1)$,

  - $D(T_{ao,i})_t = 1$ si $t = T_{ao,i}$ y 0 de otro modo,

- δ_i es el tamaño del outlier,
- n_t son los componentes determinísticos. Consideramos sólo un intercepto.

- Tres casos para δ_i:
  - $\delta_i = 0$ para $i = 1, 2, 3, 4$
  - $\delta_1 = 10$, $\delta_i = 5$ para $i = 2, 3, 4$
  - $\delta_i = 10$ para $i = 1, 2, 3, 4$

- Evaluación de los resultados: Sesgo del estimador $\widehat{d}$, MSE del estimador $\widehat{d}$, tamaño (nivel) del estadístico $t_{\widehat{d}}$.

5 Resultados

- Cuando hay outliers (sin corregir):
  - La media está distorsionada por la presencia de outliers. Más grande el tamaño del outlier, más grande la distorsión.
  - El sesgo y el MSE aumentan a medida que el tamaño de los outliers aumenta.
  - El estadístico $t_{\widehat{d}}$ presenta severas distorsiones de nivel.

- Cuando se corrige por la presencia de outliers:
  - La media de $\widehat{d}$ es próxima al verdadero valor de d. Excepto para $d > 0.5$.
  - El sesgo y el MSE de $\widehat{d}$ son reducidos y estables salvo para $d > 0.5$.
  - El nivel del estadístico $t_{\widehat{d}}$ es cercano al nivel nominal excepto para valores de d próximos a la unidad.

Table 1. Media del parámetro $\widehat{d}$, $T = 100$

d	$\delta_1 = 0.0, \delta_2 = 0.0,$ $\delta_3 = 0.0, \delta_4 = 0.0$	$\delta_1 = 10.0, \delta_2 = 5.0,$ $\delta_3 = 5.0, \delta_4 = 5.0$	$\delta_1 = 10.0, \delta_2 = 10.0,$ $\delta_3 = 10.0, \delta_4 = 10.0$
$d = 0.00$	0.008	-0.017	-0.026
$d = 0.24$	0.249	0.123	0.080
$d = 0.48$	0.496	0.29	0.221
$d = 0.72$	0.752	0.483	0.398
$d = 0.96$	0.979	0.712	0.629

Table 2. Sesgo del parámetro $\widehat{d}$, $T = 100$

d	$\delta_1 = 0.0, \delta_2 = 0.0,$ $\delta_3 = 0.0, \delta_4 = 0.0$	$\delta_1 = 10.0, \delta_2 = 5.0,$ $\delta_3 = 5.0, \delta_4 = 5.0$	$\delta_1 = 10.0, \delta_2 = 10.0,$ $\delta_3 = 10.0, \delta_4 = 10.0$
$d = 0.00$	0.008	-0.017	-0.026
$d = 0.24$	0.009	-0.117	-0.160
$d = 0.48$	0.016	-0.19	-0.259
$d = 0.72$	0.032	-0.237	-0.322
$d = 0.96$	0.019	-0.248	-0.331

Table 3. MSE del parámetro $\widehat{d}$, $T = 100$

d	$\delta_1 = 0.0, \delta_2 = 0.0,$ $\delta_3 = 0.0, \delta_4 = 0.0$	$\delta_1 = 10.0, \delta_2 = 5.0,$ $\delta_3 = 5.0, \delta_4 = 5.0$	$\delta_1 = 10.0, \delta_2 = 10.0,$ $\delta_3 = 10.0, \delta_4 = 10.0$
$d = 0.00$	0.014	0.012	0.011
$d = 0.24$	0.014	0.026	0.037
$d = 0.48$	0.015	0.049	0.079
$d = 0.72$	0.016	0.071	0.119
$d = 0.96$	0.012	0.087	0.147

Table 4. Nivel de $t_{\widehat{d}}$,$T = 100$

d	$\delta_1 = 0.0, \delta_2 = 0.0,$ $\delta_3 = 0.0, \delta_4 = 0.0$	$\delta_1 = 10.0, \delta_2 = 5.0,$ $\delta_3 = 5.0, \delta_4 = 5.0$	$\delta_1 = 10.0, \delta_2 = 10.0,$ $\delta_3 = 10.0, \delta_4 = 10.0$
$d = 0.00$	0.054	0.028	0.019
$d = 0.24$	0.056	0.131	0.174
$d = 0.48$	0.060	0.306	0.498
$d = 0.72$	0.075	0.488	0.726
$d = 0.96$	0.054	0.600	0.734

Table 5. Media del parámetro $\widehat{d}$ utilizando τ_d, $T = 100$

d	$\delta_1 = 10.0, \delta_2 = 5.0,$ $\delta_3 = 5.0, \delta_4 = 5.0$	$\delta_1 = 10.0, \delta_2 = 10.0,$ $\delta_3 = 10.0, \delta_4 = 10.0$
$d = 0.24$	0.238	0.239
$d = 0.48$	0.465	0.467
$d = 0.72$	0.651	0.651
$d = 0.96$	0.721	0.721

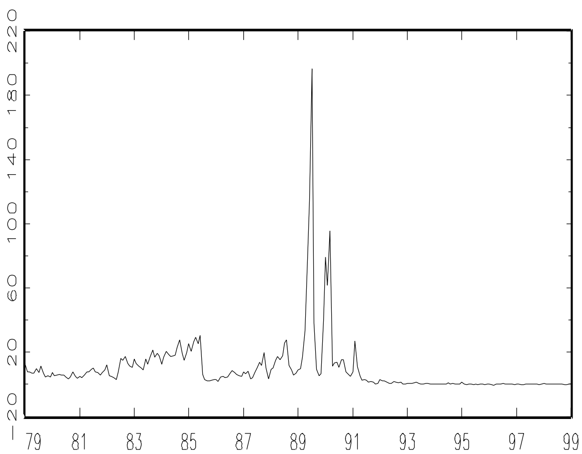
Table 6. MSE del parámetro $\widehat{d}$ utilizando τ_d, $T = 100$

d	$\delta_1 = 0.0, \delta_2 = 0.0,$ $\delta_3 = 0.0, \delta_4 = 0.0$	$\delta_1 = 10.0, \delta_2 = 5.0,$ $\delta_3 = 5.0, \delta_4 = 5.0$	$\delta_1 = 10.0, \delta_2 = 10.0,$ $\delta_3 = 10.0, \delta_4 = 10.0$
$d = 0.24$	0.017	0.014	0.014
$d = 0.48$	0.016	0.014	0.014
$d = 0.72$	0.015	0.019	0.019
$d = 0.96$	0.022	0.076	0.076

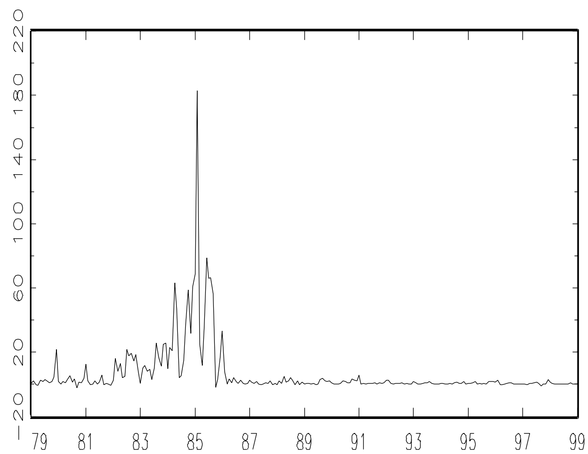
Table 7. Nivel de $t_{\widehat{d}}$ utilizando τ_d, $T = 100$

d	$\delta_1 = 0.0, \delta_2 = 0.0,$ $\delta_3 = 0.0, \delta_4 = 0.0$	$\delta_1 = 10.0, \delta_2 = 5.0,$ $\delta_3 = 5.0, \delta_4 = 5.0$	$\delta_1 = 10.0, \delta_2 = 10.0,$ $\delta_3 = 10.0, \delta_4 = 10.0$
$d = 0.00$	0.055	0.062	0.057
$d = 0.24$	0.056	0.059	0.058
$d = 0.48$	0.060	0.061	0.059
$d = 0.72$	0.074	0.105	0.106
$d = 0.96$	0.058	0.585	0.585

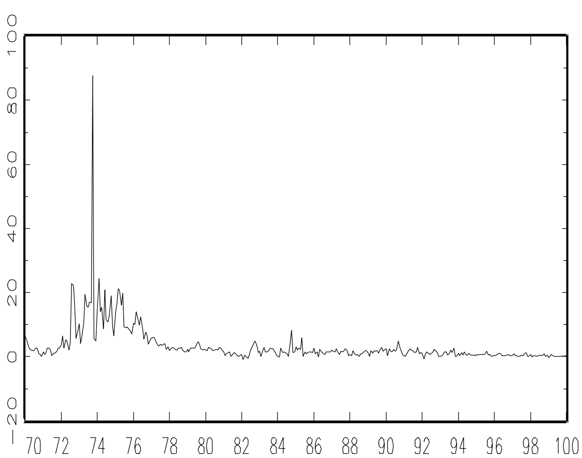
Argentina
(1979:01-1999:03)



Bolivia
(1979:01-2000:05)



Chile
(1970:01-2000:05)



Peru
(1979:01-2000:05)

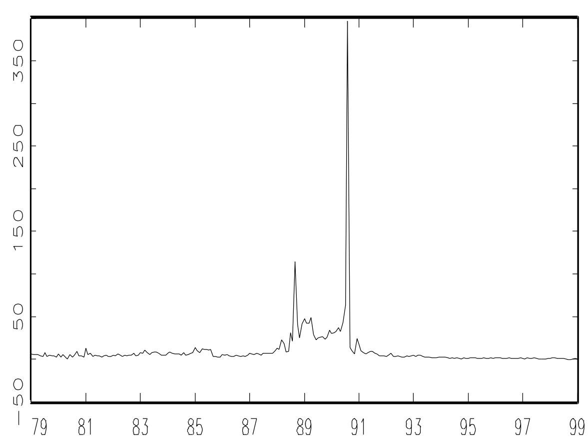


Table 8. Inflación en América Latina (Series Mensuales)

País	Estimación sin detectar outliers		Estimación detectando outliers	
	$\widehat{d}$	$t_{\widehat{d}}$	$\widehat{d}$	$t_{\widehat{d}}$
Argentina	0.614	11.459	0.736	11.052
Bolivia	0.468	9.383	0.301	5.017
Chile	0.215	5.379	0.344	6.577
Perú	0.201	8.246	0.477	12.063

Table 9. Inflación en América Latina (Series Trimestrales)

País	Estimación sin detectar outliers		Estimación detectando outliers	
	$\widehat{d}$	$t_{\widehat{d}}$	$\widehat{d}$	$t_{\widehat{d}}$
Argentina	0.522	6.436	0.734	7.089
Bolivia	0.572	7.026	0.579	6.092
Chile	0.637	11.079	0.670	4.805
Perú	0.284	7.109	0.652	4.346
Colombia	0.737	6.100	0.737	6.100
Ecuador	0.827	8.385	0.668	7.353
Paraguay	0.390	2.599	0.390	2.599
Uruguay	0.598	5.639	0.363	2.718
Venezuela	0.575	5.902	0.415	3.502

6 Conclusiones

- La presencia de outliers produce severas distorsiones en la estimación del parámetro de integración fraccional $(\widehat{d})$ y su estadístico $t_{\widehat{d}}$. En particular la media, el sesgo, el MSE de $\widehat{d}$ son afectados. El estadístico $t_{\widehat{d}}$ presenta severas distorsiones de nivel.

- La utilidad del estadístico τ_d para detectar outliers es confirmada.

- Utilizando el estadístico τ_d, la media, el sesgo y el MSE de $\widehat{d}$ son reducidos y estables excepto para valores de d cada vez más próximos a la unidad.

- El estadístico $t_{\widehat{d}}$ muestra niveles exactos muy próximos al nivel nominal excepto para valores de d cercanos al circuilo unitario.