

Puntos de Cambio y Outliers en Series de Tiempo

Gabriel Rodriguez-Yam

Universidad Autónoma Chapingo

R.A. Davis

Columbia University

T. Lee

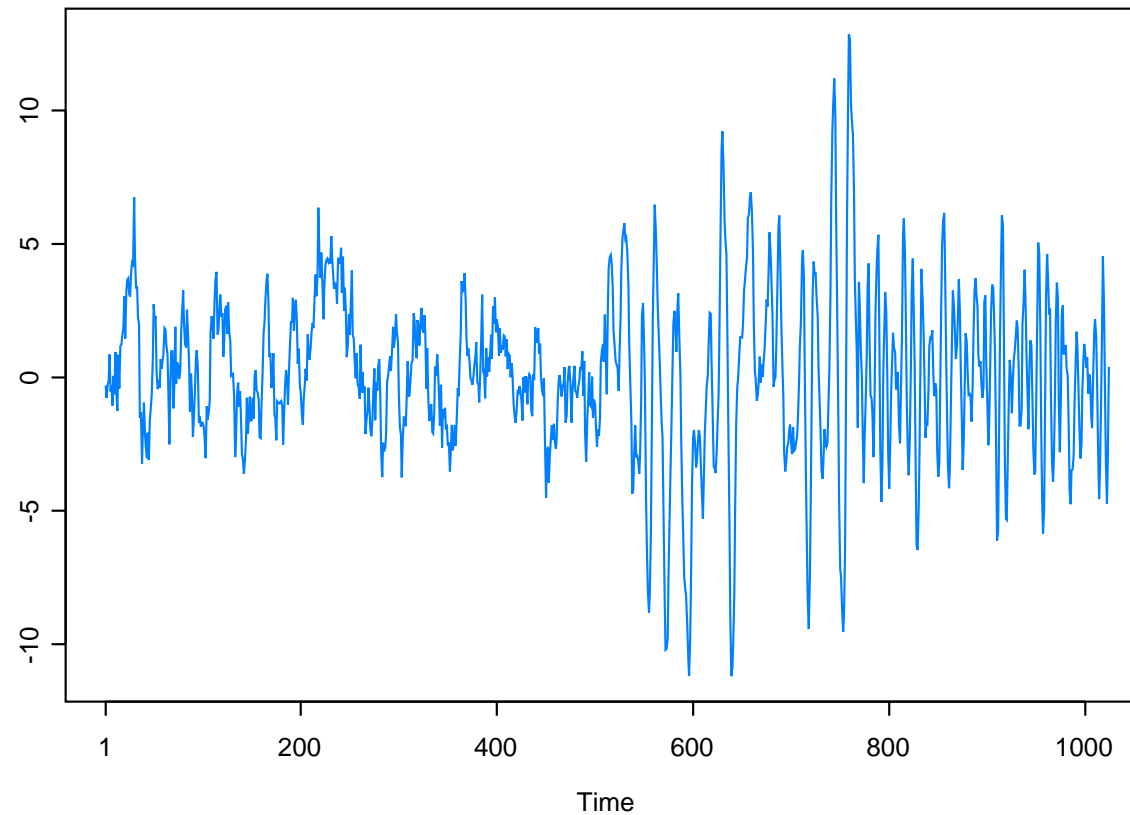
University of California, Davis

Feb 2015

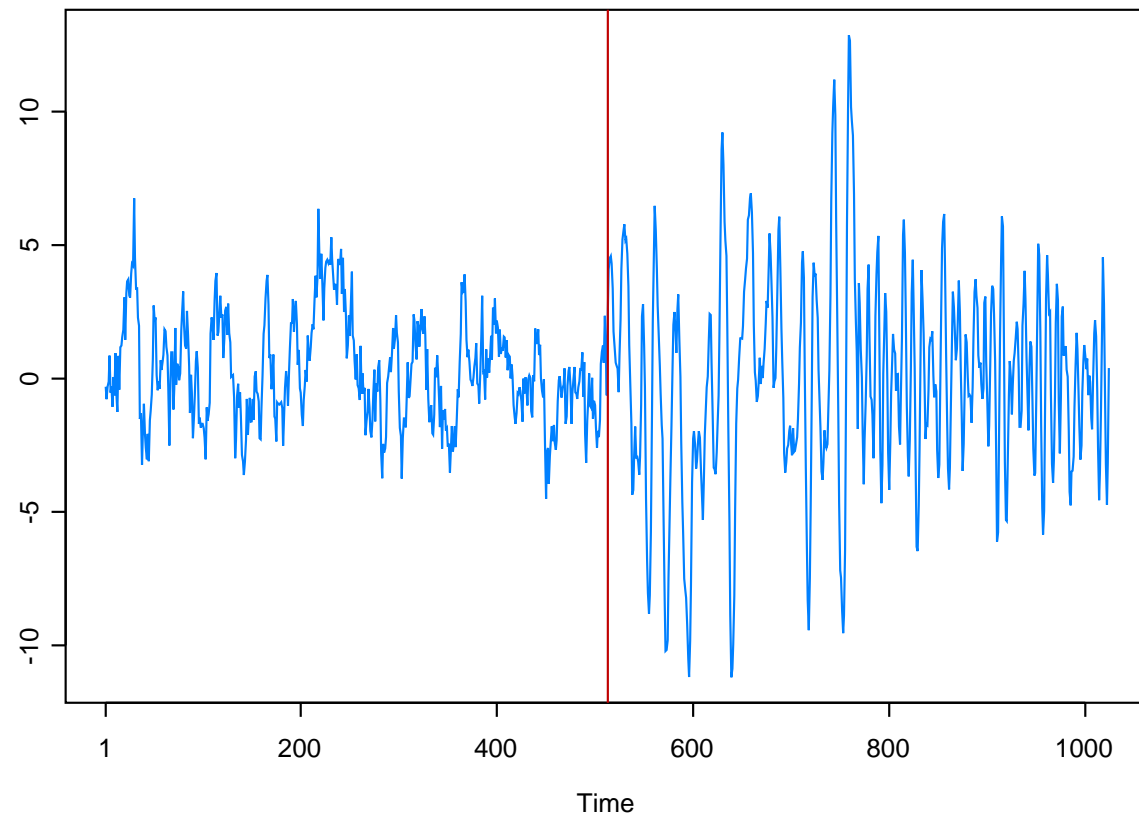
Bosquejo

0. Introducción
1. Modelo AR segmentado
2. Proceso Estacionario Local
3. Literatura de puntos de cambio
4. MDL
5. Algoritmos genéticos
6. Ejemplos
7. Outliers

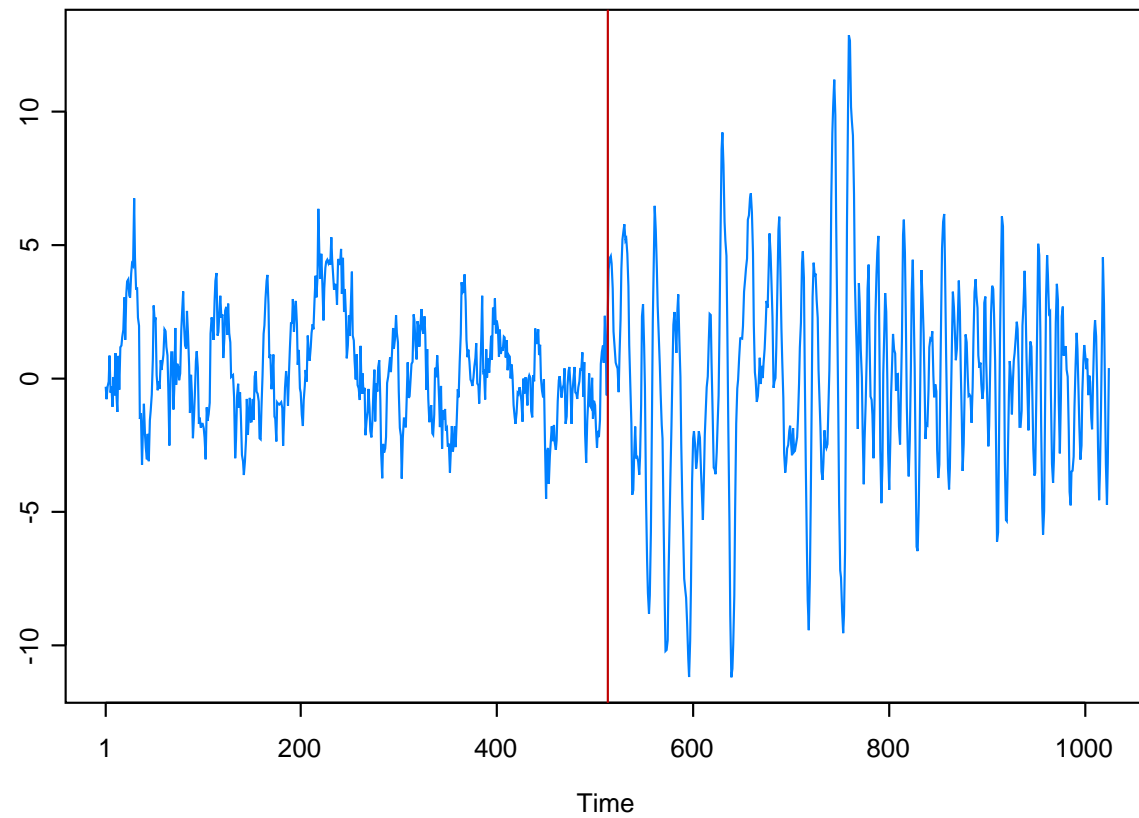
¿Cuántas “piezas” se ven en esta serie?



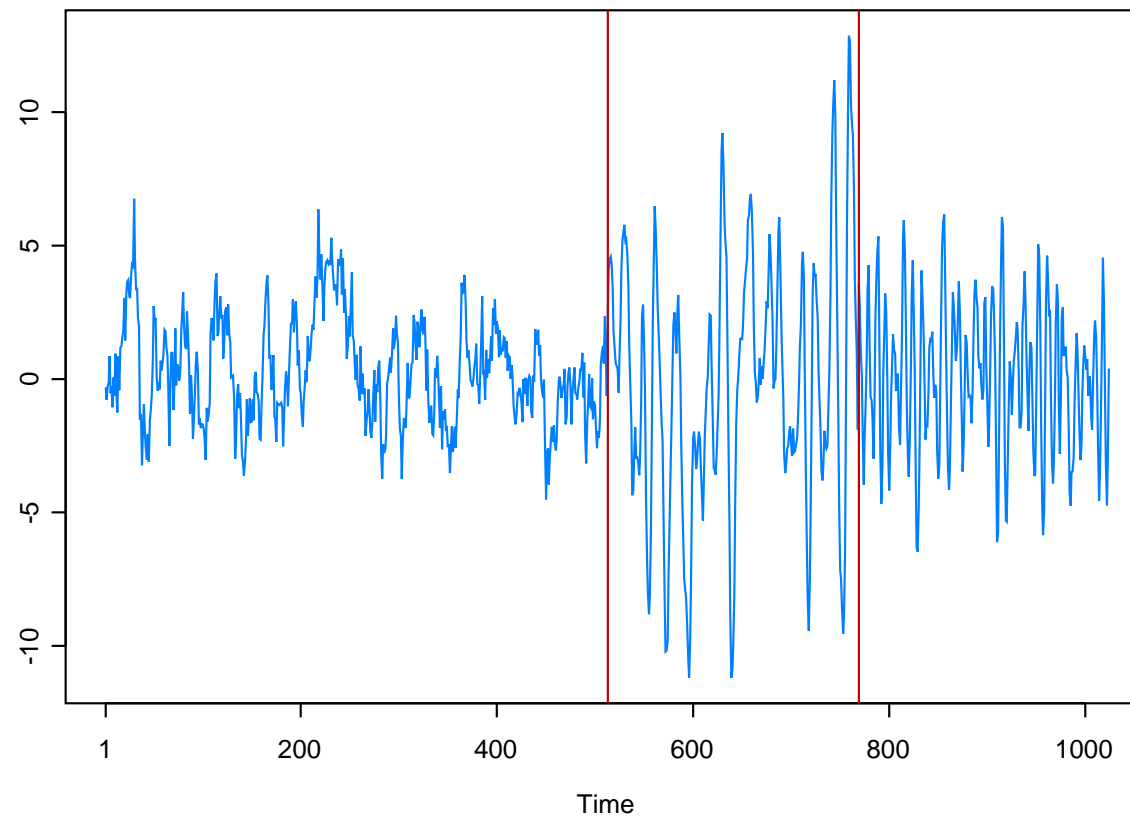
Hay un “punto de cambio” (pc) en $t=513$



¿Hay otro(s)?



Si. En $t=769$

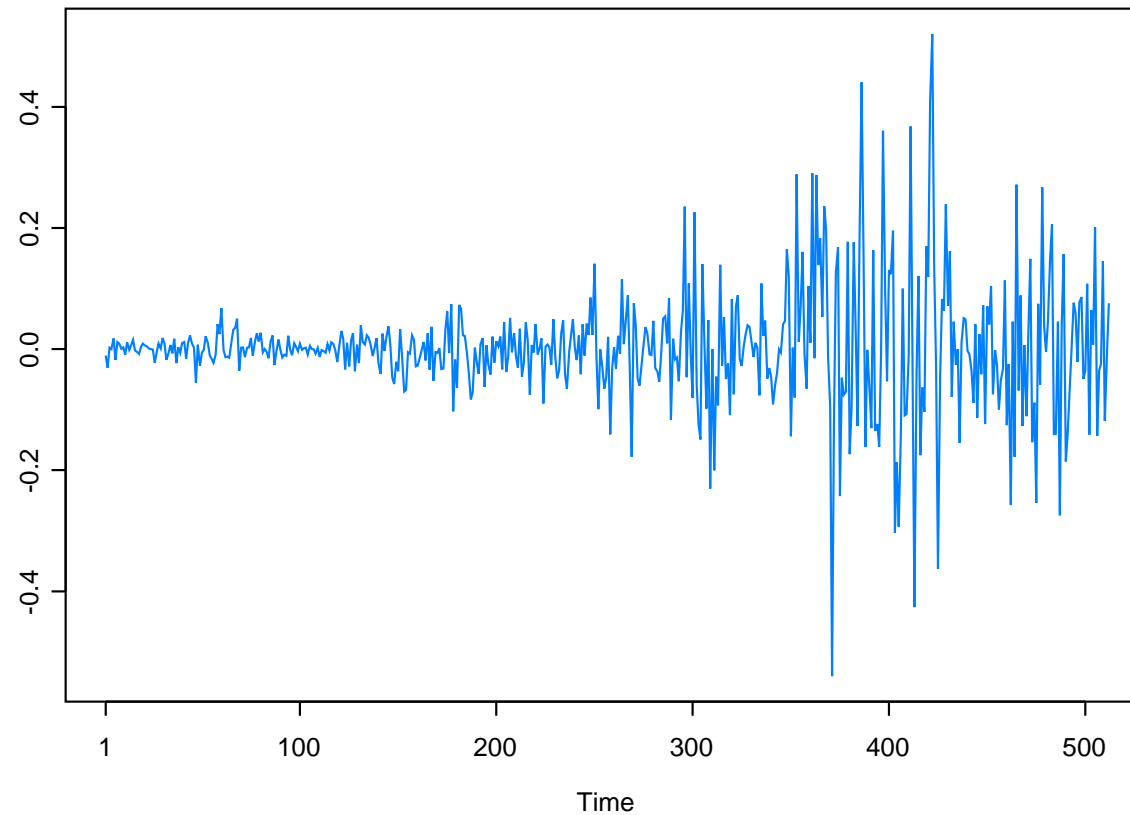


Esta serie es una realización del *proceso*

$$Y_t = \begin{cases} 0.9Y_{t-1} + \varepsilon_t, & \text{si } 1 \leq t \leq 512, \\ 1.69Y_{t-1} - 0.81Y_{t-2} + \varepsilon_t, & \text{si } 513 \leq t \leq 768, \\ 1.32Y_{t-1} - 0.81Y_{t-2} + \varepsilon_t, & \text{si } 769 \leq t \leq 1024, \end{cases}$$

$\varepsilon_t \sim \text{iid } N(0, 1)$.

¿Cuántas “piezas” tiene esta otra serie?



Sólo una!

y_1, \dots, y_{500} , es una realización del modelo:

$$Y_t | \alpha_t \stackrel{iid}{\sim} N(0, e^{\alpha_t}), \quad t = 1, \dots, 500$$

$$\alpha_t = -0.0645 + 0.9889\alpha_{t-1} + 0.3057\varepsilon_t$$

$$\{\varepsilon_t\} \stackrel{iid}{\sim} N(0, 1)$$

Modelo AR segmentado

Se desea modelar un proceso no-estacionario $\{Y_t\}$ usando segmentos AR. Es decir,

$$\Rightarrow Y_t = X_{t,j}, \quad \tau_{j-1} \leq t < \tau_j, \quad j = 1, \dots, m+1$$

$\Rightarrow \{X_{t,j}\}$ es el proceso $\text{AR}(p_j)$:

$$X_{t,j} = \gamma_j + \phi_{j1}X_{t-1,j} + \dots + \phi_{j,p_j}X_{t-p_j,j} + \sigma_j\varepsilon_t,$$

☞ $\psi_j := (\gamma_j, \phi_{j1}, \dots, \phi_{j,p_j}, \sigma_j^2)$: parámetros del j -ésimo proceso AR.

☞ $\{\varepsilon_t\}$: Errores iid con media 0 y varianza 1.

El número m de puntos de cambio (pc), los puntos de cambio τ_1, \dots, τ_m , y los órdenes p_1, \dots, p_{m+1} no se conocen.

Proceso Estacionario Local

➡ *Proceso estacionario local* (Dahlhaus 1997): En el tiempo t y frecuencia ω tiene espectro

★ $|A(t/n, \omega)|^2,$

★ $A(u, \omega)$, $u \in [0, 1]$, $\omega \in [-1/2, 1/2]$ es *continua* en u .

➡ *Proceso estacionario en segmentos* (Adak 1998)

$$\tilde{Y}_{t,n} = \sum_{j=1}^{m+1} X_{t,j} I_{[\tau_{j-1}/n, \tau_j/n)}(t/n),$$

donde $\{X_{t,j}\}, j = 1, \dots, m+1$ son procesos estacionarios.

➡ Bajo ciertas condiciones, Ombao, Raz, Von Sachs, y Malow (2001) discuten que un LSP se puede aproximar por procesos estacionarios en piezas.

Literatura de puntos de cambio

- ☞ El caso de un solo pc para observaciones independientes ha sido muy estudiado en la Literatura.
- ☞ pc múltiples, un problema más difícil, ha sido también analizado (Shaban (1980); Zacks (1983); Krishnaiah and Miao (1988); Bhattacharya (1994); and Csörgő and Horváth (1997).)
- ☞ En series de tiempo, un sólo pc ha sido también muy estudiado. En Picard (1985); Davis, Huang and Yao (1995) and Kitagawa, Takanami and Matsumoko (2001), los segmentos son procesos AR.
- ☞ El caso de pc múltiples en series de tiempo ha sido también estudiado. En Kitagawa and Akaike (1978), Fearnhead (2005), Davis, Lee and Rodriguez-Yam (2006), la serie observada es descompuesta en procesos AR.

- ➡ Procesos estacionarios en segmentos más general (AR es un caso particular) es considerado en Ombao, Raz, Von Sachs and Malow (2001).
- ➡ McCulloch and Tsay (1993); Djurić (1994), Lavielle (1998) and Punskey et al. (2002) siguen un enfoque Bayesiano.
- ➡ Csörgő and Horváth (1997) dedican un capítulo para pc en series de tiempo.

Notación: Dada la serie observada $\mathbf{y} := (y_1, \dots, y_n)$, sea

\mathcal{F} : familia de modelos AR segmentados (lámina 37) que se pueden definir para modelar el proceso $\{Y_t\}$.

Algunas restricciones (para \mathcal{F}):

P_0 : máximo orden AR permitido.

m_p : Mínimo número de observaciones en un segmento con orden p .

Objetivo: Obtener el “mejor modelo” $\mathcal{M} \in \mathcal{F}$.

En otras palabras, obtener la mejor combinación de valores

☞ $m + 1$: número de segmentos.

☞ τ_1, \dots, τ_m : localización de los pc

☞ p_1, \dots, p_{m+1} : órdenes AR.

¿Cómo encontrar este modelo en \mathcal{F} ?

- ➡ CL (objeto \mathbf{O}): Longitud codificada de un objeto \mathbf{O} (cantidad de “memoria” que se requiere para “almacenarlo”).
- ➡ De hecho, preferiríamos la memoria mínima de almacenamiento. Este es el principio de Rissanen (1989). Se le denota MDL
- ➡ CL de \mathbf{y} (relativo a $\mathcal{M} \in \mathcal{F}$) puede obtenerse como

$$CL_{\mathcal{M}}(\mathbf{y}) = CL_{\mathcal{M}}(\hat{\mathcal{M}}) + CL_{\mathcal{M}}(\hat{\mathbf{e}}|\hat{\mathcal{M}}),$$

$\hat{\mathbf{e}} : \mathbf{y} - \hat{\mathbf{y}}$

$\hat{\mathbf{y}}$: vector de valores ajustados.

Para encontrar CL de $\hat{\mathcal{M}}$

$$\begin{aligned} CL_{\mathcal{M}}(\hat{\mathcal{M}}) &= CL_{\mathcal{M}}(m) + CL_{\mathcal{M}}(\tau_1, \dots, \tau_m) + CL_{\mathcal{M}}(p_1, \dots, p_{m+1}) \\ &\quad + CL_{\mathcal{M}}(\hat{\psi}_1) + \dots + CL_{\mathcal{M}}(\hat{\psi}_{m+1}) \\ &= CL_{\mathcal{M}}(m) + CL_{\mathcal{M}}(n_1, \dots, n_{m+1}) + CL_{\mathcal{M}}(p_1, \dots, p_{m+1}) \\ &\quad + CL_{\mathcal{M}}(\hat{\psi}_1) + \dots + CL_{\mathcal{M}}(\hat{\psi}_{m+1}) \end{aligned}$$

$$n_j := \tau_j - \tau_{j-1}$$

De acuerdo a Risannen (1989):

- ☞ CL de un entero k : $\log(k)$
- ☞ CL de un entero k acotado por M : $\log(M)$
- ☞ CL parámetro estimado con n observaciones: $\frac{1}{2} \log(n)$

Se obtiene que

$$\begin{aligned} CL_{\mathcal{M}}(\hat{\mathcal{M}}) &= \log(m) + \sum_{j=0}^{m+1} \log(p_j) + (m+1) \log(n) \\ &+ \frac{1}{2} \sum_{j=1}^{m+1} (p_j + 2) \log n_j \end{aligned}$$

CL de los residuales ($CL_{\mathcal{M}}(\hat{\mathbf{e}}|\hat{\mathcal{M}})$):

$-\log(\text{Like})$ (Rissanen, 1989)

Así,

$$\begin{aligned} CL_{\mathcal{M}}(\mathbf{y}) &= \text{MDL}(m, \tau_1, \dots, \tau_m, p_1, \dots, p_{m+1}) \\ &\approx \log m + (m+1) \log n + \sum_{j=1}^{m+1} \log p_j \\ &\quad + \sum_{j=1}^{m+1} \frac{p_j+2}{2} \log n_j + \sum_{j=1}^{m+1} \frac{n_j}{2} \log(2\pi\hat{\sigma}_j^2) \end{aligned}$$

Donde $\hat{\sigma}_j^2$ es el estimador de Y-W de σ_j^2 del j-ésimo segmento AR se puede aproximar por $n_j \log(2\pi\hat{\sigma}_j^2)$

Así, se requiere encontrar $\tilde{\mathcal{M}}$, tal que

$$\tilde{\mathcal{M}} := \arg \min_{\mathcal{M} \in \mathcal{F}} \text{MDL}(m, \tau_1, \dots, \tau_m, p_1, \dots, p_{m+1}),$$

Para minimizar esta función, se emplearán *algoritmos genéticos* (*GA*).

- ☞ Estos algoritmos, propuestos por Holland (1975), son métodos de búsqueda aleatorios.
- ☞ Para optimizar una función los algoritmos genéticos “imitan” la teoría de la selección natural y han sido usados en estadística, entre otros, por Chatterjee et. al. (1996), Gaetan (2000), Lee (2002), Davis, et al (2006, 2008)

Algoritmos genéticos (Holland, 1975)

Componentes (algoritmo genético básico):

- ➡ Cromosomas: Vector (codificación de \mathcal{M})
- ➡ Condición física de los elementos de la población (función objetivo).
- ➡ Operadores cruzamiento (crossover) y mutación (mutation):
Formas de reproducción.
- ➡ Regla para detener las iteraciones.

Algoritmo Genético Canónico

1. Generar (aleatoriamente) población inicial de (popsize) individuos.
2. Crear nueva generación de descendientes (usando operadores).
e.g. j -ésimo descendiente, $j=2, \dots, \text{popsize}$:
proviene de una mutación (con probabilidad π_m) y proviene de un cruzamiento (con probabilidad $1 - \pi_m$).
3. Obtener el “lider” de la generación.
4. Si no se cumple criterio para detener iteraciones, ir a 2, de otro modo, ir a 5.
5. Lider es una (aproximación de la) solución.

Implementación

Cromosomas: $\mathbf{g} := (g_1, \dots, g_n)$,

$$g_t = \begin{cases} -1, & \text{si no hay un pc en } t, \\ p_j, & \text{si } t = \tau_{j-1} \text{ y el AR order del segmento } j \text{ es } p_j. \end{cases}$$

Ejemplo. Si $n = 10$, entonces

$$\mathbf{g} = (2, -1, -1, -1, 1, -1, -1, 1, -1, -1)$$

corresponde al modelo \mathcal{M}

☞ que tiene $m = 2$ pc

☞ $\tau_1 = 5$ y $\tau_2 = 8$.

☞ Primer segmento es un proceso AR(2)

☞ Segundo segmento es un proceso AR(1).

Distribuciones “útiles”

- ➡ Distribución selectora π^{AR} de ordenes: Variable aleatoria discreta D con probabilidades $P[D = j] = \pi_j^{AR}$, $j = 0, 1, \dots, P_0$, proporcionada por el usuario.
- ➡ Distribución selectora π en una población: Si $\mathbf{g}_1, \dots, \mathbf{g}_{popsize}$ constituye una generación (que se va reproducir), sea $\pi = (\pi_1, \dots, \pi_{popsize})$ tal que π_i es decreciente en MDL .

Población inicial: Cada cromosoma $\mathbf{g} := (g_1, \dots, g_{popsize})$ es generada (aleatoriamente) como sigue:

1. g_1 : generada de la distribución de D
2. A $g_2, g_3, \dots, g_{m_{g_1}}$ se les asigna -1
4. Hágase $j = m_{g_1}$
5. $j = j + 1$. Si $j > n$, terminar. De otro modo continuar en 6.
6. g_j es definida ya sea un pc (con probabilidad π_{break}), o -1 .
7. Si no es pc, hacer $g_j = -1$ y regresar a 5.
8. Si es pc, g_j se selecciona (independiente de selección anterior) de D .
9. A $g_{j+1}, \dots, g_{j+m_{g_j}-1}$ se les asigna -1 .
10. Haga $j = j + m_{g_j} - 1$ e ir a 5.

cruzamiento: Se seleccionan aleatoriamente un padre y una madre de acuerdo a la distribución π y los genes del descendiente en turno se generan como sigue:

- ➡ g_i se selecciona con igual probabilidad del i -ésimo gene de papá y mamá.
- ➡ En caso que g_i no sea un pc, el siguiente se selecciona de la misma manera.
- ➡ Si g_i es un pc, $m_{g_i} - 1$ siguientes genes se hacen -1, y el proceso se repite con el gene $i + m_{g_i}$.

Mutación:

Se selecciona aleatoriamente un padre de la distribución π .

El i -ésimo gene del descendiente en turno se forman de la manera que sigue:

- ➡ Con probabilidad π_p se asigna el gene del padre.
- ➡ Con probabilidad π_{nb} se le asigna -1, y
- ➡ con probabilidad $1 - \pi_p - \pi_{nb}$ se inicia un segmento nuevo con orden seleccionado aleatoriamente de la distribución de D .

Criterio para convergencia El proceso se detiene cuando el mejor individuo no ha cambiado en las últimas 10 iteraciones, o hasta que se han formado 100 generaciones. Lo que ocurra primero.

Estrategia Elitista. El mejor de una generación se incluye en generación descendiente.

Variaciones del algoritmo canónico

Modelo isla: Una estrategia para evitar “mínimos locales” es implementar el GA canónico “simultáneamente” en NI “islas”.

Migración. Después de cierto número de generaciones (M_i), en forma “circular” se permite que los mejores MN ($=5$) individuos de una isla reemplacen a los peores MN individuos de la isla vecina.

Ejemplo

Proceso estacionario en piezas con estructura “Dyadic” (Ombao, et. al. 2001)

$$Y_t = \begin{cases} 0.9Y_{t-1} + \varepsilon_t, & \text{if } 1 \leq t \leq 512, \\ 1.69Y_{t-1} - 0.81Y_{t-2} + \varepsilon_t, & \text{if } 513 \leq t \leq 768, \\ 1.32Y_{t-1} - 0.81Y_{t-2} + \varepsilon_t, & \text{if } 769 \leq t \leq 1024, \end{cases}$$

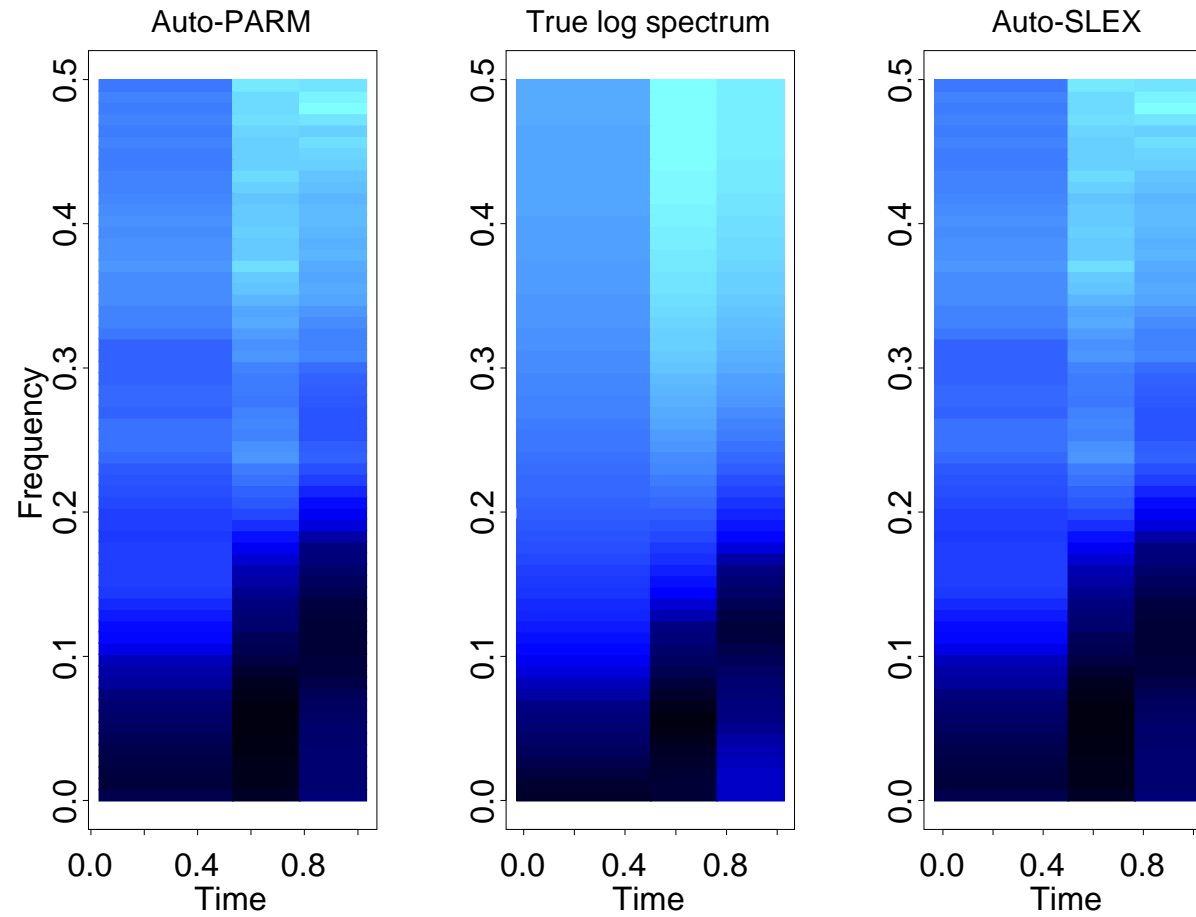
$\varepsilon_t \sim \text{iid } N(0, 1)$.

- ➡ Una realización “típica” se mostró en la Lámina 2.
- ➡ Si $\omega \in [0, 0.5)$, el espectro del j -ésimo segmento es

$$f_j(\omega) = \sigma_j^2 |1 - \phi_{j1} \exp\{-i2\pi\omega\} - \dots - \phi_{jp_j} \exp\{-i2\pi p_j \omega\}|^2,$$

- ➡ Si $t \in [\tau_{j-1}, \tau_j)$, el *espectro* de este proceso Y_t es

$$f(t/n, \omega) = f_j(\omega)$$

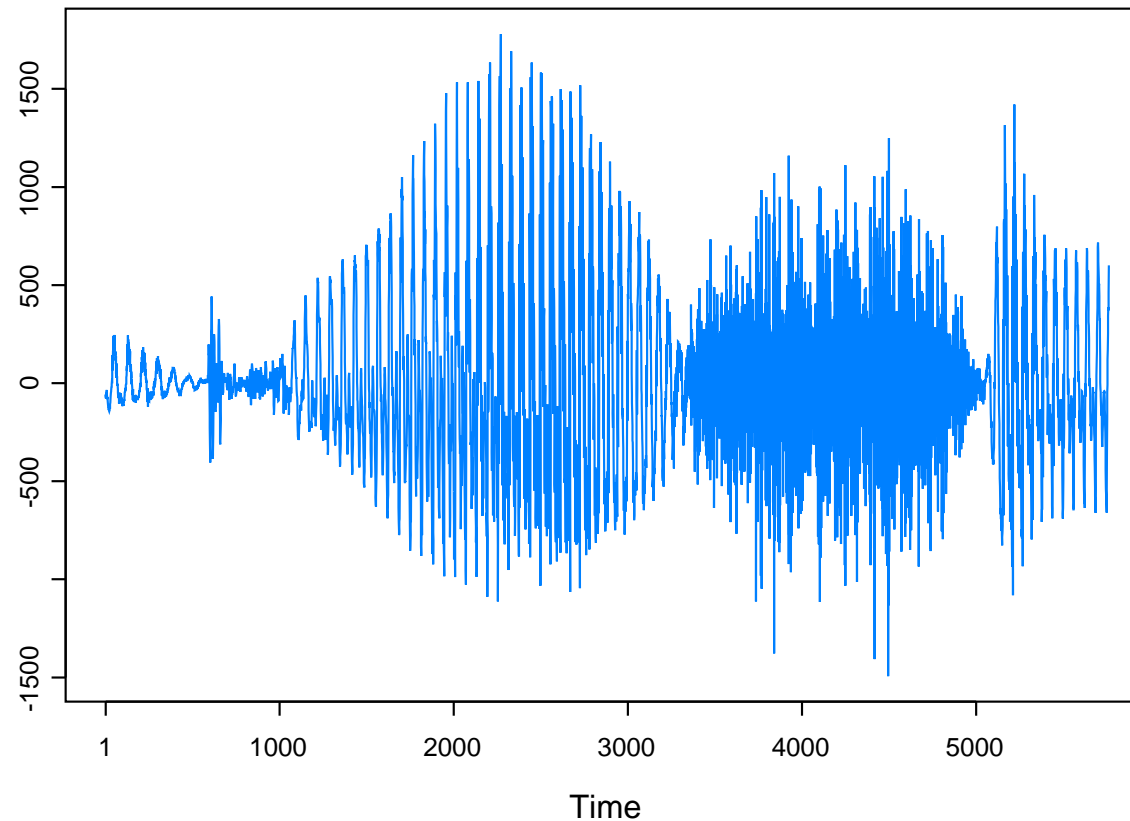


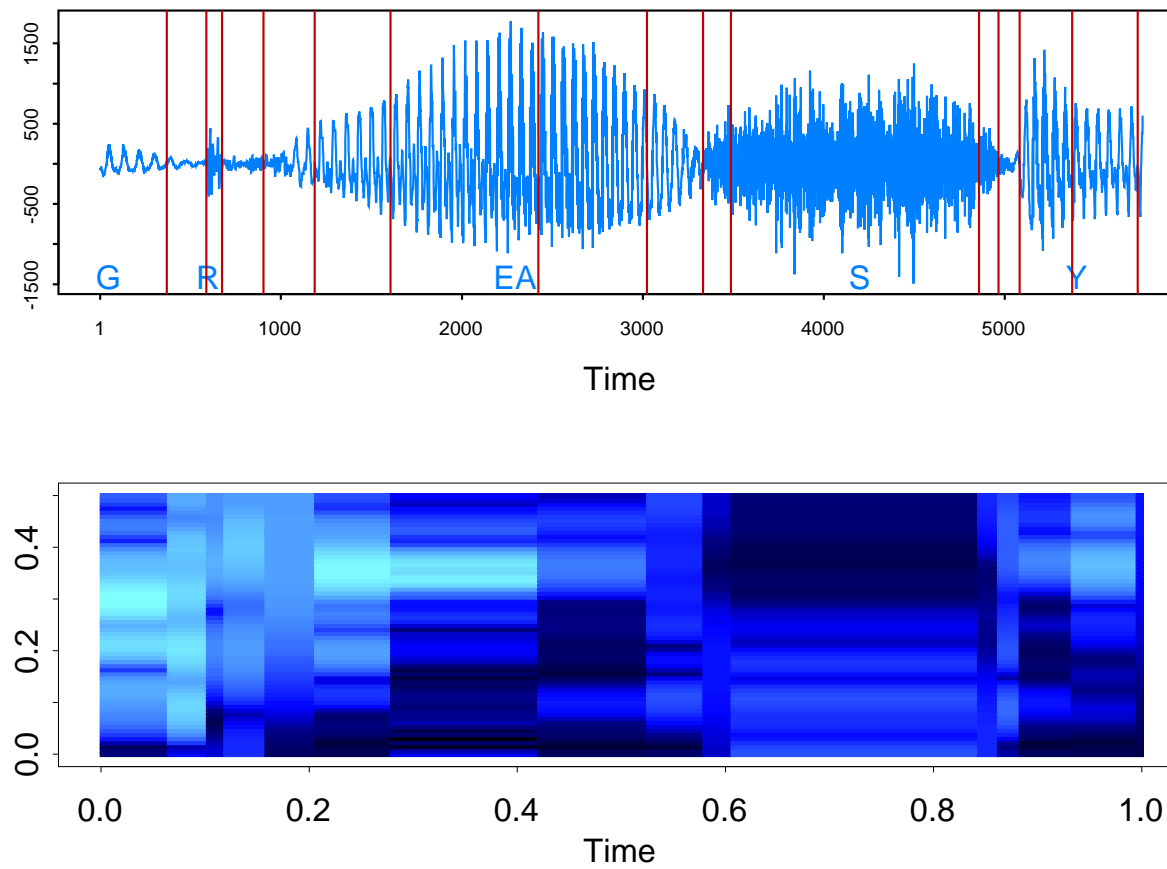
Sombreado oscuro representa frecuencias altas.

Resultados de Auto-SLEX fueron tomados de Tabla 2 de Ombao et al. (2001).

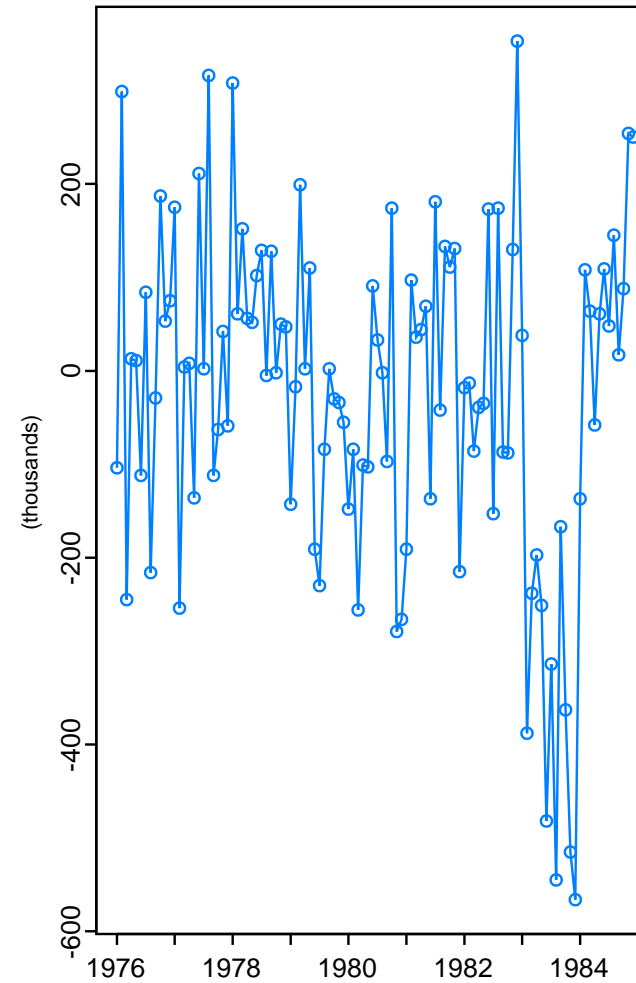
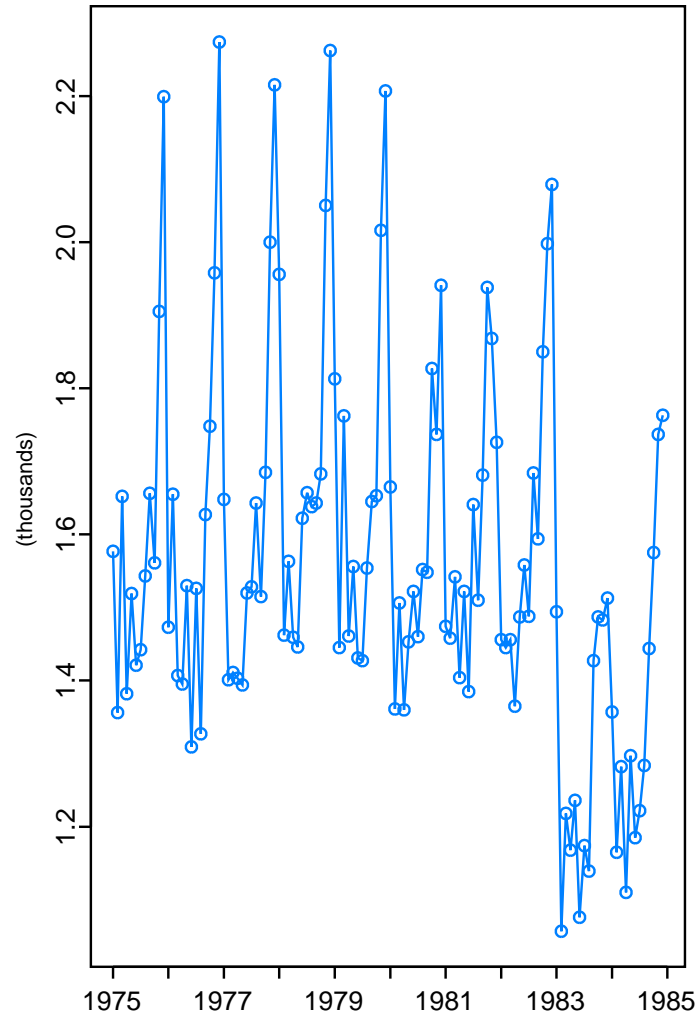
Number of segments		Auto-SLEX break points		Auto-PARM break points	
	(%)		(%)	mean	std
2	0	1/2	0		
3	60.0	2/4, 3/4	100.0	0.500	0.002
				0.742	0.007
4	34.0	1/4, 2/4, 3/4	0		
5	5.0	2/8, 4/8, 5/8, 6/8, 7/8	0		
≥ 6	1.0		0		

Ejemplo: Señal grabada de la palabra “GREASY”. $n=5762$





Ejemplo: Muertes y lesiones serias mensuales en UK, $n=120$



Auto-PARM encuentra un punto de cambio en $t = 84$

Outliers

Se considerará un proceso (no-estacionario) $\{Y_t\}$ con $m + 1$ segmentos AR.

- $\Rightarrow \tau_j, j = 1, \dots, m$: j -ésimo punto de cambio entre el segmento j y el $j + 1$
- \Rightarrow Los parámetros θ_j del j -ésimo segmento satisfacen $\theta_j \neq \theta_{j+1}$, $j = 1, \dots, m$.
- \Rightarrow Cada segmento está (posiblemente) contaminado por outliers.

Se consideran dos tipos de outliers:

- ⇒ Outlier aditivo: Afecta solamente una observación
- ⇒ Outlier innovativo: Afecta varias observaciones

Sean

- ☞ n'_j : número (desconocido) de outliers IO que contiene el j -ésimo segmento en los tiempos $I_j := \{t'_{1,j}, \dots, t'_{n'_j,j}\}$
- ☞ n_j^* : número (desconocido) de outliers AO que contiene el j -ésimo segmento en los tiempos $A_j := \{t_{1,j}^*, \dots, t_{n_j^*,j}^*\}$.

El proceso $\{Y_t\}$ se puede modelar como sigue:

$$Y_t = \begin{cases} X_{t,j} + V_{t,j} & t \in A_j, \\ X_{t,j}, & t \notin A_j, \end{cases}$$

donde

☞ $V_{t,j} \stackrel{iid}{\sim} N(0, \sigma_j^{*2})$

☞ $\{X_{t,j}\}$ es un proceso $\text{AR}(p_j)$ contaminado:
 $\{X_{t,j}\}$, i.e.,

$$X_{t,j} = \gamma_j + \phi_{j1}X_{t-1,j} + \dots + \phi_{j,p_j}X_{t-p_j,j} + \sigma_j\varepsilon_{t,j},$$

☞ Los errores $\{\varepsilon_{t,j}\}$ son independientes, tienen media cero y

varianza

$$\text{var}(\varepsilon_{t,j}) = \begin{cases} c_j^2, & t \in I_j, \\ 1, & t \notin I_j. \end{cases}$$

- ☞ Se asume que $\sigma_j^{*2} \gg \sigma_j^2$ y que $c_j^2 \gg 1$.
- ☞ Si se asume que V_t 's y $\varepsilon_{t,j}$'s son Gaussianos, entonces la j ésima pieza $\{X_{t,j}\}$ se puede representar como un modelo de espacio de estados lineal Gaussiano.
- ☞ Notación: $\psi_j := (\gamma_j, \phi_{j1}, \dots, \phi_{jp_j}, \sigma_j^2, \sigma_j^{*2}, c_j^2)$ parámetros del j –ésimo segmento.

Estimación

Sea $\mathbf{y} := (y_1, \dots, y_n)$, la serie observada, y \mathcal{F} la familia de modelos candidatos para $\mathbf{y} = (y_1, \dots, y_n)$, indexada por $m, \tau, p, n', n^*, A, I$, donde

➡ m : número de puntos de cambio

➡ $\tau := (\tau_1, \tau_2, \dots, \tau_{m+1})$

➡ $p := (p_1, p_2, \dots, p_{m+1})$

➡ $n' := (n'_1, n'_2, \dots, n'_{m+1})$

➡ $n^* := (n^*_1, n^*_2, \dots, n^*_{m+1})$

➡ $A := (A_1, A_2, \dots, A_{m+1})$

➡ $I := (I_1, I_2, \dots, I_{m+1})$

Objetivo: Seleccionar el “mejor” modelo $\mathcal{M} \in \mathcal{F}$ para \mathbf{y} .

- ➡ Para lograrlo, se usará otra vez el principio de longitud de descripción mínima (MDL) de Rissanen (1989).
- ➡ Recordando que

$$CL_{\mathcal{M}}(\mathbf{y}) = CL(\hat{\mathcal{M}}) + CL(\hat{\mathbf{e}}|\hat{\mathcal{M}}),$$

Se obtiene (Davis, et al. (2008))

$$\begin{aligned} CL(\hat{\mathcal{M}}) &= \log_2 m + (m+1) \log_2 n + \\ &\quad \sum_{j=1}^{m+1} (\log_2 n'_j + \log_2 n_j^* + (n'_j + n_j^*) \log_2 n_j + \\ &\quad \log_2 p_j + \frac{1}{2}(p_j + 3) \log_2 n_j) \\ CL_{\mathcal{M}}(\hat{\mathbf{e}}|\hat{\mathcal{M}}) &= - \sum_{j=1}^{m+1} L(\hat{\boldsymbol{\psi}}_j; \mathbf{y}_j) \end{aligned}$$

Verosimilitud

$$\begin{aligned}
 L(\boldsymbol{\psi}; \mathbf{y} | y_1, \dots, y_p, A, I) &= \int_{\mathfrak{R}^{n^*}} L(\boldsymbol{\psi}; \mathbf{y}, \mathbf{x}_A | y_1, \dots, y_p, I) d\mathbf{x}_A \\
 &= \int_{\mathfrak{R}^{n^*}} p(\mathbf{y}_A, \mathbf{y}_{A^c}, \mathbf{x}_A | y_1, \dots, y_p, I) d\mathbf{x}_A \\
 &= \int_{\mathfrak{R}^{n^*}} p(\mathbf{y}_A, \mathbf{x}_{A^c}, \mathbf{x}_A | y_1, \dots, y_p, I) d\mathbf{x}_A \\
 &= \int_{\mathfrak{R}^{n^*}} p(\mathbf{y}_A | \mathbf{x}_A) p(\mathbf{x} | x_1, \dots, x_p, I) d\mathbf{x}_A \tag{1}
 \end{aligned}$$

El factor $p(\mathbf{y}_A | \mathbf{x}_A)$ del integrando en (1) es $\phi(\mathbf{y}_A; \mathbf{x}_A, \sigma_0^2 I_{n^*})$, donde $\phi(\cdot; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ denota la densidad normal multivariada con media $\boldsymbol{\mu}$ y matriz de varianzas y covarianzas $\boldsymbol{\Sigma}$.

El segundo factor $p(\mathbf{x}|x_1, \dots, x_p, I)$ está dado por

$$p(\mathbf{x}|x_1, \dots, x_p, I) = \prod_{\substack{t > p \\ t \notin I}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{(x_t - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p})^2 / (2\sigma^2)} \times \\ \prod_{t \in I} \frac{1}{\sqrt{2\pi c^2 \sigma^2}} e^{(x_t - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p})^2 / (2c^2 \sigma^2)}$$

➡ $\hat{\psi}$ se obtiene como aquel valor que maximiza

$$L(\psi; \mathbf{y}|y_1, \dots, y_p, A, I)$$

en (1).

➡ Para obtener la integral en (1) se puede usar las recursiones de Kalman.

Defínase como el "mejor" modelo $\hat{\mathcal{M}} \in \mathcal{F}$ para \mathbf{y} como aquel que minimiza $\text{MDL}(m, \tau, p, n', n^*, A, I)$. Es decir,

$$\hat{\mathcal{M}} := \arg \min_{\mathcal{M} \in \mathcal{F}} \text{MDL}(m, \tau, p, n', n^*, I, A).$$

Cromosoma

$\mathbf{g} := (g_1, \dots, g_n)$, la codificación de $\mathcal{M} \in \mathcal{F}$, es como sigue

$$g_t = \begin{cases} -3, & \text{si hay un AO en } t, \\ -2, & \text{si hay un IO en } t, \\ -1, & \text{si en } t \text{ no hay AO, IO o pc,} \\ d_j, & \text{si } t = \tau_{j-1} \text{ y el orden del AR del segmento } j \text{ es } p_j. \end{cases}$$

Ejempl.o Si $n = 10$, el cromosoma

$$\mathbf{g} = (2, -1, -1, -1, -3, -2, 1, -2, -1, -1)$$

corresponde al modelo $\mathcal{M} \in \mathcal{F}$ que contiene

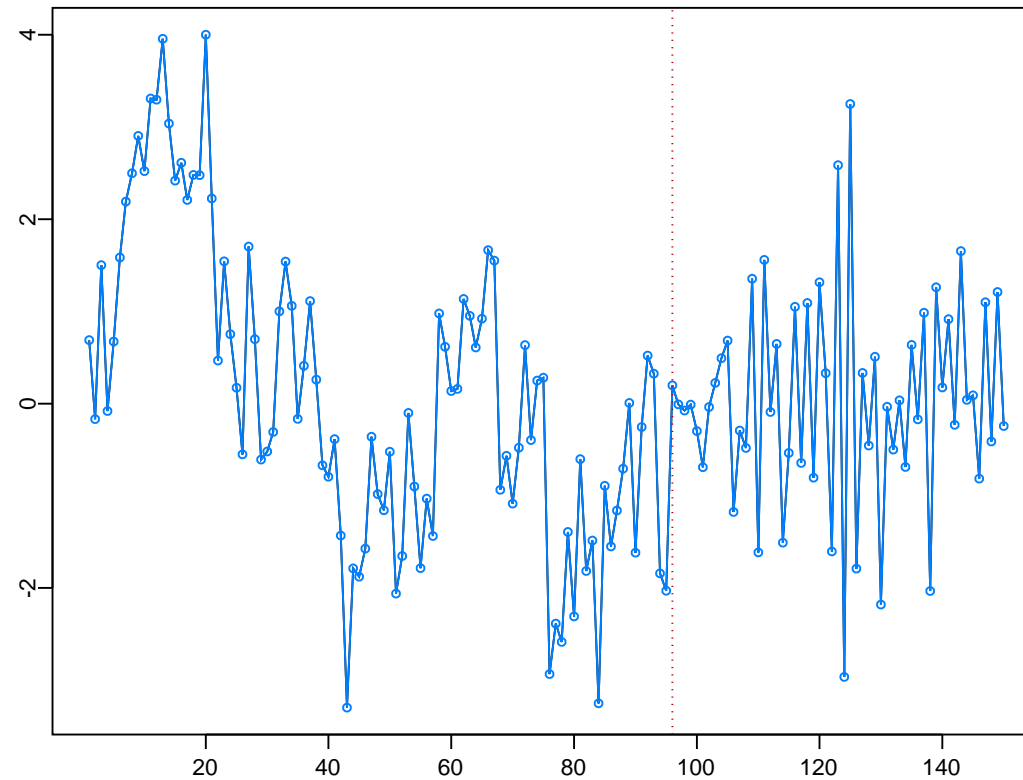
- ➡ 2 segmentos AR ($m = 1$)
- ➡ Un punto de cambio ($\tau_1 = 7$).
- ➡ Primer segmento es un proceso AR(2)
- ➡ Hay dos outliers en el primer segmento: AO en la posición 5 y IO en la 6
- ➡ Segundo segmento es un proceso AR(1).
- ➡ Hay un IO en la segunda posición de este segmento

MDL de puntos de cambio y outliers

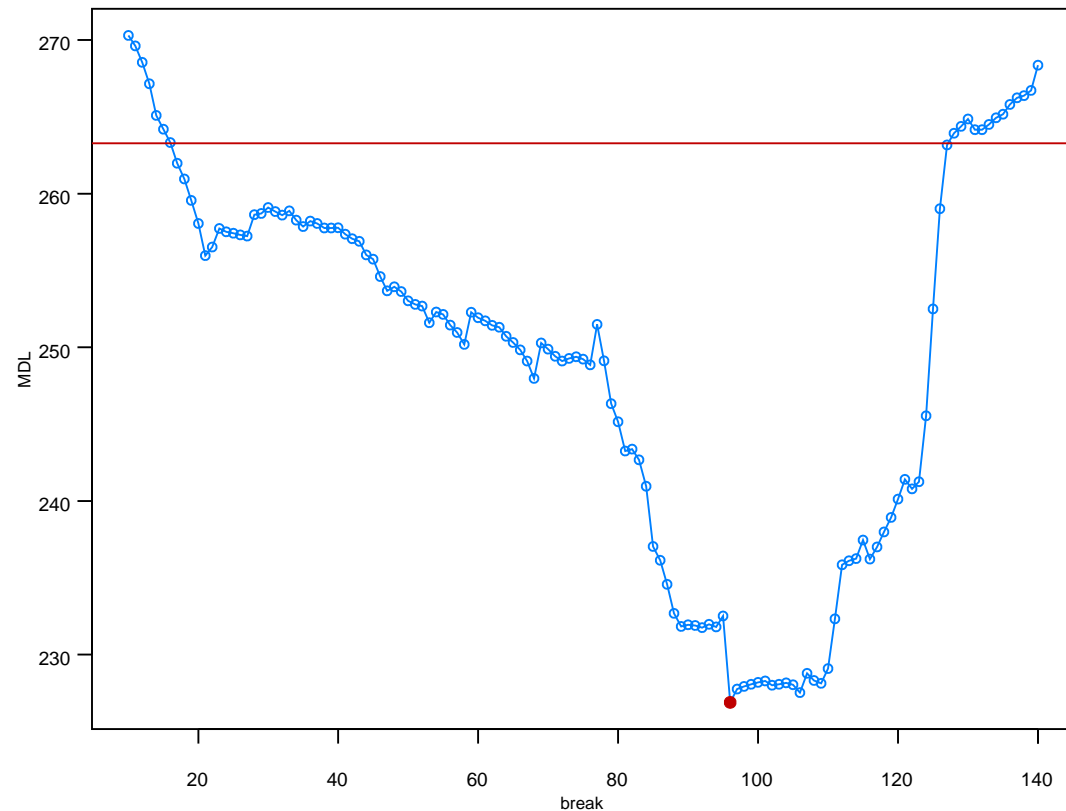
Punto de cambio:

$$Y_t = \begin{cases} 0.8Y_{t-1} + \varepsilon_t, & 1 \leq t < 101 \\ -0.8Y_{t-1} + \varepsilon_t, & 101 \leq t \leq 150. \end{cases} \quad (3)$$

donde $\varepsilon_t \stackrel{iid}{\sim} N(0, 1)$.



Una realization del proceso en (3). La línea vertical muestra el punto de cambio $\tau_1 = 101$.



puntos azules: MDL de los modelos $\mathcal{M}_{11}, \mathcal{M}_{12}, \dots, \mathcal{M}_{140}$, donde \mathcal{M}_t denota el proceso con dos segmentos AR(1) no contaminados con punto de cambio en $\tau_1 = t$.

La recta horizontal muestra el MDL del modelo \mathcal{M}_0 , el proceso AR(1) no contaminado sin ningún punto de cambio.

Punto rojo: Señala el punto de cambio estimado por Auto-PARMO

Outlier IO:

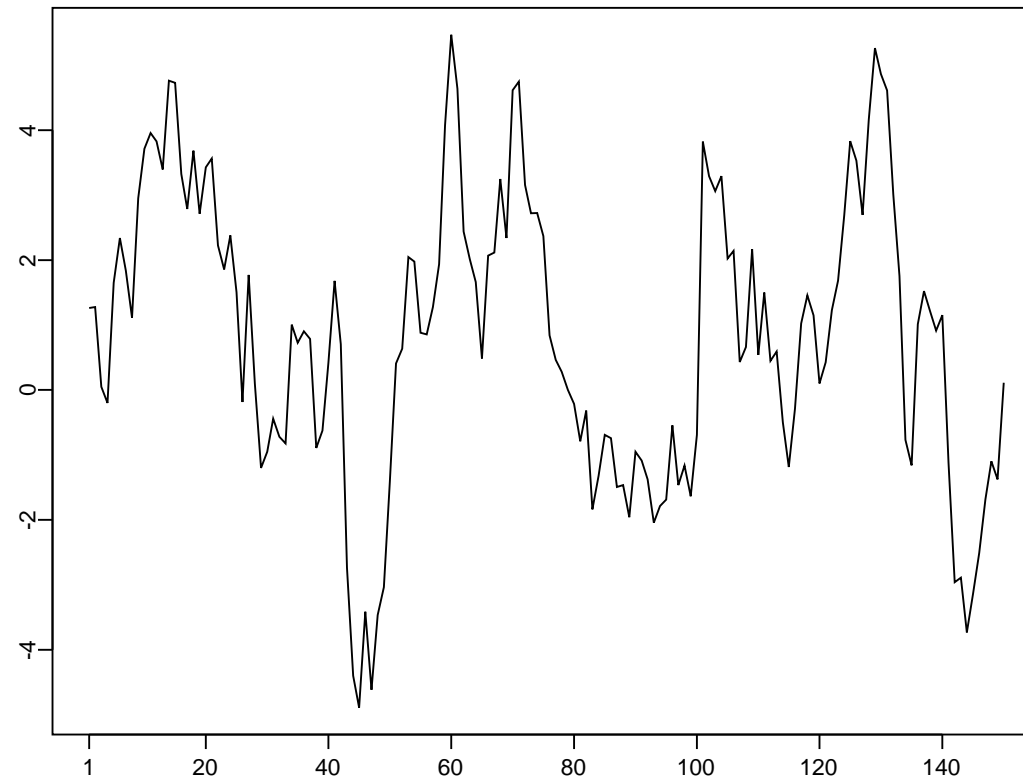
Considérese el proceso AR(1) contaminado con un IO en 101.:

$$Y_t = 0.8Y_{t-1} + \varepsilon_t^*, \quad 1 \leq t \leq 150, \quad (4)$$

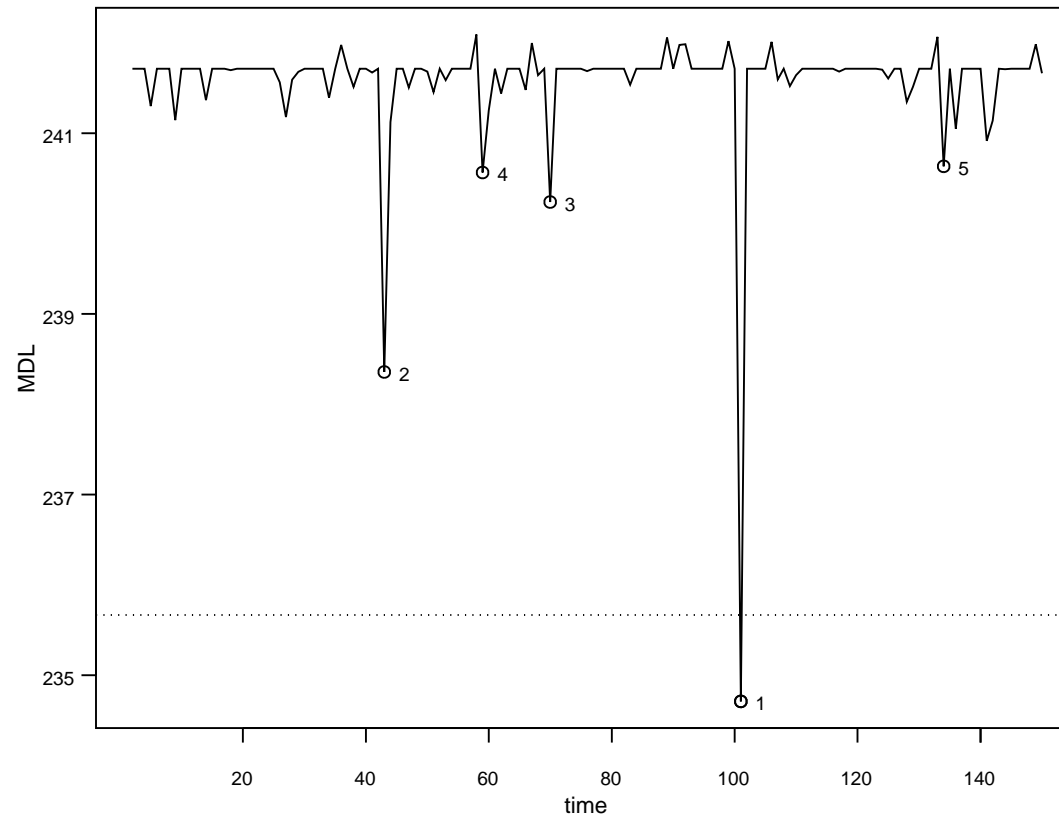
donde

☞ $\{\varepsilon_t\} \stackrel{iid}{\sim} N(0, 1),$

☞ $\varepsilon_t^* = \varepsilon_t, \text{ if } t \neq 101 \text{ and } \varepsilon_{101}^* = \varepsilon_{101} + \delta.$



Realización del proceso en (4) con $\delta = 5$.



MDL de algunos modelos CAR.

Línea horizontal: MDL de \mathcal{M}_0 .

círculos: Posiciones de los 5 residuales mayores (valor absoluto) de \mathcal{M}_0 $t_1 = 101$, $t_2 = 43$, $t_3 = 70$, $t_4 = 59$, $t_5 = 135$

Solución Auto-PARMO: \mathcal{M}_{t_1} .

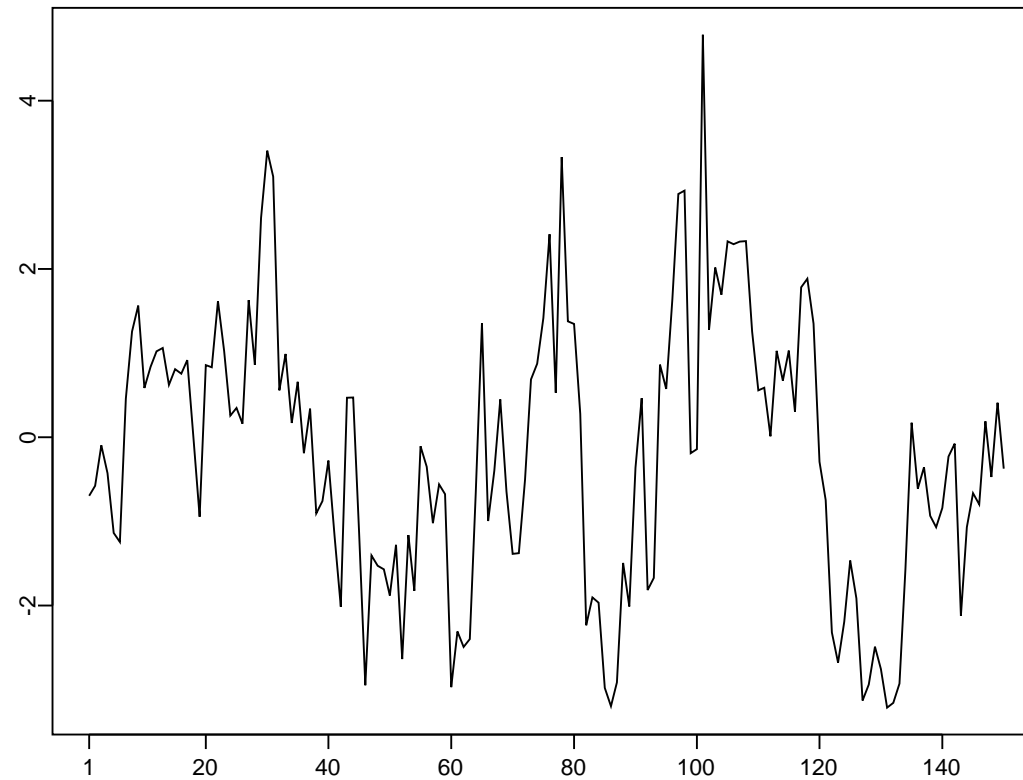
Outlier AO:

Se considera ahora un proceso AR contaminado con un outlier AO, en el tiempo 101., i.e., si se considera primero el proceso AR no contaminado siguiente

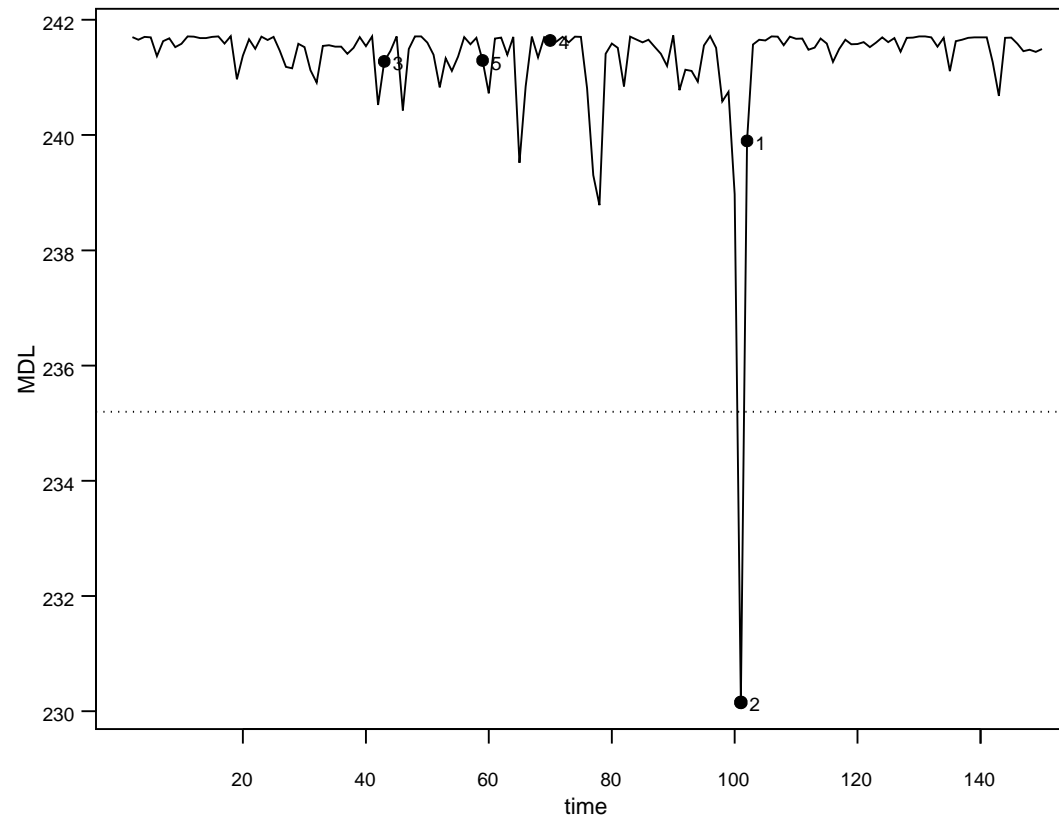
$$X_t = 0.8X_{t-1} + \varepsilon_t, \quad 1 \leq t \leq 150$$

donde $\{\varepsilon\} \stackrel{iid}{\sim} N(0, 1)$, entonces el proceso AR contaminado está dado por

$$y_t = x_t + \delta 1_{101}(t), \quad t = 1, 2, \dots, 150. \quad (5)$$



Realización del proceso en (5)



Épocas de los 5 residuales mayores de \mathcal{M}_0 : $t_1 = 102$, $t_2 = 101$, $t_3 = 43$, $t_4 = 70$, $t_5 = 59$
Solución Auto-PARMO: \mathcal{M}_{t_2} .

Desempeño de Auto-PARMO para estimar pc y outliers

El desempeño de GA es ejemplificado en los siguientes modelos:

DGP1: $\{Y_t\}$ es un proceso AR(1) (sin pc ni outliers)

$$Y_t = 0.8Y_{t-1} + \varepsilon_t, \quad t = 1, \dots, 100,$$

donde $\{\varepsilon_t\} \stackrel{iid}{\sim} N(0, 1)$.

DGP2: Como DGP1, pero contaminado con 5 IO's en 16, 32, 48, 64, 80, cada IO con $\delta = 6$.

DGP3: Como DGP1, pero contaminated con 4 AOs en 16, 32, 48 and 80, cada AO con $\delta = 5$.

DGP/Procedure	Breaks			Outliers								time
	0	1	≥ 2	0	1	2	3	4	5	6	≥ 7	
DGP1:												
Auto-PARM	98%	2%	0%	-	-	-	-	-	-	-	-	0.15
Auto-PARMO	100%	0%	0%	94%	6%							5.3
arima.rob	-	-	-									
DGP2:												
Auto-PARM	61%	36%	3%	-	-	-	-	-	-	-	-	0.15
Auto-PARMO	98%	2%	0%	1%	0%	0%	0%	4%	92%	3%		28.3
arima.rob	-	-	-	0%	0%	1%	8%	30%	31%	18%	12%	
DGP3:												
Auto-PARM	71%	23%	6%	-	-	-	-	-	-	-	-	0.15
Auto-PARMO	100%	0%	0%	0%	0%	1%	4%	95%	0%	0%		21.2
arima.rob	-	-	-	0%	0%	0%	0%	64%	24%	7%	5%	

arima.rob: Función contenida en el módulo S+FinMetrics (Zivot and Wang, 2006, Chapter 17), en el que se implementa el procedimiento en Bianco, et al., (2001) para detectar outliers de tipos AO, IO, y LS en modelos arima (sin puntos de cambio).

ubicación	Tipo	DGP2		DGP3	
		arima.rob	Auto-PARMO	arima.rob	Auto-PARMO
16	IO	0%	92%	0%	1%
	AO	58%	4%	99%	97%
	LS	27%	-	0%	-
	Total:	85%	96%	99%	98%
32	IO	0%	96%	0%	1%
	AO	78%	3%	98%	98%
	LS	17%	-	0%	-
	Total:	95%	99%	98%	99%
48	IO	0%	96%	0%	0%
	AO	70%	1%	99%	98%
	LS	20%	- 0%	-	
	Total:	90%	97%	99%	98%
64	IO	0%	97%		
	AO	79%	2%		
	LS	20%	-		
	Total:	86%	99%		
80	IO	0%	98%	0%	3%
	AO	79%	0%	100%	94%
	LS	12%	-	0%	-
	Total:	91%	98%	100%	97%

Tipo	DGP2		DGP3	
	arima.rob	Auto-PARMO	arima.rob	Auto-PARMO
False outlier locations				
IO	0	3	0	1
AO	27	2	46	1
LS	27	-	9	-
False outliers	91	5	62	2

Conclusiones

- ➡ El problema de segmentar una serie contaminada consiste en encontrar aquél proceso que minimiza la función MDL
- ➡ Los algoritmos genéticos se pueden usar para resolver este problema (difícil) de optimización.
- ➡ Cuando el verdadero modelo es un proceso AR segmentado, este algoritmo puede estimar (con poco error) el número de segmentos y sus órdenes, así como los puntos donde ocurren los cambios
- ➡ El procedimiento se puede generalizar fácilmente a otras familias de segmentos estacionarios
- ➡ Para un proceso AR (sin pcs) contaminado con outliers, MDL parece ofrecer un procedimiento superior que arima.rob.
- ➡ Este procedimiento se puede generalizar a otros procesos.

GRACIAS