

# ***ESTIMACIÓN DEL CICLO REFERENCIAL EN EL PERÚ: 1992-2012***

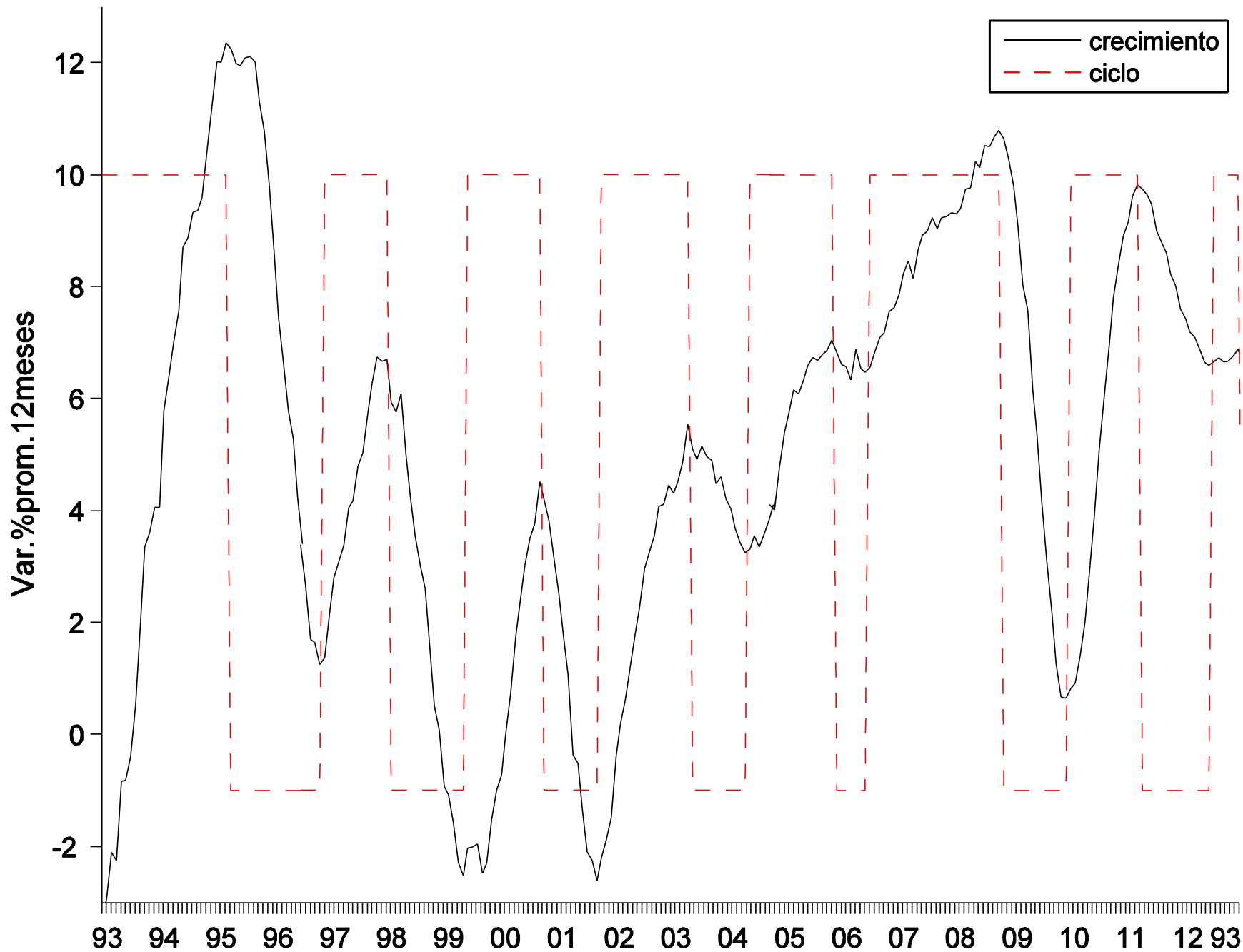
Carlos Barrera & Diego Winkelried

29 de octubre del 2013

# CONTENIDO

- Objetivo
- Motivación
- Propuesta y datos
- Algoritmo *wKhmeans* y criterio *Silhouette*
- Resultados
- Índices compuestos
- Conclusiones

# Crecimiento PBINP y su ciclo discreto (H&P)



## Objetivo

- Metodología alternativa para estimar las fechas de la ***cronología del ciclo referencial*** (de sus puntos de giro, picos y valles) usando un conjunto amplio de información (esquema de 'fechar, luego agregar').
- Metodología de Stock & Watson (2010a,b) {**S&Wab**} requiere pre-determinar las fechas candidatas a fechas referenciales (mediante el procedimiento de Harding & Pagan (2006) {**H&P**} ).
- S&Wab, siguiendo la literatura reciente, no mencionan explícitamente que también se requiere los grupos asociados a cada una de estas fechas (p.ej., mediante el procedimiento propuesto en Barrera(2012)).

## Motivación

S&Wab retoman el método tradicional ‘fechar, luego agregar’ aplicado desde antes de Bry & Boschan (1971) y proponen un **modelo de panel desbalanceado** aplicado a las fechas de puntos de quiebre (picos o fosas o ambas) de todas las variables consideradas como relevantes para definir el ciclo referencial.

Específicamente, se tiene  $N$  indicadores indexados por  $i$ ,  $i \in I \equiv \{1, 2, \dots, N\}$ , así como  $S$  episodios definidos en la vecindad de  $S$  fechas candidatas a fechas referenciales  $D_s^0$ ,  $s \in \{1, 2, \dots, S\}$ . Cada uno de estos **episodios, o intervalos disjuntos de fechas**, contiene  $N_s$  fechas de los puntos de giro de sendos indicadores: las fechas  $\tau_{js}$

## Motivación

de los puntos de giro del grupo  $G_s \equiv \{j_1, j_2, \dots, j_{N_s}\}$  con  $N_s$  indicadores,  $j \in G_s \subseteq I$ .

S&Wab proponen el siguiente modelo de panel para estas fechas,

$$\tau_{js} = k_j + D_s + \varepsilon_{js} \quad \begin{array}{l} j \in G_s \subseteq I \\ s \in \{1, 2, \dots, S\} \end{array}$$

y los parámetros a estimar son los  $\{k_j\}$  y los  $\{D_s\}$ .  $D_s$  es la fecha del punto de giro referencial que corresponde al episodio  $s$ .  $k_j$  es el rezago medio de la variable  $j$  respecto al ciclo de referencia, es decir, respecto a todas las  $D_s$ . Por un tema de identificación,

## Motivación

$\sum_i k_i = 0$ . Los grupos  $G_s$  indican que el panel es desbalanceado: no todos los  $N$  indicadores presentan un punto de giro en la **vecindad** de todas las fechas  $D_s$  en todos los  $S$  episodios (no participan en todos los grupos).

La **motivación**: el ancho de los intervalos disjuntos (en meses) –cerca de las fechas referenciales candidatas  $D_s^0$  que definen los episodios- determina qué fechas (qué indicadores) pertenecen a cada grupo  $G_s$  y afecta las desv.estánd. de los estimados de las fechas referenciales  $D_s$ . Un procedimiento no estadístico para definir estos grupos  $G_s$  sería *ad hoc*, elevando las desv.estánd. de cada grupo.

## Motivación

Esto se menciona de manera indirecta en S&Wa:

***“...once it has been established that a turning point has occurred. This allows us to partition the data into  $S$  non-overlapping episodes, each of which contains a single turning point of unknown date. Conditioning on the knowledge that an episode contains a single turning point introduces potential two-step or pretest bias, but it allows a useful simplification, and (we suspect) this two-step assumption could be relaxed by iterating on the definitions of the episodes.”***

Stock & Watson (2010a) *Indicators for dating business cycles: cross-history selection and comparisons*, AER P&P, Vol. 100 (mayo), pgs. 16-17.



## Motivación

La mención al problema es más directa en S&Wb:

***“We consider the problem of dating a reference cycle turning point (peak or trough), conditional on the event that a single turning point occurred in a given episode covering a known time span. This corresponds to a situation in which it is known that a recession occurred during a particular time interval and all that remains is to date the peak within the interval.”***

Stock & Watson (2010b) *Estimating turning points using large data sets*, NBER Working Paper No. 16532, pg. 5.

## Propuesta y datos

Se propone:

- (a) aplicar el algoritmo estadístico de agrupación ***wkhmeans*** para corregir estos grupos  $G_s$  (y sus desviaciones estándar) si se mantuviera fijo el número predeterminado de fechas referenciales  $D_s^0$  ;
- (b) aplicar sucesivamente el mismo algoritmo modificando el número de grupos  $S$  para hallar la mejor agrupación según el criterio ***Silhouette***;
- (c) depurar (si fuera necesario imponer la alternancia de puntos de giro -picos y valles-), todo punto de giro candidato que resulte 'redundante' por tener un menor ***Silhouette*** grupal, lo que también

## Propuesta y datos

depura al grupo asociado y modifica  $S$ ; y

d) utilizar los ponderadores disponibles para los distintos indicadores individuales considerados (así como los ponderadores que se calcularon con el propio algoritmo *wkhmeans* para la agrupación más idónea) en la estimación de una versión ponderada del modelo de panel desbalanceado de S&Wa, **wS&W**, con las fechas referenciales corregidas  $D_s^0$  y grupos corregidos según (a) y (b) así como con grupos depurados según (c).

## Propuesta y datos

El conjunto de datos en niveles corresponde a la desagregación más amplia de los índices del PBI sectorial, disponibles desde Enero 1992 hasta Diciembre 2012, al cual se añade otros sub-conjuntos de información complementaria (el empleo sectorial, las importaciones reales y los agregados crediticios y monetarios en términos reales). Se dispone así de un total de 99 indicadores.

Cada una de las variables en este conjunto es tratada con el procedimiento TRAMO-SEATS, luego se obtiene las variaciones semestrales suavizadas y finalmente se extrae las fechas de sus puntos de giro.

## Algoritmo *wKmeans* y criterio *Silhouette*

En general, un algoritmo de agrupación no jerárquico requiere como inputs a 2 variables:  $K$ , el número de grupos que se formará, y los  $n$  puntos (fechas) que serán agrupados.

El algoritmo optimizará estocásticamente una función objetivo (ver sgte. diapositiva) que depende de las distancias de cada punto a los denominados *centroides* (los puntos medios de cada grupo). En cada iteración, calculará las distancias de cada punto a cada *centroide*, alguna forma de asignación de los puntos a los *centroides* (membresía), así como los propios centroides (a partir del grupo de puntos que cada uno tiene asignado).

# Algoritmo *wKhmean* y criterio *Silhouette*

En el caso de *wkhmeans*, la función a optimizar es

$$F(X, C) \equiv \sum_{i=1}^n \bar{\omega}_i \frac{K}{\sum_{j=1}^K \|x_i - c_j\|^{-p}}$$

con función de membresía  $\mu(c_h | x_i) \equiv \frac{\|x_i - c_h\|^{-p-2}}{\sum_{j=1}^K \|x_i - c_j\|^{-p-2}}$

función de ponderadores  $\omega(x_i) \equiv \frac{\sum_{j=1}^K \|x_i - c_j\|^{-p-2}}{\left(\sum_{j=1}^K \|x_i - c_j\|^{-p}\right)^2}$

y centros  $c_h \equiv \frac{\sum_{i=1}^n \mu(c_h | x_i) \omega(x_i) \bar{\omega}_i x_i}{\sum_{i=1}^n \mu(c_h | x_i) \omega(x_i) \bar{\omega}_i}$

Una vez agrupados todos los puntos (fechas), se puede determinar cuán bien está agrupado

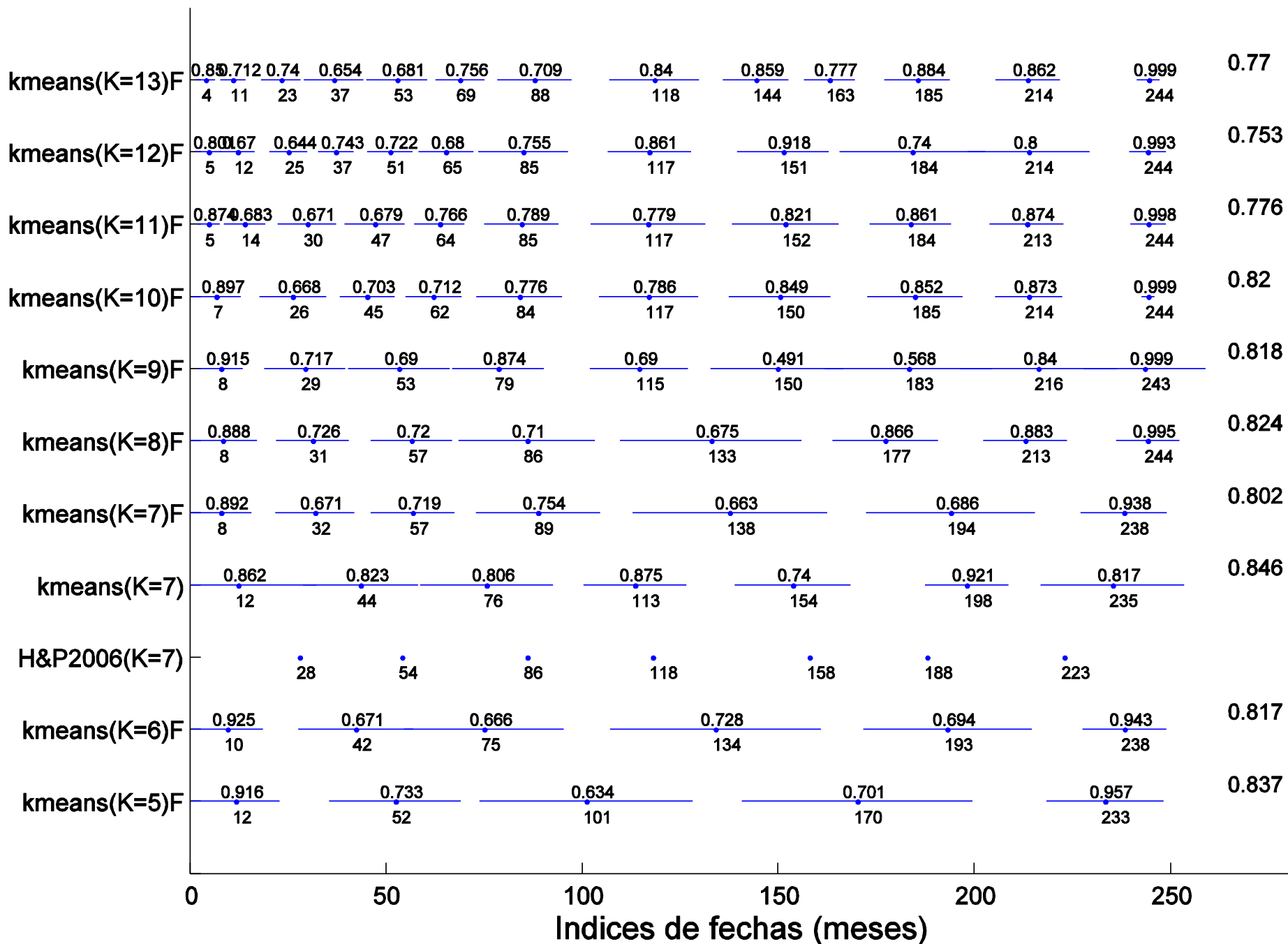
## Algoritmo *wKhmeans* y criterio *Silhouette*

está agrupado el punto  $x_i$  mediante su medida *Silhouette*,  $s(i)$ , definida como

$$s(i) \equiv \frac{b(i) - a(i)}{\max(a(i), b(i))}$$

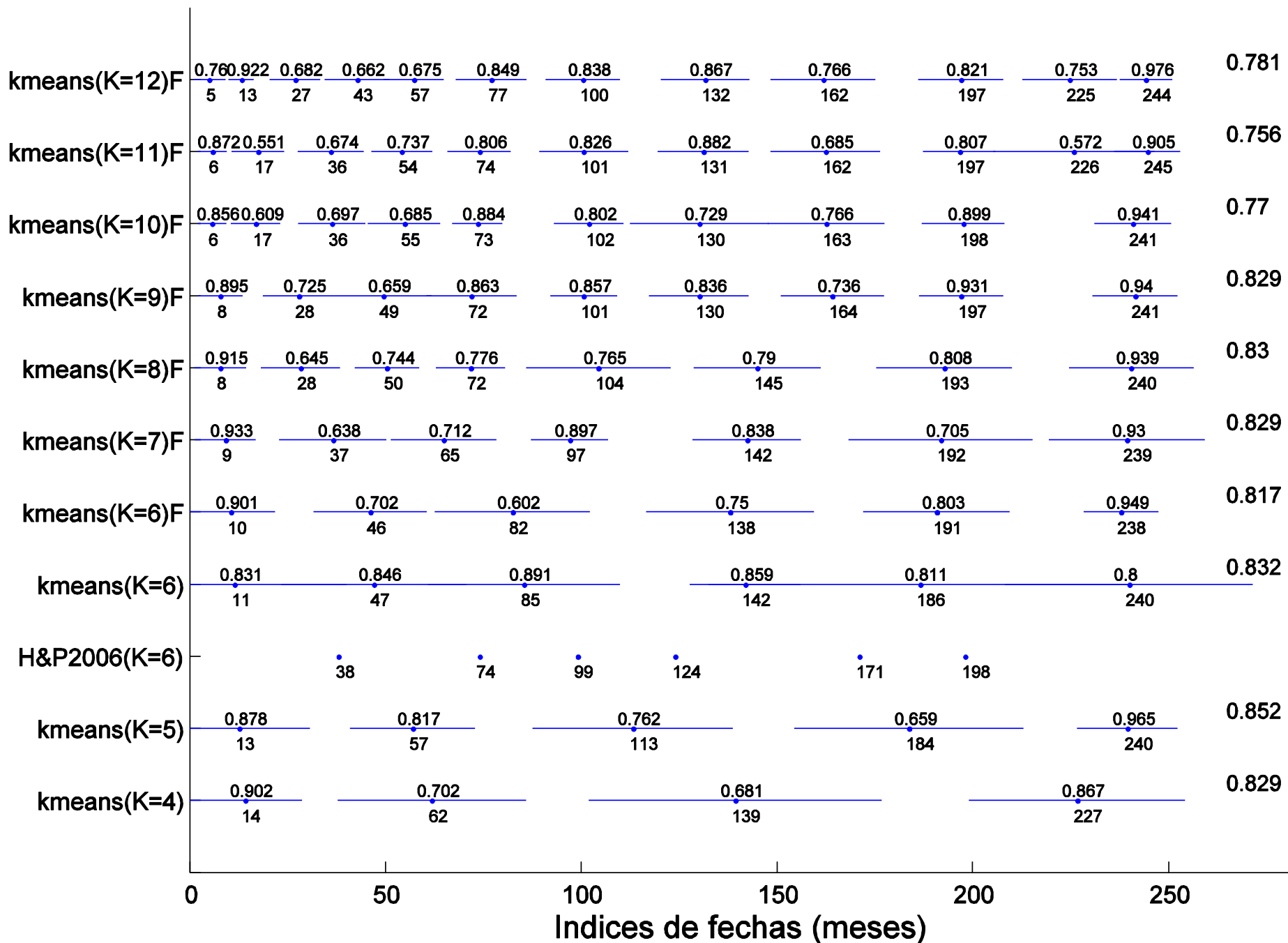
donde  $a(i)$  es la distancia (disimilaridad) **promedio** de las distancias entre  $i$  y todos los otros puntos  $j$  en el grupo al que pertenece  $i$ . Luego se calcula todas las otras  $K-1$  distancias promedio entre  $i$  y todos los puntos en cada uno de los otros  $K-1$  grupos a los que no pertenece  $i$  y se obtiene  $b(i)$  como mínima distancia de ellas, es decir,  $b(i)$  es la distancia **promedio** de las distancia entre  $i$  y todos los puntos  $h$  en el grupo más cercano.

# Centroides de "clusters" de Picos con silhouettes promedio y desv. estándar. (all)





# Centroides de "clusters" de Valles con silhouettes promedio y desv. estándar. (all)



# Resultados

Con el criterio *Silhouette*, se obtiene sugerencias para el número de grupos de picos y valles con base Y/YNP.

Conjunto	No. picos	<i>Silhou</i> .prom.	Conjunto	No. valles	<i>Silhou</i> .prom.
YEMR	10/10	0.762/0.772	YEMR	6/6	0.781h/0.777
	5/5	0.801/0.811		4/4 ó 5	0.793/0.812
YEMR(w)	7/7	<b>0.846/0.839</b>	YEMR(w)	6/6	<b>0.832h/0.819</b>
	5/5	<b>0.837/0.835</b>		5/4	<b>0.852/0.842</b>

La versión ponderada del criterio da sugerencias menos extremas. Se usa la sugerencia 5 picos y 5 valles.

## Resultados del método tradicional: H&P + S&Wa

Fecha	$D_s^0$	Estim.	Desv.	Fecha	$D_s^0$	Estim.	Desv.
Feb.95	28	35.64	1.56	Dic.95	38	42.73	1.19
Abr.97	54	55.59	1.28	Dic.98	74	73.24	1.04
Dic.99	86	85.10	0.97	Ene.01	99	100.17	1.13
Ago.02	118	118.70	0.98	Feb.03	124	125.95	1.18
Dic.05	158	158.10	1.03	Ene.07	171	170.06	1.11
Jun.08	188	185.54	0.95	Abr.09	198	200.59	0.75
May.11	223	216.12	0.70				

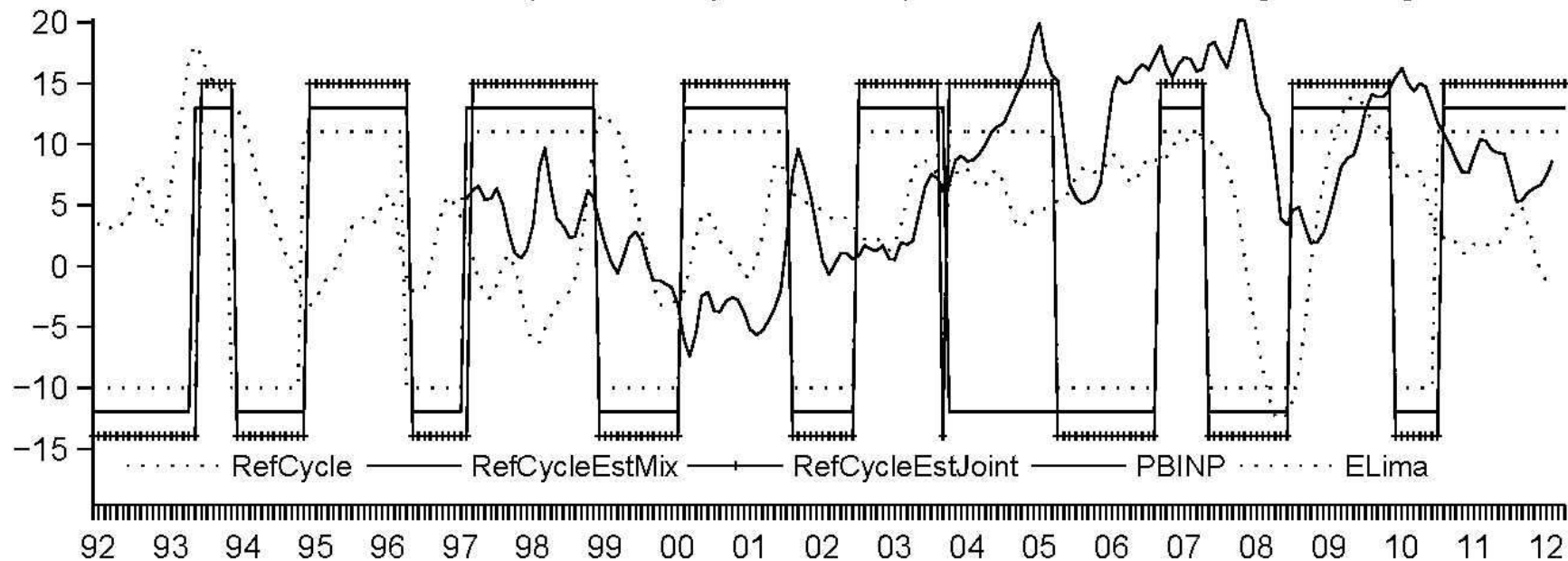
## Resultados del método propuesto: $wKhmeans + wS\&W$

Fecha	$D_s^0$	Estim.	Desv.	Fecha	$D_s^0$	Estim.	Desv.
Feb.95	28	29.78	0.90	Dic.95	38	44.47	2.35
Abr.97	54	54.51	1.33	Dic.98	74		
Dic.99	86			Ene.01	99	101.83	1.56
Ago.02	118	120.55	0.83	Feb.03	124	130.06	1.03
Dic.05	158	146.54	1.19			155.09	0.74
		161.69	1.23	Ene.07	171	175.46	1.56
Jun.08	188	185.12	0.72	Abr.09	198	199.23	0.61
May.11	223	214.22	1.17			226.95	1.82

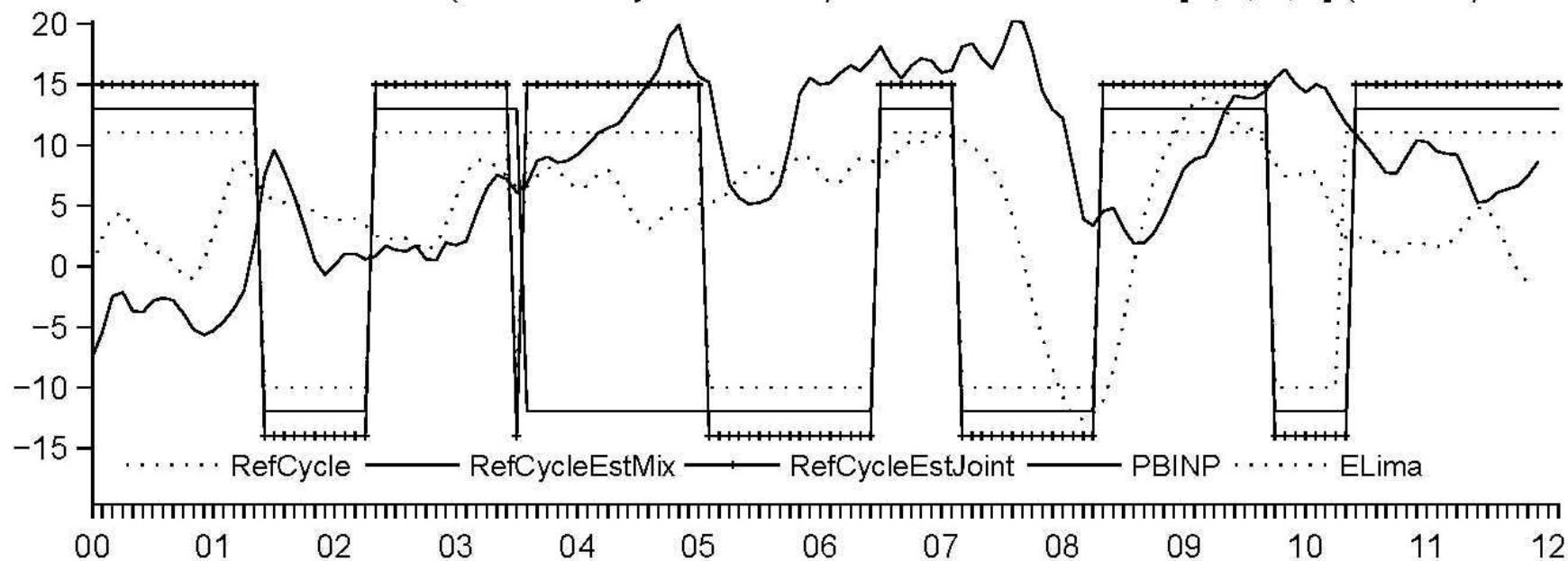
## Resultados

- En el caso del método tradicional, que usa H&P para las fechas candidatas y S&Wa, todas las fechas estimadas tienen desviaciones estándares entre 0.7 y 1.6 meses y se encuentran muy cerca de las de H&P.
- En el caso del método propuesto, que usa ***wKhmeans*** para las fechas candidatas y una versión ponderada de S&Wa, todas las fechas estimadas, excepto la del valle más reciente, tienen desviaciones estándares que toman valores entre 0.6 y 2.4 meses.
- Sin embargo, hay una gran distancia respecto a las fechas H&P (que se obtienen con una media geométrica ponderada).

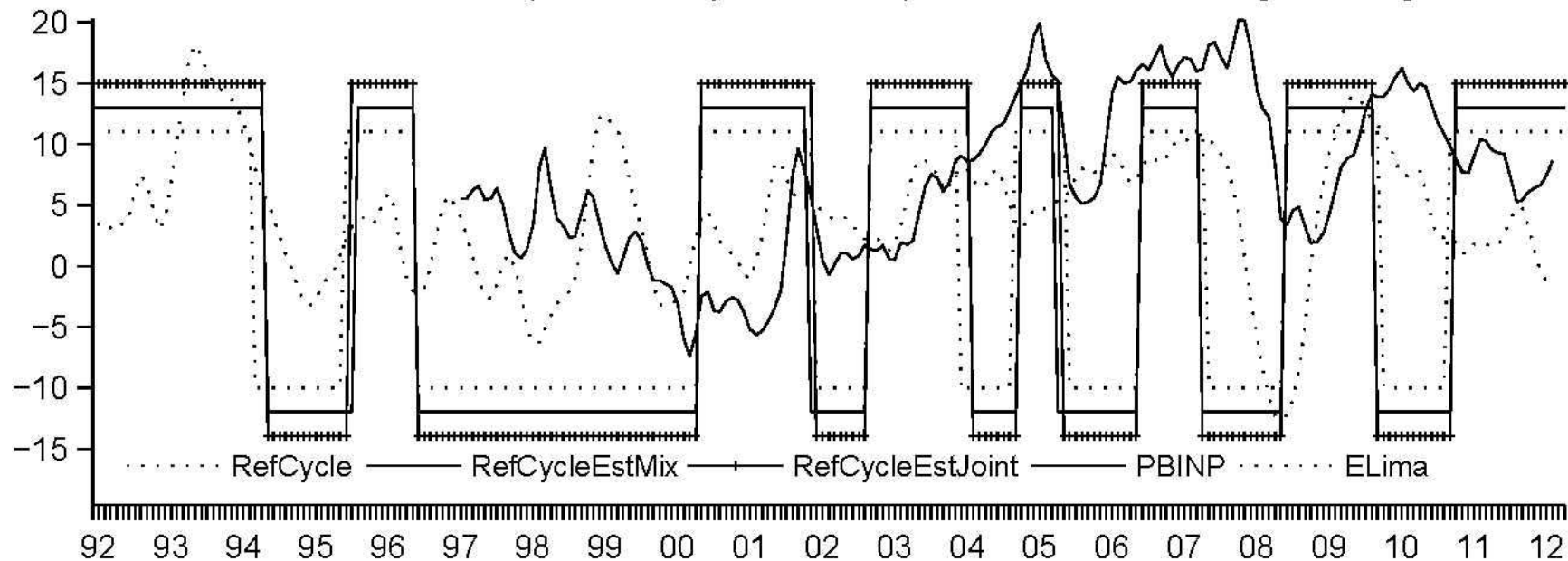
Ciclo referencial (tradicional y estimados) vs. 2 indicadores en [Y,E,M,R]



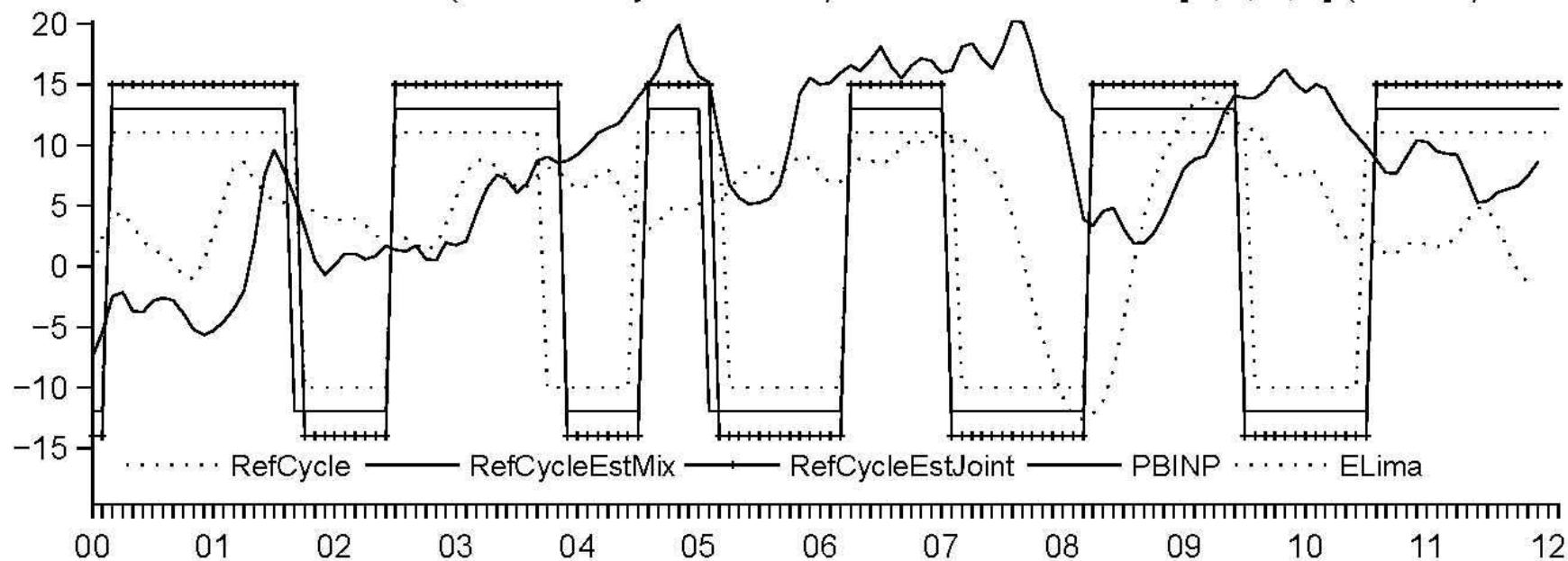
Ciclo referencial (tradicional y estimados) vs. 2 indicadores en [Y,E,M,R] (últ.obs.)



Ciclo referencial (tradicional y estimados) vs. 2 indicadores en [Y,E,M,R]



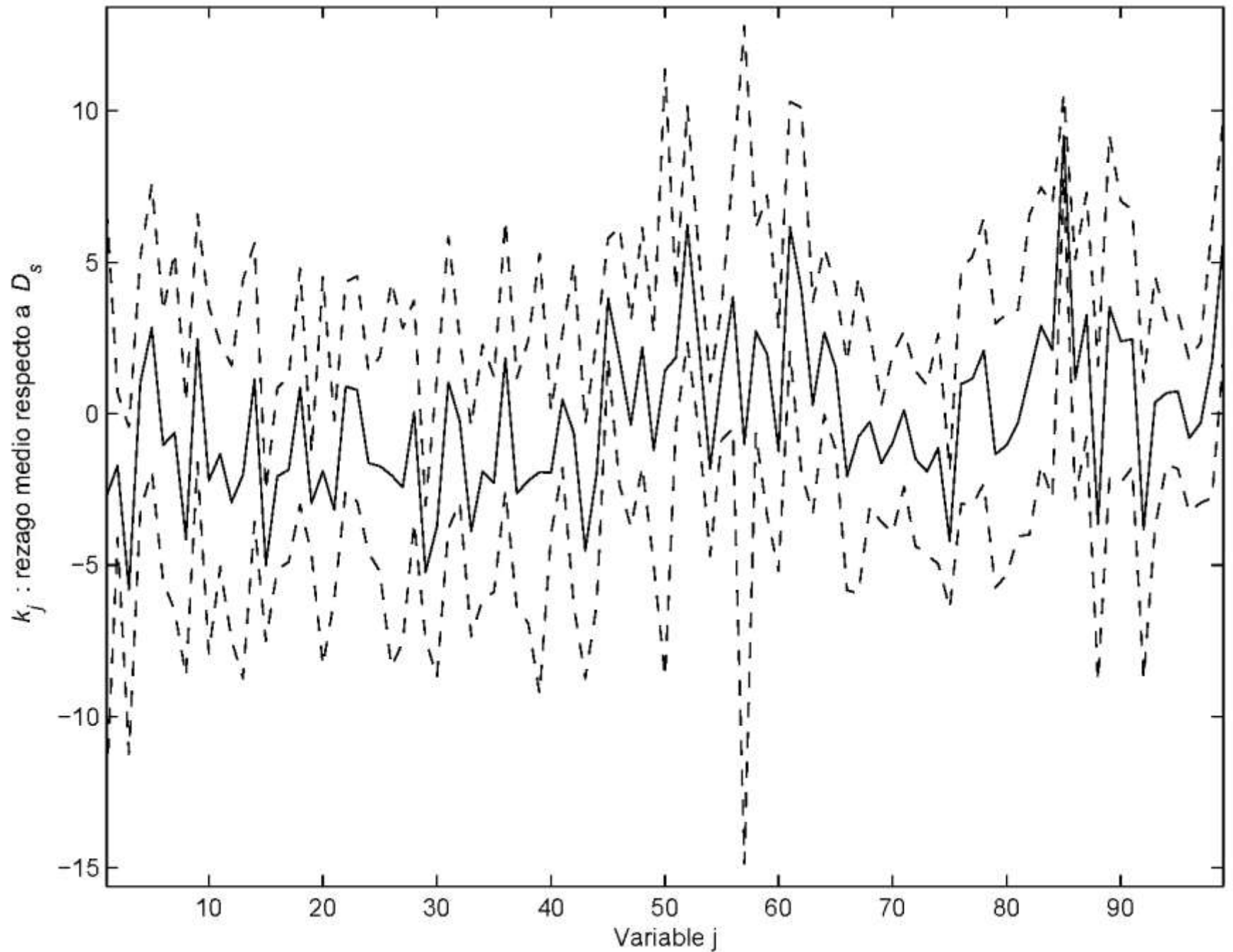
Ciclo referencial (tradicional y estimados) vs. 2 indicadores en [Y,E,M,R] (últ.obs.)



## Indices compuestos

- Se ha presentado los estimados de las fechas referenciales,  $D_s$ . Pasamos ahora a los numerosos estimados de los efectos fijos  $k_i$ .
- Esta proliferación de parámetros no sería problemática si se dispusiera de un número mayor de episodios (picos y valles). La mayoría de los 99 parámetros estimados para picos y para valles resultan estadísticamente iguales a cero.
- Este resultado es otra motivación para la metodología alternativa aquí propuesta. Se esperaba que parte de la elevada incertidumbre de estos estimados con la propuesta original de S&Wa se reduciría.

Efectos fijos  $k_j$  de los PGs en conjunto universo [Y,E,M,R]: IC(90%)

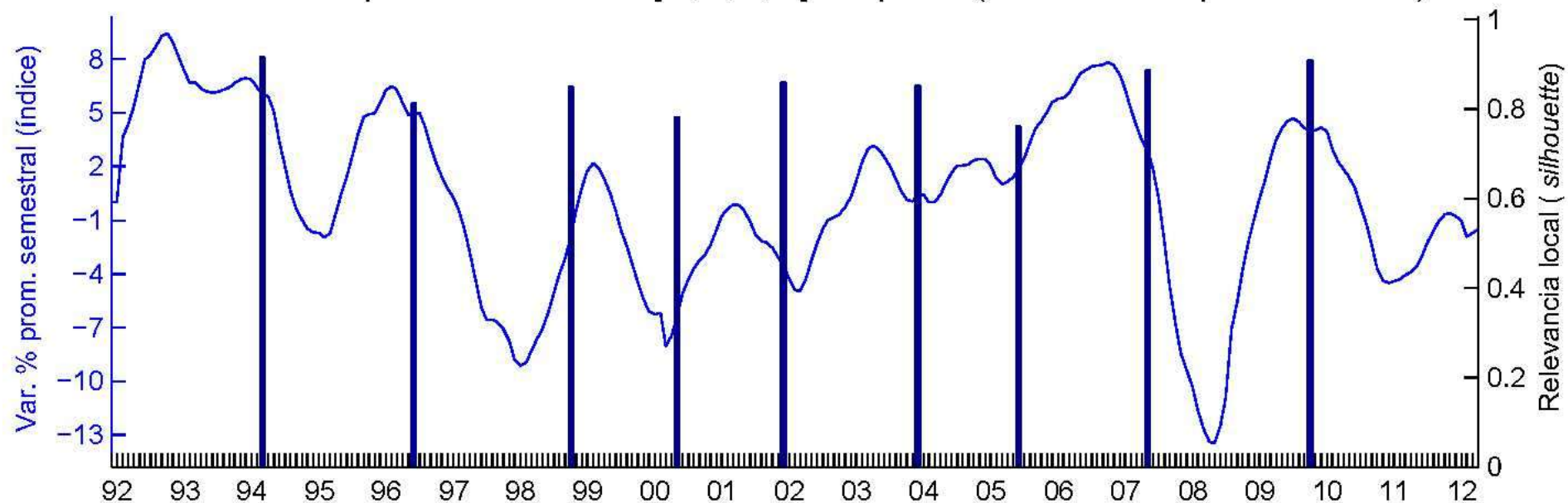




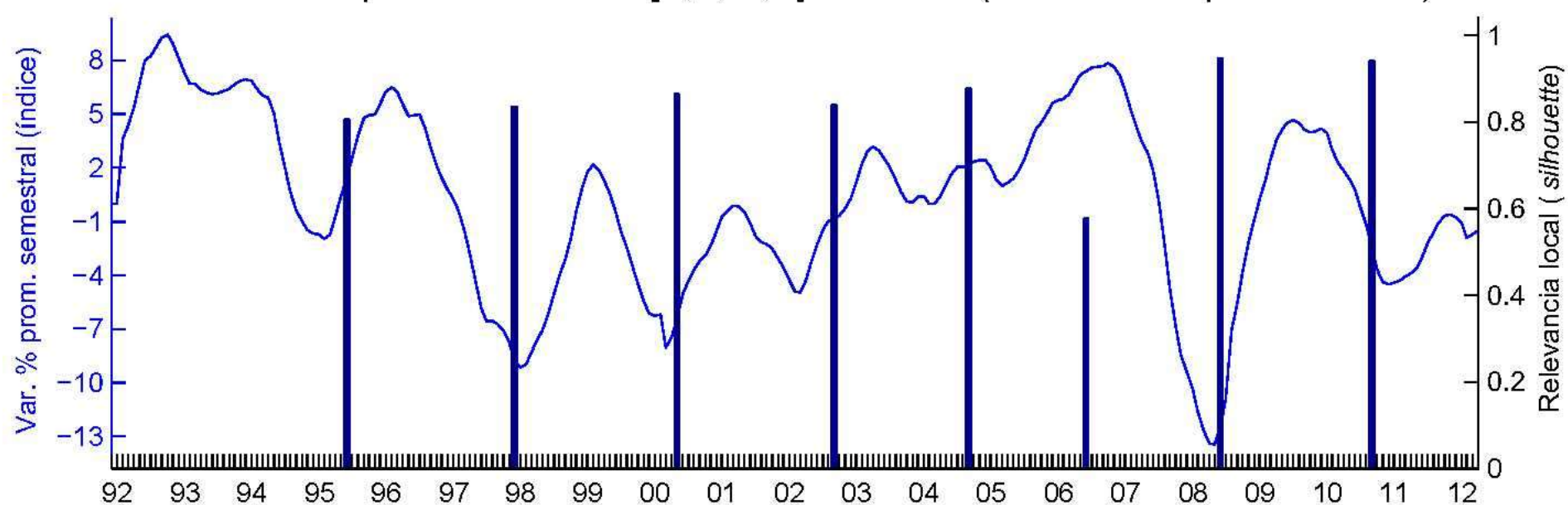
## Indices compuestos adelantados

- Por esta razón se ordenó estos coeficientes y se seleccionó los 5 indicadores con mayores rezagos medios  $k_i$  y se construyó un índice compuesto adelantado para los picos y otro índice análogo para los valles.
- Efectivamente, a pesar de su significancia estadística, las variaciones semestrales suavizadas del índice para los picos brindan señales muy claras y anticipadas de los picos en la muestra, aunque no necesariamente en términos de los picos en estas variaciones.
- Similarmente, estos mismos resultados se obtienen con el índices para los valles así como para el combinado.

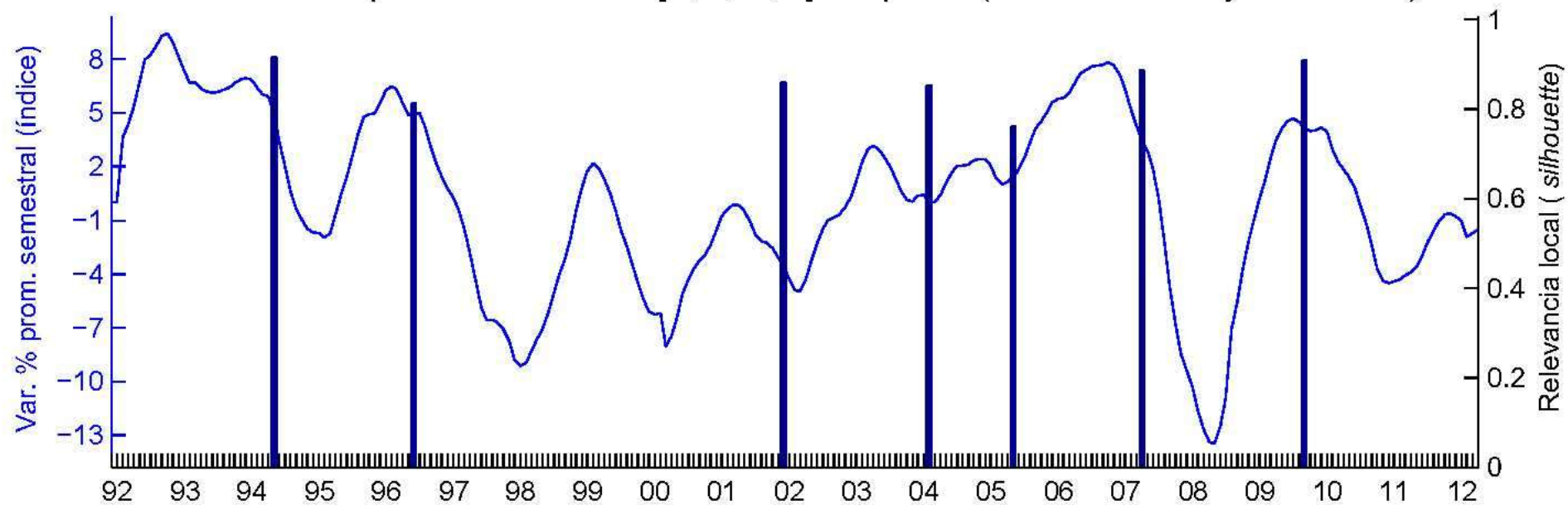
Indice anticipador de PGs en [Y,E,M,R] vs. picos ('clustered' separadamente)



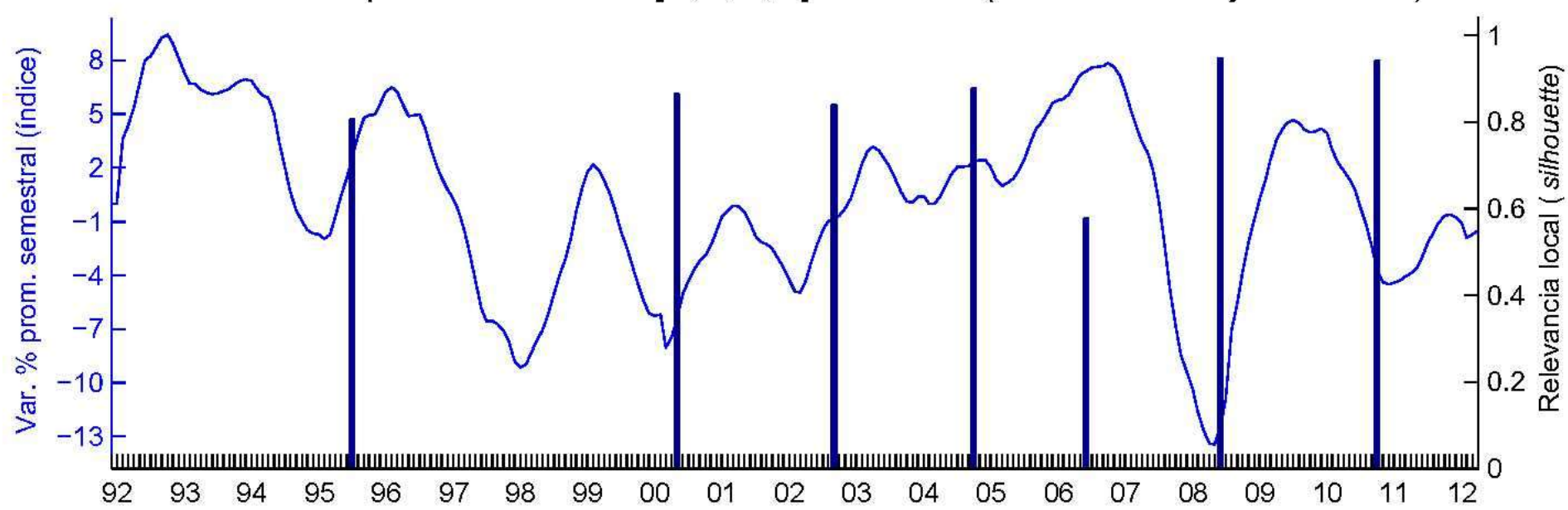
Indice anticipador de PGs en [Y,E,M,R] vs. valles ('clustered' separadamente)



Indice anticipador de PGs en [Y,E,M,R] vs. picos (estimados conjuntamente)



Indice anticipador de PGs en [Y,E,M,R] vs. valles (estimados conjuntamente)



## Conclusiones

- Se ha utilizado un procedimiento estadístico de agrupación que corrige las fechas candidatas y sus grupos.
- Este procedimiento es robusto ante la presencia de *outliers*, es menos sujeto a problemas de óptimos locales y permite usar los ponderadores presentes en los componentes de diferentes agregados macroeconómicos.
- La disponibilidad de estos ponderadores es natural en las estadísticas macroeconómicas y su importancia no puede minimizarse bajo cualquier método.
- La versión ponderada del modelo de panel desbalanceado de S&Wa favorece la construcción de índices compuestos, especialmente si se dispone de un número adecuado de episodios.

# ***ESTIMACIÓN DEL CICLO REFERENCIAL EN EL PERÚ: 1992-2012***

Carlos Barrera & Diego Winkelried

29 de octubre del 2013