

5 AUTOCORRELACIÓN

En este capítulo se presenta una herramienta de análisis, el correlograma, o representación gráfica de la función de autocorrelación, que tiene una doble utilidad. Por una parte, puede servir para confirmar la presencia de estacionalidad y determinar su período; por otra, indica cuántas previsiones son admisibles, a partir del último tiempo de recogida de información.

El concepto de autocorrelación es bien simple; supongamos que se dispone de la serie cronológica $Y_1, Y_2, \dots, Y_{1+k}, \dots, Y_N$, y se desplaza dicha serie k unidades de tiempo; se pueden formar las parejas $(Y_1; Y_{1+k}), (Y_2; Y_{2+k}), (Y_3; Y_{3+k}), \dots, (Y_{N-k}; Y_N)$.

El coeficiente de correlación entre ambas series, es decir, de las parejas citadas, se denota por ρ_k y recibe el nombre de coeficiente de autocorrelación de orden k ; el desplazamiento k también se denomina retardo, y representando gráficamente ρ_k en función del retardo k , se obtiene el autocorrelograma de la serie. De la estructura del planteamiento se deduce que $\rho_k = \rho_{-k}$.

5.1 Correlograma

Un valor no nulo de ρ_k indica que existe correlación entre informaciones separadas k unidades de tiempo, es decir, la historia se transmite k unidades de tiempo más allá. En consecuencia, si el último valor del tiempo del que se dispone de datos es el T , será admisible hacer previsiones para un tiempo igual a $T+k$. Evidentemente, si ρ_k fuese nulo, sería inadmisibile una predicción para $T+k$, ya que los datos disponibles no transmiten ninguna información relevante a una distancia como la considerada.

Sea que se dispone de una serie cronológica de datos $y_1, y_2, \dots, y_1, \dots, y_N$, para elaborar el correlograma o gráfico de la función de autocorrelación. Se estiman las siguientes características:

■ Media: $\hat{m} = \bar{y} = \frac{\sum_{i=1}^N y_i}{N}$

■ Autocovariancia: $\hat{\gamma}_k = \frac{\sum_{i=1}^{N-k} (y_i - \bar{y})(y_{i+k} - \bar{y})}{N} \quad k = 0, 1, \dots, N-1$

■ Autocorrelación: $\hat{\rho}_k = r_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$

Para poder estimar la autocovariancia, γ_k , el número de componentes de la serie debe ser tal que $N > k+1$, y es recomendable $N \geq 50$ y $k \leq N/4$.

Para identificar los coeficientes de autocorrelación que sean significativamente distintos de cero, es necesario estudiar el comportamiento estadístico de los estimadores. Bartlett ha estudiado el estadístico r_k y, bajo el supuesto de que $\rho_k = 0$ para todo $k \geq K$, obtiene que

$$V(r_k) \cong \frac{1}{N} \sum_{-(K-1)}^{K-1} \rho_i^2 \quad \forall k \geq K$$

al sustituir ρ_i por su estimador, r_i , y, dado que $\rho_0 = 1$ y $\rho_i = \rho_{-i}$, resulta que la estimación de la variancia de r_k es igual a

$$\hat{V}(r_k) \cong \frac{1}{N} \sum_{-(K-1)}^{K-1} r_i^2 \Rightarrow \begin{cases} V(r_k) \cong \frac{1}{N} & k \geq K \quad K = 1 \\ V(r_k) \cong \frac{1}{N} \left(1 + 2 \sum_1^{K-1} r_i^2 \right) & k \geq K \quad K > 1 \end{cases}$$

Anderson indica que para valores de k tales que $\rho_k = 0$ y N suficientemente grande, r_k se distribuye aproximadamente $N(0; V(r_k))$. De esta forma, con una probabilidad del orden del 95%, si $\rho_k = 0$, su estimador r_k se encontrará en el intervalo $\pm 2 S(r_k)$, donde $S(r_k)$ representa la desviación tipo estimada de r_k , es decir, $S(r_k) = \sqrt{\hat{V}(r_k)}$.

El intervalo $\pm 2 S(r_k)$ se denomina intervalo de no significación de ρ_k , y es el conjunto de valores que puede tomar r_k para que, con un riesgo del 5%, se pueda admitir la ausencia de correlación entre valores de la serie, desplazados k unidades de tiempo. Por todo ello, al calcular la función de autocorrelación de una serie, es bueno representarla gráficamente junto al intervalo $\pm 2 S(r_k)$, con objeto de considerar únicamente como coeficientes de autocorrelación no nulos aquellos cuya estimación esté fuera del citado intervalo.

Las autocorrelaciones suelen disponerse matricialmente, lo que da lugar a la matriz de autocorrelaciones P_N , simétrica, definida positiva, cuya estructura es

$$P_N = \begin{pmatrix} 1 & \rho_1 & \rho_2 & \dots & \rho_{N-1} \\ \rho_1 & 1 & \rho_1 & \dots & \rho_{N-2} \\ \rho_2 & \rho_1 & 1 & \dots & \rho_{N-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho_{N-1} & \rho_{N-2} & \rho_{N-3} & \dots & 1 \end{pmatrix}$$

Como aplicación se va a analizar la serie cronológica de la tabla 5.1, que corresponde al valor de los *stocks* en I.C.I. (Serie D* de *Time Series Analysis and Forecasting*, O.D. Anderson, ed. Butterworths, 1977). Su evolución se muestra en la figura 5.1.

t	Y	t	Y	t	Y	t	Y	t	Y	t	Y
1	304	19	278	37	282	55	273	73	291	91	282
2	303	20	277	38	283	56	272	74	288	92	286
3	307	21	279	39	279	57	273	75	288	93	286
4	299	22	278	40	280	58	271	76	290	94	287
5	296	23	270	41	280	59	272	77	293	95	284
6	293	24	268	42	279	60	271	78	288	96	283
7	301	25	272	43	278	61	273	79	289	97	286
8	293	26	273	44	283	62	277	80	291	98	282
9	301	27	279	45	278	63	274	81	293	99	287
10	295	28	279	46	270	64	274	82	293	100	286
11	284	29	280	47	275	65	272	83	290	101	287
12	286	30	275	48	273	66	280	84	288	102	292
13	286	31	271	49	273	67	282	85	287	103	292
14	287	32	277	50	272	68	292	86	289	104	294
15	284	33	278	51	275	69	295	87	292	105	291
16	282	34	279	52	273	70	295	88	288	106	288
17	278	35	283	53	273	71	294	89	288	107	289
18	281	36	284	54	272	72	290	90	285		

Tabla 5.I.- Valor de los stocks en I.C.I

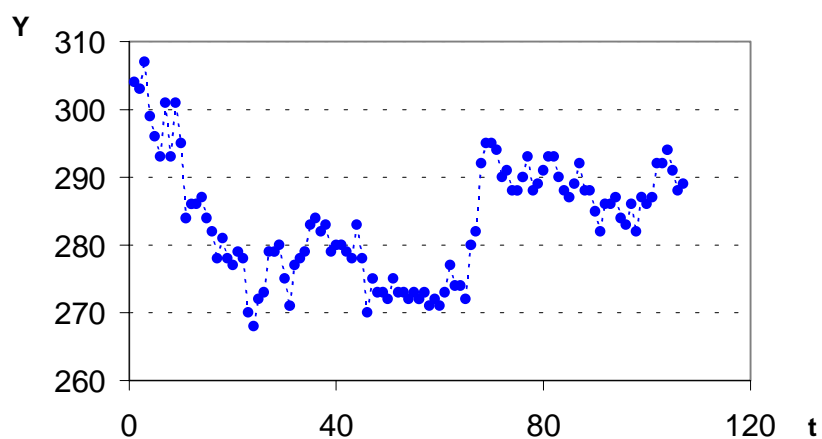


Fig. 5.1.- Evolución de la serie de la tabla 5.I

En la tabla 5.II se presenta el detalle del cálculo de las autocorrelaciones para los casos de $k=1$ y $k=2$, de los valores de la tabla 5.I. En primer lugar es necesario calcular la media de todos los datos

$$\hat{m} = \bar{y} = \frac{1}{10} (304 + \dots + 295) = 299,2$$

	t	1	2	3	4	5	...	105	106	107
k = 0	y	304	303	307	299	296		291	288	289
	y - \bar{y}	20,44	19,44	23,44	15,44	12,44	...	7,44	4,44	5,44
k = 1	y		304	303	307	299	...	294	291	288
	y - \bar{y}		20,44	19,44	23,44	15,44	...	10,44	7,44	4,44
k = 2	y			304	303	307	...	292	294	291
	y - \bar{y}			20,44	19,44	23,44	...	8,44	10,44	7,44

Tabla 5.II.- Detalle del cálculo de las autocorrelaciones de la tabla 5.I

Las estimaciones se obtienen como

$$\hat{\gamma}_0 = \frac{1}{107} \sum_1^{107} (y_i - \bar{y})^2 = \frac{20,44^2 + \dots + 5,44^2}{107} = 74,695$$

$$\hat{\gamma}_1 = \frac{1}{107} \sum_2^{107} (y_i - \bar{y})(y_{i-1} - \bar{y}) = \frac{19,44 \times 20,44 + \dots + 5,44 \times 4,44}{107} = 66,123$$

$$r_1 = \hat{\rho}_1 = \frac{\hat{\gamma}_1}{\hat{\gamma}_0} = \frac{66,123}{74,695} = 0,885$$

$$\hat{\gamma}_2 = \frac{1}{107} \sum_3^{107} (y_i - \bar{y})(y_{i-2} - \bar{y}) = \frac{23,44 \times 20,44 + \dots + 5,44 \times 7,44}{107} = 59,775$$

$$r_2 = \hat{\rho}_2 = \frac{\hat{\gamma}_2}{\hat{\gamma}_0} = \frac{59,775}{74,695} = 0,800$$

etc.

Según Bartlett las desviaciones tipo estimadas para r_1 y r_2 son:

$$S(r_1) = \sqrt{\frac{1}{N}} = \sqrt{\frac{1}{107}} = 0,097$$

$$S(r_2) = \sqrt{\frac{1}{N} (1 + 2 r_1^2)} = \sqrt{\frac{1 + 2 \times 0,885^2}{107}} = 0,155$$

etc.

Y los intervalos $\pm 2 S(r_k)$ son, respectivamente, $\pm 0,194$ y $\pm 0,310$ para $k=1$ y $k=2$. En consecuencia, con un riesgo del 5%, ρ_1 y ρ_2 pueden ser significativamente distintos de cero.

Analizando los 107 valores de la serie completa para $k = 1, 2, \dots, 10$, se obtienen los valores mostrados en la tabla 5.III y presentados en la figura 5.2. De estos resultados, se verifica que a partir de $k = 7$ ya se puede considerar ρ_k como nulo, es decir, no es admisible hacer previsiones separadas en más de 7 unidades de tiempo del último momento de recogida de datos.

k	1	2	3	4	5	6	7	8	9	10
r_k	0,89	0,80	0,70	0,63	0,58	0,55	0,48	0,40	0,31	0,23
$S(r_k)$	0,10	0,15	0,19	0,21	0,23	0,24	0,25	0,26	0,27	0,27

Tabla 5.III.- Valores de las autocorrelaciones y sus desviaciones tipo

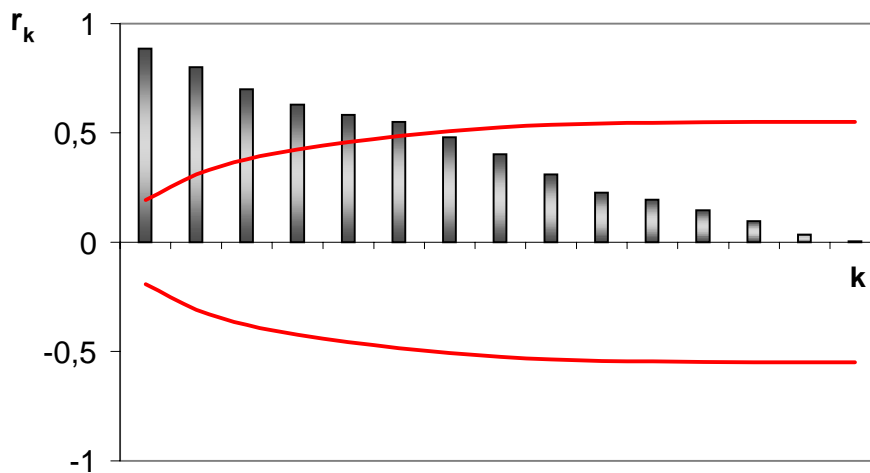


Fig. 5.2.- Correlograma

Con los valores de la tabla 5.III, se puede escribir la matriz de autocorrelaciones que, para $k = 3$, adquiere la siguiente forma:

$$\hat{P}_3 = \begin{pmatrix} 1 & 0,885 & 0,800 & 0,699 \\ 0,885 & 1 & 0,885 & 0,800 \\ 0,800 & 0,885 & 1 & 0,885 \\ 0,699 & 0,800 & 0,885 & 1 \end{pmatrix}$$

5.2 Interpretación de los correlogramas

En la figura 5.3 se muestran los correlogramas de las series analizadas hasta ahora.

El primero, que corresponde a los datos de la figura 1.2 de la evolución del índice IBEX35, muestra que sólo son significativos los tres primeros coeficientes de autocorrelación; por tanto, las previsiones dejan de ser válidas a partir de tres unidades de tiempo después del último dato. Es decir, lo que ocurra en un instante se transmite hasta tres unidades de tiempo más adelante.

El segundo, de los datos de la tabla 3.1 y de la figura 3.1 sobre las ventas trimestrales de material deportivo, confirma la estacionalidad de período cuatro, ya que cada cuatro barras de autocorrelación se repite la misma estructura de comportamiento. En este caso es posible hacer previsiones a cuatro trimestres vista, ya que para $k = 4$ el coeficiente de autocorrelación es significativamente distinto de cero, aunque no lo sean los de $k = 1, 2$ y 3 . Este hecho se puede interpretar como que la información de un trimestre se transmite directamente hasta una distancia temporal de cuatro trimestres, sin que afecte el comportamiento de los tres trimestres intermedios. Así, por ejemplo, una vez conocidas las ventas de invierno, se puede hacer la previsión para el invierno próximo puesto que lo que ocurra en primavera, verano y otoño no afectará al invierno siguiente.

El tercero y el cuarto corresponden, respectivamente, a las temperaturas medias mensuales (tabla 3.X y figura 3.6), y a los usuarios mensuales de un transporte público (tabla 3.XIV y figura 3.12). En ambos casos se confirma la estacionalidad de período 12, y se pueden hacer previsiones a 18 meses para la temperatura, y a 13 para los usuarios. A diferencia del segundo caso, en el último la información de un mes afecta directamente al resto de meses del mismo grupo de 12; es decir, si bien es cierto que se puede predecir el número de usuarios para dentro de un año, lo que ocurra en los meses venideros puede afectar esta previsión; por tanto interesa incorporar los datos disponibles lo más rápidamente posible al modelo, para una mayor fiabilidad de las previsiones.

El último de los correlogramas es el de los usuarios de un teléfono de atención al cliente, (tabla 4.V y figura 4.7). Además de confirmar la estacionalidad de período 5, pone de manifiesto que es aceptable hacer previsiones para los próximos 10 días, es decir, a dos semanas vista.

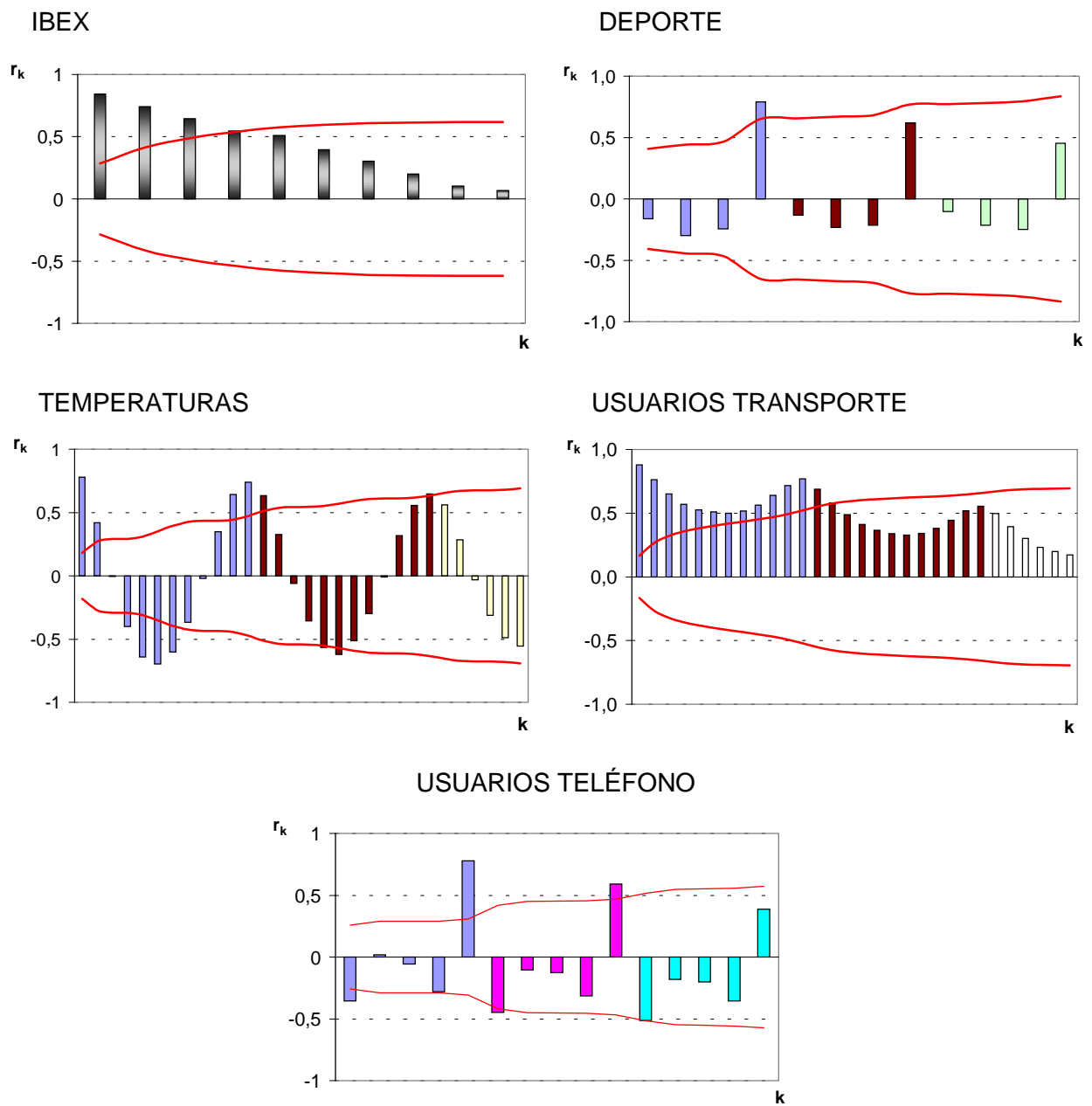


Fig. 5.3.- Correlogramas

6 OTRAS TÉCNICAS DE PREVISIÓN: PONDERACIÓN EXPONENCIAL

Cuando la serie presenta componente estacional y tendencia que se mantienen de forma sostenida a lo largo de todo el período de recogida de datos, se han expuesto dos formas de modelizarla y poder hacer previsiones: la descomposición clásica y las variables categóricas.

Sin embargo, son frecuentes las situaciones en que la tendencia, caso de existir, puede ser difícil modelizarla a través de un simple modelo polinómico de menor o mayor grado. Podría entonces pensarse en un modelo de evolución que cambiase a lo largo del tiempo; en estos casos las técnicas asociadas a la metodología de la ponderación exponencial son útiles para hacer previsiones sobre la evolución futura.

6.1 Suavizado exponencial

La ponderación exponencial, o suavizado exponencial, es otra técnica destinada también a estabilizar la serie, eliminando en lo posible la influencia del componente aleatorio. Para ello se construye una nueva serie, la serie suavizada S_t , a partir de los datos iniciales, Y_t , de manera que

$$S_t = \lambda Y_t + (1 - \lambda) S_{t-1} \quad \text{con } 0 < \lambda < 1$$

Para que la serie suavizada quede definida, es necesario concretar los valores de S_0 , que generalmente se considera igual a Y_1 , y el del coeficiente de ponderación λ . En la selección del valor de λ se pueden emplear distintos criterios de minimización de errores, que se expondrán a continuación.

Sustituyendo repetitivamente S_{t-1} , S_{t-2} ,... por su expresión de S_t , se obtiene

$$S_t = \lambda Y_t + \lambda(1-\lambda)Y_{t-1} + \lambda(1-\lambda)^2 Y_{t-2} + \dots + \lambda(1-\lambda)^{t-1} Y_1 + \lambda(1-\lambda)^t S_0$$

El valor de S_t es la previsión para el tiempo siguiente, es decir, $\tilde{Y}_{t+1} = S_t$.

El análisis de la expresión anterior permite interpretar este tipo de suavizado, de forma que el valor de Y previsto para el período $t+1$, es decir S_t , se obtenga como promedio ponderado de los valores reales que ha presentado la serie cronológica desde el inicio de la recogida de información. La discrepancia entre los valores obtenidos y los previstos, $Y_{t+1} - S_t$, es atribuible en parte al componente aleatorio y, posiblemente, a cambios bruscos en el comportamiento de la serie.

El coeficiente de ponderación λ juega el siguiente papel: cuanto mayor sea su valor, tanto más peso se dará a los valores recientes, en detrimento de los antiguos; mientras que valores de λ próximos a cero dan gran peso a la historia y poca importancia a los valores próximos.

Así, si la serie se mantiene estable, serán interesantes valores pequeños del coeficiente de ponderación ya que amortiguarán fuertemente la oscilación aleatoria, mientras que si la serie presentara cambios bruscos, la serie suavizada tardaría mucho en detectarlos si su λ fuese pequeña, mientras que respondería prontamente a ellos con valores altos del coeficiente λ .

Analizando la expresión del valor suavizado, para distintos valores de λ , se puede escribir, por ejemplo,

$$\lambda = 0,10 \quad \Rightarrow \quad \tilde{Y}_5 = S_4 = 0,10 Y_4 + 0,09 Y_3 + 0,081 Y_2 + 0,729 Y_1$$

$$\lambda = 0,50 \quad \Rightarrow \quad \tilde{Y}_5 = S_4 = 0,50 Y_4 + 0,25 Y_3 + 0,125 Y_2 + 0,125 Y_1$$

$$\lambda = 0,90 \quad \Rightarrow \quad \tilde{Y}_5 = S_4 = 0,90 Y_4 + 0,09 Y_3 + 0,009 Y_2 + 0,001 Y_1$$

Es decir, con un valor del factor de ponderación de 0,10, la previsión para $t = 5$ está constituida por un 10% del valor observado en $t = 4$, un 9% del de $t = 3$, un 8,1% del de $t = 2$ y un 72,9 % del de $t = 1$; o sea, con un valor pequeño de λ , la previsión está constituida mayoritariamente por el valor más antiguo.

Cuando λ es igual a 0,50, los pesos aplicados a cada valor recogido están más uniformemente repartidos y, cuando λ es grande, por ejemplo 0,90, el mayor componente de la previsión es el último valor observado; los demás tendrán un valor de ponderación tanto más pequeño cuanto más alejados estén en el tiempo.

El suavizado exponencial puede verse como un método alternativo a las medias móviles, con sus ventajas e inconvenientes.

Entre las primeras hay que citar que con la ponderación exponencial no se pierde ninguna información, al contrario que con las medias móviles, pues cuanto mayor era la longitud del período a promediar, tanta más información se perdía, en el inicio y en el fin de la serie. Además una serie con cambios de tendencia, más o menos bruscos, se puede modelizar por suavizado exponencial y no podría hacerse ni por descomposición ni por variables categóricas. Por el contrario, si la serie presenta estacionalidad con las medias móviles, siempre que se escoja correctamente el período, ésta desaparece totalmente y da lugar a una serie estabilizada que permite modelizar directamente la tendencia, hecho que no ocurre con la ponderación exponencial simple, que no es capaz de suavizar la oscilación debida a la estacionalidad.

Para solucionar este inconveniente, se han desarrollado técnicas basadas en el suavizado exponencial, que permiten incorporar un modelo de tendencia o bien una componente estacionaria; éstas son las técnicas de Brown, para el primer caso, o de Winters para el segundo.

6.2 Selección del factor de ponderación

Tal como se ha expuesto, en función del valor de λ , se puede dar mayor o menor peso a la historia, y detectar con más o menos rapidez cambios bruscos en la serie; es por ello que la selección del valor más adecuado para el factor de ponderación es crucial en el éxito de la modelización de la serie y la previsión de valores futuros.

Todos los métodos utilizados para esta selección se basan en minimizar alguna función de los errores de ponderación.

Recordando que $\tilde{Y}_t = S_{t-1}$, los errores más destacables son:

- *Error medio*: promedio de los errores de previsión; atendiendo a que para hacer previsiones hay que disponer de datos, el primer valor previsto posible será el de $t = 2$:

$$ME = \frac{\sum_{t=2}^n (Y_t - S_t)}{n - 1}$$

- *Error cuadrático medio*: promedio de los cuadrados de los errores de previsión:

$$MSE = \frac{\sum_{t=2}^n (Y_t - S_t)^2}{n - 1}$$

- *Error absoluto medio*: promedio de los valores absolutos de los errores de previsión:

$$MAE = \frac{\sum_{t=2}^n |Y_t - S_t|}{n - 1}$$

- *Media del porcentaje del error*: promedio de los porcentajes de los errores relativos de previsión:

$$MPE = \frac{\sum_{t=2}^n \frac{Y_t - \tilde{Y}_t}{Y_t} \times 100}{n - 1}$$

- *Media del porcentaje de error absoluto*: promedio de los porcentajes de los valores absolutos de los errores de previsión relativos:

$$MAPE = \frac{\sum_{t=2}^n \frac{|Y_t - \tilde{Y}_t|}{Y_t} \times 100}{n - 1}$$

Hay que insistir en que en una serie en la que el tiempo es $t = 1, 2, \dots, n$, el suavizado exponencial no ofrece ninguna previsión para $t = 1$, y, por tanto, no existe error de previsión

en este punto; consecuentemente, en este caso los errores siempre son promedios de $n - 1$ valores.

De los errores expuestos, aquellos que no toman valor absoluto, ME y MPE, tienen poco interés ya que, a causa de la compensación de valores positivos y negativos, pueden dar valores de los promedios muy próximos a cero aun cuando existan errores de previsión muy grandes. En general, se selecciona aquel valor de λ para el cual los valores del error absoluto medio y del cuadrático medio, MAE y MSE, alcancen los valores más bajos.

Como ejemplo consideremos los datos de la tabla 6.I, serie cronológica de 50 valores, cuya representación gráfica puede verse en la figura 6.1.

t	Y_t	t	Y_t	t	Y_t	t	Y_t	t	Y_t
1	9,958	11	16,510	21	26,267	31	25,217	41	28,448
2	10,096	12	12,674	22	20,401	32	24,653	42	35,726
3	11,552	13	17,504	23	18,748	33	28,062	43	30,602
4	9,113	14	13,462	24	20,800	34	27,317	44	31,011
5	13,898	15	16,945	25	21,683	35	26,122	45	31,732
6	11,487	16	18,653	26	27,069	36	29,837	46	31,538
7	11,114	17	18,942	27	23,728	37	28,854	47	32,175
8	9,505	18	15,084	28	24,890	38	27,129	48	35,543
9	17,934	19	16,568	29	26,132	39	30,194	49	35,534
10	12,339	20	20,733	30	24,663	40	34,104	50	37,336

Tabla 6.I.- 50 valores de una serie cronológica

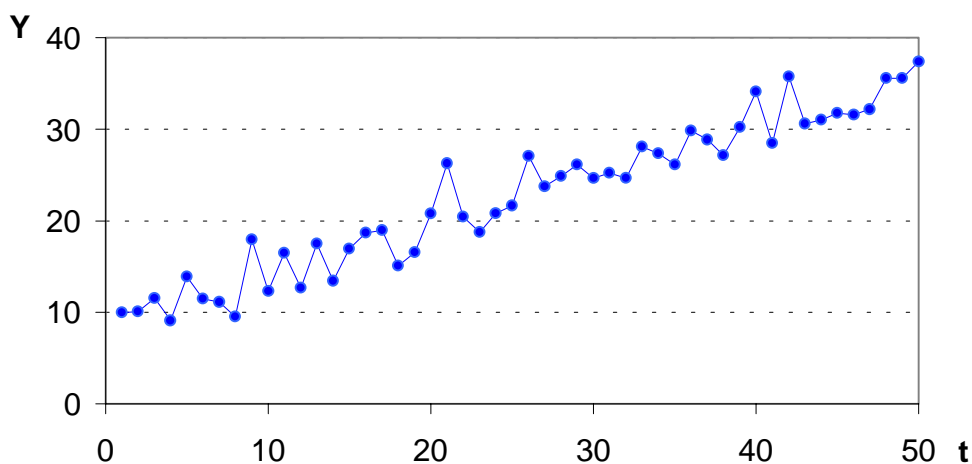


Fig. 6.1.- Evolución cronológica de los datos de la tabla 6.I

Observando el gráfico, difícilmente se aprecia una estacionalidad, mientras que es evidente una tendencia creciente y casi seguramente lineal. La figura 6.2 es el correlograma de la serie que pone de manifiesto la ausencia de estacionalidad, junto con una autocorrelación significativa hasta un retardo de 4 unidades de tiempo.

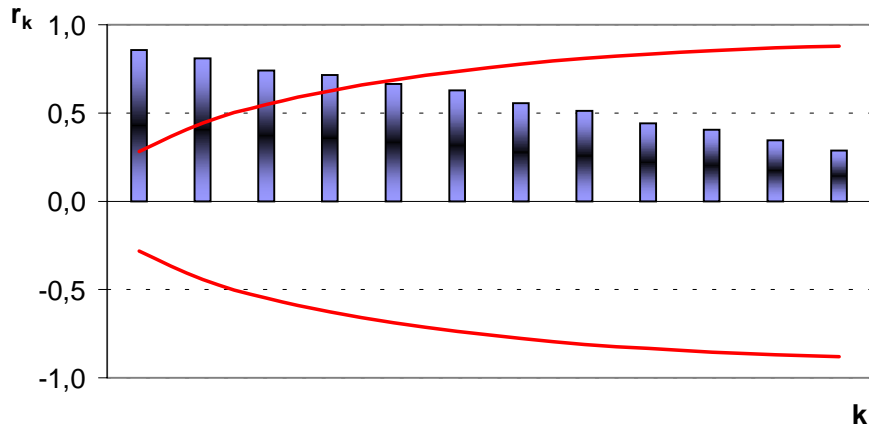


Fig. 6.2.- Correlograma de la serie de la tabla 6.I

Aplicando la ponderación exponencial a estos datos, en función del valor de λ , los errores evolucionan según muestra la tabla 6.II.

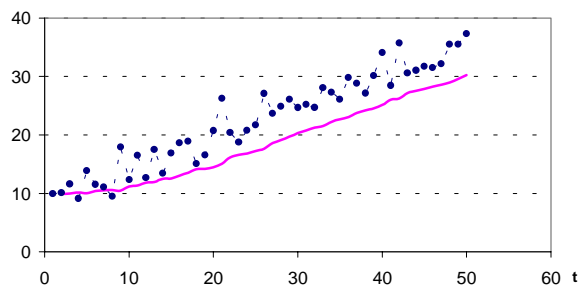
λ	ME	MSE	MAE	MPE	MAPE
0,100	4,278	25,408	4,363	17,577	18,494
0,150	3,092	15,628	3,196	12,719	13,812
0,200	2,416	11,731	2,627	9,817	11,622
0,250	1,984	9,904	2,351	7,896	10,632
0,300	1,685	8,975	2,217	6,531	10,214
0,350	1,467	8,496	2,174	5,510	10,157
0,400	1,300	8,270	2,171	4,716	10,271
0,450	1,169	8,200	2,183	4,079	10,449
0,500	1,062	8,233	2,214	3,556	10,696
0,550	0,974	8,341	2,251	3,119	10,960
0,600	0,899	8,507	2,289	2,746	11,233
0,650	0,835	8,722	2,330	2,425	11,508
0,700	0,780	8,981	2,371	2,145	11,786
0,750	0,731	9,283	2,418	1,898	12,082
0,800	0,689	9,628	2,470	1,677	12,396
0,850	0,651	10,019	2,525	1,479	12,723
0,900	0,617	10,461	2,581	1,299	13,052
0,950	0,586	10,961	2,643	1,133	13,406

Tabla 6.II.- Evolución de los errores en función de λ .

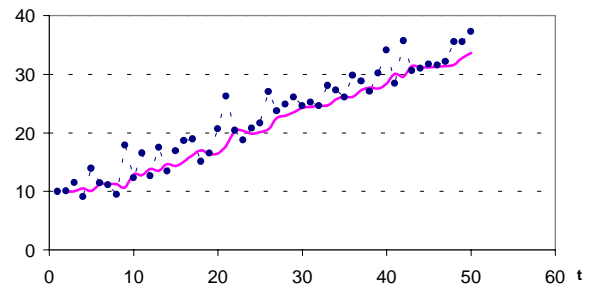
Analizando detenidamente la evolución de errores mostrada en la citada tabla, se observa que, tanto el error medio (ME) como la media del porcentaje del error (MPE) disminuyen al aumentar λ , mientras que los otros errores (MSE, MAE y MAPE) alcanzan el mínimo para el intervalo $0,35 \leq \lambda \leq 0,45$. Dado que el error cuadrático medio (MSE) es el que, en general, presenta mejores propiedades, se va a tener en cuenta para seleccionar definitivamente el coeficiente de ponderación; así, en el caso del ejemplo, se va a trabajar con $\lambda = 0,45$.

Para entender mejor lo que representa la selección de uno u otro valor de λ , en la figura 6.3 se pueden comparar las evoluciones de las series ponderadas respecto a los datos cronológicos para distintos valores del parámetro de ponderación.

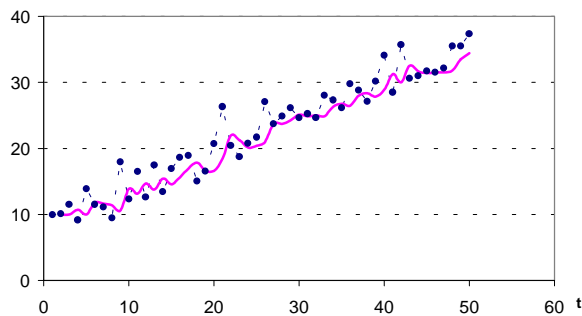
$\lambda=0,10$



$\lambda=0,30$



$\lambda=0,45$



$\lambda=0,95$

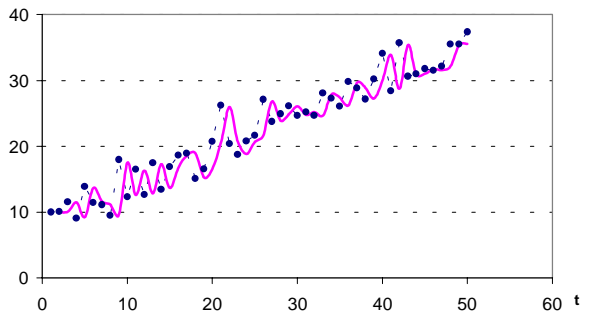


Fig. 6.3.- Serie original (•) y suavizada (—) para distintas λ .

Directamente se observa que, en este caso, para valores pequeños de λ la serie suavizada va por detrás de la real, es decir, tarda mucho en responder a la evolución. Sin embargo, cuando $\lambda = 0,95$, la suavizada está totalmente ligada a la oscilación aleatoria de la serie, es decir, la previsión para el tiempo inmediato siguiente es prácticamente igual al último valor medido. Cuando $\lambda = 0,45$, valor para el que ha resultado un error cuadrático medio mínimo, la serie suavizada exponencialmente, sigue más claramente el esqueleto de la serie cronológica y queda amortiguada la oscilación aleatoria.

La figura 6.4 muestra los residuos, $R_t = Y_t - \tilde{Y}_t = Y_t - S_{t-1}$, para tres valores de λ . En ella se observa que para $\lambda = 0,10$ la mayoría de los residuos son positivos, es decir, la previsión va por detrás del valor real, mientras que para $\lambda = 0,45$ están, casi siempre, entre los de $\lambda = 0,10$ y los de $\lambda = 0,90$, es decir, son más próximos a cero, y por tanto, más pequeños.

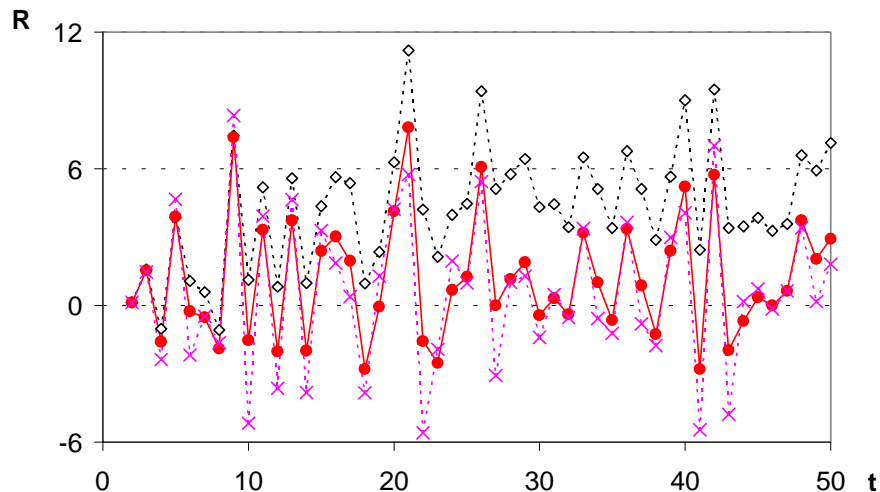


Fig. 6.4.- Residuos en función de λ . $\lambda = 0,10$ (···◇···) $\lambda = 0,45$ (—●—) $\lambda = 0,90$ (···×···)

Para el valor del coeficiente de ponderación seleccionado ($\lambda = 0,45$), se calculan los valores de la serie suavizada, ($S_t = 0,45 \times Y_t + 0,55 \times S_{t-1}$), las previsiones ($\tilde{Y}_t = S_{t-1}$) y los residuos ($R_t = Y_t - \tilde{Y}_t$), cuyos valores se muestran parcialmente en la tabla 6.III. En dicha tabla figuran también los valores previstos para los tiempos 51, 52 y 53, de los que ya no se dispone de datos.

La previsión para cualquier valor de $t = 51$, se ha calculado como

$$\hat{Y}_{51} = S_{50} = \lambda Y_{50} + (1-\lambda) S_{49} = 0,45 \times 37,3364 + 0,55 \times 34,4142 = 35,7292$$

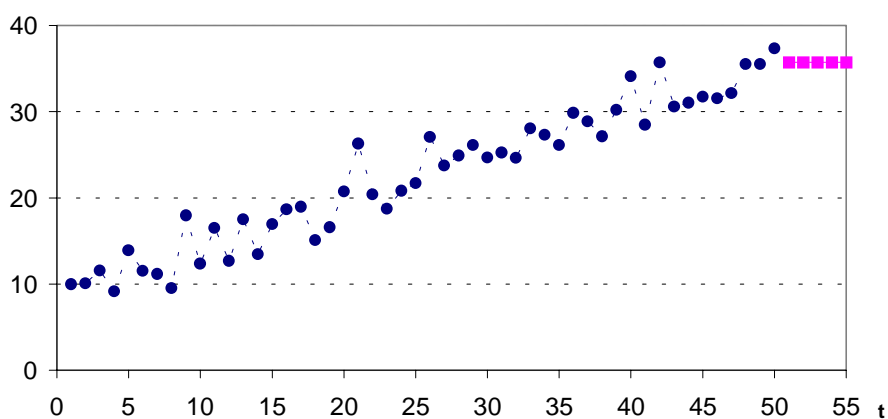
La estimación para cualquier otro valor de t superior a éste, se tendrá que hacer tomando como Y_t el valor de la previsión, ya que no se dispone de datos reales. Así

$$\hat{Y}_{51} = S_{50} = \lambda Y_{50} + (1-\lambda) S_{49} = 0,45 \times 37,3364 + 0,55 \times 34,4142 = 35,7292$$

es decir, con este sistema la previsión es idéntica para cualquier tiempo futuro, tal como se aprecia en las últimas filas de la tabla 6.III. Ello evidencia que la previsión no concuerda con la evolución cronológica presente (figura 6.5), aunque dentro del período estudiado la serie suavizada sigue de forma muy razonable a los datos disponibles.

t	Y_t	S_t	\tilde{Y}_t	R_t
1	9,958	9,958	—	—
2	10,096	10,020	9,958	0,138
3	11,552	10,709	10,020	1,532
.....
48	35,543	33,498	31,825	3,718
49	35,534	34,414	33,498	2,036
50	37,336	35,729	34,414	2,922
51	—	35,729	35,729	—
52	—	35,729	35,729	—
53	—	35,729	35,729	—

Tabla XXIII.- Datos, previsiones y residuos

Fig. 6.5.- Suavizado exponencial (●) y predicción (■), con $\lambda = 0,45$

Aquí se manifiesta la necesidad de incorporar de alguna forma la tendencia al suavizado exponencial, tal como hace el método de Brown que se expondrá en el apartado siguiente.

6.3 Método de Brown

Cuando la serie cronológica presenta tendencia, el suavizado exponencial simple no es capaz de incorporarla para poder hacer previsiones. Este problema fue abordado por Brown, que elaboró la metodología necesaria para aunar a la ponderación exponencial la estimación de la tendencia. Así, supongamos una tendencia lineal tipo

$$Y_t = a + b t + \varepsilon$$

que puede interpretarse como un componente aleatorio (ε) unido a un modelo o predicción (\tilde{Y}_t), es decir, $\tilde{Y}_t = a + b t$.

En el apartado 6.1. se ha expuesto que la serie suavizada exponencialmente puede escribirse como

$$S_t = \lambda \sum_{i=0}^{t-1} (1-\lambda)^i Y_{t-i} + (1-\lambda)^t S_0$$

y sustituyendo en ella Y_{t-i} por su expresión de tendencia, $Y_{t-i} = a + b(t-i)$, se obtiene

$$\begin{aligned} S_t &= \lambda \sum_{i=0}^{t-1} (1-\lambda)^i [a+b(t-i)] + (1-\lambda)^t S_0 = \\ &= \lambda (a+bt) \sum_{i=0}^{t-1} (1-\lambda)^i - \lambda b \sum_{i=0}^{t-1} i (1-\lambda)^i + (1-\lambda)^t S_0 \end{aligned}$$

Considerando que se dispone de suficiente información como para considerar que t es grande, la convergencia de las series anteriores es tal que

$$\sum_{i=0}^{t-1} (1-\lambda)^i \rightarrow \frac{1}{1-\lambda} \quad \sum_{i=0}^{t-1} i (1-\lambda)^i \rightarrow \frac{1-\lambda}{(1-\lambda)^2} \text{ y } (1-\lambda)^t \rightarrow 0$$

En consecuencia,

$$S_t = (a+bt) - \lambda b \frac{1-\lambda}{(1-\lambda)^2} = \tilde{Y}_t - \frac{1-\lambda}{1-\lambda} b$$

Se observa que la serie ponderada de unos datos cronológicos con tendencia lineal es una recta paralela a los datos con un desplazamiento igual a $-\frac{1-\lambda}{1-\lambda} b$.

Análogamente, la serie resultante de volver a suavizar S_t , será

$$S_t^{(2)} = \lambda S_t + (1-\lambda) S_{t-1}^{(2)}$$

que, por desarrollo análogo con el del primer suavizado, se puede expresar como

$$S_t^{(2)} = S_t - \frac{1-\lambda}{1-\lambda} b = \tilde{Y}_t - 2 \frac{1-\lambda}{1-\lambda} b$$

Restando las expresiones de S_t y $S_t^{(2)}$ se obtiene la estimación, asociada al instante t , de la pendiente de la recta que ajusta la tendencia:

$$\hat{b}_t = \frac{\lambda}{1-\lambda} (S_t - S_t^{(2)})$$

Si se dispone del valor de la serie para el tiempo t , se puede calcular la pendiente estimada en ese instante, es decir, \hat{b}_t , que representa el incremento del valor de la serie por unidad de tiempo. En ese momento, la *previsión* para un valor del tiempo igual a $t+T$ se puede obtener como el valor previsto para el tiempo t , más T veces \hat{b}_t , es decir, $\tilde{Y}_{t+T} = \tilde{Y}_t + \hat{b}_t T$.

En la ecuación anterior \tilde{Y}_t hace las veces de ordenada cuando se toma como origen del tiempo el valor t , es decir, equivale a \hat{a}_t .

A partir de esta consideración y de las expresiones del primer y segundo suavizado, se puede escribir

$$2S_t - S_t^{(2)} = 2\left[\tilde{Y}_t - \frac{1 - \lambda}{\lambda} b\right] - \left[\tilde{Y}_t - 2 \frac{1 - \lambda}{\lambda} b\right] = \tilde{Y}_t = \hat{a}_t$$

Como consecuencia, a partir de los datos disponibles hasta un cierto instante se puede predecir el inmediato siguiente. De esta manera la serie cronológica formada por las previsiones (estimaciones) de Y, según el modelo lineal suavizado, estará constituida por los valores

$$\tilde{Y}_t = \hat{a}_{t-1} + \hat{b}_{t-1} \times 1 = \hat{a}_{t-1} + \hat{b}_{t-1}$$

y los residuos, o errores de ponderación, se podrán evaluar como

$$R_t = Y_t - \tilde{Y}_t$$

Cuando la última información disponible es la del tiempo t, y se desea hacer la previsión para T unidades de tiempo a partir de este instante, suponiendo que se mantenga el mismo comportamiento de la serie, la previsión será

$$\tilde{Y}_{t+T} = \hat{a}_t + \hat{b}_t T$$

Como ejemplo, se va a aplicar esta metodología a los datos de la tabla 6.I. Para ello hay que dar valores a λ y, para cada valor de t, calcular S_t , $S_t^{(2)}$, \hat{a}_t , \hat{b}_t , \tilde{Y}_t y R_t . La evolución de los errores en función de λ se muestra en la figura 6.6.

λ	MSE	MAE
0,10	7,164	1,995
0,15	6,563	2,004
0,20	6,848	2,094
0,30	7,979	2,266
0,40	9,466	2,443
0,50	11,325	2,683
0,60	13,684	2,956
0,70	16,752	3,258
0,80	20,879	3,577
0,90	26,643	4,047

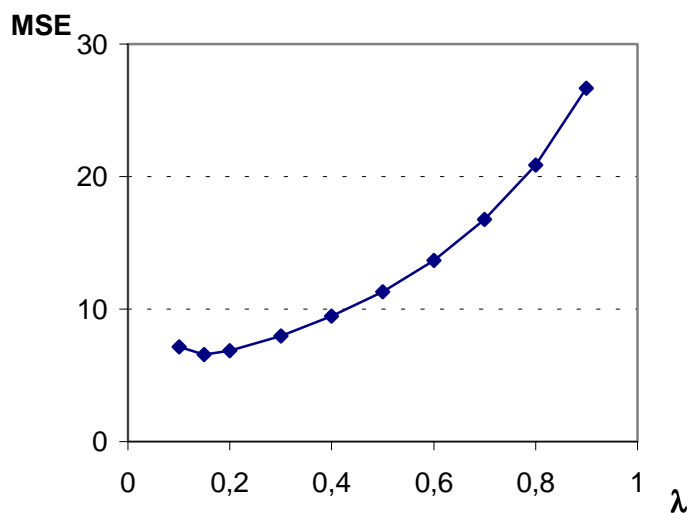


Fig. 6.6.- Selección de λ con tendencia lineal

El factor de ponderación seleccionado es $\lambda = 0,15$; con él la aplicación del método de Brown conduce a las ponderaciones que numéricamente se detallan, para los últimos datos en la tabla 6.IV.

Las previsiones desde $t = 51$ hasta $t = 55$, ($T = 1, \dots, 4$), que son las aceptables según indicó el correlograma de la figura 6.2, se obtienen a partir de la expresión de las previsiones, es decir,

$$\tilde{Y}_{t+T} = \hat{a}_t + \hat{b}_t T = 35,794 + 0,549 \times T$$

La evolución gráfica de las series suavizada y prevista se muestran en la figura 6.7, donde se observa una muy buena concordancia entre los datos reales y los suavizados, y se aprecia que la previsión sigue la tendencia marcada por la serie cronológica real.

t	Y_t	S_t	$S_t^{(2)}$	\hat{a}_t	\hat{b}_t	\tilde{Y}_t
...
48	35,543	31,216	28,526	33,906	0,475	33,278
49	35,534	31,864	29,027	34,701	0,501	34,381
50	37,336	32,685	29,576	35,794	0,549	35,202

Tabla 6.IV.- Ponderaciones con $\lambda=0,15$ y tendencia lineal

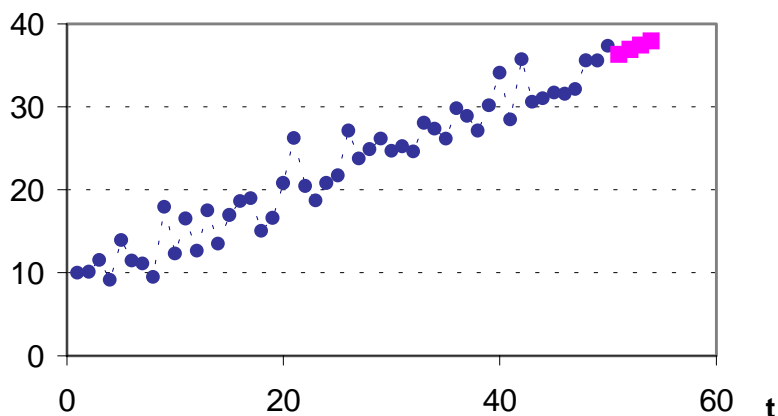


Fig. 6.7.- Suavizado exponencial de Brown (●) y previsión (■), con $\lambda = 0,15$

La figura 6.8 contiene los residuos del modelo, o sea $R_t = Y_t - \tilde{Y}_t$, y da idea de la buena concordancia entre los datos reales y el modelo resultante del suavizado exponencial de Brown. Este hecho avala la veracidad de las previsiones siempre y cuando no se modifique el patrón de comportamiento que regía durante el período de recogida de datos.

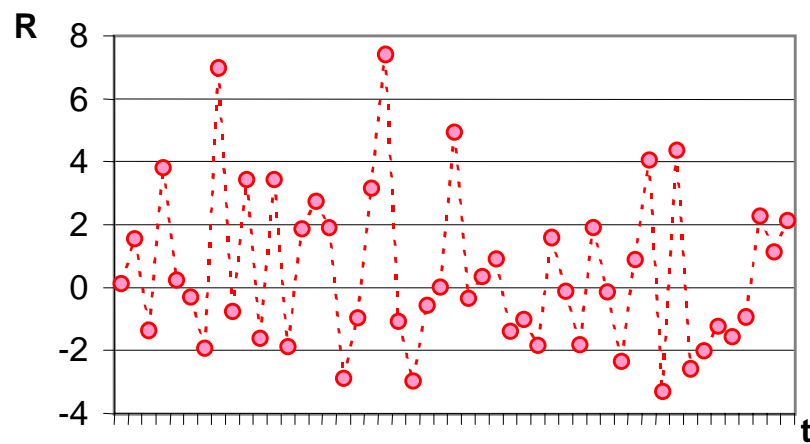


Fig. 6.8.- Residuos

En el próximo capítulo se expondrán otros casos prácticos de análisis de series temporales para efectuar su modelización y realizar las previsiones oportunas.

7 OTROS EJEMPLOS

En este capítulo se van a desarrollar algunos casos prácticos de aplicación de las técnicas propuestas anteriormente.

7.1 Ventas de papel

La tabla 7.I contiene las ventas mensuales de papel de impresión, en cientos de francos. (Forecasting. Methods and Applications.; Makridakis, Wheelwright, McGee; página 433.)

En todos los ejemplos desarrollados hasta ahora, se disponía de unos datos, se procedía a su modelización y a hacer previsiones, pero en ningún caso se han podido contrastar dichas previsiones. De los 120 valores disponibles en el presente ejemplo se van a hacer dos grupos, los 108 primeros, 9 años, serán utilizados para modelizar la serie y hacer previsiones y los últimos 12 datos, un año, se utilizarán para validar las previsiones efectuadas

t	Y	t	Y	t	Y	t	Y	t	Y
1	562,674	25	646,783	49	747,636	73	843,038	97	895,217
2	599,000	26	658,442	50	773,392	74	847,000	98	856,075
3	668,516	27	712,906	51	813,788	75	941,952	99	893,268
4	597,798	28	687,714	52	766,713	76	804,309	100	875,000
5	579,889	29	723,916	53	728,875	77	840,307	101	835,088
6	668,233	30	707,183	54	749,197	78	871,528	102	934,595
7	499,232	31	629,000	55	680,954	79	656,330	103	832,500
8	215,187	32	237,530	56	241,424	80	370,508	104	300,000
9	555,813	33	613,296	57	680,234	81	742,000	105	791,443
10	586,935	34	730,444	58	708,326	82	847,152	106	900,000
11	546,136	35	734,925	59	694,238	83	731,675	107	781,729
12	571,111	36	651,812	60	772,071	84	898,527	108	880,000
13	634,712	37	676,155	61	795,337	85	778,139	109	875,024
14	639,283	38	748,183	62	788,421	86	856,075	110	992,968
15	712,182	39	810,681	63	889,968	87	938,833	111	976,804
16	621,557	40	729,363	64	797,393	88	813,023	112	968,697
17	621,000	41	701,108	65	751,000	89	783,417	113	871,675
18	675,989	42	790,079	66	821,255	90	828,110	114	1006,852
19	501,322	43	594,621	67	691,605	91	657,311	115	832,037
20	220,286	44	230,716	68	290,655	92	310,032	116	345,587
21	560,727	45	617,189	69	727,147	93	780,000	117	849,528
22	602,530	46	691,389	70	868,355	94	860,000	118	913,871
23	626,379	47	701,067	71	812,390	95	780,000	119	868,746
24	605,508	48	705,777	72	799,556	96	807,993	120	993,733

Tabla 7.I.- Ventas mensuales de papel

Las figuras 7.1 y 7.2 muestran, respectivamente, la evolución cronológica de los datos y el correlograma. De ellas se deduce la existencia de una estacionalidad de período 12, naturalmente ligada a la evolución mensual de los datos, y una tendencia creciente en los primeros años y más estables en los últimos; esto hace pensar en la posibilidad de un

modelo parabólico. Las previsiones, según el correlograma, son aceptables a tres años vista.

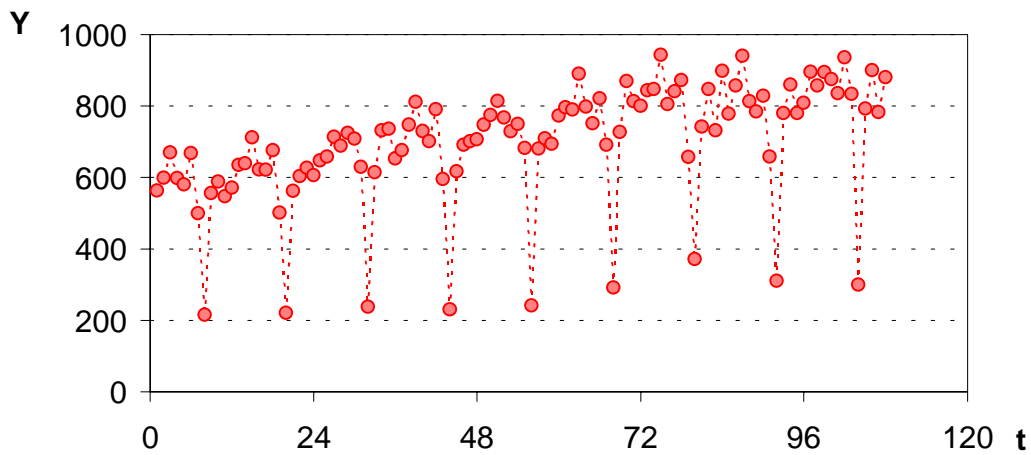


Fig. 7.1.- Evolución cronológica de los datos de los primeros 9 años

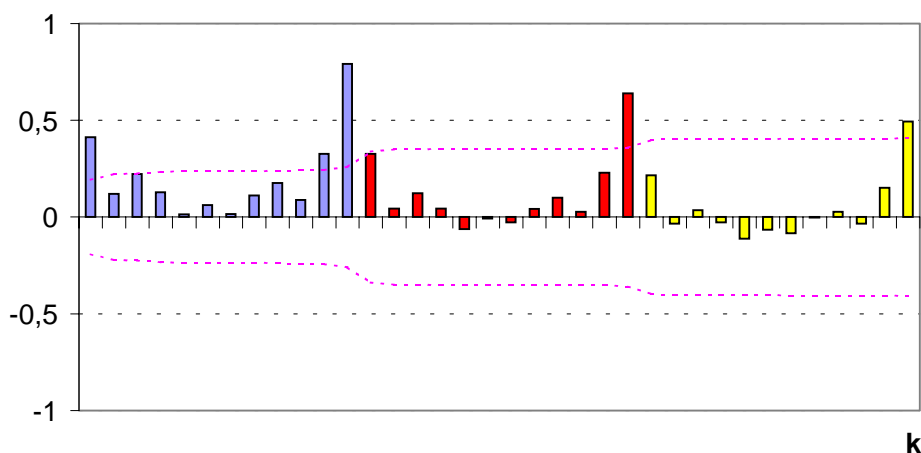


Fig. 7.2.- Correlograma

Dada la generalidad del método, para tratar modelos aditivos, multiplicativos o mixtos, se optará por la modelización con variables categóricas.

Inicialmente se plantea el modelo que incluye el término lineal y el cuadrático del tiempo, las 11 variables categóricas y su conjunción con el tiempo, es decir:

$$\begin{aligned}
 Y = & \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 Q_4 + \beta_5 Q_5 + \beta_6 Q_6 + \beta_7 Q_7 + \beta_8 Q_8 \\
 & + \beta_9 Q_9 + \beta_{10} Q_{10} + \beta_{11} Q_{11} + \beta_{12} Q_{12} + \gamma_2 Q_2 t + \gamma_3 Q_3 t + \gamma_4 Q_4 t + \gamma_5 Q_5 t \\
 & + \gamma_6 Q_6 t + \gamma_7 Q_7 t + \gamma_8 Q_8 t + \gamma_9 Q_9 t + \gamma_{10} Q_{10} t + \gamma_{11} Q_{11} t + \gamma_{12} Q_{12} t + \varepsilon
 \end{aligned}$$

Después de eliminar todos los términos no significativos, el modelo definitivamente ajustado se presenta en la tabla 7.II donde es destacable el buen ajuste ($R^2 = 95,59\%$) obtenido.

El modelo definitivo es

$$\hat{Y} = 579,34 + 4,12 t - 0,011 t^2 + 80,1576 Q_3 - 113,66 Q_7 - 399,06 Q_8 - 83,26 Q_9 - 1,55 t Q_8 - 0,83 t Q_{11}$$

es decir, se trata de un modelo mixto, en donde la ecuación de la tendencia es distinta para algunos meses. Así los meses de enero (Q_1), febrero (Q_2), abril (Q_4), mayo (Q_5), junio (Q_6), octubre (Q_{10}) y diciembre (Q_{12}) tienen el mismo comportamiento puesto que sus variables categóricas no aparecen en el modelo, hecho que los haría distinguibles entre sí. Los meses de marzo (Q_3), julio (Q_7), agosto (Q_8) y septiembre (Q_9) tienen distinta ordenada en el origen, es decir, su volumen de ventas tiene la misma evolución parabólica con el tiempo pero con distinto valor inicial de salida. Además agosto (Q_8) y noviembre (Q_{11}) tienen distinta pendiente que los demás; al ser su coeficiente negativo, indica que su velocidad de crecimiento es sistemáticamente inferior a la del resto de meses del año.

	nu	S.C.	C.M.	F	p-val
Regresión	8	2782737,3	347842,16	268,54	1,5011E-63
Residuos	99	128237,3	1295,33		
Total	107	2910974,6			

	Coeficientes	Error típico	t	p-val
Ord. Origen	570,3432	10,9786	51,9503	0,0000
t	4,1152	0,4487	9,1704	0,0000
t ²	-0,0113	0,0040	-2,8242	0,0057
Q3	80,1576	12,8019	6,2614	0,0000
Q7	-113,6589	12,8057	-8,8756	0,0000
Q8	-399,0648	25,9335	-15,3880	0,0000
Q9	-83,2560	12,8143	-6,4971	0,0000
tQ8	-1,5549	0,4051	-3,8384	0,0002
tQ11	-0,8270	0,1932	-4,2797	0,0000

$$R^2 = 0,9559$$

Tabla 7.II.- Modelo resultante

La figura 7.3. muestra el ajuste del modelo obtenido sobre los datos reales de los nueve años estudiados, y la 7.4 los residuos, diferencia entre el valor real y el modelizado, observándose un buen ajuste general a todos los puntos. En la figura 7.3, a través de los triángulos de los valores modelizados, se ve claramente la homogeneidad de comportamiento de un conjunto de meses, frente a la disparidad de los otros

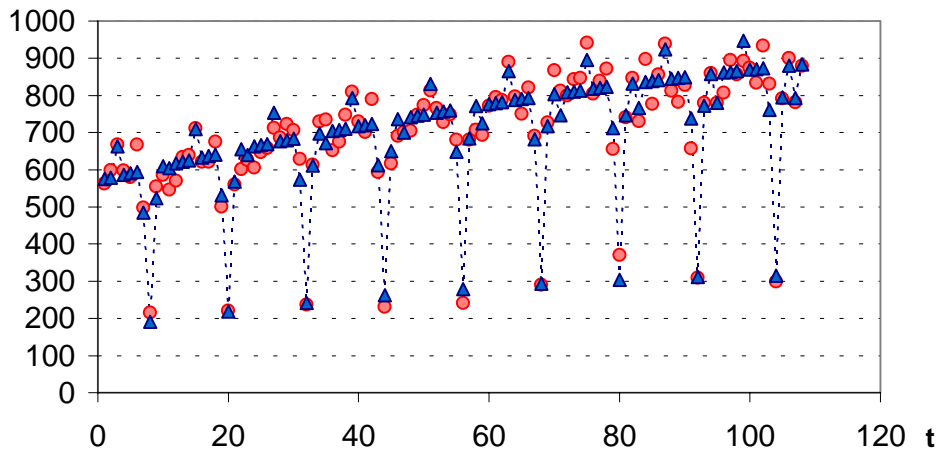


Fig. 7.3.- Datos reales (●) y modelizados (▲)

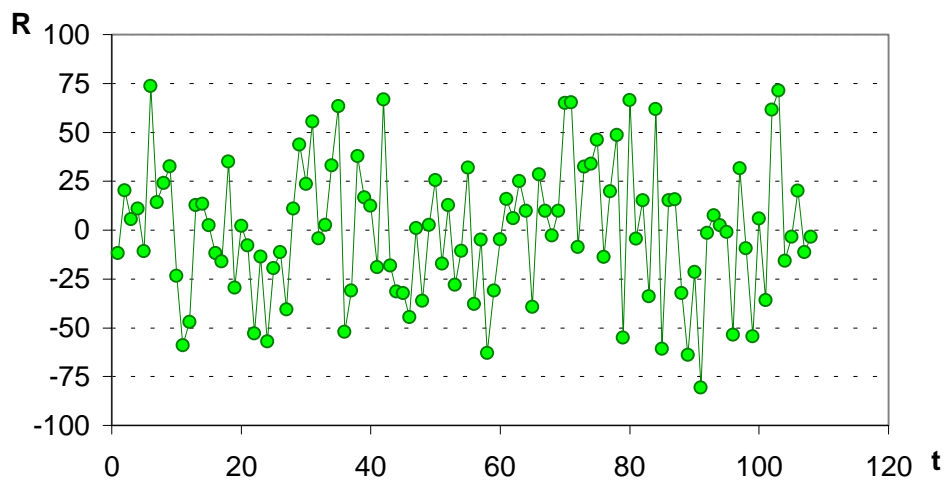


Fig. 7.4.- Residuos

Con el modelo disponible se puede proceder a hacer previsiones, por ejemplo para los próximos doce meses. En la tabla 7.III se presentan los valores previstos junto a los que realmente se obtuvieron en estos meses y que han sido separados y guardados a la hora de hacer la modelización anterior; también se han calculado las diferencias entre ellas y el porcentaje de error de previsión sobre el valor real. Hay que resaltar que las pequeñas discrepancias entre la previsión y el valor real, en parte, deben ser atribuidas al componente aleatorio, cuya presencia en cada momento no hay que olvidar. La figura 7.5 muestra ambos conjuntos de datos, de los que sólo hay que comentar la gran concordancia entre la previsión y la realidad, lo cual valida lo precedente de la aplicación de la metodología del análisis mediante variables categóricas.

t	Y	Previsión	Diferencia	%
109	875,024	884,998	-9,974	-1,14
110	992,968	886,645	106,323	10,71
111	976,804	968,430	8,374	0,86
112	968,697	889,872	78,825	8,14
113	871,675	891,451	-19,776	-2,27
114	1006,852	893,008	113,844	11,31
115	832,037	780,882	51,155	6,15
116	345,587	316,626	28,961	8,38
117	849,528	814,283	35,245	4,15
118	913,871	899,010	14,861	1,63
119	868,746	802,041	66,705	7,68
120	993,733	901,876	91,857	9,24

Tabla 7.III. - Previsiones y valores reales obtenidos en el mismo período

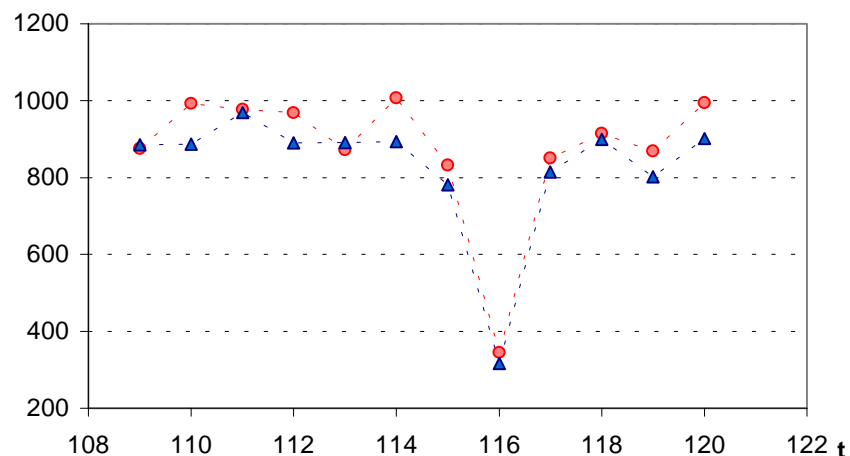


Fig. 7.5.- Previsiones (▲) y valores reales (●) del año siguiente

7.2 Generación de electricidad

En la tabla 7.IV se presenta la generación mensual de electricidad, por la industria eléctrica de Estados Unidos, durante 8 años, (Forecasting. Methods and Applications.; Makridakis, Wheelwright, McGee; página 469).

Como en el caso anterior, de los 96 valores disponibles en el presente ejemplo se van a hacer dos grupos, los 84 primeros, 7 años, serán utilizados para modelizar la serie y hacer previsiones y los últimos 12 datos, resaltados en cursiva en la tabla y que corresponden al último año, se utilizarán para validar las previsiones efectuadas

t	Y	t	Y	t	Y	t	Y
1	144,58	25	164,33	49	196,37	73	209,69
2	137,3	26	147,08	50	162,73	74	186,35
3	140,06	27	155,48	51	169,16	75	182,85
4	132,14	28	146,22	52	156,85	76	169,96
5	137,75	29	153,23	53	169,33	77	178,07
6	145,52	30	162,44	54	180,79	78	186,68
7	147,85	31	176,82	55	198,92	79	202,25
8	162,82	32	179,72	56	196,09	80	204,85
9	147,36	33	155,22	57	176,26	81	180,75
10	143,74	34	154,94	58	166,39	82	179,71
11	143,87	35	152,79	59	167,07	83	177,5
12	154,35	36	169,35	60	184,21	84	188,71
13	157,24	37	178,31	61	197,83	85	200
14	142,46	38	156,67	62	173,5	86	188,72
15	150,02	39	164,16	63	173,19	87	187,47
16	142,02	40	153,15	64	159,74	88	168,72
17	153,49	41	157,35	65	175,24	89	175,73
18	156,13	42	173,36	66	188,31	90	189,43
19	177,91	43	186,41	67	202,68	91	216,78
20	173,81	44	186,38	68	206,41	92	215,39
21	152,16	45	164,97	69	185,57	93	191,48
22	151,87	46	163,63	70	175,8	94	178,56
23	149,73	47	168,99	71	176,17	95	178,55
24	159,6	48	183,09	72	191,87	96	195,59

Tabla 7.IV.- Generación mensual de electricidad

Las figuras 7.6 y 7.7 corresponden, respectivamente, a la evolución cronológica de los datos de los siete años utilizados en la modelización, y su correlograma. Al ser los datos mensuales, parece que de haber alguna estacionalidad, ésta debería ser de período 12, es decir anual, sin embargo no es esta la situación del caso estudiado. El correlograma muestra de forma inequívoca que existe una estacionalidad de período 6, la cual, aunque menos evidente, también se detecta en la evolución cronológica. Intentando buscar una explicación a dicha periodicidad, quizás sería posible pensar que el consumo eléctrico, y por tanto la electricidad generada por las centrales, tiene similar comportamiento en los meses cálidos que en los fríos como consecuencia del consumo por los aires acondicionados, mientras que la parte del consumo atribuible a la industria en general, no tiene porque presentar distinto comportamiento de un mes a otro. En cuanto a la tendencia, es posible que sea cuadrática, tal como se detecta en la evolución cronológica de los datos.

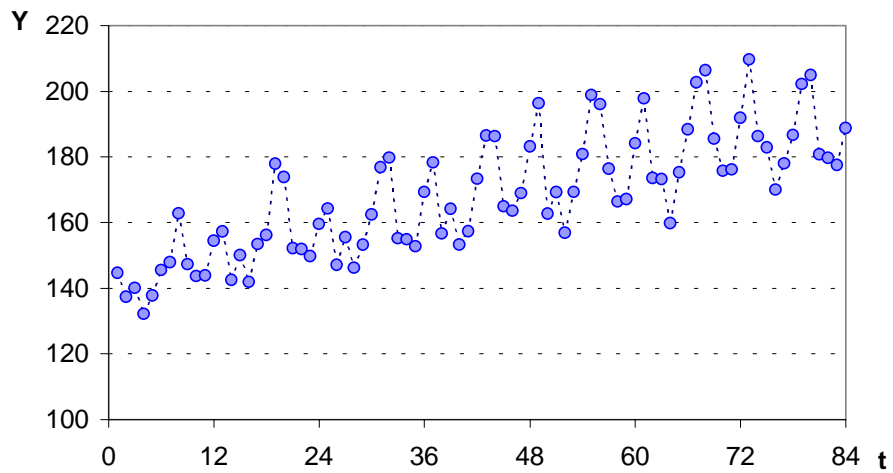


Fig.7.6.- Evolución cronológica de la electricidad generada en siete años

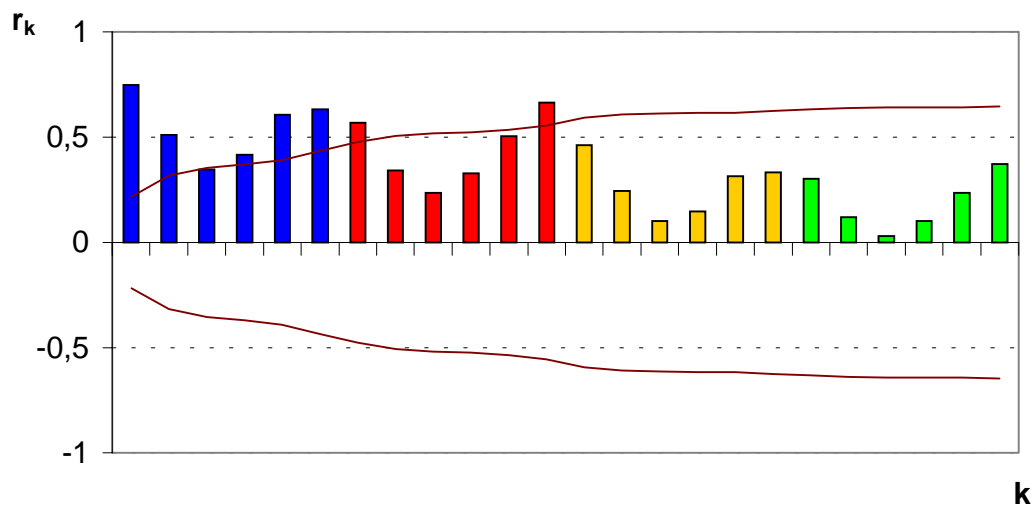


Fig.7.7.- Correlograma

El modelo inicial por variables categóricas debe incluir la tendencia, la estacionalidad, reflejada en las variables Q, y sus posibles cambios en el tiempo, términos $Q_j t$, es decir:

$$Y = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \beta_2 Q_2 + \beta_3 Q_3 + \beta_4 Q_4 + \beta_5 Q_5 + \beta_6 Q_6 + \gamma_2 Q_2 t + \gamma_3 Q_3 t + \gamma_4 Q_4 t + \gamma_5 Q_5 t + \gamma_6 Q_6 t + \varepsilon$$

Después de eliminar los términos no significativos, regresión paso a paso, el modelo resultante se presenta en la tabla 7.V. Hay que destacar especialmente el buen ajuste, reflejado en un coeficiente de determinación, R^2 , del 86,64 %.

	nu	S.C.	C.M.	F	p-val
Regresión	7	24674,8398	3524,9771	70,4316	1,2626E-30
Residuos	76	3803,6637	50,0482		
Total	83	28478,5035			

	Coefficientes	Error típico	t	p - val
Ord. Origen	143,8383	1,7040	84,4105	0,0000
t	0,9072	0,0511	17,7608	0,0000
Q4	-8,7533	4,2164	-2,0760	0,0413
tQ2	-0,2042	0,0567	-3,6025	0,0006
tQ3	-0,4053	0,0562	-7,2125	0,0000
tQ4	-0,4006	0,0934	-4,2904	0,0001
tQ5	-0,4800	0,0553	-8,6822	0,0000
tQ6	-0,2707	0,0549	-4,9352	0,0000

$$R^2 = 0,8664$$

Tabla 7.V.- Modelo resultante

En consecuencia, el modelo es

$$\hat{Y} = 143,838 + 0,907 t - 8,753 Q_4 - 0,204 Q_2 t - 0,405 Q_3 t - 0,401 Q_4 t - 0,480 Q_5 t - 0,271 Q_6 t$$

y de él se deduce que el término cuadrático del tiempo no ha resultado significativo, pero la serie presente una tendencia lineal y creciente, pendiente igual a 0,907. La figura 7.8 muestra la modelización de la energía eléctrica producida para cada uno de los 6 meses que componen un período. El comportamiento estacional no es fácil de interpretar; el cuarto mes de cada período pierde 8,753 unidades (coeficiente de Q_4) respecto a todos los demás, que tienen una ordenada en el origen común. En cada mes la evolución con el tiempo es distinta, pero frente al primer mes al que correspondería un crecimiento asociado a una pendiente de 0,907, se podrían hacer dos agrupaciones, los meses 2 y 6, con una pérdida de pendiente del orden de $0,2 \div 0,3$, (rectas casi coincidentes en la figura 7.8) y los 3, 4 y 5 con una pérdida, respecto al primero, de $0,4 \div 0,5$ unidades en la pendiente, que dan lugar a una evolución similar, tal como se observa en el gráfico.

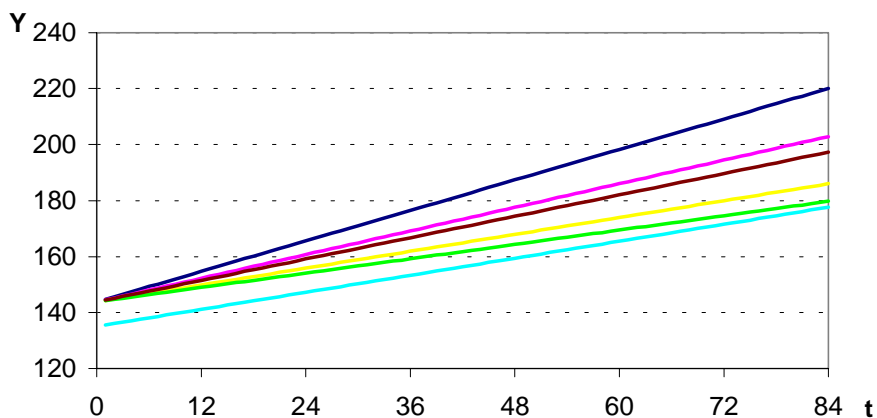


Fig. 7.8.- Modelo para cada componente del período estacional

La figura 7.9 contiene el gráfico del modelo ajustado sobre los datos iniciales, y la 7.10 los residuos. En ambas se puede apreciar la buena adecuación del modelo a los datos. Quizás se podría destacar que en el inicio de los datos (figura 7.9), hay algunos valores reales de producción de energía eléctrica claramente distintos, en cuanto a su comportamiento del resto. Esto incide en un mayor valor del residuo, y si no es posible conocer las causas de estas producciones más altas que lo previsto deberían atribuirse al azar.

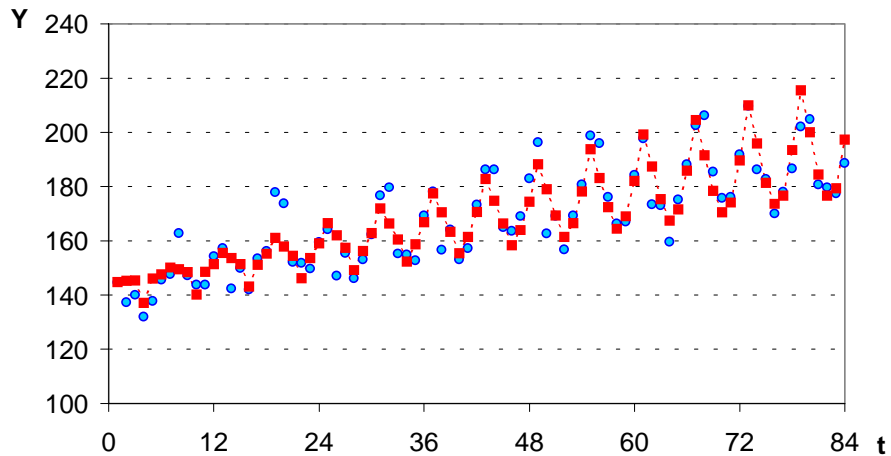


Fig. 7.8.- Datos reales ● y modelizados ■

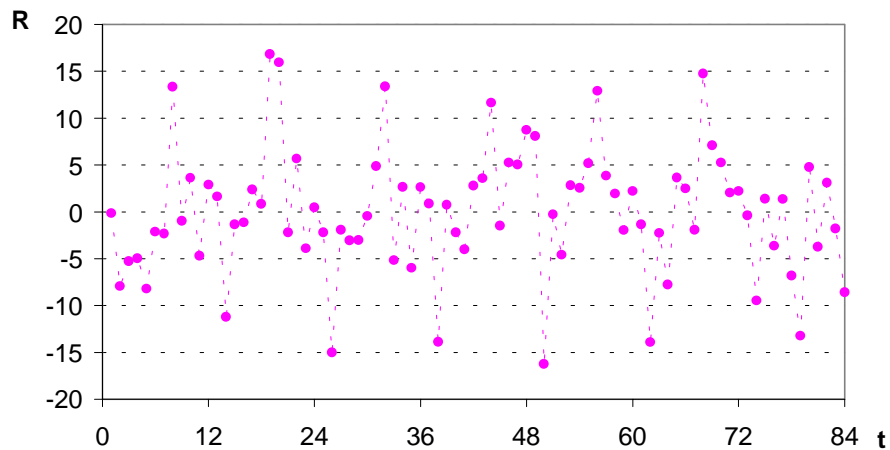


Fig. 7.9.- Residuos

Con el modelo resultante, se pueden hacer previsiones para un período no superior al año, según evidencia el correlograma. Esto es suficiente para comprobar la adecuación de la previsión, según el modelo obtenido, con los 12 datos reales de los que se dispone y que han sido reservados en la modelización para comprobar la fiabilidad de las predicciones. La tabla 7.VI contiene los valores reales las previsiones según el modelo obtenido y el error respecto al valor medido, salvo el primero, los demás presentan una buena aproximación

t	Y	Previsión	Diferencia	%
85	200	220,95	-20,95	-10,48
86	188,72	204,30	-15,58	-8,25
87	187,47	187,50	-0,03	-0,02
88	168,72	179,67	-10,95	-6,49
89	175,73	181,86	-6,13	-3,49
90	189,43	201,12	-11,69	-6,17
91	216,78	226,40	-9,62	-4,44
92	215,39	208,52	6,87	3,19
93	191,48	190,51	0,97	0,50
94	178,56	182,70	-4,14	-2,32
95	178,55	184,00	-5,45	-3,05
96	195,59	204,31	-8,72	-4,46

Tabla 7.VI.- Previsiones y valores reales obtenidos en el mismo período

La figura 7.10 muestra ambos conjuntos de valores, en donde a pesar de la buena aproximación, es destacable que los valores reales, casi siempre son inferiores a los previstos.

En esta situación cabría la posibilidad de pensar en un posible cambio en el comportamiento del modelo, es decir, quizás la tendencia al crecimiento ha sufrido un frenazo en el último año; esto se confirmaría con la evolución futura, sin embargo si se examina la historia, se puede detectar que en los últimos 12 meses incluidos en la modelización, los residuos negativos predominan sobre los positivos, figura 7.9, o sea que, en muchos casos, los valores del modelo superan a los valores reales. La figura 7.11, une a los datos reales y las previsiones, las líneas de evolución del modelo en cada componente del período, mostrando el mismo hecho de una situación de los valores reales por debajo de la previsión.

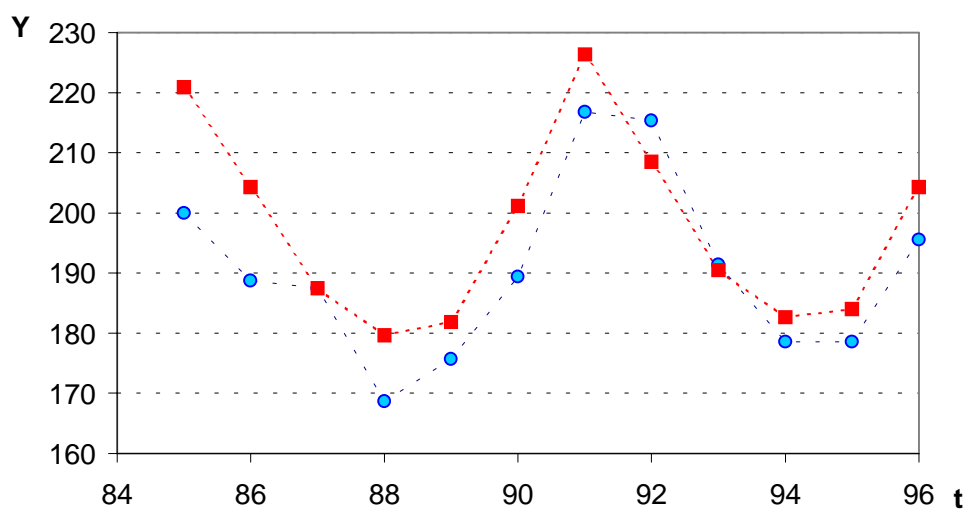


Fig. 7.10.- Previsiones ■ y valores reales ● del año siguiente

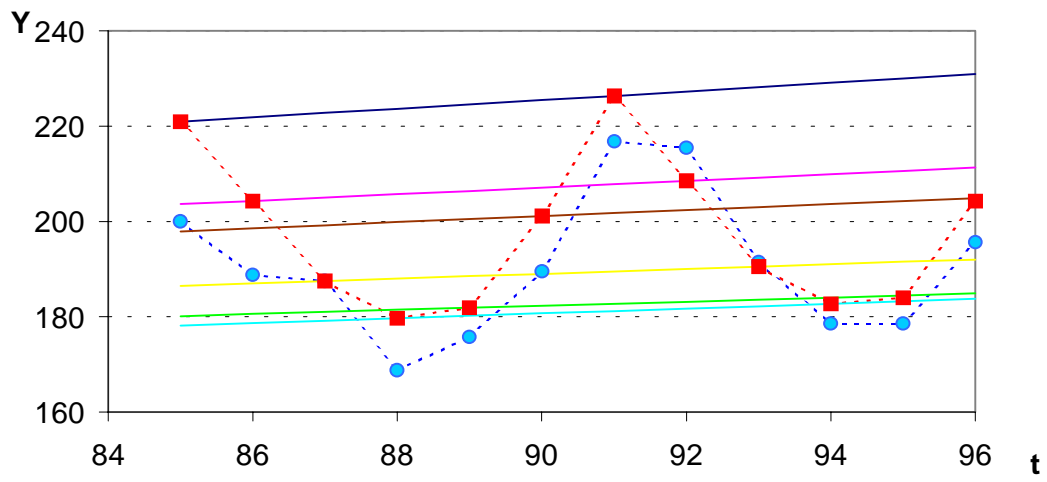


Fig. 7.11.- Previsiones y valores reales junto al modelo