

Identificación, diagnóstico y predicción con series temporales



Metodología de predicción con series temporales

- ▶ Siguiendo la metodología clásica de Box y Jenkins consideramos las tres fases previas que ellos consideran aunque ahora teniendo en cuenta los nuevos elementos aportados posteriormente.
- ▶ Se trabaja las series por un proceso repetitivo de:
 - Identificación
 - Estimación
 - Diagnóstico del modelo
- ▶ y una vez bien diagnosticada se pasa a la fase final que es la predicción.

Identificación

- ▶ Dada la composición sospechada de la serie en
$$X_t = \{(tendencia + ciclos) + \text{proceso aleatorio estacionario (PAE)}\}$$
- ▶ lo primero será conocer cada uno de los elementos que componen la serie, empezando por la tendencia y los ciclos, que son los que producen la no estacionariedad de la serie.
- ▶ La tendencia marca siempre la principal evolución de la serie, especialmente la tendencia estocástica.
 - ▶ Cuando es solo tendencia estocástica, influye en la variabilidad de la serie
 - ▶ Cuando es una estacionalidad estocástica, se descompone en dos elementos:
 - ▶ La tendencia estocástica: $(1-L)$
 - ▶ Y la estacionalidad: $(1+L+L^2+\dots+L^{s-1})$
- ▶ En cualquier caso la parte tendencial indica la principal evolución de la serie, por tanto esa es la primera en ser analizada. Luego será la parte estacional, por si aparecen efectos de largo plazo y, finalmente, la parte estacionaria, cuando se hayan eliminado las anteriores

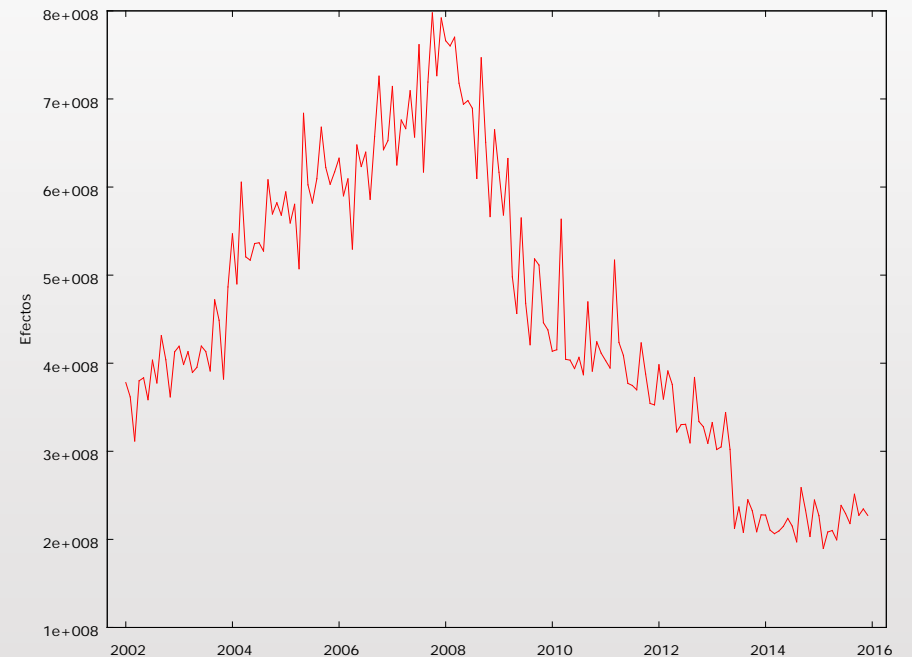


Proceso e instrumentos de Identificación

- ▶ Por tanto el proceso de identificación será el siguiente:
 1. Identificación y eliminación de la tendencia
 2. Identificación y eliminación de la estacionalidad
 3. Identificación del orden del proceso ARMA que aproxima las perturbaciones aleatorias.
- ▶ Los instrumentos que utilizaremos serán
 - ▶ Gráficos
 - ▶ Criterios estadísticos
 - ▶ Test de hipótesis
 - ▶ Criterios de selección o descriptivos

Ejemplo: Efectos de comercio impagados de Pontevedra

- ▶ A modo de ejemplo vamos a analizar la serie de la cantidad en euros de efectos de comercio impagados en la Pontevedra en todas las entidades desde el año 2002.
- ▶ El grafico muestra la evolución de la serie. Aparentemente muestra una tendencia, posiblemente estocástica y un ciclo que sigue el ciclo económico y posiblemente no se detecte al tener pocos años.
- ▶ Vamos a ir aplicando las diferentes técnicas de identificación estimación y diagnosis definidas en este tema.





Identificación de la tendencia

- ▶ **Métodos gráficos**

1. Representación gráfica de la serie: Consiste en dibujar la serie y hacer un examen visual, para ver si aparece creciente, decreciente, constante, con muchas variaciones, etc....
2. Representación del correlograma: ver si decae linealmente o si presenta autocorrelaciones altas para retardos mayores de $N/5$

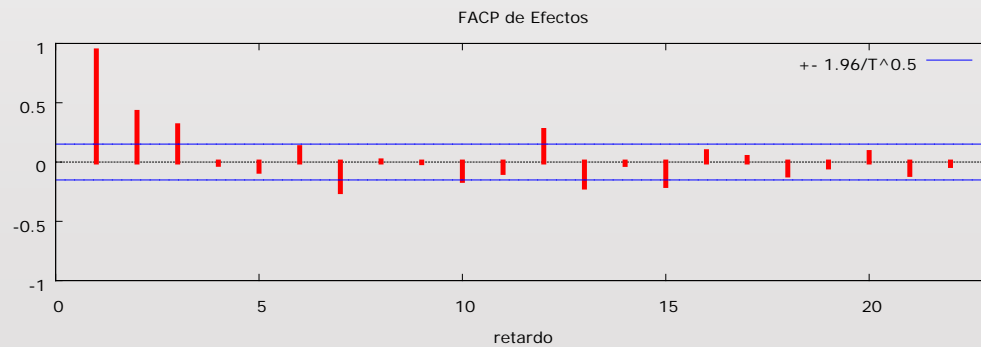
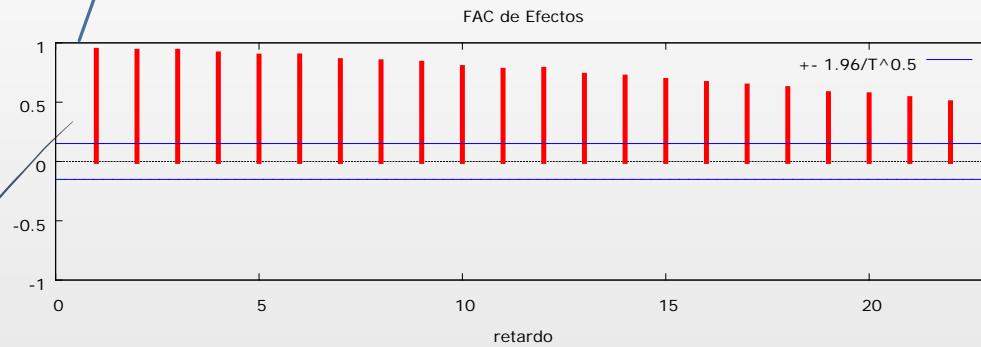
- ▶ **Métodos de Test de hipótesis**

- ▶ Se utilizan los test de raíces unitarias indicados previamente. Con ellos se comprueba si la tendencia es determinística o estocástica.

Correlogramas

Correlación cercana a 1 y significativa: posible raíz unitaria

Decaimiento lento. Posible tendencia estocástica



Función de autocorrelación para Efectos

***, ** y * indica significatividad a los niveles del 1%, 5% y 10%

Utilizando desviación típica $1/T^{0.5}$

RETARDO	FAC		FACP		Estad-Q.	[valor p]
1	0.9372	***	0.9372	***	150.2173	[0.000]
2	0.9293	***	0.4187	***	298.7954	[0.000]
3	0.9294	***	0.3060	***	448.3033	[0.000]
4	0.9064	***	-0.0195		591.3890	[0.000]
5	0.8884	***	-0.0777		729.6910	[0.000]
6	0.8902	***	0.1213		869.3872	[0.000]
7	0.8503	***	-0.2495	***	997.6550	[0.000]
8	0.8407	***	0.0098		1123.8098	[0.000]
9	0.8284	***	-0.0055		1247.0691	[0.000]
10	0.7921	***	-0.1546	**	1360.4689	[0.000]
11	0.7688	***	-0.0879		1467.9814	[0.000]
12	0.7777	***	0.2663	***	1578.7048	[0.000]
13	0.7269	***	-0.2112	***	1676.0505	[0.000]
14	0.7117	***	-0.0204		1769.9877	[0.000]
15	0.6833	***	-0.1979	**	1857.1375	[0.000]
16	0.6573	***	0.0870		1938.3106	[0.000]
17	0.6362	***	0.0386		2014.8745	[0.000]
18	0.6147	***	-0.1092		2086.8089	[0.000]
19	0.5716	***	-0.0415		2149.4434	[0.000]
20	0.5622	***	0.0803		2210.4285	[0.000]
21	0.5298	***	-0.1046		2264.9610	[0.000]
22	0.4950	***	-0.0293		2312.8935	[0.000]

Test de raíces unitarias

No se rechaza en los dos primeros y se rechaza el tercero indicando la existencia de raíz unitaria

Contraste aumentado de Dickey-Fuller para v_1

incluyendo 12 retardos de $(1-L)v_1$
(el máximo fue 13, el criterio AIC)
tamaño muestral 155
hipótesis nula de raíz unitaria: $a = 1$

contraste sin constante

modelo: $(1-L)y = (a-1)y(-1) + \dots + e$

valor estimado de $(a - 1)$: -0.0044545

Estadístico de contraste: $\tau_{nc}(1) = -0.679934$

valor p asintótico 0.4228

Coef. de autocorrelación de primer orden de e : -0.008

diferencias retardadas: $F(12, 142) = 13.777 [0.0000]$

contraste con constante

modelo: $(1-L)y = b_0 + (a-1)y(-1) + \dots + e$

valor estimado de $(a - 1)$: -0.0175812

Estadístico de contraste: $\tau_c(1) = -0.82962$

valor p asintótico 0.8102

Coef. de autocorrelación de primer orden de e : -0.010

diferencias retardadas: $F(12, 141) = 13.304 [0.0000]$

con constante y tendencia

modelo: $(1-L)y = b_0 + b_1*t + (a-1)y(-1) + \dots + e$

valor estimado de $(a - 1)$: -0.0633353

Estadístico de contraste: $\tau_{ct}(1) = -2.29862$

valor p asintótico 0.4341

Coef. de autocorrelación de primer orden de e : -0.009

diferencias retardadas: $F(12, 140) = 12.930 [0.0000]$

Contraste aumentado de Dickey-Fuller (GLS) para v_1

incluyendo 12 retardos de $(1-L)v_1$
(el máximo fue 13, el criterio AIC modificado, Perron-Qu)
tamaño muestral 155
hipótesis nula de raíz unitaria: $a = 1$

con constante y tendencia

modelo: $(1-L)y = b_0 + b_1*t + (a-1)y(-1) + \dots + e$

valor estimado de $(a - 1)$: -0.0220301

Estadístico de contraste: $\tau = -1.08516$

	10%	5%	2.5%	1%
Valores críticos:	-2.64	-2.93	-3.18	-3.46
Coef. de autocorrelación de primer orden de e :	-0.010			
diferencias retardadas: $F(12, 142) = 13.371 [0.0000]$				

Contraste KPSS para v_1

$T = 168$

Parámetro de truncamiento de los retardos = 4

Estadístico de contraste = 1.81008

	10%	5%	1%
Valores críticos:	0.349	0.462	0.738
Valor p < .01			

El resultado nos sugiere la existencia de tendencia estocástica

Eliminación de la tendencia

- ▶ Según sea la tendencia cambia el proceso de eliminación
- ▶ Tendencia determinística
 - ▶ Para eliminarla tendremos que hacer un ajuste temporal, según sea el tipo de tendencia. Este hecho, va a afectar a la estimación del modelo de las perturbaciones, pues estaremos trabajando con coeficientes estimados.
- ▶ **TENDENCIAS ESTOCÁSTICA O DETERMINÍSTICA Y ESTOCÁSTICA**
 - ▶ Para eliminarlas tendremos que hacer diferencias. Se diferencia una vez y se analiza el comportamiento para ver si sigue teniendo tendencia estocástica. Así sucesivamente hasta que la serie sea estacionaria.
 - ▶ La tendencia determinística es secundaria en este caso, puesto que si $X_t = a + bt + Z_t$, donde Z_t es un proceso $I(1)$, esto es con tendencia estocástica, diferenciando, se obtiene que $\Delta X_t = b + \Delta Z_t$ donde ΔZ_t será ahora $I(0)$, luego el resultado es una serie estacionaria con impulso o constante.
 - ▶ Si solo es estocástica, b será nulo.

Variación de los efectos de comercio impagados de Pontevedra

- Diferenciamos la serie a la vista de la existencia de tendencia estocástica
- Y de nuevo analizamos si aun queda tendencia

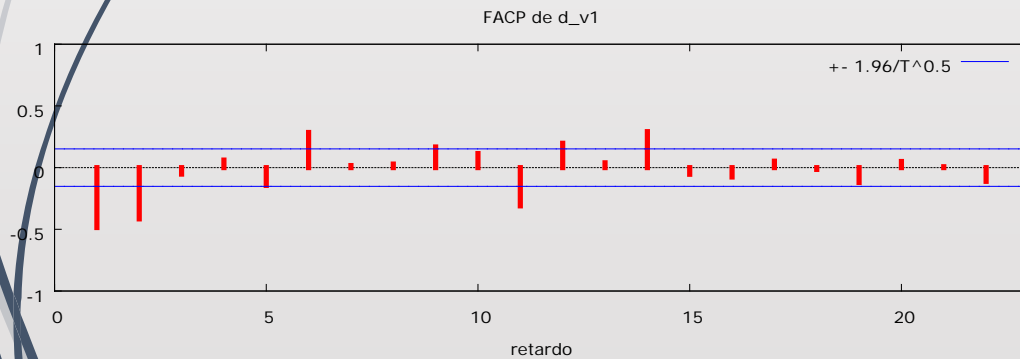
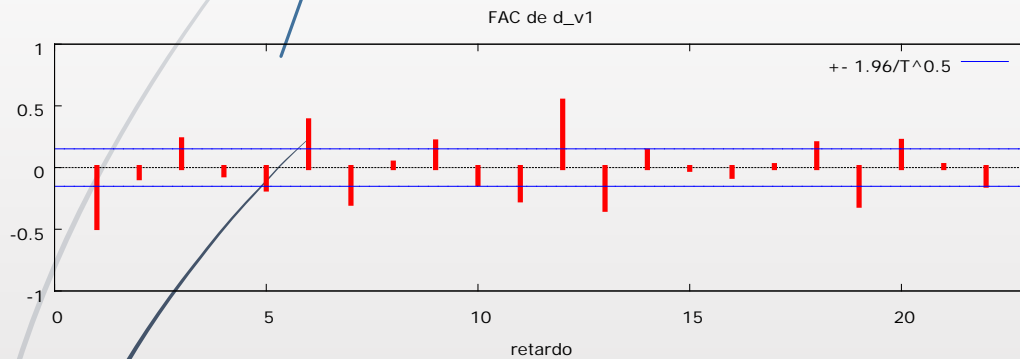


Correlogramas

Aparece significativa en muchos valores y el mas alto es el 12, pero ninguno cercano a 1. No se observa tendencia

Ninguna correlación es cercana a 1. No se sospecha raíz unitaria

Función de autocorrelación para d_v1
 ***, ** y * indica significatividad a los niveles del 1%, 5% y 10%
 Utilizando desviación típica $1/T^{0.5}$



RETARDO	FAC		FACP		Estad-Q.	[valor p]
1	-0.4855	***	-0.4855	***	40.0765	[0.000]
2	-0.0818		-0.4155	***	41.2220	[0.000]
3	0.2246	***	-0.0527		49.9047	[0.000]
4	-0.0582		0.0610		50.4921	[0.000]
5	-0.1737	**	-0.1443	*	55.7466	[0.000]
6	0.3786	***	0.2848	***	80.8732	[0.000]
7	-0.2888	***	0.0167		95.5863	[0.000]
8	0.0358		0.0288		95.8137	[0.000]
9	0.2076	***	0.1673	**	103.5137	[0.000]
10	-0.1305	*	0.1142		106.5737	[0.000]
11	-0.2614	***	-0.3099	***	118.9335	[0.000]
12	0.5378	***	0.1974	**	171.5895	[0.000]
13	-0.3380	***	0.0391		192.5254	[0.000]
14	0.1328	*	0.2917	***	195.7767	[0.000]
15	-0.0139		-0.0538		195.8126	[0.000]
16	-0.0708		-0.0762		196.7487	[0.000]
17	0.0160		0.0521		196.7967	[0.000]
18	0.1926	**	-0.0146		203.8202	[0.000]
19	-0.3049	***	-0.1202		221.5456	[0.000]
20	0.2121	***	0.0498		230.1816	[0.000]
21	0.0161		0.0073		230.2320	[0.000]
22	-0.1426	*	-0.1111		234.1898	[0.000]

Test de raíces unitarias

No se rechaza en los dos primeros y se rechaza el tercero indicando la existencia de raíz unitaria

Contraste aumentado de Dickey-Fuller para d_{v1}

incluyendo 13 retardos de $(1-L)d_{v1}$
(el máximo fue 13, el criterio AIC)
tamaño muestral 153
hipótesis nula de raíz unitaria: $a = 1$

contraste sin constante

modelo: $(1-L)y = (a-1)*y(-1) + \dots + e$
valor estimado de $(a - 1)$: -0.819324
Estadístico de contraste: $\tau_{nc}(1) = -1.87169$
valor p asintótico 0.05844
Coef. de autocorrelación de primer orden de e : 0.019
diferencias retardadas: $F(13, 139) = 8.847 [0.0000]$

contraste con constante

modelo: $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$
valor estimado de $(a - 1)$: -0.846949
Estadístico de contraste: $\tau_c(1) = -1.90771$
valor p asintótico 0.329
Coef. de autocorrelación de primer orden de e : 0.019
diferencias retardadas: $F(13, 138) = 8.784 [0.0000]$

con constante y tendencia

modelo: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$
valor estimado de $(a - 1)$: -1.05059
Estadístico de contraste: $\tau_{ct}(1) = -1.98854$
valor p asintótico 0.607
Coef. de autocorrelación de primer orden de e : 0.021
diferencias retardadas: $F(13, 137) = 8.608 [0.0000]$

Contraste aumentado de Dickey-Fuller (GLS) para d_{v1}

incluyendo 13 retardos de $(1-L)d_{v1}$
(el máximo fue 13, el criterio AIC modificado, Perron-Qu)
tamaño muestral 153
hipótesis nula de raíz unitaria: $a = 1$

con constante y tendencia

modelo: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$
valor estimado de $(a - 1)$: -0.521193
Estadístico de contraste: $\tau = -1.68453$

	10%	5%	2.5%	1%
Valores críticos:	-2.64	-2.93	-3.18	-3.46
Coef. de autocorrelación de primer orden de e :	0.016			
diferencias retardadas: $F(13, 139)$	9.462	[0.0000]		

Contraste KPSS para d_{v1}

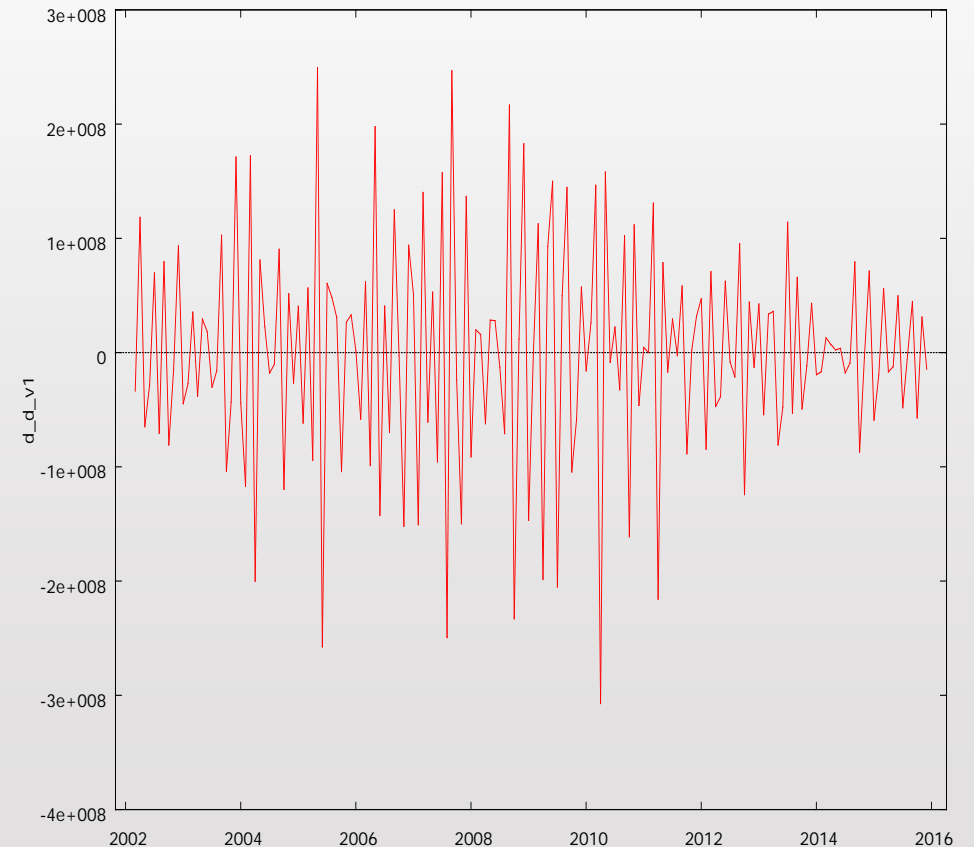
$T = 167$
Parámetro de truncamiento de los retardos = 4
Estadístico de contraste = 0.304782

	10%	5%	1%
Valores críticos:	0.349	0.462	0.738
Valor p > .10			

Este resultado nos sugiere que sigue habiendo una cierta tendencia estocástica

Variación de la variación de los efectos de comercio impagados de Pontevedra

- Diferenciamos la serie a la vista de la existencia de tendencia estocástica
- Sin embargo se observa que la varianza se ha incrementado, pues ahora varia entre -400000000 y $+300000000$, casi el doble que antes, esto nos hace sospechar de sobre diferenciación, por tanto nos quedamos con la serie diferenciada una vez.



Identificación de la estacionalidad

► *Métodos gráficos*

1. Representación gráfica de la serie sin tendencia: Consiste en dibujar la serie y hacer un examen visual, para ver si aparecen ciclos claros, repeticiones, etc...
2. Representación del correlograma de la serie sin tendencia: ver si decae linealmente o si presenta autocorrelaciones altas para retardos mayores de $N/5$

► *Métodos de Test de hipótesis*

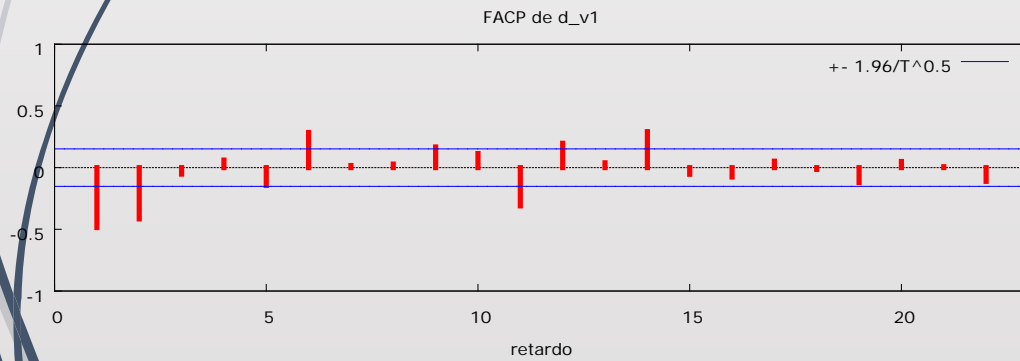
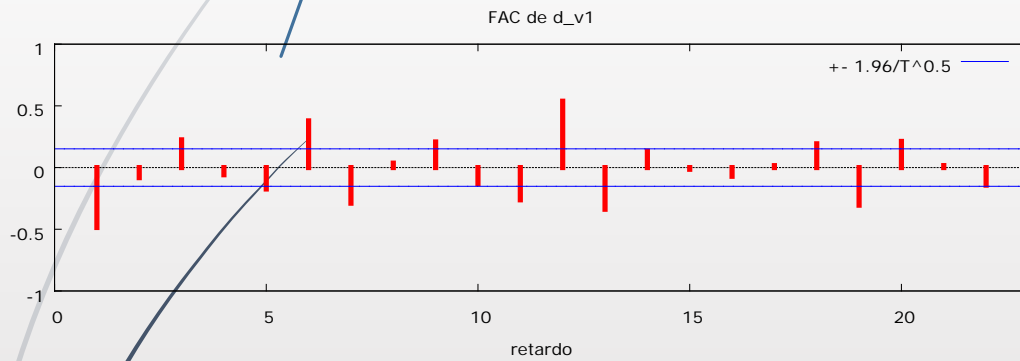
- Se utilizan los test de raíces unitarias indicados previamente con la serie sin tendencia. Con ellos se comprueba si la tendencia es determinística o estocástica.

Correlogramas

Aparece significativa en muchos valores y el mas alto es el 12. Nos hace sospechar una cierta estacionalidad anual

El valor mas alto es la correlación de orden 12, luego puede existir estacionalidad

Función de autocorrelación para d_v1
 ***, ** y * indica significatividad a los niveles del 1%, 5% y 10%
 Utilizando desviación típica $1/T^{0.5}$



RETARDO	FAC		FACP		Estad-Q.	[valor p]
1	-0.4855	***	-0.4855	***	40.0765	[0.000]
2	-0.0818		-0.4155	***	41.2220	[0.000]
3	0.2246	***	-0.0527		49.9047	[0.000]
4	-0.0582		0.0610		50.4921	[0.000]
5	-0.1737	**	-0.1443	*	55.7466	[0.000]
6	0.3786	***	0.2848	***	80.8732	[0.000]
7	-0.2888	***	0.0167		95.5863	[0.000]
8	0.0358		0.0288		95.8137	[0.000]
9	0.2076	***	0.1673	**	103.5137	[0.000]
10	-0.1305	*	0.1142		106.5737	[0.000]
11	-0.2614	***	-0.3099	***	118.9335	[0.000]
12	0.5378	***	0.1974	**	171.5895	[0.000]
13	-0.3380	***	0.0391		192.5254	[0.000]
14	0.1328	*	0.2917	***	195.7767	[0.000]
15	-0.0139		-0.0538		195.8126	[0.000]
16	-0.0708		-0.0762		196.7487	[0.000]
17	0.0160		0.0521		196.7967	[0.000]
18	0.1926	**	-0.0146		203.8202	[0.000]
19	-0.3049	***	-0.1202		221.5456	[0.000]
20	0.2121	***	0.0498		230.1816	[0.000]
21	0.0161		0.0073		230.2320	[0.000]
22	-0.1426	*	-0.1111		234.1898	[0.000]

Eliminación de la estacionalidad

- Caben dos opciones

- Estacionalidad determinística

- Para eliminarla tendremos que hacer un ajuste temporal, según sea el tipo de estacionalidad. Este hecho, va a afectar a la estimación del modelo de las perturbaciones, pues estaremos trabajando con coeficientes estimados.

- Definimos variables ficticias midiendo la estacionalidad

- Y recalculamos el modelo

$$X_t = \sum_{j=1}^s \gamma_j I_{tj} + Z_t$$

$$I_{tj} = \begin{cases} 1 & \text{si } t \in \text{estaci-}j \\ 0 & \text{en otro caso} \end{cases}$$

- Quedando Z como variable desestacionalizada

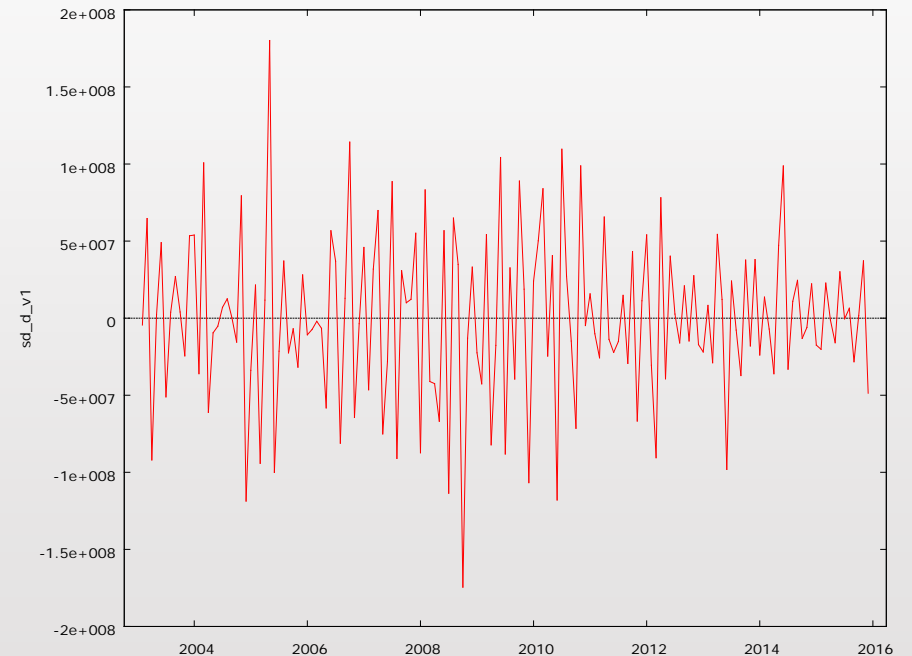
- Estacionalidad estocástica o determinística y estocástica

- Para eliminarla tendremos que hacer diferencias de periodo s, Si s es el orden de la estacionalidad. Al hacerlo se eliminan los efectos determinísticos reduciéndolos a una constante

- Si después de diferenciar sigue habiendo estacionalidad se repite el proceso

Variación anual de la variación de los efectos impagados

- ▶ Hacemos la diferencia estacional de la serie, a la vista de que los test de raíces unitarias daban que todavía existía tendencia.
- ▶ Se observa que la serie no incrementa su varianza, por lo que posiblemente sea ya estacionaria.
- ▶ No obstante lo comprobamos con los gráficos y los test .

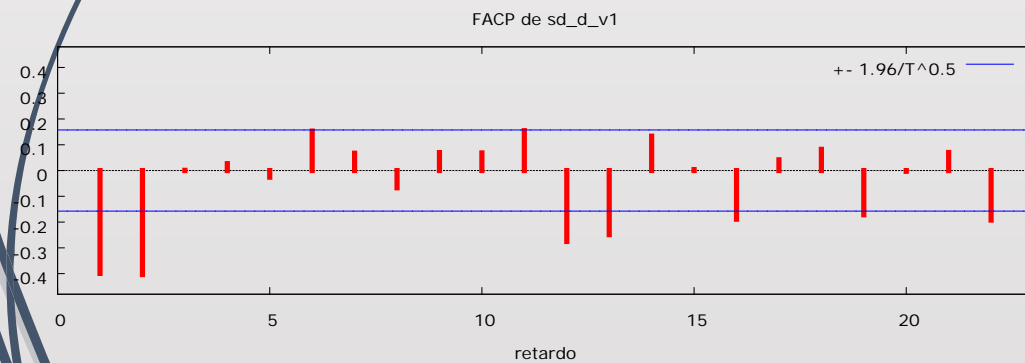
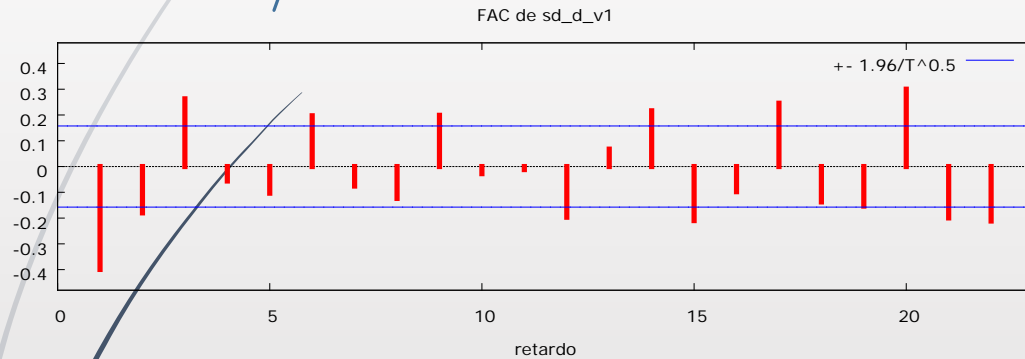


Correlogramas

Aparece significativa en muchos valores y muy irregulares. No sugiere ningún comportamiento específico estacional

No hay sospecha de estacionalidad ni tendencia

Función de autocorrelación para sd_d_v1
 ***, ** y * indica significatividad a los niveles del 1%, 5% y 10%
 Utilizando desviación típica $1/T^{0.5}$



RETARDO	FAC		FACP		Estad-Q. [valor p]
1	-0.3993	***	-0.3993	***	25.1939 [0.000]
2	-0.1799	**	-0.4037	***	30.3406 [0.000]
3	0.2627	***	0.0010		41.3890 [0.000]
4	-0.0561		0.0268		41.8969 [0.000]
5	-0.1034		-0.0259		43.6326 [0.000]
6	0.1971	**	0.1531	*	49.9769 [0.000]
7	-0.0758		0.0672		50.9215 [0.000]
8	-0.1237		-0.0669		53.4553 [0.000]
9	0.1988	**	0.0696		60.0422 [0.000]
10	-0.0277		0.0683		60.1714 [0.000]
11	-0.0119		0.1547	*	60.1954 [0.000]
12	-0.1967	**	-0.2753	***	66.7772 [0.000]
13	0.0676		-0.2489	***	67.5608 [0.000]
14	0.2164	***	0.1334	*	75.6402 [0.000]
15	-0.2098	***	0.0038		83.2912 [0.000]
16	-0.0977		-0.1889	**	84.9617 [0.000]
17	0.2456	***	0.0417		95.6019 [0.000]
18	-0.1374	*	0.0822		98.9562 [0.000]
19	-0.1535	*	-0.1717	**	103.1729 [0.000]
20	0.3002	***	-0.0031		119.4218 [0.000]
21	-0.1994	**	0.0701		126.6419 [0.000]
22	-0.2115	***	-0.1923	**	134.8302 [0.000]

Test de raíces unitarias

Se rechaza en los dos primeros y no se rechaza el tercero indicando la no existencia de raíz unitaria

Contraste aumentado de Dickey-Fuller para sd_d_v1

incluyendo 13 retardos de $(1-L)sd_d_v1$
(el máximo fue 13, el criterio AIC)
tamaño muestral 141
hipótesis nula de raíz unitaria: $a = 1$

contraste sin constante

modelo: $(1-L)y = (a-1)*y(-1) + \dots + e$
valor estimado de $(a - 1)$: -1.725
Estadístico de contraste: $\tau_{nc}(1) = -3.69084$
valor p asintótico 0.0002221
Coef. de autocorrelación de primer orden de e: 0.019
diferencias retardadas: $F(13, 127) = 5.654 [0.0000]$

contraste con constante

modelo: $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$
valor estimado de $(a - 1)$: -1.73805
Estadístico de contraste: $\tau_c(1) = -3.70563$
valor p asintótico 0.004057
Coef. de autocorrelación de primer orden de e: 0.018
diferencias retardadas: $F(13, 126) = 5.638 [0.0000]$

con constante y tendencia

modelo: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$
valor estimado de $(a - 1)$: -1.82561
Estadístico de contraste: $\tau_{ct}(1) = -3.86397$
valor p asintótico 0.0135
Coef. de autocorrelación de primer orden de e: 0.016
diferencias retardadas: $F(13, 125) = 5.729 [0.0000]$

Contraste aumentado de Dickey-Fuller (GLS) para sd_d_v1

incluyendo 0 retardos de $(1-L)sd_d_v1$
(el máximo fue 13, el criterio AIC modificado, Perron-Qu)
tamaño muestral 154
hipótesis nula de raíz unitaria: $a = 1$

con constante y tendencia

modelo: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + e$
valor estimado de $(a - 1)$: -1.38758
Estadístico de contraste: $\tau = -18.5788$

	10%	5%	2.5%	1%
Valores críticos:	-2.64	-2.93	-3.18	-3.46
Coef. de autocorrelación de primer orden de e:	-0.145			

Contraste KPSS para sd_d_v1

T = 155

Parámetro de truncamiento de los retardos = 4

Estadístico de contraste = 0.0503209

	10%	5%	1%
Valores críticos:	0.349	0.462	0.738
Valor p > .10			

Este resultado nos sugiere que sigue habiendo una cierta tendencia estocástica

Identificación del proceso ARMA

► **Métodos gráficos**

1. Representación gráfica de la serie sin tendencia y desestacionalizada: Consiste en dibujar la serie sin tendencia y desestacionalizada y hacer un examen visual, para ver si aparece algún tipo de proceso ya conocido , etc....
2. Representación del correlograma de la serie sin tendencia y desestacionalizada: ver en que valores aparece más significativos las autocorrelaciones estimadas, y la forma que presentan, según decaiga exponencialmente a partir de un valor q (parte AR) o aparezcan sólo valores aislados (parte MA).
3. Representación del correlograma parcial de la serie sin tendencia y desestacionalizada: ver en que valores aparece más significativos las autocorrelaciones parciales estimadas, y la forma que presentan, según decaiga exponencialmente a partir de un valor p (parte MA) o aparezcan sólo valores aislados (parte AR).

► **Métodos estadísticos**

► **de Test de hipótesis**

- Test de Box-Pierce o de Ljung-Box: para comprobar cuantas de las autocorrelaciones son significativas, conjuntamente.
- Test de Mann-Wald: para comprobar que autocorrelaciones son significativas, individualmente.
- Test de Wald de los coeficientes: para comprobar si el coeficiente obtenido es significativo o no, en cuyo caso se eliminaría.

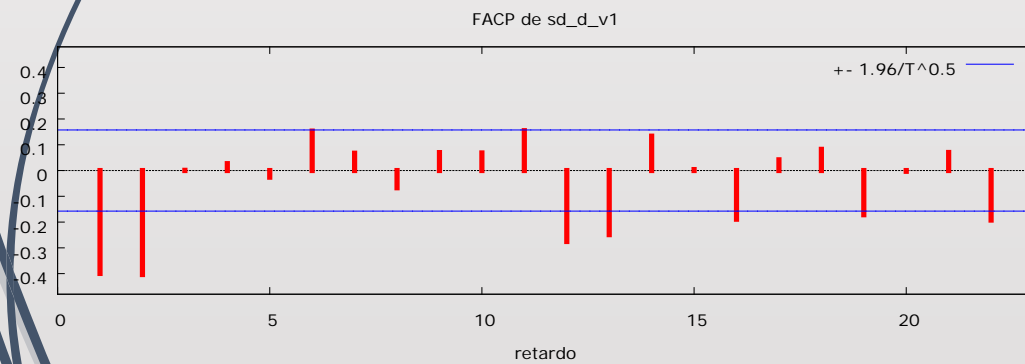
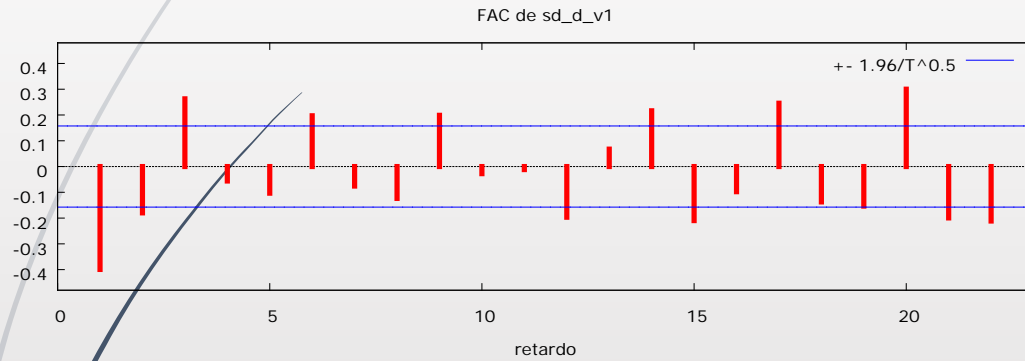
► **Criterios de selección**

- Criterios AIC, BIC o similares

Correlogramas

La FAC decae exponencialmente, mientras que el FACP presenta valores altos en los retardos 1 y 2. Probamos un modelo AR1

Aparece alto los valore alrededor del 12, lo que hace sospechar también un efecto estacional, lo incluimos en la parte MA, por ser mas fuerte el FACP



Función de autocorrelación para sd_d_v1

***, ** y * indica significatividad a los niveles del 1%, 5% y 10%

Utilizando desviación típica $1/T^{0.5}$

RETARDO	FAC		FACP		Estad-Q.	[valor p]
1	-0.3993	***	-0.3993	***	25.1939	[0.000]
2	-0.1799	**	-0.4037	***	30.3406	[0.000]
3	0.2627	***	0.0010		41.3890	[0.000]
4	-0.0561		0.0268		41.8969	[0.000]
5	-0.1034		-0.0259		43.6326	[0.000]
6	0.1971	**	0.1531	*	49.9769	[0.000]
7	-0.0758		0.0672		50.9215	[0.000]
8	-0.1237		-0.0669		53.4553	[0.000]
9	0.1988	**	0.0696		60.0422	[0.000]
10	-0.0277		0.0683		60.1714	[0.000]
11	-0.0119		0.1547	*	60.1954	[0.000]
12	-0.1967	**	-0.2753	***	66.7772	[0.000]
13	0.0676		-0.2489	***	67.5608	[0.000]
14	0.2164	***	0.1334	*	75.6402	[0.000]
15	-0.2098	***	0.0038		83.2912	[0.000]
16	-0.0977		-0.1889	**	84.9617	[0.000]
17	0.2456	***	0.0417		95.6019	[0.000]
18	-0.1374	*	0.0822		98.9562	[0.000]
19	-0.1535	*	-0.1717	**	103.1729	[0.000]
20	0.3002	***	-0.0031		119.4218	[0.000]
21	-0.1994	**	0.0701		126.6419	[0.000]
22	-0.2115	***	-0.1923	**	134.8302	[0.000]



Estimación

- ▶ Una vez eliminadas la tendencia y la estacionalidad y averiguado el orden del proceso, y la sospecha del tipo de modelo pasaríamos a estimar los coeficientes siguiendo los métodos indicados previamente.
- ▶ Una vez realizada la estimación debe hacerse la diagnosis para comprobar que el modelo esta bien identificado.
- ▶ En caso contrario debería restimarse el modelo buscando alguna otra alternativa

Estimación de impagados

Evaluaciones de la función: 19

Evaluaciones del gradiente: 8

Modelo 1: ARIMA, usando las observaciones 2003:02-2015:12 (T = 155)

Estimado usando el filtro de Kalman (MV exacta)

Variable dependiente: (1-Ls) d_v1

Desviaciones típicas basadas en el Hessiano

	Coefficiente	Desv. Típica	z	Valor p
const	-997205	750995	-1.328	0.1842
phi_1	-0.432624	0.0721370	-5.997	2.01e-09 ***
Theta_1	-0.797870	0.155596	-5.128	2.93e-07 ***
Media de la vble. dep.	-378795.5	D.T. de la vble. dep.	54082109	
media innovaciones	883603.1	D.T. innovaciones	43693901	
Log-verosimilitud	-2952.973	Criterio de Akaike	5913.945	
Criterio de Schwarz	5926.119	Crit. de Hannan-Quinn	5918.890	

- Se observa que la constante no es significativa, por lo que podría eliminarse.
- Los otros dos coeficientes son significativos.
- Realizamos la diagnosis para tomar decisiones

	Real	Imaginaria	Módulo	Frecuencia
AR				
Raíz 1	-2.3115	0.0000	2.3115	0.5000
MA (estacional)				
Raíz 1	1.2533	0.0000	1.2533	0.0000



Diagnosis

- ▶ Para evaluar la validez del modelo se hace uso de criterios gráficos y estadísticos de modo similar a como vimos anteriormente.
- ▶ Dichos criterios se aplican a los coeficientes y sobre todo a los residuos que nos dan la información que queda en el modelo
- ▶ Los coeficientes:
 - ▶ Normalmente los coeficientes de la serie deben ser menores que 1 puesto que si no el modelo no sería estacionario o no sería invertible, condiciones ambas necesarias para la estimación.
 - ▶ Por tanto debemos observar comportamientos sospechosos en los coeficiente
- ▶ Los residuos
 - ▶ Cuando el modelo está bien identificado, los residuos deberían comportarse como un ruido blanco.
 - ▶ Por tanto si los residuos se alejan de ese comportamiento es síntoma de modelo mal especificado

Diagnosis basada en los coeficientes

1. Si alguno de los coeficientes $AR(L)$ (o $AR_s(L)$) son próximos a la unidad, es síntoma de posible subdiferenciación
 2. Si alguno de los coeficientes $MA(L)$ (o $MA_s(L)$) son próximos a la unidad es posible síntoma de sobrediferenciación.
 3. Si existen coeficientes AR y MA parecidos en el mismo orden $p=q$, posiblemente se puede reducir en ambos.
 4. Si existen coeficientes AR y MA no significativos por los test de Wald, posiblemente se pueden eliminar.
- Los correlogramas y los test correspondientes al igual que los test de raíces unitarias comprueban esos efectos

Estimación de impagados

Evaluaciones de la función: 19

Evaluaciones del gradiente: 8

Modelo 1: ARIMA, usando las observaciones 2003:02-2015:12 (T = 155)

Estimado usando el filtro de Kalman (MV exacta)

Variable dependiente: (1-Ls) d_v1

Desviaciones típicas basadas en el Hessiano

	Coefficiente	Desv. Típica	z	Valor p
const	-997205	750995	-1.328	0.1842
phi_1	-0.432624	0.0721370	-5.997	2.01e-09 ***
Theta_1	-0.797870	0.155596	-5.128	2.93e-07 ***
Media de la vble. dep.	-378795.5	D.T. de la vble. dep.	54082109	
media innovaciones	883603.1	D.T. innovaciones	43693901	
Log-verosimilitud	-2952.973	Criterio de Akaike	5913.945	
Criterio de Schwarz	5926.119	Crit. de Hannan-Quinn	5918.890	

- El coeficiente del proceso MA es cercano a -1, es posible que haya sobrediferenciación
- De todas formas analizamos los residuos

	Real	Imaginaria	Módulo	Frecuencia
AR				
Raíz 1	-2.3115	0.0000	2.3115	0.5000
MA (estacional)				
Raíz 1	1.2533	0.0000	1.2533	0.0000



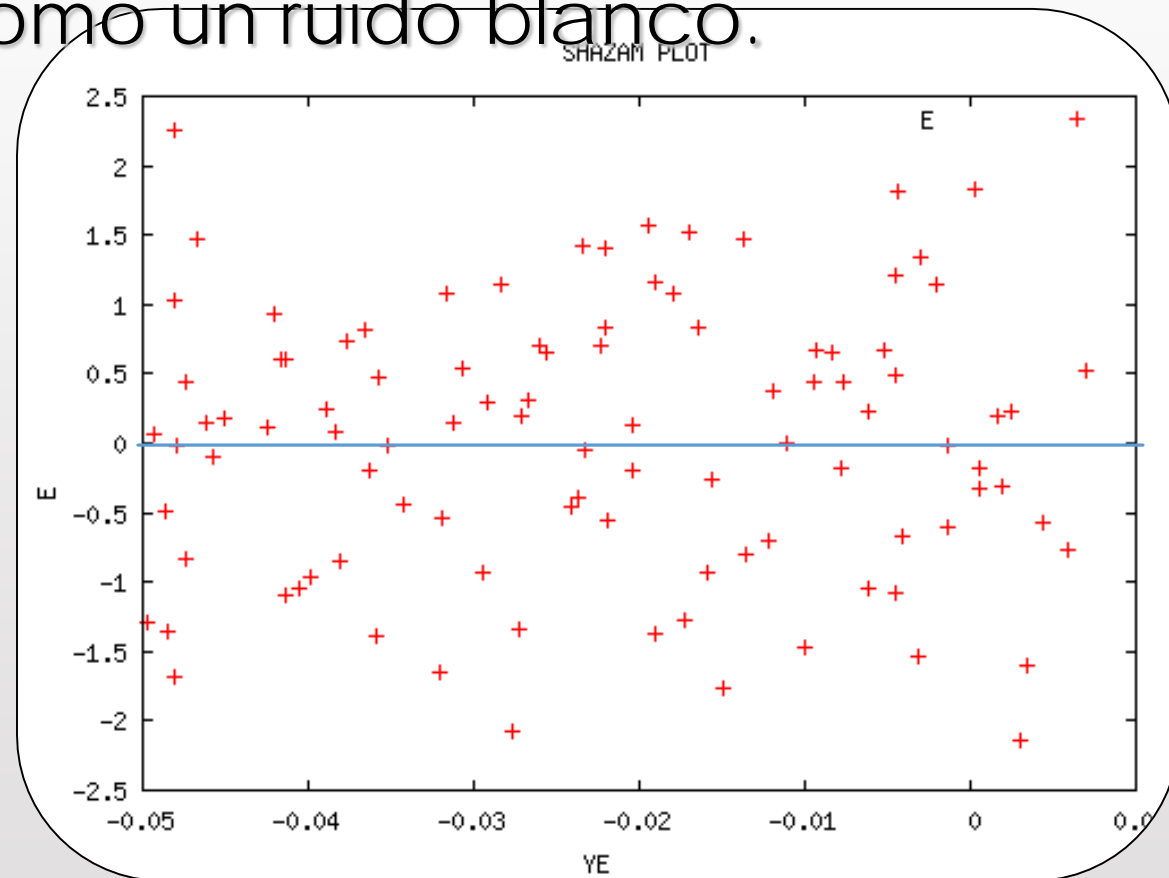
Diagnosia basada en residuos: métodos gráficos

- ▶ Representación gráfica de la serie de residuos: Consiste en dibujar la serie y hacer un examen visual, para ver si aparece un ruido blanco o presenta heterocedasticidad, autocorrelación, etc...
 - ▶ Para comprobar mejor esto se deben utilizar todos los gráficos definidos para diagnosticar el modelo lineal general, por ejemplo los Q-Q Plot para ver la normalidad, etc...
- ▶ Representación del correlograma de los residuos: ver si todas las autocorrelaciones están dentro de las bandas.
- ▶ Representación del correlograma parcial de los residuos: ver si todas las autocorrelaciones parciales están dentro de las bandas.

Ruido Blanco

El modelo bien especificado tiene residuos que se comportan como un ruido blanco.

- ▶ Sucesión de variables aleatorias independientes:
 - ▶ Media cero
 - ▶ Varianza constante
 - ▶ Simétrica
 - ▶ independencia





Comportamiento de los residuos

- Las perturbaciones del MRLN deberían comportarse como un ruido blanco, puesto que son independientes e igualmente distribuidas con media 0 y simétricas.
- Los residuos de la regresión son aproximaciones a las perturbaciones y aunque no son independientes ni con varianza constante son bastante cercanos a ese comportamiento, por lo que suponemos que bajo todas las suposiciones del modelo deberían comportarse casi como ruido blanco.
- En todos los gráficos donde se representen los residuos respecto a alguna variable, estos deben verificar que bajo las suposiciones del MRLN debería ser cercanos a un ruido blanco.

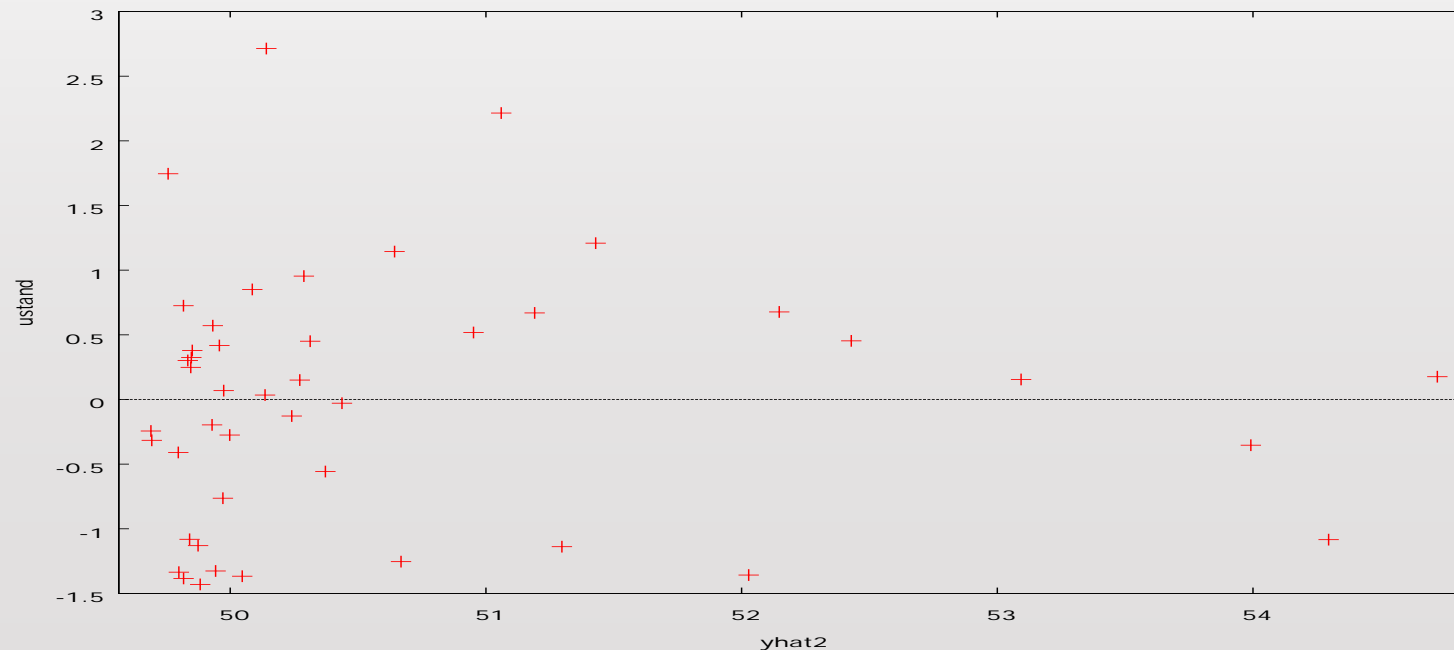
Gráficos de residuos

- ▶ Por tanto, los residuos debe verificar:
 - ▶ $E(e)=0$;
 - ▶ los residuos deben oscilar alrededor del 0, no deben presentar ninguna forma funcional
 - ▶ $Var(e)=$ constante aproximadamente.
 - ▶ Los residuos deben dispersarse de forma homogénea, estar comprendidos entre dos bandas
 - ▶ Independencia:
 - ▶ Los residuos no deben presentar relaciones entre ellos, formas graficas encadenadas.
 - ▶ Normalidad
 - ▶ Los residuos deben estar cercanos al 0, simétricos a cada lado y las bandas no muy alejadas del 0 (alrededor de dos veces la desviación estándar).
- ▶ Eso significa que cualquier forma en que se aleje de esas características existe un fallo de alguna suposición.
- ▶ El residuo se introduce en el eje de ordenadas. La variable introducida en el eje de abscisas nos dice algo sobre ese fallo.

Gráfico resumen de calidad del modelo

- Se obtiene representando residuos respecto a valores estimados de la dependiente. Por tanto, en GRETL, se deben guardar los valores estimados y los residuos y luego ver el grafico de ambos

Residuos respecto a valores estimados



Conceptos asociados al gráfico

- ▶ Este gráfico representa los residuos frente a los valores estimados de la variable dependiente.
- ▶ Los residuos deben verificar las condiciones del ruido blanco.
- ▶ La variable independiente en este caso es el valor estimado de la dependiente, por tanto recoge los posibles efectos de esta variable sobre el comportamiento de los residuos.
 - ▶ Esa variable contiene información sobre:
 - ▶ La variable dependiente, puesto que es la mejor aproximación lineal en función de las independientes
 - ▶ La buena estructura del modelo, puesto que este interviene en la estimación.
 - ▶ Eso significa que nos da información sobre como influyen tanto la variable dependiente como las independientes sobre los residuos. En ese sentido se habla de resumen de la calidad del modelo puesto que integra todos los componentes de este.



Interpretación del gráfico

Este gráfico permite observar una serie de suposiciones del modelo, puesto que deben manifestar el ruido blanco de los residuos y la independencia con las variables

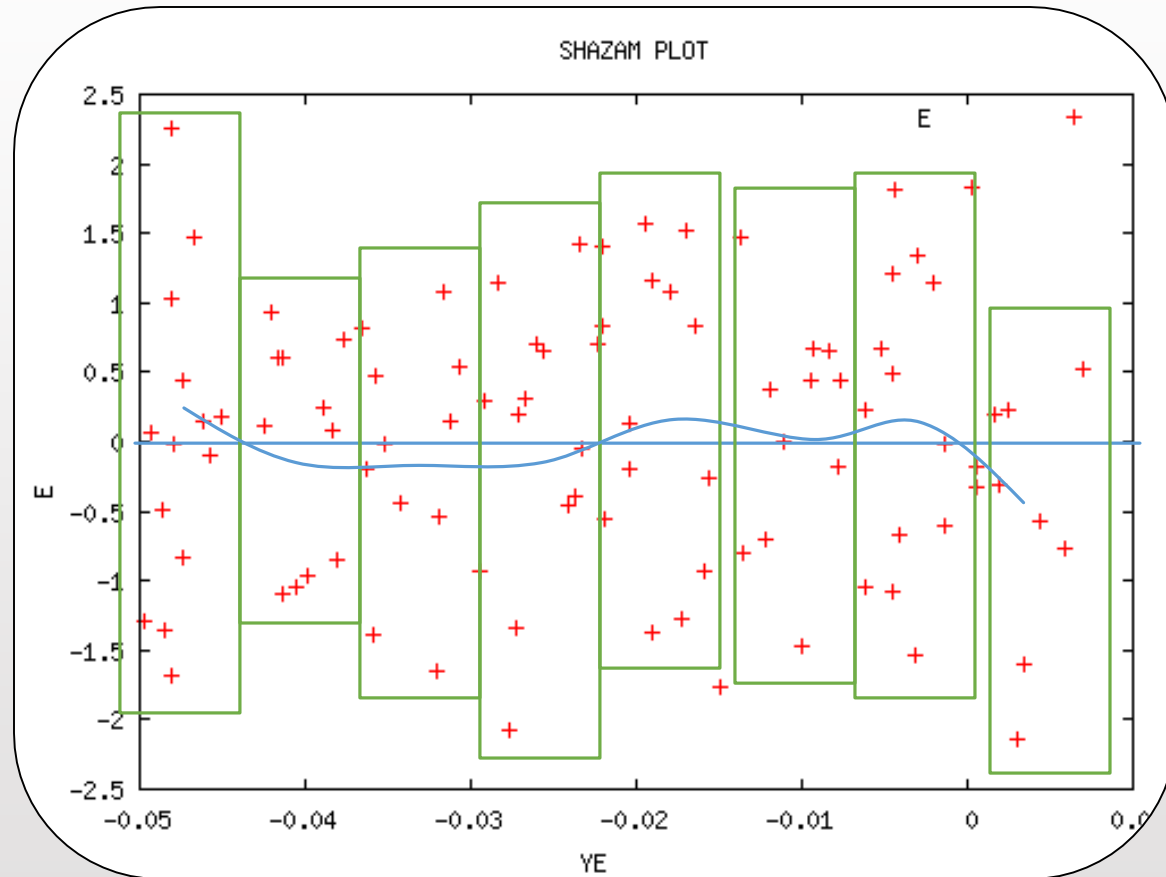
Se observa si existe:

- ▶ Linealidad,
- ▶ Normalidad
- ▶ Homocedasticidad
- ▶ Estabilidad
- ▶ Independencia

No se observan el resto de las suposiciones, puesto que las suposiciones sobre las variables independientes no se relacionan con los residuos

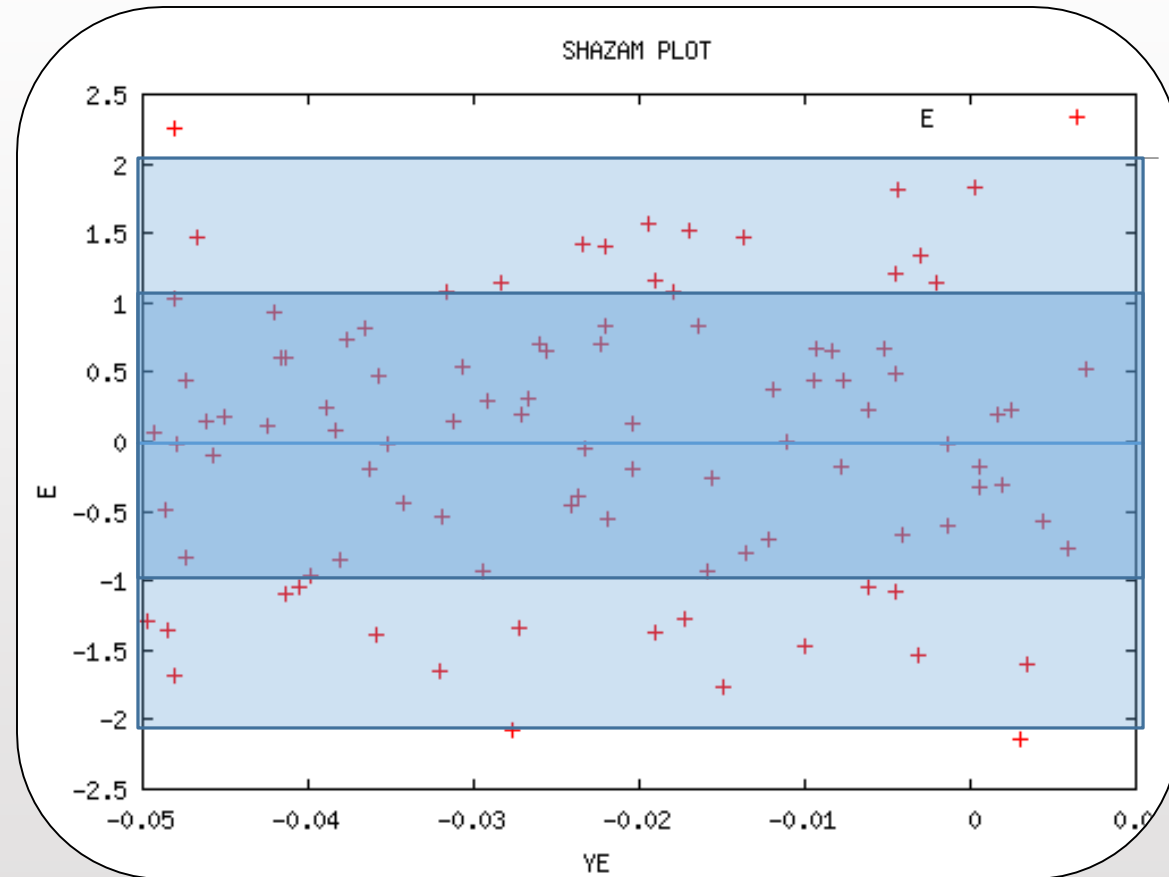
Linealidad,

- Se trazan rodajas verticales y se unen sus centros para ver la forma funcional empírica
- no se debe observar ninguna forma funcional.
- La media es constante e igual a cero en todos los casos



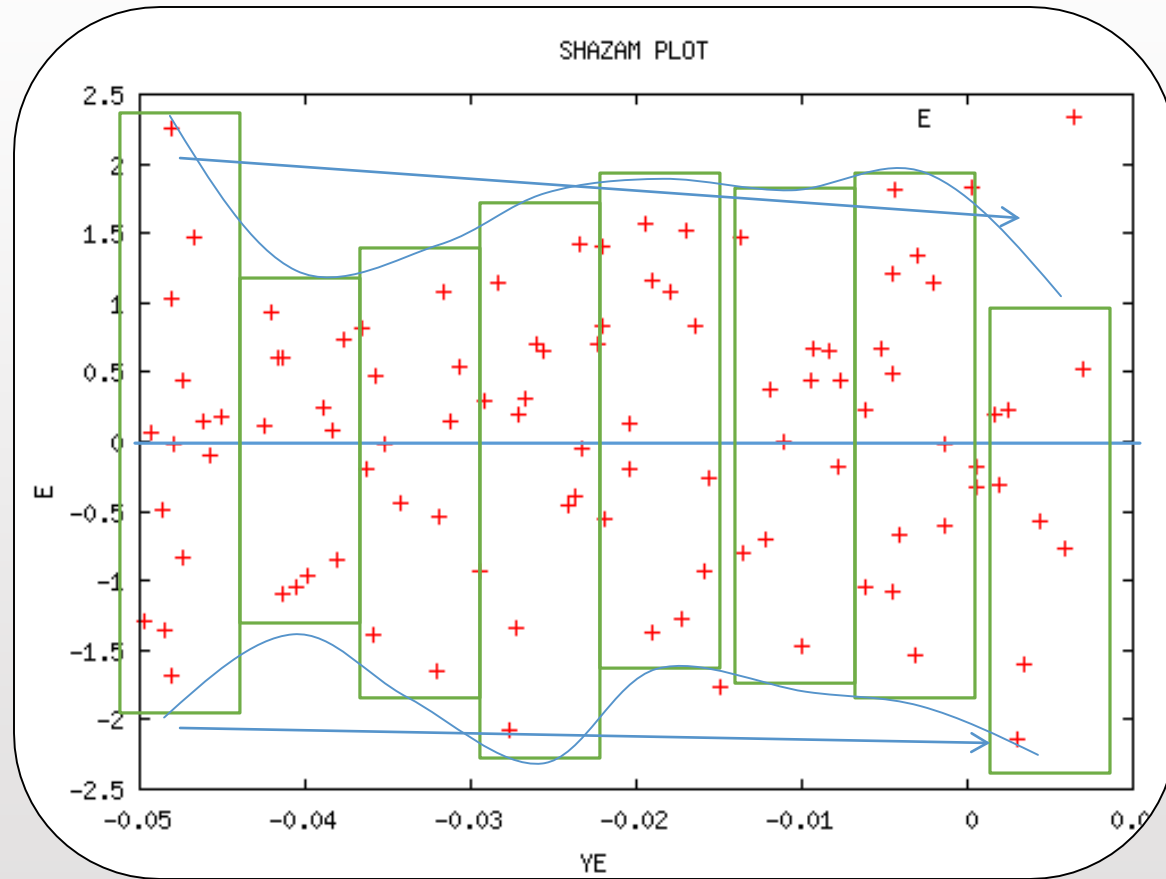
Normalidad

- Los datos se distribuyen de forma simétrica y más concentrados alrededor del 0.
- Se traza una banda central y una banda ampliada
- En la banda entre $-s$ y $+s$ deben estar la mayor parte de los valores (un 75%) y entre $-2s$ y $+2s$ el resto. Los que están fuera son posibles valores raros



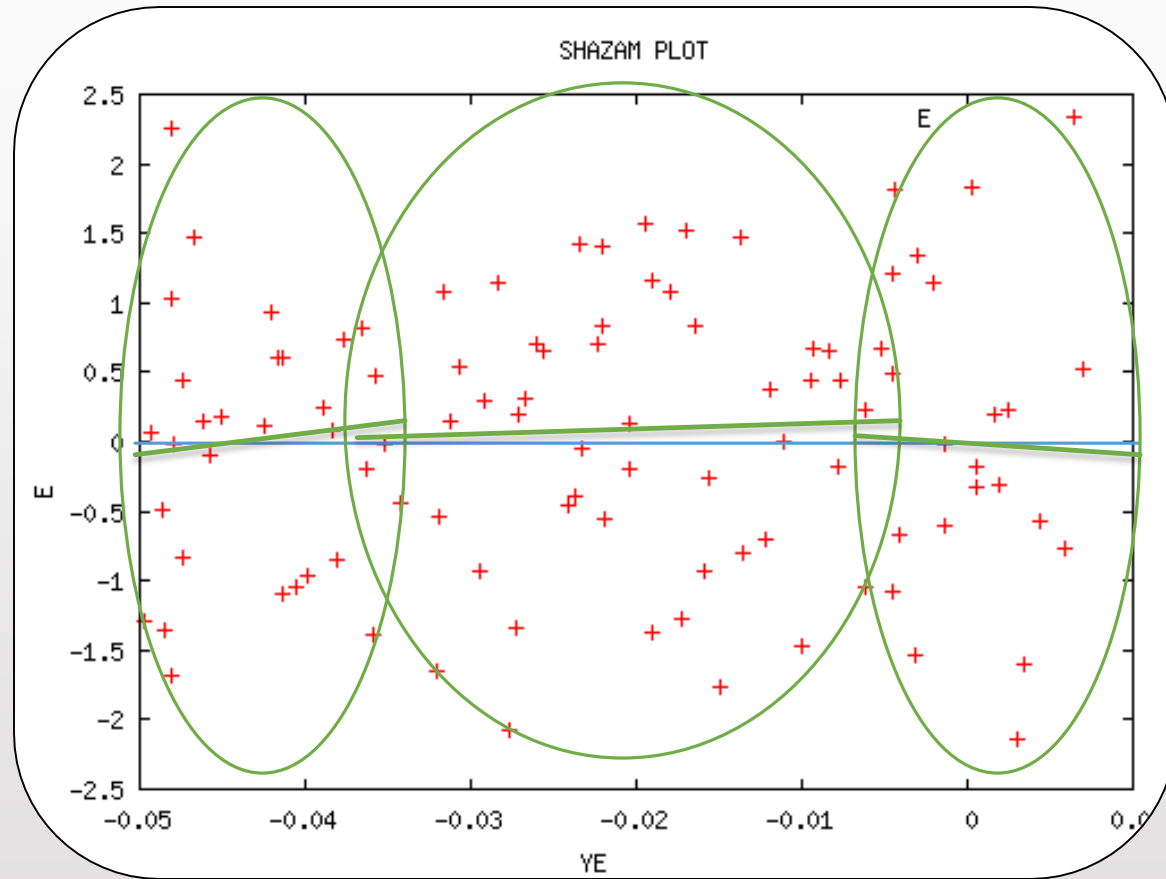
Homocedasticidad

- Se trazan rodajas verticales y se unen sus extremos para ver la variación de la amplitud de esas rodajas
- La varianza es constante, dispersión igual para todos la serie.
- Las amplitudes de ambos lados deben ser mas o menos constantes



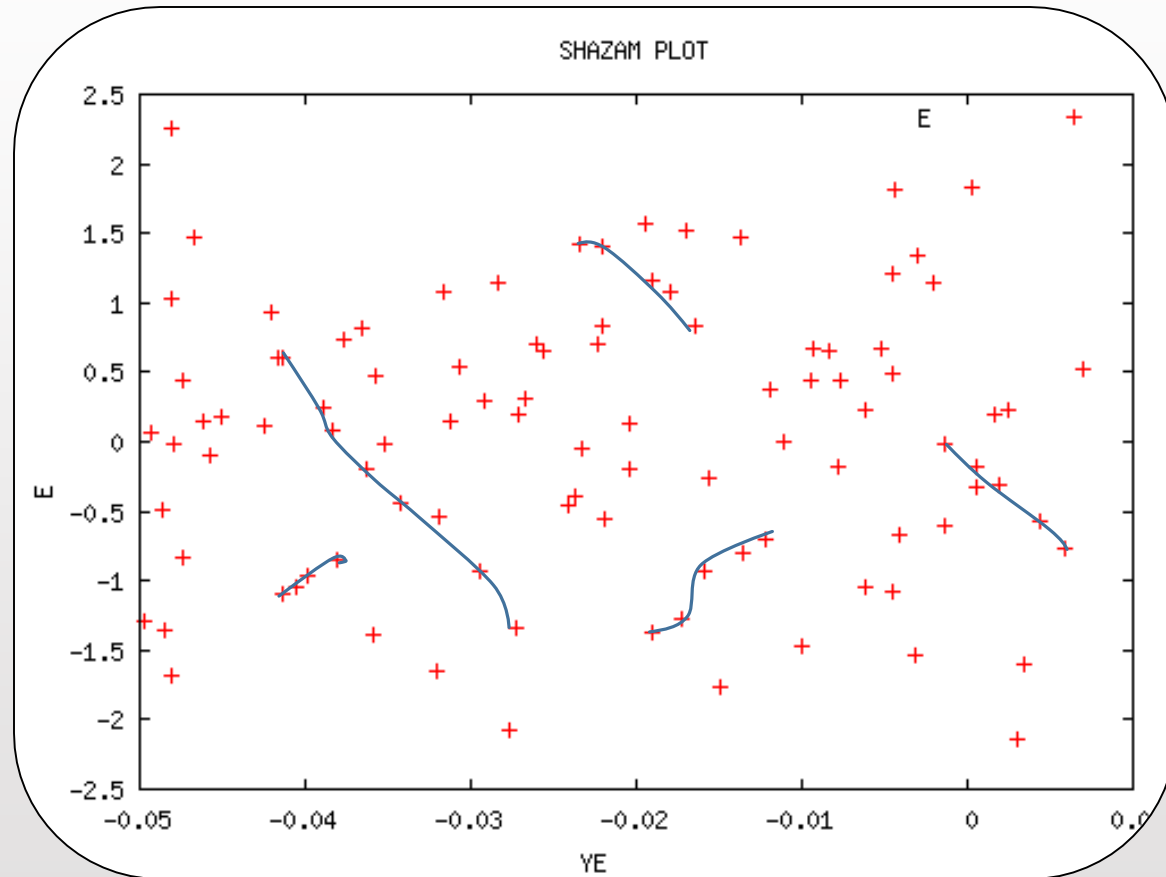
Estabilidad

- ▶ No se producen cambios repentinos de la serie de residuos
- ▶ Se buscan las nubes de puntos existentes, dibujando los cortes que se intuyan
- ▶ Si aparece más de una con diferente tendencia, indica fallo de estabilidad
- ▶ Si hay varias pero la tendencia de las nubes de puntos no varía, hay estabilidad

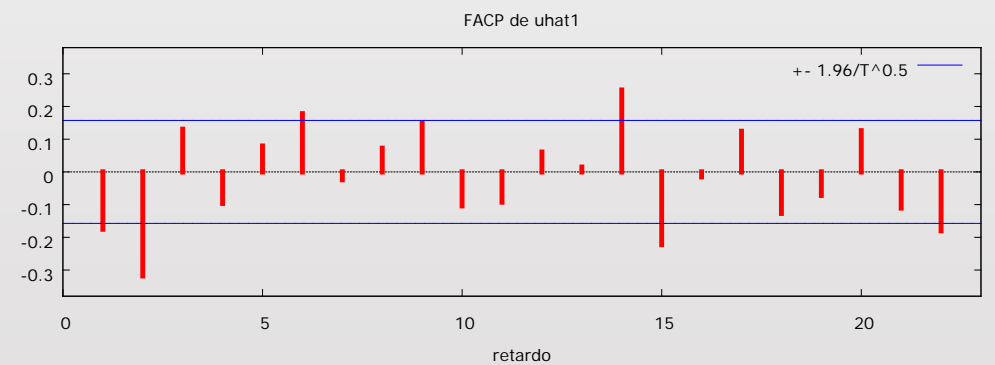
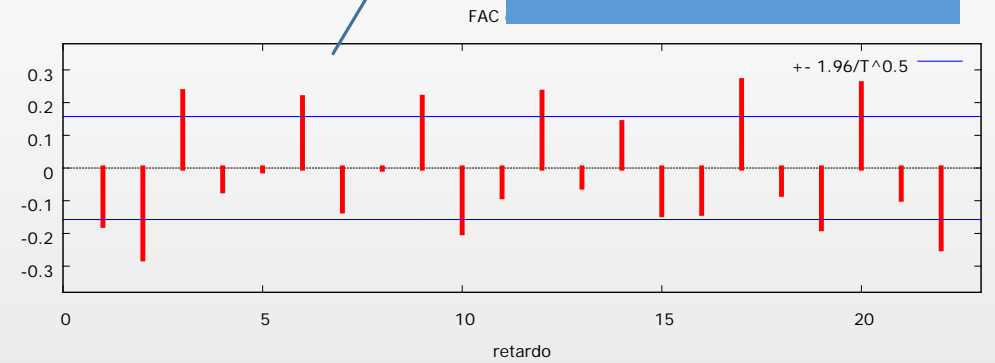
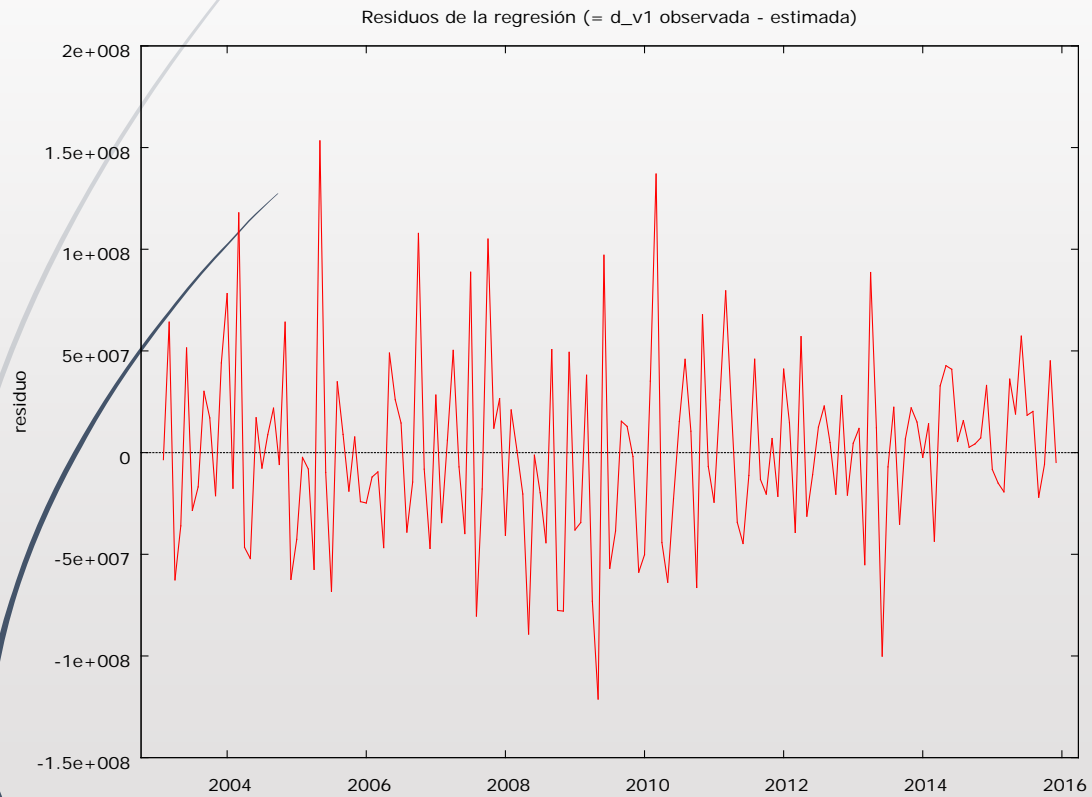


Independencia

- No se observan relaciones entre residuos o grupos de residuos en la evolución de los datos
- Se espera un comportamiento aleatorio, la forma en que quedarían unas monedas al tirarlas al azar al aire.
- Por consiguiente no deben verse "gusanitos" o grandes "huecos"



Residuos de impagados: Gráficos



Aparecen efectos de orden trimestral inducidos, por lo que la estacionalidad incluida no esta clara



Diagnosia basada en residuos: métodos estadísticos

- ▶ Se utiliza los mismos test definidos previamente, pero como ahora los residuos deben ser ruido blanco, todos ellos deben ser no significativos.
 - ▶ Si alguno de los tests es significativo se introduce ese parámetro en el modelo, MA si es ACF o AR si es PACF o en general aparece en ambos, y se calculan los nuevos residuos.
- ▶ Para comprobar la normalidad de las perturbaciones y la homocedasticidad, se utilizan los mismos test definidos en el caso del modelo lineal general.
 - ▶ La corrección en este caso se hace de modo similar a como se hacia en el e modelo de regresión lineal
- ▶ El proceso termina cuando los residuos son ruido blanco.

Residuos de impagados: test

Función de autocorrelación de los residuos

***, ** y * indica significatividad a los niveles del 1%, 5% y 10%

Utilizando desviación típica $1/T^{0.5}$

RETARDO	FAC		FACP		Estad-Q. [valor p]
1	-0.1750	**	-0.1750	**	
2	-0.2775	***	-0.3179	***	
3	0.2333	***	0.1304		25.8051 [0.000]
4	-0.0694		-0.0961		26.5809 [0.000]
5	-0.0086		0.0790		26.5929 [0.000]
6	0.2149	***	0.1779	**	34.1361 [0.000]
7	-0.1311		-0.0238		36.9617 [0.000]
8	-0.0032		0.0723		36.9634 [0.000]
9	0.2161	***	0.1495	*	44.7470 [0.000]
10	-0.1975	**	-0.1037		51.2904 [0.000]
11	-0.0874		-0.0923		52.5822 [0.000]
12	0.2316	***	0.0606		61.7079 [0.000]
13	-0.0581		0.0146		62.2865 [0.000]
14	0.1388	*	0.2503	***	65.6101 [0.000]
15	-0.1421	*	-0.2222	***	69.1198 [0.000]
16	-0.1383	*	-0.0149		72.4666 [0.000]
17	0.2669	***	0.1242		85.0264 [0.000]
18	-0.0801		-0.1267		86.1653 [0.000]
19	-0.1852	**	-0.0717		92.3022 [0.000]
20	0.2576	***	0.1259		104.2619 [0.000]
21	-0.0951		-0.1103		105.9041 [0.000]
22	-0.2466	***	-0.1798	**	117.0347 [0.000]

Contraste de normalidad de uhat1:

Contraste de Doornik-Hansen = 6.3894, con valor p 0.0409788

W de Shapiro-Wilk = 0.984289, con valor p 0.075817

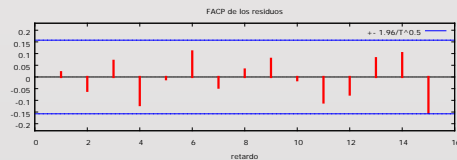
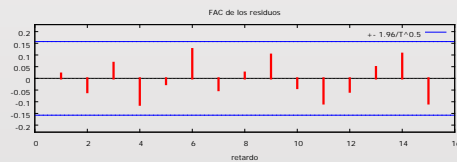
Contraste de Lilliefors = 0.0627011, con valor p ≈ 0.14

Contraste de Jarque-Bera = 9.07672, con valor p 0.010691

Fallan autocorrelacion y normalidad, no es buen modelo

Re-estimación del modelo

- Observando la varianza nos sugiere trabajar con el logaritmo
- Aplicamos las diferencias al logaritmo
- Siguiendo el proceso anterior, llegamos a un modelo $ARIMA(0,1,3)(1,1,0)_{12}$
- Que mas o menos se ajusta



Evaluaciones de la función: 62
Evaluaciones del gradiente: 16

Modelo 7: ARMA, usando las observaciones 2003:02-2015:12 (T = 155)
Estimado usando el filtro de Kalman (MV exacta)
Variable dependiente: sd_ld_v1
Desviaciones típicas basadas en el Hessiano

	Coefficiente	Desv. Típica	z	Valor p	
const	-0.00112460	0.00360754	-0.3117	0.7552	
Phi_1	-0.271865	0.0791067	-3.437	0.0006	***
theta_1	-0.581901	0.0724263	-8.034	9.40e-016	***
theta_3	0.171192	0.0604346	2.833	0.0046	***

Media de la vble. dep.	-0.001145	D.T. de la vble. dep.	0.115317
media innovaciones	0.000844	D.T. innovaciones	0.095212
Log-verosimilitud	143.8792	Criterio de Akaike	-277.7585
Criterio de Schwarz	-262.5413	Crit. de Hannan-Quinn	-271.5776

	Real	Imaginaria	Módulo	Frecuencia
AR (estacional)				
Raíz 1	-3.6783	0.0000	3.6783	0.5000
MA				
Raíz 1	1.2063	0.9829	1.5560	0.1088
Raíz 2	1.2063	-0.9829	1.5560	-0.1088
Raíz 3	-2.4125	0.0000	2.4125	0.5000



Limitaciones

- ▶ El manejo de la estacionariedad de orden s en el modelo puede afectar a la estructura estocástica del proceso, produciendo parámetros ficticios.
- ▶ Las tendencias no polinómicas son aproximadas, y por tanto la diferenciación es imprecisa, pasando parte del efecto de la tendencia al error.
- ▶ La aproximación ARMA de las perturbaciones tampoco es exacta y esto puede afectar a las estimaciones de los parámetros.



Predicción en series temporales

Un análisis de los modelos estocásticos de predicción



Introducción

- Existen muchas variables económicas cuyo comportamiento a lo largo del tiempo no es estable, en el sentido de que sus valores presentan una tendencia, o ciclos, etc... de forma que no se puede suponer que es una muestra de una misma variable. Eso ocurre, por ejemplo, con la renta, el IPC, ... Para tratar de modelar estos comportamientos económicos vamos a definir un nuevo concepto estadístico:



Predicción



- ▶ Denominamos predicción a la estimación de valores futuros de la variable en función del comportamiento pasado de la serie.
- ▶ la predicción mediante modelos basados en la teoría de series temporales, puede servir para una buena planeación de recursos sanitarios, en función de la demanda que se espera en el futuro, prevista por el modelo

Predicción econométrica

- ▶ Predecir significa obtener el valor mas probable de una o mas variables basándonos en la información del pasado
- ▶ $\text{Pred}(Y_{T+k}) = \text{Est}(Y_{T+k} / \text{Informacion}_T)$

$$\hat{Y}_{T+k/T} = \hat{Y}_{T,k} = E\left(\frac{Y_{T+k}}{\mathfrak{I}_T}\right)$$

- ▶ Predecir exige suposiciones sobre la evolución de la variable y sobre la información disponible en el pasado



Suposiciones de la Predicción

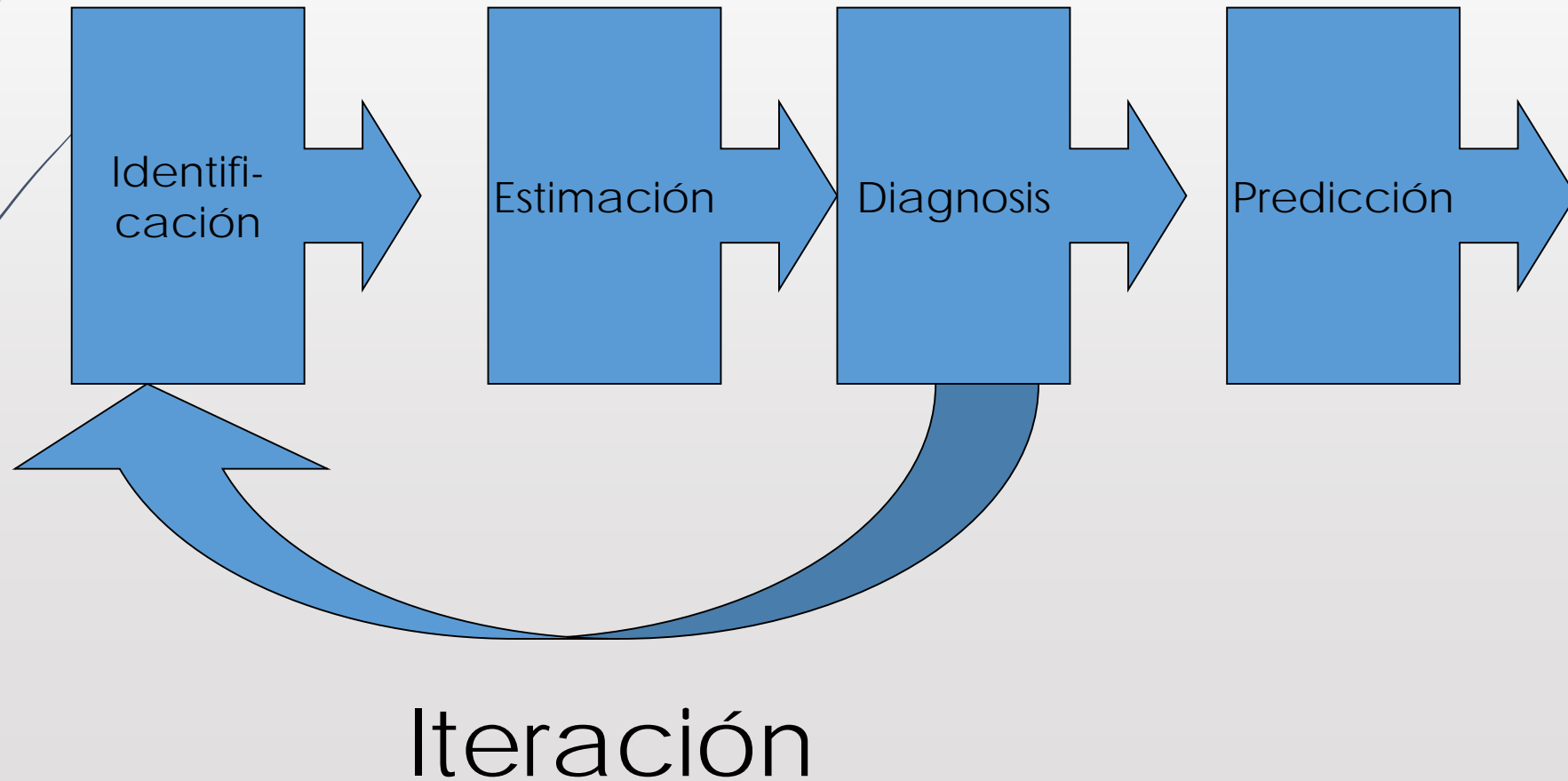
- ▶ La primera suposición que se hace siempre es que la variable es un proceso estocástico, es decir una sucesión de variables aleatorias en el tiempo
- ▶ La segunda es que ese proceso es lineal, es decir, la información del pasado afecta de forma lineal a la información actual
- ▶ La tercer se refiere a la fuente de información relevante, que es de dos tipos:
 - ▶ Contenida e la propia variable
 - ▶ Contenida en otras variables



Tipos de predicción

- ▶ Cuando la información se contiene en la misma variable se habla de **proceso estocástico lineal**
- ▶ Cuando la información se halla contenida en otras variables se habla de **regresión multivariante**
- ▶ Cuando se combinan ambos se habla de **regresión dinámica multivariante**
- ▶ Centraremos este apartado en el primero, que se basa en la metodología Box-Kenkins. Introduciremos los otros métodos en cuanto sea posible

Proceso de la metodología Box-Jenkins



PREDICCIÓN CON MODELOS ARMA

- Sea una variable que sigue un proceso ARMA(p,q)

$$x_t - \phi_1 x_{t-1} - \dots - \phi_p x_{t-p} = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}$$

- Si queremos predecir el valor x_{T+h} h-veces adelante, denominaremos a T origen de la predicción, a h tiempo de adelanto y a la predicción predicción h pasos adelante en T. El valor de la predicción vendrá dado por

$$x_{T,h} = E(x_{T+h} / x_T, x_{T-1}, \dots)$$

- Tomando

$$\alpha_0(L) = -(1-L)^d \phi(L) + 1 = \sum_{j=1}^{p+d} \alpha_j L^j$$

Información del pasado debido a la propia variable

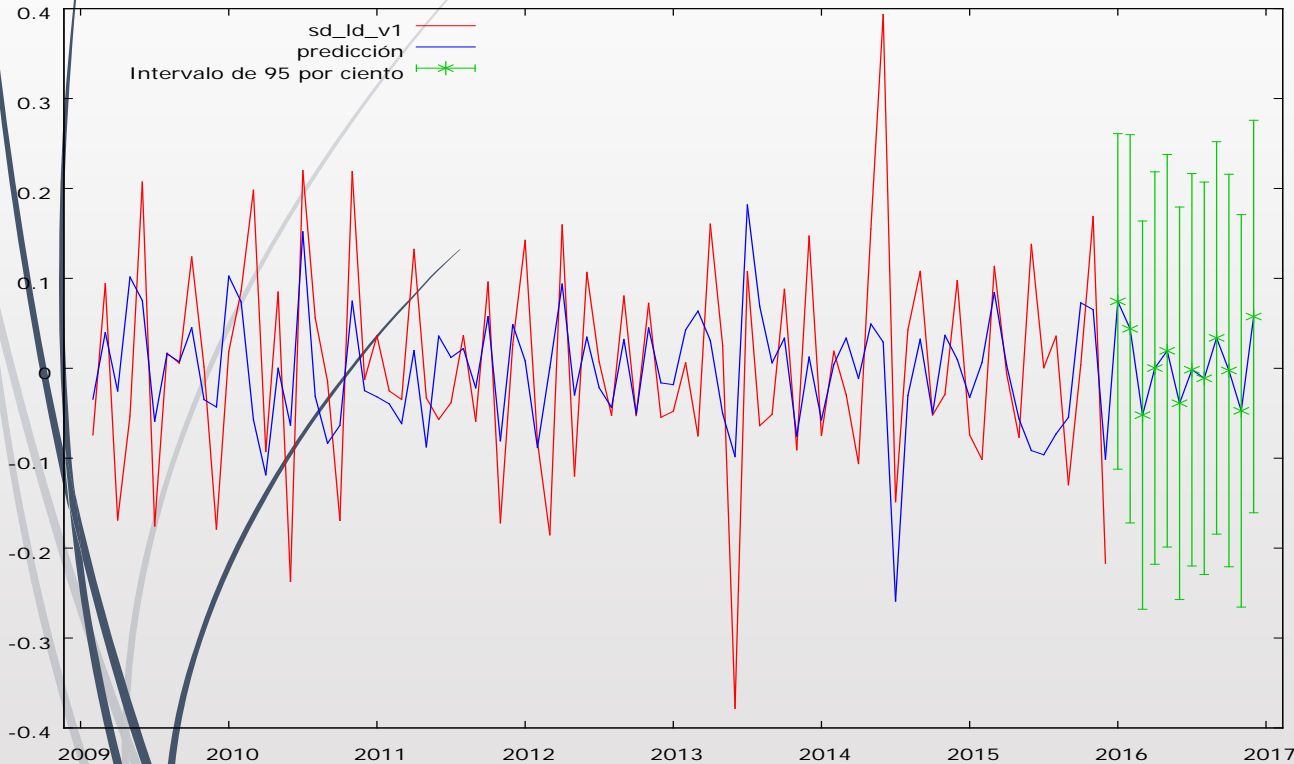
- La predicción vendrá dada por

$$x_{T,h} = \sum_{i=0}^{h-1} \alpha_{h-i} x_{T,i} + \sum_{i=0}^{p+d-h} \alpha_{h+i} x_{T-i} + \sum_{j=0}^{q-h} \theta_{j+h} \varepsilon_{T-j}$$

Información del pasado debido a los errores

Información debido a las predicciones

Predicción de la variación interanual de la tasa de crecimiento de impagados



Para intervalos de confianza 95%, $z(0.025) = 1.96$

	sd_ld_v1	predicción	Desv. Típica	Intervalo de
confianza 95%				
2015:04	-0.006305	0.003375		
2015:05	-0.076895	-0.056351		
2015:06	0.137867	-0.091502		
2015:07	0.000633	-0.096205		
2015:08	0.035710	-0.072937		
2015:09	-0.129572	-0.054703		
2015:10	0.003820	0.072856		
2015:11	0.168864	0.065219		
2015:12	-0.217016	-0.101082		
2016:01		0.074359	0.095212	-0.112254 - 0.260972
2016:02		0.043879	0.110159	-0.172029 - 0.259787
2016:03		-0.052090	0.110159	-0.267998 - 0.163818
2016:04		0.000284	0.111359	-0.217975 - 0.218542
2016:05		0.019475	0.111359	-0.198784 - 0.237733
2016:06		-0.038912	0.111359	-0.257170 - 0.179347
2016:07		-0.001602	0.111359	-0.219861 - 0.216656
2016:08		-0.011139	0.111359	-0.229397 - 0.207120
2016:09		0.033796	0.111359	-0.184463 - 0.252055
2016:10		-0.002469	0.111359	-0.220728 - 0.215790
2016:11		-0.047339	0.111359	-0.265597 - 0.170920
2016:12		0.057569	0.111359	-0.160690 - 0.275828

Actualización de las predicciones

- ▶ Si se incluye un nuevo valor x_{T+1} , la actualización es inmediata a partir de la ecuación anterior

$$x_{T+1,h-1} = x_{T,h} + \psi_{h-1}(x_{T+1} - x_{T,1})$$

Una parte es debida a la predicción anterior y otra a la corrección del valor predicho en el momento $T+1$ por el valor observado.

- ▶ Conclusiones

1. La parte MA de un proceso desaparece cuando $h > q$, luego a largo plazo no hay efecto de los errores.
- 2 A largo plazo el efecto estacionario de la parte AR va disminuyendo hasta desaparecer totalmente

Predicción en modelos ARIMA

- ▶ Sea una variable que sigue un proceso ARIMA(p,d,q)
- ▶ Si descomponemos el polinomio $\alpha(L)$ en su parte estacionaria y su parte tendencial tendremos que:

$$\alpha(L) = (1 - L)^d \Phi(L)$$

- ▶ Por tanto la predicción vendrá dada por

$$\Delta^d x_{T,h} = \sum_{i=0}^{h-1} \alpha_{h-i} \Delta^d x_{T,i} + \sum_{i=0}^{p+d-h} \alpha_{h+i} \Delta^d x_{t-i} + \sum_{j=0}^{q-h} \theta_{j+h} \varepsilon_{T-j}$$

- ▶ De las conclusiones anteriores, la predicción a largo plazo depende solo de la tendencia

Predicción en modelos ARIMA estacionales

- ▶ La introducción de partes estacionales afectan a la predicción de modo semejante a los otros términos ARIMA, por tanto las predicciones tendrán en cuenta esos términos introduciendo los polinomios correspondientes
- ▶ Sin embargo, es de interés estudiar el efecto a largo plazo, para lo cual, sabemos que la parte Ma desaparece para $h > \max(q, Qs)$ y la parte AR verificará que

$$\Phi_P(L^S)\phi_p(L) = \prod_{i=1}^P(1 - g_{is}L^S) \prod_{j=1}^p(1 - g_jL)$$

- ▶ y como $|g_{is}| < 1$ y $|g_j| < 1$, al tender h a infinito r_{k+h} tiende a 0, y sólo aquellas raíces estacionales o tendenciales $g_{is}=1$ o $g_{j=1}$ permanecen.

Consecuencia de la predicción de modelos ARIMA estacionales

1. En el proceso ARMA la parte MA solo influyen en los primeros $\max\{q, Q_s\}$ valores, luego el comportamiento depende solo de la parte AR, por tanto tenderá a decrecer linealmente.
2. Si existe alguna diferenciación podemos suponer que alguna de las $g_i = 1$ y si es estacional, alguna de las $g_{is} = 1$ y se ve que entonces $r_k = 1$ todos (en el primer caso) o $r_{ks} = 1$, luego es de orden s en el 2º, por tanto el decrecimiento es casi nulo.
3. En el PACF del modelo $MA(q, Q_s)$ ocurre algo similar a las PACF del AR y en el ARMA se superponen los efectos, pero tiende a decaer linealmente el ACF, y el PACF depende del comportamiento MA.
4. Si las raíces MA son cercanas a 1 -procesos no invertibles o sobrediferenciados- entonces el comportamiento es simétrico ACF dependiendo del AR y PACF decreciendo linealmente.

Varianza de las predicciones

- ▶ Si tenemos un modelo ARMA y lo representamos en forma compacta como un proceso MA infinito, nos quedaría

$$x_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \dots + \psi_q \varepsilon_{t-q} + \dots$$

- ▶ Por tanto, la predicción se escribiría como

$$\hat{x}_{T,k} = \varepsilon_{T+k} + \psi_1 \varepsilon_{T+k-1} + \dots$$

- ▶ Por tanto el error de predicción sería

$$e_{T,k} = x_{T+k} - \hat{x}_{t,k} = \varepsilon_{T+k} + \psi_1 \varepsilon_{T+k-1} + \dots + \psi_{k-1} \varepsilon_{T+1}$$

- ▶ Por tanto, la varianza será

$$\text{Var}(e_{T,k}) = \sigma^2 (1 + \psi_1^2 + \dots + \psi_{k-1}^2)$$

- ▶ La varianza crece al incrementarse el horizonte de predicción

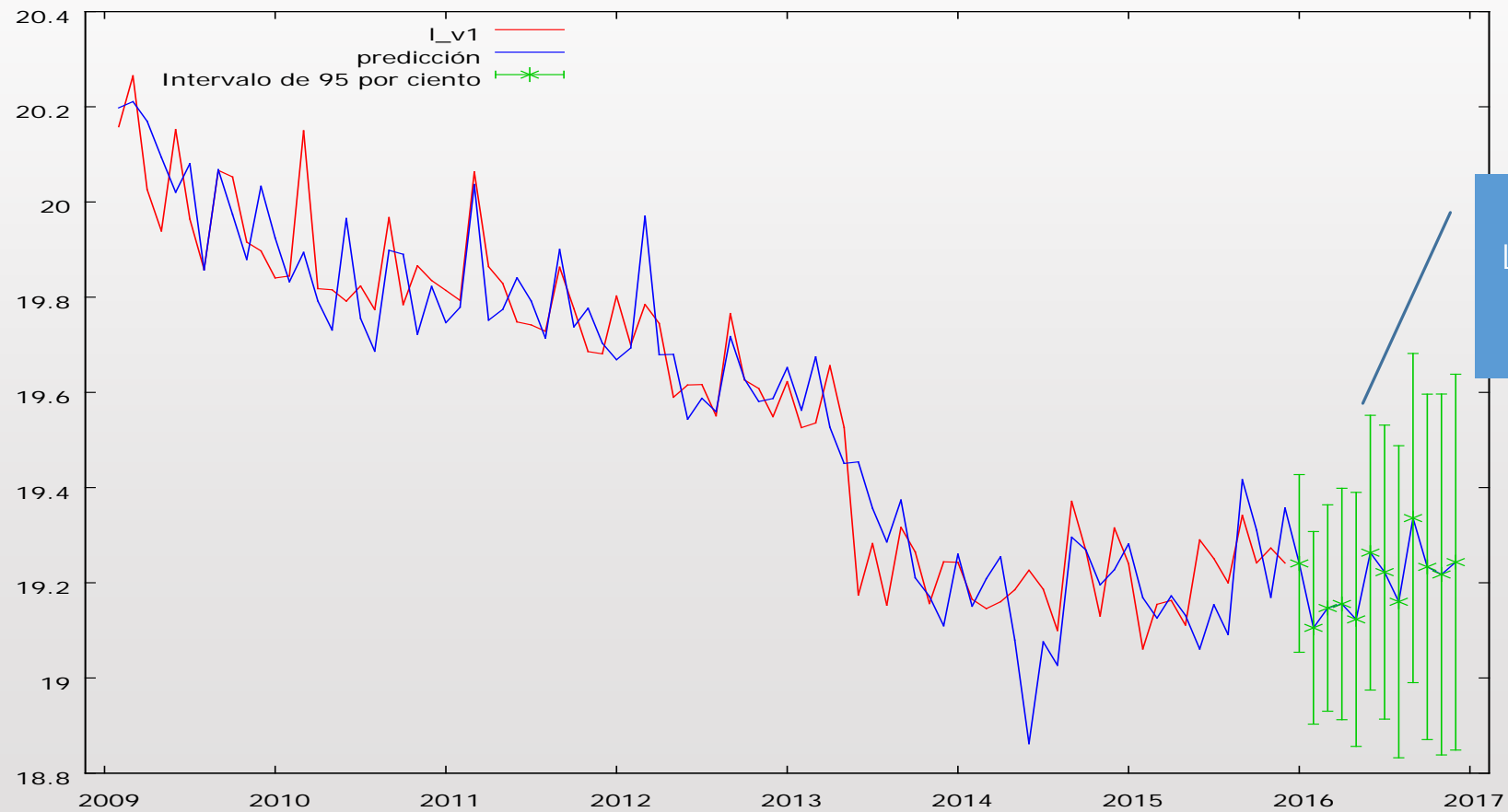
Varianza en modelos ARIMA Estacionales

- La varianza de la predicción se construye de modo similar pero ahora interviene mas factores a la hora de definir el polinomio $\Psi(L)$

$$\Psi(L) = \frac{\Theta(L)\Theta_s(L)}{(1-L)^d(1-L^s)^D\Phi(L)\Phi_s(L)}$$

- Una vez definida de esa forma la varianza seria la misma, pero los coeficientes Ψ serian diferentes

Predicción del logaritmo de impagados



La varianza se incrementa

A decorative graphic on the left side of the slide. It features a solid blue arrow pointing to the right, positioned horizontally. Behind the arrow and extending upwards and to the right are several thin, curved lines in shades of blue and grey, creating a sense of movement and depth.

Análisis de intervención en modelos ARIMA

El concepto de intervención

- ▶ Sea una variable dicotómica que toma dos valores 0 y 1, diferentes según estén en el valor t_0 o no
- ▶ A esa variable se le denomina impulso en el instante t_0

$$A = \begin{cases} 1 & \text{si } t = t_0 \\ 0 & \text{si } t \neq t_0 \end{cases}$$

- ▶ la variable impulso representa una intervención puntual en el instante t
- ▶ *Ejemplo 1*
 - ▶ *En una serie de producción conocemos que se ha producido una huelga en un momento t . Se puede incluir ese efecto a través de una variable ficticia como la anterior*

El concepto de intervención

- Sea una variable dicotómica que toma dos valores 0 y 1 diferentes a partir de valor t_0
- A esa variable se le denomina escalón en el instante t_0

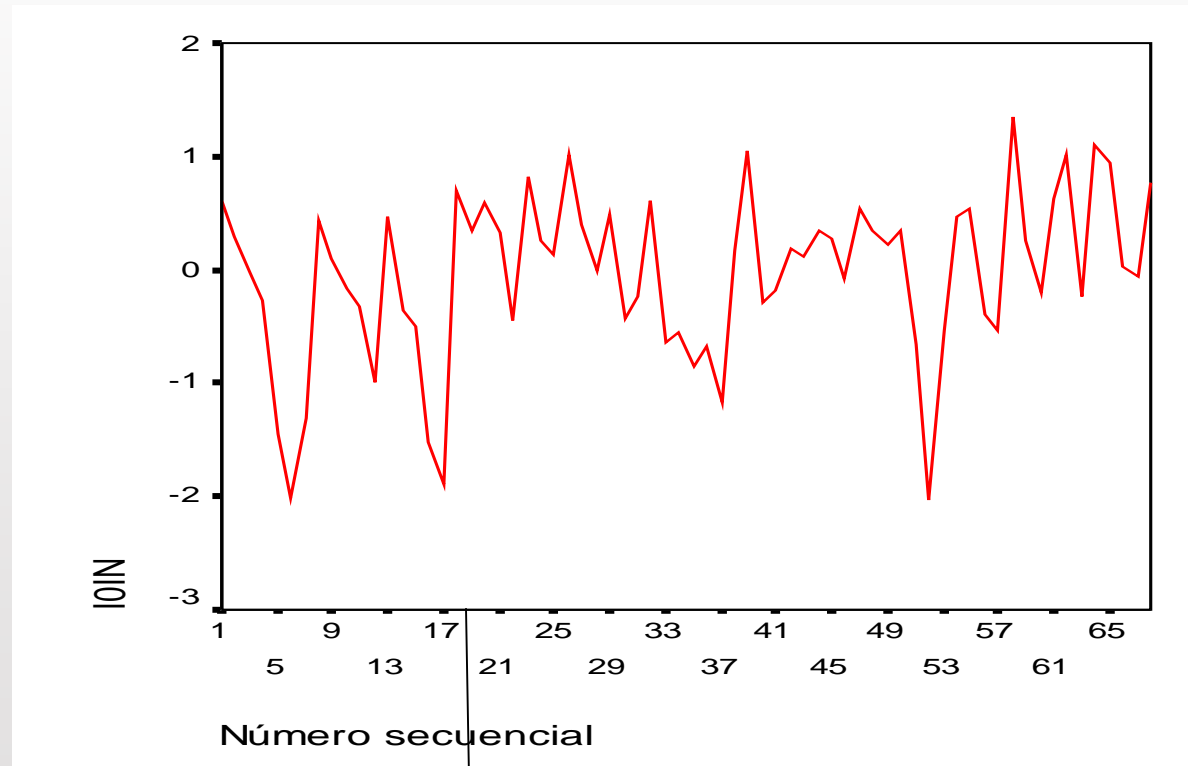
$$A_{t_0} = \begin{cases} 1 & \text{si } t \geq t_0 \\ 0 & \text{si } t < t_0 \end{cases}$$

- la variable escalón representa una intervención puntual en el instante t_0 que se mantiene en el tiempo
- *Ejemplo*
 - *Se produce un cambio legal que modifica la definición de desempleo a partir de un determinado momento*

Efecto de la intervención en un modelo ARIMA

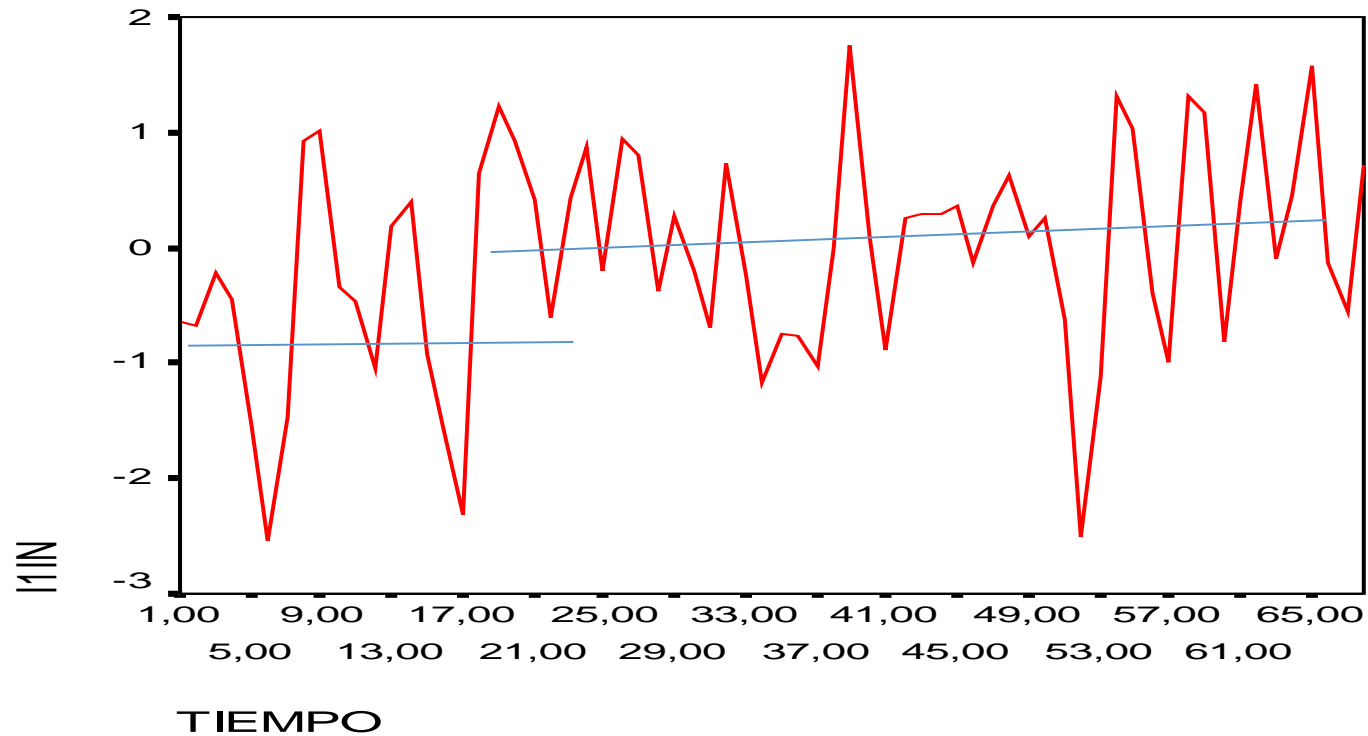
- ▶ Supongamos una variable cuyo comportamiento es $y_t = \Psi(L)\varepsilon_t$ donde y sigue un proceso ARIMA
- ▶ Sea $I_{t,h}$ una intervención en el momento h , por tanto, la nueva variable $Z_t = W_0 I_t^h + y_t = W_0 I_t^h + \Psi(L)\varepsilon_t$
- ▶ La parte ARI del modelo genera que el efecto de la intervención se continúe en el futuro, aunque si es estacionaria ese efecto tiende a desaparecer. La parte MA afecta indirectamente a la intervención en los errores que genera

Efecto de un impulso en una serie I(0)



Intervención en el periodo 18. Casi no afecta.

Efecto de una impulso en una serie I(1)



Intervención en el periodo 18. A partir de ese momento se mantiene un nivel superior



Detección de la intervención en una serie temporal

- ▶ Suponiendo un modelo ARIMA, buscamos detectar la existencia de intervención en una serie, lo cual señalara si ha tenido efecto una determinada política.
- ▶ La intervención se detecta cuando aparecen valores extraños en el comportamiento de la serie respecto a su modelo previsto.
- ▶ Caben varios tipos de valores extraños, que denominaremos atípicos
 - ▶ Atípico aditivo
 - ▶ Atípico innovativo
 - ▶ Atípico de nivel
 - ▶ Atípico transitorio

Atípico aditivo

- ▶ Diremos que ha ocurrido un AO sobre una serie temporal en el momento h si el valor de la serie se genera en ese instante diferente que el resto.
- ▶ El modelo que seguirá la series Z_t observada si ha sido afectada por un AO en t es $Z_t = W_A I_t^h + \Psi(L)\varepsilon_t$
- ▶ La alteración es puntual y solo se refleja en el cambio en el punto , que se notará en un residuo especialmente alto en ese periodo y su efecto sobre el futuro será reflejo de la parte AR del modelo

$$\varepsilon_t = \Pi(L)(Z_t - W_A I_t^h) \text{ siendo } \Pi(L) = \Psi(L)^{-1}$$

- ▶ Representa una intervención puntual en un momento dado

Atípico innovativo

- Diremos que ha ocurrido un atípico innovativo en la serie en el momento $t = h$ cuando la innovación en ese punto es directamente afectado por una cantidad desconocida debida al suceso atípico.

- El modelo que seguirá la series Z_t observada si ha sido afectada por un AI en t es

$$Z_t = \Psi(L)(W_I I_t^h + \varepsilon_t)$$

- La alteración es puntual pero su efecto depende del modelo, y se notará tanto en la serie como en el residuo
- Se propaga siguiendo el mismo patrón dinámico de la serie
 - En $t = h$ el valor de la serie es igual al valor original mas la magnitud del atípico
 - Si la serie es estacionaria y sigue un modelo $MA(q)$ los q valores siguientes de la serie estarán afectados por la magnitud que depende de los coeficientes de la MA
 - Si la serie es estacionaria y tiene un componente AR , todos los valores posteriores a $t = h$ estarán afectados , aunque con pesos decrecientes ya que los Ψ_j van a caer al crecer j .
- Representa un alteración puntual en las condiciones externas

Atípico de cambio de nivel

- ▶ Diremos que la serie temporal en el momento h ha sufrido un cambio de nivel si el valor de la serie se ve afectado por un escalón a partir de ese instante.
- ▶ El modelo que seguirá la series Z_t observada si ha sido afectada por un AL en t es $Z_t = W_L S_t^h + \Psi(L)\varepsilon_t$ siendo S una variable escalón
- ▶ La alteración es continua y se refleja en el cambio en el punto h , que se notará en un residuo especialmente alto en ese periodo y su efecto sobre el futuro será reflejo de la parte AR del modelo

$$\varepsilon_t = \Pi(L)(Z_t - W_L S_t^h) \text{ siendo } \Pi(L) = \Psi(L)^{-1}$$

Representa una alteración permanente en un momento dado sobre la serie original

Atípico de cambio transitorio

- ▶ Un atípico transitorio es un suceso cuyo efecto en la serie observada perdura pero no permanece
- ▶ El modelo que seguirá la series Z_t observada si ha sido afectada por un AT en t es $Z_t = \frac{W_L}{1-\delta L} I_t^h + \Psi(L)\varepsilon_t$ siendo I una variable impulso
 - ▶ Si $\delta=1$ daría lugar a un modelo de cambio de nivel
 - ▶ Si $\delta=0$ daría lugar a un modelo de atípico aditivo
- ▶ La alteración es transitoria y se refleja en el cambio en el punto , que se notará en un residuo especialmente alto en ese periodo y su efecto sobre el futuro ira disminuyendo hasta desaparecer

$$\varepsilon_t = \Pi(L)\left(Z_t - \frac{W_C}{1-\delta L} I_t^h\right) \text{ siendo } \Pi(L) = \Psi(L)^{-1}$$

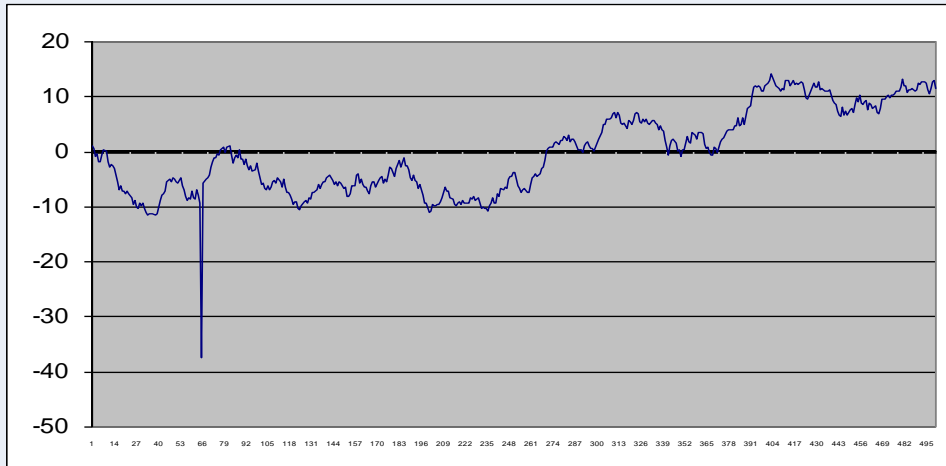
- ▶ Luego del impacto inicial, el efecto del atípico cae gradualmente. El parámetro d determina la velocidad con la que el efecto del atípico desaparece en el nivel de la serie.
- ▶ Representa una alteración en la serie original cuyo efecto no es instantáneo ni permanente

Efecto rampa

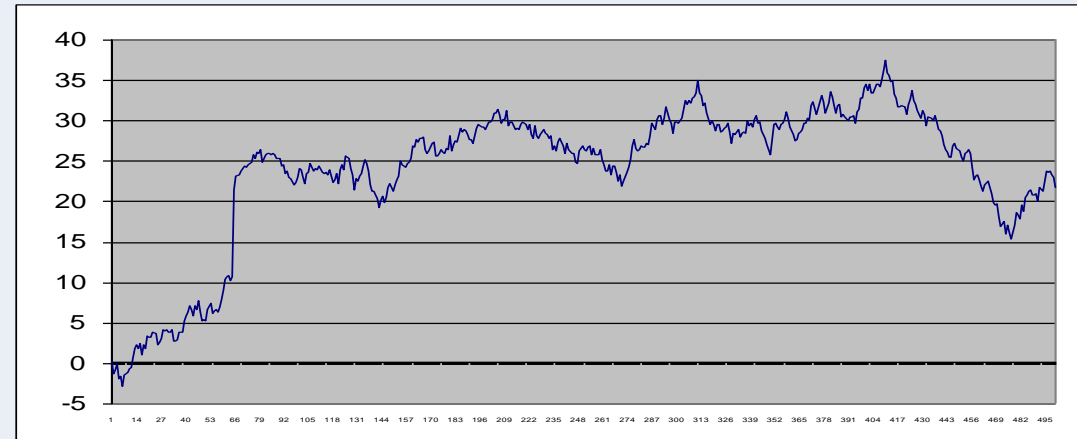
- ▶ Se produce cuando hay una intervención generando una tendencia determinística.
- ▶ El modelo sería $Z_t = W_R R_t^h + \Psi(L)\varepsilon_t$ siendo $R_t^h = \begin{cases} 0 & \text{si } t < h \\ t + 1 - h & \text{si } t \geq h \end{cases}$
- ▶ Introduce una tendencia determinística con pendiente WR en la serie a partir de h.
- ▶ Representa una alteración permanente acumulativa en el tiempo: por ejemplo decisión de ahorrar (fondo de pensiones)

Efectos de los diferentes atípicos

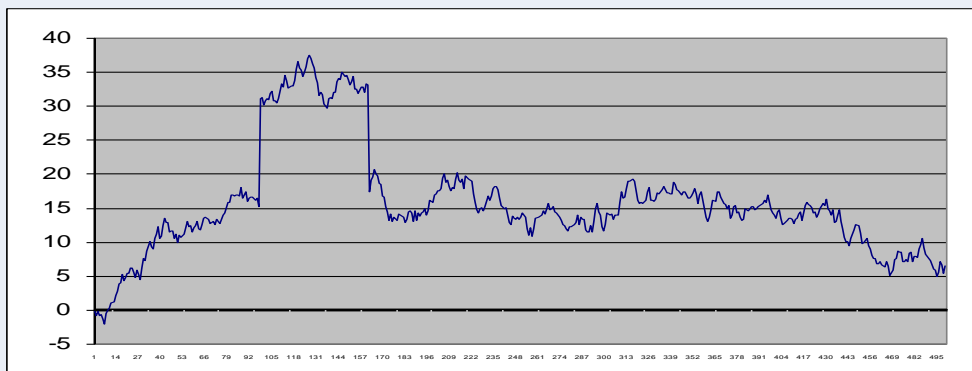
Impulso: El acontecimiento es puramente puntual afectando a una única observación



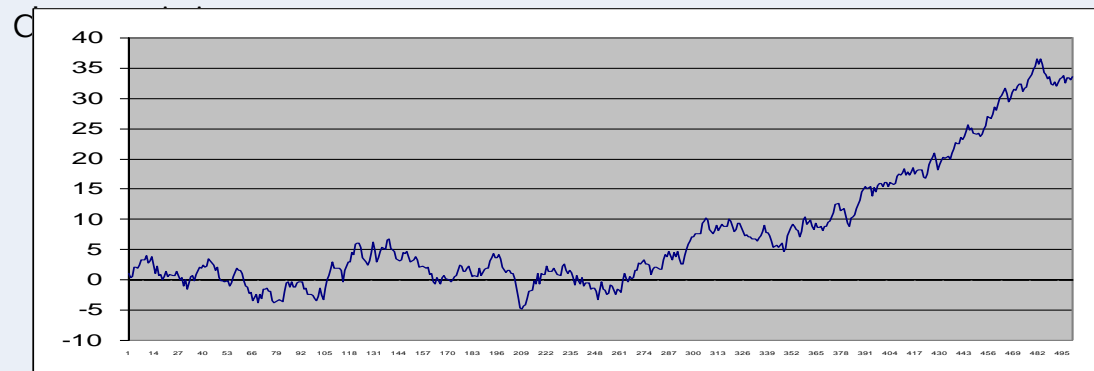
Escalón: Se produce un cambio de nivel (media) en la serie a partir de un determinado acontecimiento



Meseta: Una variante del atípico de escalón pero de duración determinada



Tendencia (o rampa): El acontecimiento impacta progresivamente en la serie generando una tendencia

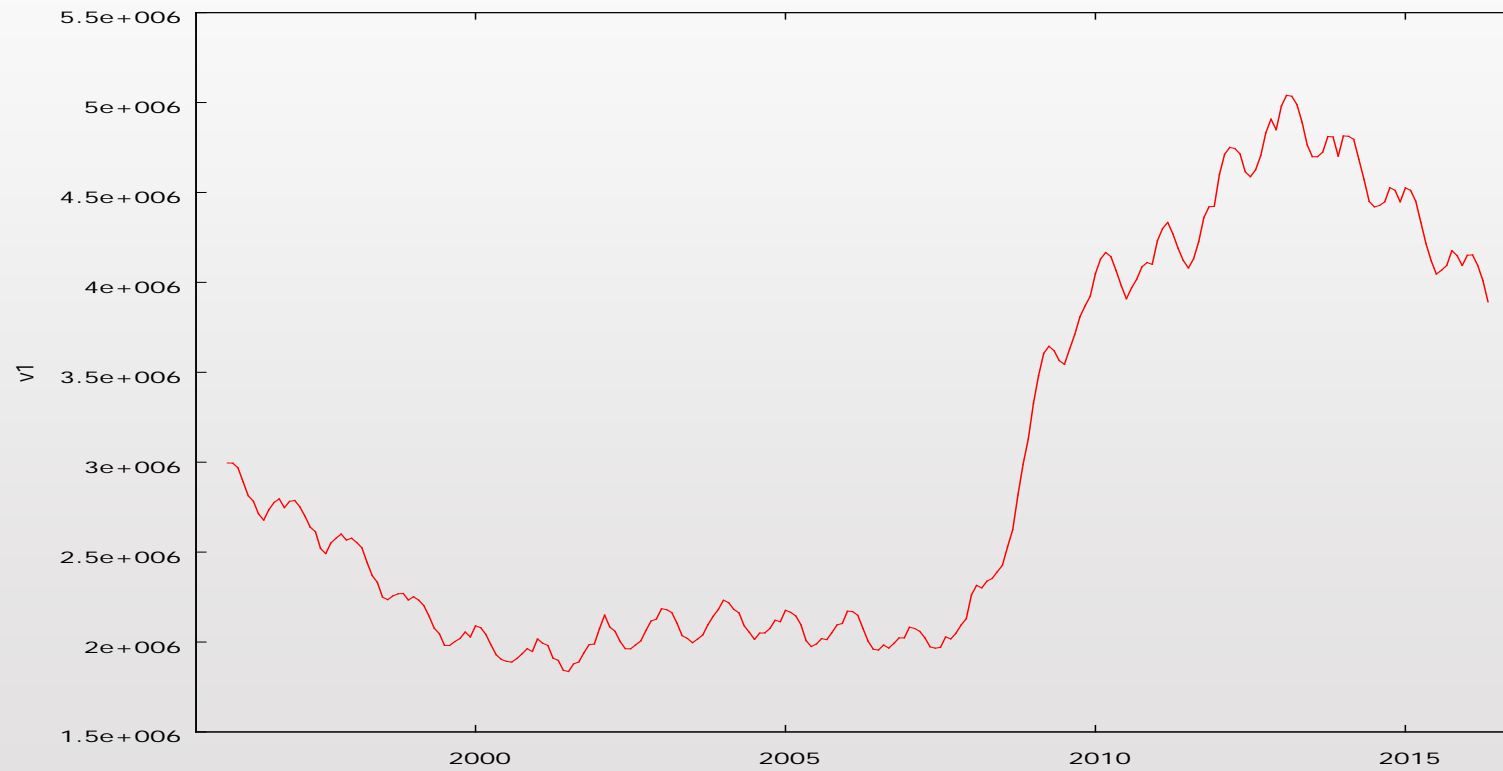




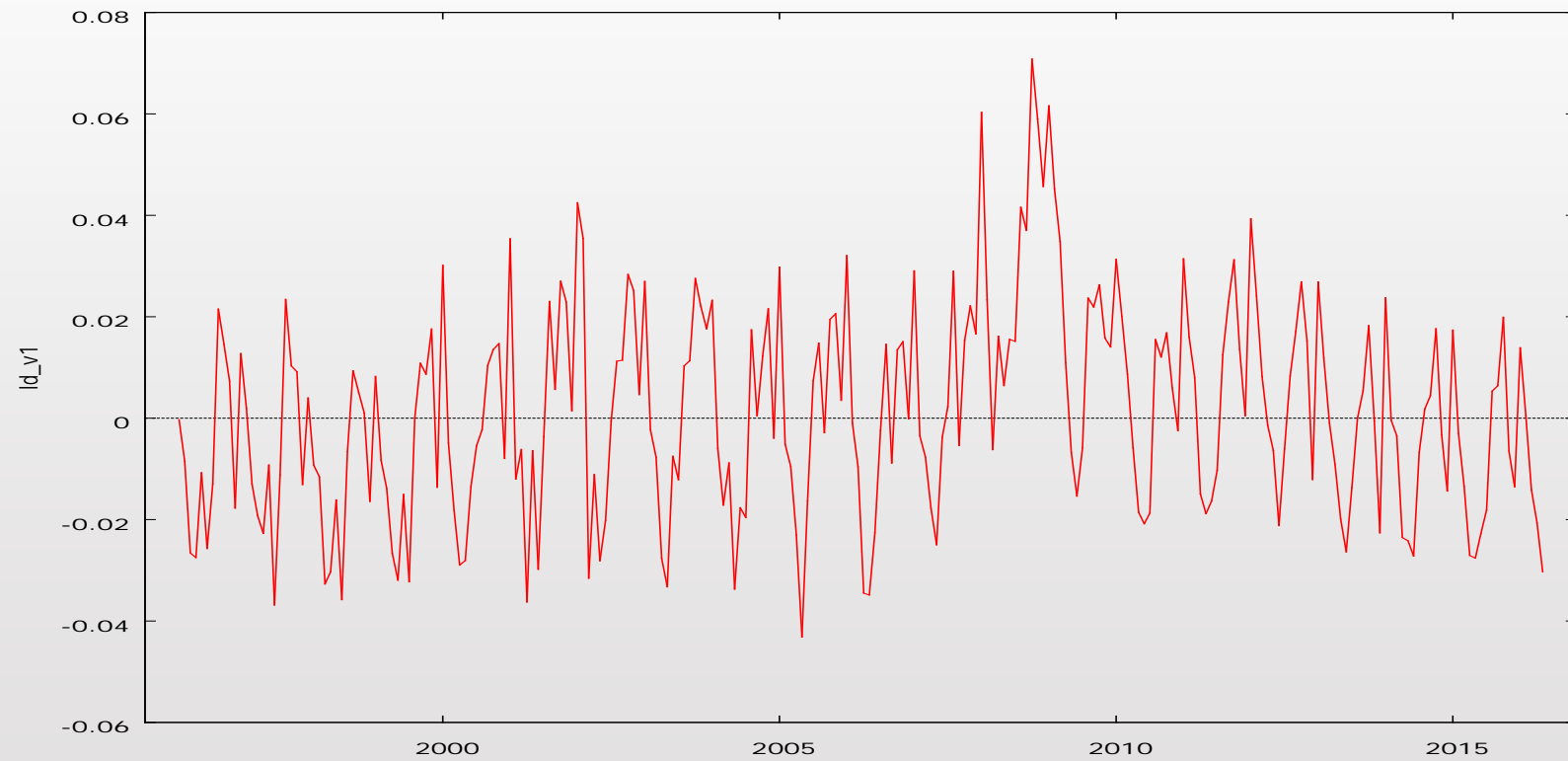
Evaluación de políticas mediante atípicos

- ▶ Las políticas establecidas mediante leyes deberían tener un efecto permanente, por lo que debemos definir las como variables escalón, mientras que los decretos puntuales o convocatorias son variables impulso.
- ▶ Para evaluar el efecto de una política sobre una determinada variable económica se siguen los siguientes pasos:
 1. Estudiar el comportamiento de la variable
 2. Introducir la variable relativa a la política como variable independiente en el modelo
 3. Desarrollar el modelo de series temporales con la variable de política
 4. Analizar si su efecto es significativo

Paro en España



Tasa de crecimiento del Paro en España

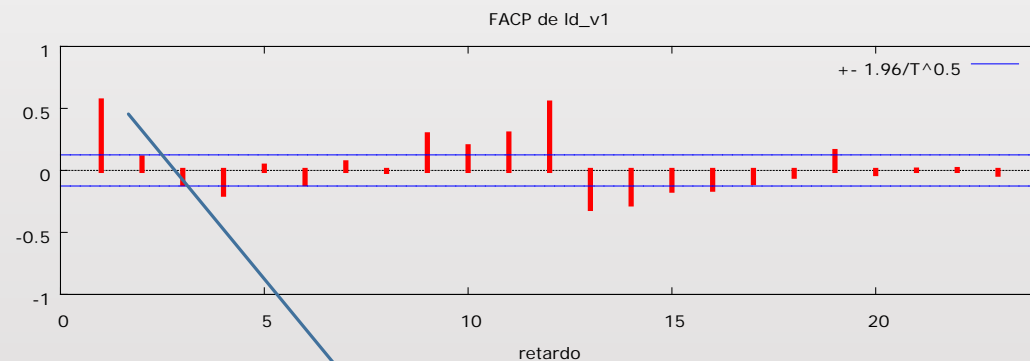
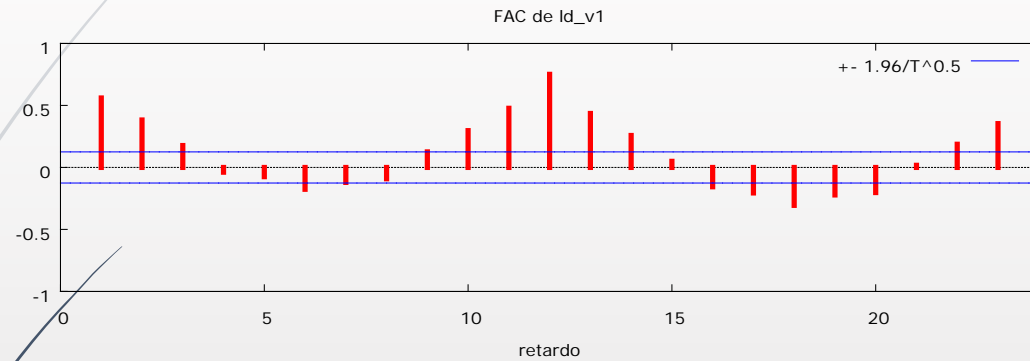


Análisis de la serie de crecimiento del paro

Función de autocorrelación para ld_v1

***, ** y * indica significatividad a los niveles del 1%, 5% y 10%

Utilizando desviación típica $1/T^{0.5}$



Efecto del periodo anterior y estacional

RETARDO	FAC		FACP		Estad-Q. [valor p]
1	0.5602	***	0.5602	***	77.5285 [0.000]
2	0.3828	***	0.1004		113.8706 [0.000]
3	0.1767	***	-0.1067	*	121.6484 [0.000]
4	-0.0384		-0.1915	***	122.0172 [0.000]
5	-0.0745		0.0346		123.4121 [0.000]
6	-0.1762	***	-0.1064	*	131.2388 [0.000]
7	-0.1207	*	0.0610		134.9293 [0.000]
8	-0.0912		-0.0078		137.0451 [0.000]
9	0.1257	**	0.2872	***	141.0786 [0.000]
10	0.2971	***	0.1907	***	163.7254 [0.000]
11	0.4772	***	0.2939	***	222.3926 [0.000]
12	0.7513	***	0.5434	***	368.4439 [0.000]
13	0.4356	***	-0.3065	***	417.7465 [0.000]
14	0.2580	***	-0.2699	***	435.1131 [0.000]
15	0.0496		-0.1592	**	435.7572 [0.000]
16	-0.1561	**	-0.1511	**	442.1692 [0.000]
17	-0.2057	***	-0.0975		453.3618 [0.000]
18	-0.3060	***	-0.0467		478.2267 [0.000]
19	-0.2218	***	0.1525	**	491.3504 [0.000]
20	-0.2024	***	-0.0249		502.3291 [0.000]
21	0.0173		0.0030		502.4097 [0.000]
22	0.1872	***	0.0071		511.8863 [0.000]
23	0.3538	***	-0.0301		545.8916 [0.000]

Raíces unitarias

Contraste aumentado de Dickey-Fuller para ld_v1
incluyendo 14 retardos de $(1-L)ld_v1$
(el máximo fue 14, el criterio AIC)
tamaño muestral 229

hipótesis nula de raíz unitaria: $a = 1$

contraste sin constante

modelo: $(1-L)y = (a-1)*y(-1) + \dots + e$

valor estimado de $(a - 1)$: -0.178313

Estadístico de contraste: $\tau_{nc}(1) = -2.45198$

valor p asintótico 0.01375

Coef. de autocorrelación de primer orden de e : -0.035

diferencias retardadas: $F(14, 214) = 26.102 [0.0000]$

contraste con constante

modelo: $(1-L)y = b_0 + (a-1)*y(-1) + \dots + e$

valor estimado de $(a - 1)$: -0.183777

Estadístico de contraste: $\tau_c(1) = -2.48282$

valor p asintótico 0.1197

Coef. de autocorrelación de primer orden de e : -0.035

diferencias retardadas: $F(14, 213) = 25.954 [0.0000]$

con constante y tendencia

modelo: $(1-L)y = b_0 + b_1*t + (a-1)*y(-1) + \dots + e$

valor estimado de $(a - 1)$: -0.192241

Estadístico de contraste: $\tau_{ct}(1) = -2.44139$

valor p asintótico 0.3579

Coef. de autocorrelación de primer orden de e : -0.036

diferencias retardadas: $F(14, 212) = 25.776 [0.0000]$

Contraste KPSS para ld_v1

$T = 244$

Parámetro de truncamiento de los
retardos = 4

Estadístico de contraste = 0.4808

10% 5%

1%

Valores críticos: 0.348 0.462
0.740

Valor p interpolado 0.047

Se acepta la
existencia de una
raíz unitaria

Estimación del modelo

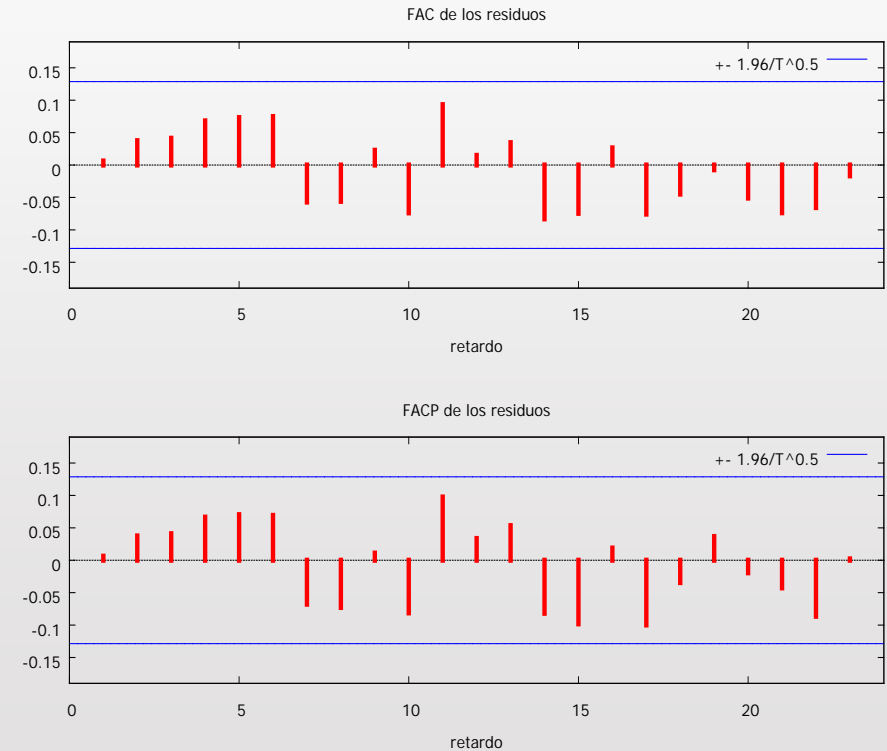
Evaluaciones de la función: 48
Evaluaciones del gradiente: 14

Modelo 11: ARMA, usando las observaciones 1997:02-2016:05 (T = 232)
Estimado usando el filtro de Kalman (MV exacta)
Variable dependiente: sd_ld_v1
Desviaciones típicas basadas en el Hessiano

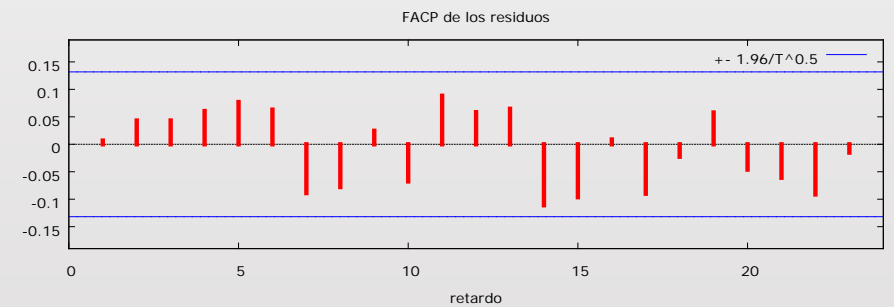
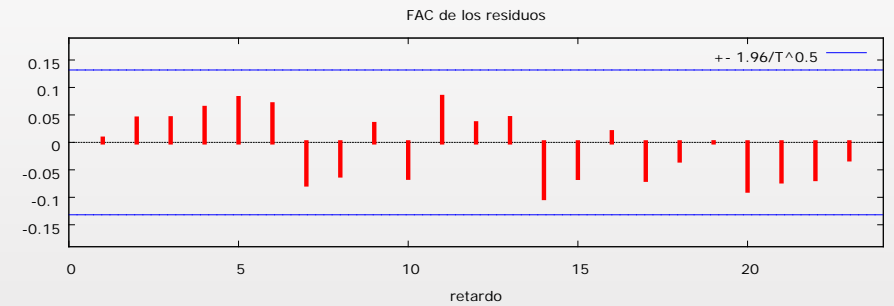
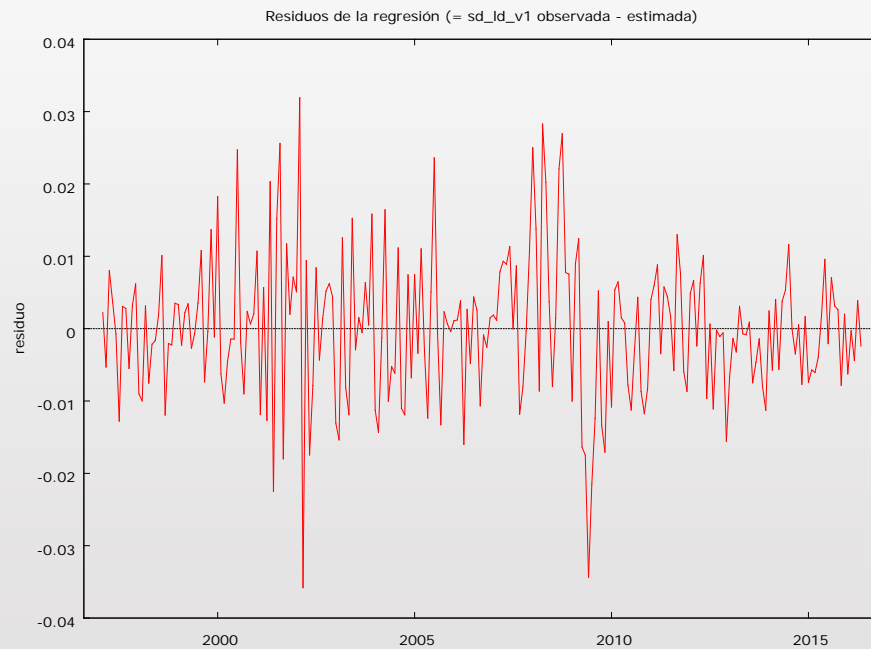
	Coefficiente	Desv. Típica	z	Valor p	
phi_2	0.245648	0.0655423	3.748	0.0002	***
phi_3	0.136484	0.0630253	2.166	0.0303	**
phi_4	0.121104	0.0643793	1.881	0.0600	*
phi_5	0.129379	0.0628037	2.060	0.0394	**
phi_6	0.176486	0.0650532	2.713	0.0067	***
theta_1	0.226188	0.0646148	3.501	0.0005	***
Theta_1	-0.584654	0.0656065	-8.912	5.03e-019	***
Media de la vble. dep.	-0.000026	D.T. de la vble. dep.	0.013900		
media innovaciones	-0.000021	D.T. innovaciones	0.009894		
Log-verosimilitud	739.0368	Criterio de Akaike	-1462.074		
Criterio de Schwarz	-1434.500	Crit. de Hannan-Quinn	-1450.953		

Real Imaginaria Módulo Frecuencia

Correlograma de los residuos



Comportamiento de los residuos



¿Afecta la reforma laboral de 2012?

- ▶ Introducimos una variable escalón a partir de febrero de 2012
- ▶ Los resultados señalan que no es significativa. Parece que la reforma no tuvo un impacto tan significativo en la generación de empleo considerada independientemente

Evaluaciones de la función: 71

Evaluaciones del gradiente: 19

Modelo 2: ARMAX, usando las observaciones 1997:02-2016:05 (T = 232)

Estimado usando el filtro de Kalman (MV exacta)

Variable dependiente: sd_ld_v1

Desviaciones típicas basadas en el Hessiano

	Coeficiente	Desv. Típica	z	Valor p	
const	0.000445474	0.00176369	0.2526	0.8006	
phi_2	0.242614	0.0657739	3.689	0.0002	***
phi_3	0.129700	0.0639492	2.028	0.0425	**
phi_4	0.114081	0.0652926	1.747	0.0806	*
phi_5	0.123269	0.0635956	1.938	0.0526	*
phi_6	0.174020	0.0651431	2.671	0.0076	***
theta_1	0.225584	0.0646148	3.491	0.0005	***
Theta_1	-0.586094	0.0655800	-8.937	4.00e-019	***
RL12	-0.00242984	0.00385305	-0.6306	0.5283	