

## Capítulo 2

# Análisis descriptivo de una serie temporal

**Sir Arthur Schuster** (1851–1934)

Físico inglés, nacido alemán. Fue profesor de matemáticas y física en la Universidad de Manchester. Hizo contribuciones fundamentales a la espectroscopía y al estudio de las periodicidades e inventó el periodograma. Dirigió en 1975 la expedición de la Royal Society para observar el eclipse solar en Tailandia y fue nombrado Lord por sus méritos científicos.

## 2.1 Introducción

En este capítulo presentamos los procedimientos descriptivos desarrollados entre 1940 y 1970 para analizar series temporales. El objetivo que se pretende es explicar sus valores pasados y prever sus valores futuros. Estos métodos tienen poca utilidad directa hoy en día, ya que disponemos de métodos más generales y eficaces que se exponen en los capítulos siguientes y que los incluyen como caso particular. De todas formas, hemos considerado interesante introducirlos por dos razones. En primer lugar, los métodos actuales se entienden mejor si se conocen las limitaciones de otros métodos más simples desarrollados para obtener predicciones de una serie. En segundo lugar, los métodos que se presentan en este capítulo están todavía disponibles como opciones de análisis en muchos programas de ordenador de uso frecuente, y es conveniente que el lector conozca sus bases y sus limitaciones.

Los ejemplos de series presentados en el capítulo 1 ponen claramente de manifiesto que los procedimientos estudiados para describir variables estáticas no son en general apropiados. Por ejemplo, la media de una variable estática, como las estaturas de los estudiantes en una clase, estima la media de la población de estudiantes y tiene una interpretación clara. Sin embargo, si hacemos la media de los datos de una serie con tendencia, esta media no refleja ninguna característica constante y tiene difícil interpretación. Esto llevó a plantear los modelos de análisis de tendencias deterministas, que constituye una extensión inmediata de los métodos de regresión. La aplicación de estos métodos para obtener previsiones de una serie conduce en general a altos errores de predicción, lo que mostró enseguida sus limitaciones. El problema principal es que en una serie temporal la observación hoy,  $z_t$ , depende en general de sus valores previos,  $z_{t-1}, z_{t-2}, \dots$ , siendo esta dependencia más fuerte con los datos más recientes y más débil con los más alejados. Esta característica no se tiene en cuenta en los modelos de análisis de tendencias deterministas. Una alternativa más fructífera fueron los métodos de alisado, que se establecen para imponer una estructura de dependencia que disminuye con el tiempo. Estos métodos se introdujeron en los años 60 y se extendieron gracias a la aparición del ordenador digital.

Para series estacionales se desarrollaron procedimientos basados en modelar separadamente la evolución de la serie y el efecto estacional. La extensión de los análisis de tendencias deterministas para estas series conduce a los métodos de decomposición, que se presentan a continuación. La extensión de los métodos de alisado condujo a métodos de doble alisado, uno para la evolución de la serie y otro para el efecto

estacional, como el método de Holt-Winters. Paralelamente a estos métodos se desarrollaron procedimientos para series periódicas, como lo son las series climáticas donde domina el efecto estacional. La extensión de estos procedimientos permitió representar cualquier serie estable como una suma de efectos cíclicos, y dió lugar al periodograma, que es una herramienta muy útil para detectar ciclos periódicos deterministas en una serie temporal.

## 2.2 El análisis de tendencias deterministas

### 2.2.1 Tipos de modelos

Comenzaremos presentando el análisis de tendencias deterministas en series sin estacionalidad, como las de las figuras 1.1 a 1.5 del capítulo anterior. Este método es una aplicación inmediata de las ideas de regresión para relacionar variables estáticas. Consiste en representar una serie observada,  $z_t$ , donde  $t = 1, \dots, T$ , mediante

$$z_t = \mu_t + a_t \quad (2.1)$$

donde

$$\mu_t = f(t, \beta)$$

es el nivel de la serie que es una función conocida determinista del tiempo que depende del instante de tiempo considerado y de un vector de parámetros,  $\beta$ . Estos parámetros deben estimarse a partir de los datos, como veremos a continuación. La ecuación (2.1) establece que el modelo descompone la variable observada,  $z_t$ , como suma de dos componentes. El primero,  $\mu_t$ , es el *nivel de la serie* que se supone es una función determinista del tiempo. El segundo,  $a_t$ , suele denominarse la *innovación* y es un componente aleatorio que recoge todos los demás efectos que actúan sobre la serie. Se supone que las variables aleatorias  $a_t$  tienen una estructura constante a lo largo del tiempo: para cualquier instante, media cero, varianza constante y distribución normal. Además las observaciones correspondientes a dos periodos de tiempo distintos son independientes, es decir, conocer el valor de  $a_t$  no proporciona ninguna información sobre el posible valor de  $a_{t+1}$ . Sea cual sea el valor de  $a_t$  la variable  $a_{t+1}$  proviene al azar de una distribución normal con media cero y varianza constante. En este sentido si conocemos la distribución de la que provienen los valores particulares que observemos de las variables  $a_t$  no contienen información, ya que son valores independientes al azar de una distribución conocida que no proporcionan información sobre los posibles valores futuros.

La predicción de la serie con este modelo en un periodo futuro,  $T + k$  se obtiene extrapolando el nivel de la serie  $\mu_t$ . Llamando  $\widehat{z}_T(k)$  a la predicción realizada desde el origen  $T$  para  $k$  periodos en adelante, es decir la predicción del valor  $z_{T+k}$  con la información disponible hasta el momento  $T$ , tenemos que

$$\widehat{z}_T(k) = \mu_{T+k} = f(T + k, \beta). \quad (2.2)$$

La forma que establezcamos para la evolución del nivel de la serie a lo largo del tiempo determina el modelo concreto utilizado. Por ejemplo, el modelo más simple posible supone que el nivel de la serie es constante en el tiempo, es decir  $\mu_t = \mu_{t-1} = \mu$ , y se conoce como el modelo de nivel constante o sin *tendencia*. Viene dado por:

$$z_t = \mu + a_t \quad (2.3)$$

y la serie oscilará alrededor de su media que es constante. Las series de las figuras 1.1 y 1.2 no tienen tendencia y podrían explicarse por este modelo. Al ser el nivel de la serie constante, y no depender de  $t$ , la predicción con este modelo para cualquier horizonte será  $\mu$ .

Un modelo más general, que se aplica a series que tienen tendencia, creciente o decreciente, aproximadamente constante, es el modelo de *tendencia lineal*. La figura 1.3 presenta una serie que podría tener esta propiedad. El modelo para la serie es entonces:

$$z_t = \mu_t + a_t$$

donde ahora el nivel no es constante y viene dado por:

$$\mu_t = f(t, \beta) = \beta_0 + \beta_1 t \quad (2.4)$$

donde ahora  $\beta_1$  representa la pendiente de la recta que describe el nivel de la serie y es el crecimiento esperado entre dos periodos de la serie. La predicción con este modelo del valor de la serie en el instante  $T + k$  con información hasta  $T$  será

$$\widehat{z}_T(k) = \beta_0 + \beta_1(T + k) \quad (2.5)$$

Estos dos modelos son casos particulares del modelo más general de *tendencia polinómica*, donde el nivel de la serie evoluciona según un polinomio de orden  $r$

$$z_t = \mu_t + a_t$$

y ahora

$$\mu_t = \beta_0 + \beta_1 t + \dots + \beta_r t^r \quad (2.6)$$

Los dos casos estudiados anteriormente corresponden a  $r = 0$ , con lo que obtenemos el modelo de nivel constante, y a  $r = 1$ , que conduce al modelo de tendencia lineal. En la práctica  $r \leq 2$ .

### 2.2.2 Estimación

El ajuste de estos modelos a una serie temporal requiere la estimación del vector de parámetros,  $\beta = (\beta_0, \dots, \beta_r)$ . Los parámetros se estiman por el criterio de mínimos cuadrados, es decir, minimizando las diferencias entre los valores observados y los previstos a horizonte uno por el modelo. Este criterio se escribe como

$$\text{minimizar } \sum_{t=1}^T (z_t - f(t, \beta))^2 \quad (2.7)$$

Los valores de los parámetros que minimizan esta expresión se obtienen fácilmente en el caso de modelos de tendencias polinómicas. Vamos a ilustrar su estimación en el caso del modelo lineal. Para simplificar la exposición vamos a suponer que la variable  $t$  se construye para que tenga media cero. Esto es especialmente simple si suponemos que existen  $T = 2n + 1$  datos (períodos observados), de manera que podemos definir  $t$  como  $t = (-n, -n + 1, \dots, -1, 0, 1, \dots, n - 1, n)$ . El criterio es:

$$\text{minimizar } \sum_{t=1}^T (z_t - \beta_0 - \beta_1 t)^2 \quad (2.8)$$

y derivando respecto a  $\beta_0$  y  $\beta_1$  e igualando a cero las derivadas, se obtienen las ecuaciones habituales de mínimos cuadrados. Comenzando con la ecuación obtenida al igualar a cero la derivada de (2.8) respecto a  $\beta_0$  y llamando  $\widehat{\beta}_0$  y  $\widehat{\beta}_1$  a los valores que verifican esta ecuación, se obtiene:

$$\sum_{t=1}^T (z_t - \widehat{\beta}_0 - \widehat{\beta}_1 t) = 0 \quad (2.9)$$

Llamando residuos a las estimaciones de las innovaciones, que según las ecuaciones (2.1) y (2.4) vienen dadas por

$$\widehat{a}_t = z_t - \widehat{\beta}_0 - \widehat{\beta}_1 t$$

La ecuación (2.9) puede interpretarse como que los residuos estimados deben de tener media cero. Entonces, como  $\bar{t} = 0$ , llamando  $\bar{z}_t = \sum z_t / T$  a la media de los datos observados, la ecuación (2.9) se reduce a:

$$\widehat{\beta}_0 = \bar{z}_t \quad (2.10)$$

Observemos que este es también el resultado de estimar  $\beta_0 = \mu$  en el modelo de nivel constante (2.3). Para estimar la pendiente, derivando en (2.8) respecto a  $\beta_1$  e igualando a cero el resultado, tenemos que:

$$\sum_{t=1}^T (z_t - \widehat{\beta}_0 - \widehat{\beta}_1 t)t = 0 \quad (2.11)$$

y esta ecuación puede interpretarse como la covarianza entre los residuos y el tiempo es cero, es decir, los residuos no tienen relación lineal con el tiempo. Sustituyendo (2.10) y despejando  $\beta_1$  se obtiene, dado que  $\bar{t} = 0$ , la expresión:

$$\hat{\beta}_1 = \frac{\sum_{t=1}^T (z_t - \bar{z}_t)t}{\sum_{t=1}^T t^2} \quad (2.12)$$

A continuación presentamos dos ejemplos de la utilización de estos modelos.

### Ejemplo 2.1

Vamos a ajustar el modelo de nivel o media constante (2.3) a la serie de las leguas diarias recorridas por la flota de Colón. La media de los datos observados de la serie es

$$\bar{z}_t = \frac{9 + \dots + 49}{34} = 31.82$$

que es, además, según el modelo (2.3) la predicción de la distancia que se recorrerá el día siguiente. Los errores de esta predicción dentro de la muestra son los residuos,  $a_t$  estimados como diferencia entre el valor de la serie y su media, que se presentan en la figura 2.1. El promedio de estos residuos es una medida del error esperable con el modelo constante utilizado. Como los errores pueden ser positivos y negativos y nos interesa su magnitud absoluta, calcularemos, como es habitual en estadística, el promedio de los errores al cuadrado y tomaremos después la raíz cuadrada. Se obtiene entonces

$$\hat{\sigma}_a = \sqrt{\frac{(-22.82)^2 + \dots + (17.18)^2}{34}} = 16.6$$

que indica que el error promedio en la predicción con este modelo es de 16.6 leguas diarias.

Figura 2.1: Residuos del modelo de tendencia constante para la serie de leguas recorridas por Colón

### Ejemplo 2.2

Vamos a ajustar el modelo de tendencia lineal a los datos de la población española de mayores de 16 años de la figura 1.4. Para crear una variable temporal de media cero, como hay 96 datos creamos primero una variable que tome los valores 1,..., 96 y a continuación restamos a cada uno de estos valores la media de la secuencia, de modo que el conjunto de datos tenga media cero. Esta operación proporciona los valores de  $t = (-47.5, -46.5, \dots, -1, 0, 1, \dots, 47.5)$ . Utilizando las fórmulas (2.10) y (2.12), se obtiene la ecuación de predicción

$$\hat{z}_t = 29693 + 79.7t$$

Esta recta indica que, en el periodo estudiado, el número de mayores de 16 años es en promedio 29.693 millones de personas y que cada año este número se incrementa en 79.700 personas aproximadamente. El modelo parece tener un excelente ajuste ya que el coeficiente de correlación es .995.

Sin embargo, la figura 2.2, que muestra los datos y el modelo ajustado, indica que un ajuste de crecimiento lineal no es muy bueno, porque la tendencia no es exactamente constante y ha ido cambiando suavemente en el tiempo. En particular, las predicciones generadas en el año 2000 para los dos años siguientes van considerablemente por arriba de los datos observados. Este problema aparece más claramente en la figura 2.3, que presenta los residuos de este modelo calculados mediante

$$\hat{a}_t = z_t - 29693 - 79.7t$$

Podemos observar que estos residuos contienen claramente información, lo que lleva a concluir que el modelo esta mal especificado. Podría pensarse que el problema es que la tendencia sigue un polinomio de segundo grado y que la forma de los residuos, primero negativos, luego positivos y de nuevo negativos, indica esta necesidad de una ecuación de segundo grado para reflejar esta curvatura. Si estimamos un modelo de segundo grado la ecuación resultante es:

$$z_t = 29910 + 79.7t - 0.2823t^2.$$

Los residuos de este modelo se presentan en la figura 2.4, donde se observa que aumentar el grado del polinomio no resuelve el problema. Los errores de predicción para los últimos datos muestrales son muy altos, y puede comprobarse que las predicciones de este modelo para los datos de 2001 son muy malas. El problema es que no estamos teniendo en cuenta que la predicción de un dato futuro debe basarse principalmente en los últimos valores observados de la serie, lo que no es posible tener en cuenta con los modelos de tendencias deterministas.

Figura 2.2: Población mayor de 16 años en España y el modelo lineal ajustado

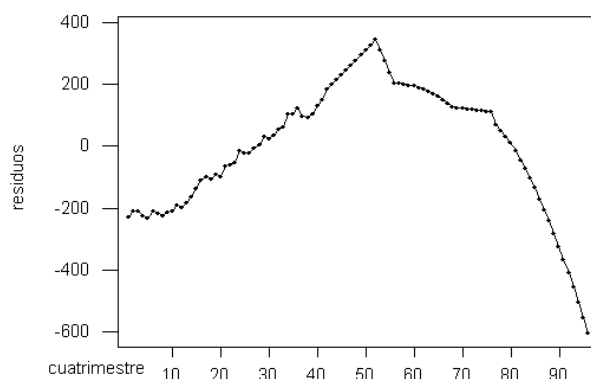


Figura 2.3: Residuos del modelo lineal para la población española de mayores de 16 años.

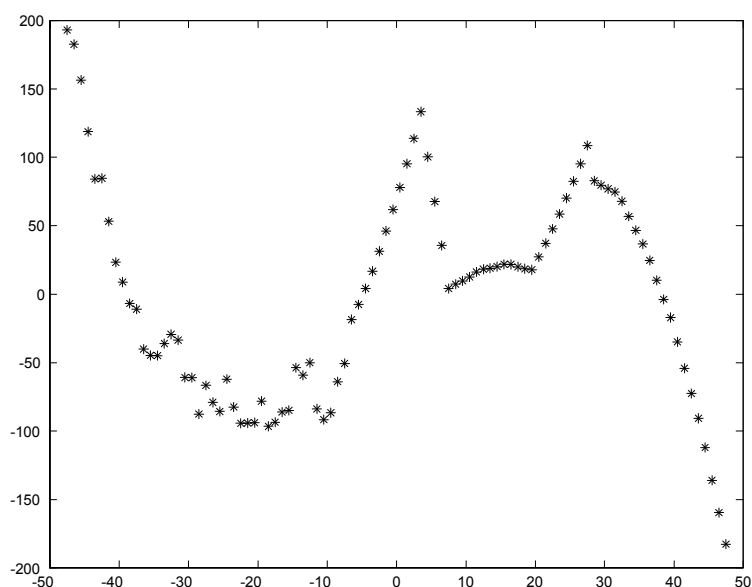


Figura 2.4: Residuos del ajuste cuadrático a los datos de la población mayor de 16 años

### 2.2.3 Limitaciones de los modelos de tendencias deterministas

Como se ha visto en los ejemplos anteriores, la limitación principal de estos modelos es que, aunque las series con nivel constante son frecuentes, es muy raro que una serie real tenga una tendencia determinista lineal, o en general polinómica con  $r \geq 1$ . Esto sólo suele ocurrir en series construidas para que tengan esta propiedad, pero no en series reales que evolucionan como resultado de un conjunto complejo de causas. Por ejemplo, la evolución de la serie de nacimientos en la figura 1.4 o la de los precios del trigo en la figura 1.5 no pueden representarse por tendencias polinómicas. Una posibilidad es intentar ajustar tendencias lineales

por tramos, es decir cortar la serie en tramos que tengan, aproximadamente, tendencia constante y ajustar en cada tramo un modelo lineal o constante. Este procedimiento puede servir para describir la evolución histórica de algunas series, pero es poco útil para predecir los valores futuros, ya que no sabemos que modelo aplicar para el futuro.

Un inconveniente adicional de ajustar una tendencia lineal en un intervalo es que el modelo de crecimiento implícito es poco razonable. Para justificar esta afirmación, veamos cuales son las implicaciones de que una serie se haya generado por un modelo con tendencia lineal determinista. En este caso, la predicción del crecimiento de la serie entre  $z_t$  y  $z_{t+1}$ , medido por  $z_{t+1} - z_t$  será la pendiente estimada,  $\hat{\beta}_1$ . Hemos visto que la pendiente se estima por la ecuación (2.12), y puede demostrarse (véase Peña, 1995) que esta estimación puede siempre escribirse como una media ponderada de los crecimientos que la serie ha experimentado en el periodo muestral donde estamos ajustando la recta. Los pesos de los crecimientos pasados para prever el crecimiento futuro son tales que el mayor peso se atribuye al crecimiento en el centro del periodo observado, y el peso mínimo al crecimiento en los extremos del intervalo. Es decir, llamando  $\nabla z_t = z_t - z_{t-1}$  al crecimiento observado de la serie en el instante  $t$ , se demuestra que

$$\hat{\beta}_1 = \sum_{t=2}^T w_t \nabla z_t \quad (2.13)$$

donde los pesos  $w_t$  son números positivos que suman uno y que son simétricos respecto al centro de la muestra. Los valores máximos de  $w_t$  corresponden a los incrementos centrales y van disminuyendo simétricamente hasta alcanzar el valor mínimo en los extremos.

Por ejemplo, supongamos una serie  $z_t$  con 5 datos en los instantes,  $t = -2, -1, 0, 1, 2$ , donde hemos construido la variable tiempo de manera que  $\bar{t} = 0$ , para simplificar la presentación. Llamemos entonces a los valores de la serie  $z_{-2}, z_{-1}, z_0, z_1, z_2$ . Aplicando las fórmulas (2.10) y (2.12) para el cálculo de la pendiente, se obtiene que

$$\begin{aligned} \sum_{t=-2}^2 t^2 &= (-2)^2 + (-1)^2 + (0)^2 + (1)^2 + (2)^2 = 10 \\ \sum t z_t &= -2z_{-2} - z_{-1} + z_1 - 2z_2 = 2(z_{-1} - z_{-2}) + 3(z_0 - z_{-1}) \\ &\quad + 3(z_1 - z_0) + 2(z_2 - z_1) \end{aligned}$$

y por tanto:

$$\hat{\beta}_1 = 0.2\nabla z_{-1} + 0.3\nabla z_0 + 0.3\nabla z_1 + 0.2\nabla z_2$$

que es un caso particular de (2.13). Esta ecuación indica que la predicción del crecimiento futuro es una media ponderada de los crecimientos observados en cada uno de los periodos, con pesos simétricos respecto al centro y tales que el peso mínimo corresponde a los crecimientos en los periodos extremos.

La implicación de este resultado es la siguiente: si admitimos que una serie tiene una tendencia determinista lineal estamos diciendo que la predicción futura de su crecimiento debe hacerse ponderando los crecimientos observados, pero dando peso mínimo al último valor observado. Este peso es, además, igual al que se atribuye al crecimiento más alejado en el tiempo, es decir, al primer crecimiento observado en la muestra. Además, si aumentamos el tamaño muestral, el peso de los crecimientos en los últimos años disminuye, pero siempre manteniéndose igual al peso de los primeros crecimientos observados. Esto es claramente poco razonable en muchas ocasiones. Por ejemplo, si admitiésemos que la población crece linealmente y tuviésemos sólo los últimos cinco años de datos, el incremento de población con respecto al último valor observado se obtendría sumando un crecimiento previsto que se obtiene dando pesos 20%, 30%, 30% y 20% a los cuatro crecimientos anuales observados en los cinco años. Sin embargo, si aumentamos el tamaño muestral y pasamos a tener una muestra de 100 datos anuales, el peso del último año pasa a tener un valor muy pequeño, e igual al crecimiento observado hace 100 años!

## 2.3 Métodos de alisado

Para resolver las limitaciones de los modelos con tendencias deterministas se introdujeron en los años 60 los métodos de alisado. La idea de estos métodos es permitir que los últimos datos de la serie tengan más peso en las predicciones que los valores más antiguos. Esto se consigue permitiendo que los parámetros del modelo de tendencias deterministas no sean constantes, sino que puedan ir variando en el tiempo. De esta manera vamos a ver que se obtiene modelos que generan predicciones dando más peso a los valores más recientes que a los más alejados en el tiempo.

### 2.3.1 El modelo de alisado simple

Supongamos que hemos hecho una predicción del valor de una variable para el periodo  $T$  que llamaremos  $\hat{z}_T$  y observamos después un valor  $z_T$ . ¿Cómo generar la próxima predicción? Holt propuso hacer una combinación lineal de la última predicción y el último valor observado de manera que

$$\hat{z}_{T+1} = \theta \hat{z}_T + (1 - \theta) z_T \quad (2.14)$$

donde  $0 < \theta < 1$  determina el peso que damos a cada componente para generar predicciones. Este método fué propuesto por Holt (1956) para la predicción de inventarios y se comprobó que funcionaba bien en muchos casos en la práctica. Si suponemos que  $\hat{z}_T$  se ha generado de la misma forma, tenemos que

$$\hat{z}_{T+1} = \theta(\theta \hat{z}_{T-1} + (1 - \theta) z_{T-1}) + (1 - \theta) z_T = \theta^2 \hat{z}_{T-1} + (1 - \theta)(z_T + \theta z_{T-1})$$

y repitiendo este proceso, tenemos que

$$\hat{z}_{T+1} = \theta^T \hat{z}_1 + (1 - \theta)(z_T + \theta z_{T-1} + \theta^2 z_{T-2} + \dots)$$

Suponiendo que  $T$  es grande, como  $\theta < 1$  el primer término será muy pequeño y podemos escribir la predicción como

$$\hat{z}_{T+1} = (1 - \theta)(z_T + z_{T-1}\theta + \theta^2 z_{T-2} + \dots)$$

que es una media ponderada de las observaciones previas con pesos decrecientes que suman uno, ya que

$$(1 + \theta + \theta^2 + \dots) = \frac{1}{1 - \theta}.$$

Un modelo que justifica esta forma de generar predicciones es el siguiente:

$$z_t = \mu_t + a_t$$

donde  $\mu_t$  es la media de  $z_t$ , que ahora no es constante, sino que puede evolucionar con el tiempo. Supongamos que disponemos de una muestra de tamaño  $T$ . Llamaremos  $\hat{\mu}_T(1)$  a la estimación del nivel futuro de la serie en el instante  $T + 1$ , dados los datos hasta  $T$ . Esta notación pone de manifiesto que  $T$  es la información disponible y el (1) que se trata de la predicción a un horizonte, es decir, la predicción  $\mu_{T+1}$ . Esta estimación se realiza minimizando la función

$$\sum_{t=1}^T (z_t - \mu_{T+1})^2 w_t \quad (2.15)$$

donde los  $w_t$  son pesos, es decir números positivos que suman la unidad. Si elegimos estos pesos disminuyendo de forma geométrica con el tiempo, de manera que la última observación en el instante  $T$  tiene peso máximo,  $c$ , la siguiente, en  $T - 1$ , tiene peso  $c\theta$ , donde  $0 < \theta < 1$ , la siguiente, en  $T - 2$ , peso  $c\theta^2$ , y así sucesivamente, se obtiene la ecuación de predicción anterior. En efecto, la constante  $c$  debe ser tal que los pesos sumen la unidad, y utilizando la fórmula de una progresión geométrica indefinida, resulta

$$1 = c(1 + \theta + \theta^2 + \dots) = \frac{c}{1 - \theta}$$

con lo que  $c = 1 - \theta$ . De esta forma, los pesos vienen dados por:

$$w_t = (1 - \theta)\theta^{T-t}$$



La estimación de  $\mu_{T+1}$  resulta, sustituyendo esta expresión de los pesos en (2.15). Derivando respecto a este parámetro e igualando a cero se obtiene la ecuación:

$$\hat{\mu}_T(1) = (1 - \theta) \sum_{t=1}^T z_t \theta^{T-t} \quad (2.16)$$

que implica que la predicción de  $z_{T+1}$  con información hasta  $T$  es

$$\hat{z}_T(1) = \hat{z}_{T+1} = \hat{\mu}_T(1) = (1 - \theta)(z_T + z_{T-1}\theta + \theta^2 z_{T-2} + \dots)$$

donde vamos a utilizar la notación  $\hat{z}_{T+1}$  para indicar la predicción de  $z_{T+1}$  con información hasta el instante anterior. Con esta notación, la predicción de  $z_T$  con información hasta  $T - 1$  es

$$\hat{z}_{T-1}(1) = \hat{z}_T = \hat{\mu}_{T-1}(1) = (1 - \theta)(z_{T-1} + z_{T-2}\theta + z_{T-3}\theta^2 + \dots)$$

y multiplicando la última ecuación por  $\theta$  y restándola de la anterior, se obtiene

$$\hat{z}_{T+1} - \theta \hat{z}_T = (1 - \theta)z_T$$

que conduce a la ecuación de predicción (2.14). Esta ecuación puede también escribirse:

$$\hat{z}_{T+1} = \hat{z}_T + (1 - \theta)(z_T - \hat{z}_T). \quad (2.17)$$

que nos indica que la predicción de la próxima observación se realiza modificando la última predicción por una fracción  $(1 - \theta)$  del último error de predicción cometido.

La utilización de este modelo requiere determinar el parámetro  $\theta$ . Este parámetro en las primeras aplicaciones del método se fijaba a priori, habitualmente entre .70 y .99, pero progresivamente se obtuvieron mejores resultados permitiendo una mayor amplitud de valores posibles y estimando su magnitud a partir de los datos de la serie con el criterio de minimizar los errores de predicción. Esto puede hacerse probando una rejilla de valores, como 0.1, 0.2, ..., 0.9 para  $\theta$ , calculando los errores de predicción dentro de la muestra  $\hat{a}_t = z_t - \hat{z}_t$  y tomando el valor de  $\theta$  que conduzca a un valor menor de  $\sum \hat{a}_t^2$ , la suma cuadrática de los residuos o errores de predicción. Si el mejor valor obtenido en la búsqueda anterior es  $\theta_0$  podemos refinar la búsqueda y probar ahora con valores de  $\theta$  en el intervalo  $\theta_0 - .05$  y  $\theta_0 + .05$ . Por ejemplo, si  $\theta_0 = .7$  probaremos los valores del conjunto .65, .66, ..., .70, .71, ..., .75. Normalmente esta precisión es suficiente en las aplicaciones.

Observemos que si  $\theta$  es muy próximo a uno el modelo de alisado es muy parecido al modelo de nivel constante. En efecto, según (2.17) la predicción en este caso prácticamente no se modifica de una observación a la siguiente. Por el contrario, si  $\theta$  es pequeño, próximo a cero, el peso de las observaciones pasadas desaparece rápidamente y las predicciones serán muy distintas de las del modelo con nivel constante.

### Ejemplo 2.3

Vamos a determinar el mejor valor para el parámetro de alisado  $\theta$  para la serie las leguas recorridas por la flota de Colón. Utilizando la opción de alisado simple con el programa SPSS, que prueba para los valores de  $\theta$  desde 0.1 hasta .1 se obtiene la tabla siguiente para la suma de los residuos al cuadrado con cada valor de  $\theta$ :

$\theta$	1.0	.9	.8	.7	.6	.5	.4	.3	.2
SSE	8414.9	8478.6	8574.0	8686.9	8798.8	8891.3	8957.1	9018.8	9141.6

Como la suma de cuadrados de los residuos mínima se obtiene con  $\theta = 1$ , que es equivalente al modelo de tendencia constante, concluimos que la serie oscila alrededor de una media constante y que no hay evidencia de que tomar  $\theta$  menor que uno mejore las predicciones.

### 2.3.2 El método de alisado doble de Holt

Las ideas anteriores pueden aplicarse a modelos con tendencia lineal. En lugar de suponer que los parámetros son fijos podemos admitir que evolucionan en el tiempo y estimarlos dando un peso decreciente a las observaciones. El modelo es ahora

$$z_t = \mu_t + a_t$$

pero ahora en lugar de suponer una tendencia determinista permitimos que el nivel evolucione linealmente en el tiempo, pero con una pendiente que puede ser distinta en los distintos periodos. Esto se consigue escribiendo el nivel de la serie en el instante  $t$  como

$$\mu_t = \mu_{t-1} + \beta_{t-1}$$

de manera que la diferencia entre los niveles de dos instantes consecutivos,  $t-1$  y  $t$ , es  $\beta_{t-1}$  la pendiente en el momento  $t-1$ . Observemos que si  $\beta_{t-1} = \beta$ , constante en el tiempo, este modelo es idéntico al de tendencia lineal determinista. Al permitir que la pendiente sea variable este modelo es mucho más flexible.

La predicción de  $z_t$ , con información hasta  $t-1$ , que llamaremos  $\hat{z}_{t-1}(1)$ , se obtiene como

$$\hat{z}_{t-1}(1) = \hat{\mu}_{t-1} + \hat{\beta}_{t-1}$$

donde la estimación del nivel de la serie en el instante  $t$  se obtiene como suma de las últimas estimaciones del nivel y de la pendiente con información hasta  $t-1$ . Para simplificar la notación hemos llamado  $\hat{\mu}_{t-1}$  a la predicción  $\hat{\mu}_{t-1}(0)$ , es decir la estimación de  $\mu_{t-1}$  con información que incluye la del instante  $t-1$ , es decir el dato  $z_{t-1}$ .

Para construir las predicciones con este modelo vamos a generalizar lo aprendido en el modelo anterior. Las predicciones del alisado simple se obtienen con la fórmula (2.17), que indica que la predicción futura es igual a la última predicción realizada más un factor de corrección que es el producto de un coeficiente (uno menos el factor de descuento) por el último error cometido. Como ahora tenemos que estimar dos parámetros, el método de Holt generaliza esta idea introduciendo dos factores de descuento. Supongamos que la última predicción se realiza en el periodo  $T$ , después de haber observado  $z_T$  y corresponde a la estimación de  $z_{T+1}$ , el valor de la serie en el instante  $T+1$ . La predicción será  $\hat{\mu}_T(1) = \hat{\mu}_T + \hat{\beta}_T$ , donde  $\hat{\mu}_T$  y  $\hat{\beta}_T$  son las estimaciones del nivel y el crecimiento con información hasta el instante  $T$ . Al observar el valor  $z_{T+1}$  podemos calcular el error de predicción ( $z_{T+1} - \hat{\mu}_T - \hat{\beta}_T$ ), y, como en el método de alisado simple, corregir la estimación anterior por una fracción del error cometido. En consecuencia, la estimación de  $\hat{\mu}_{T+1}$  con información hasta  $T+1$ , será

$$\begin{aligned}\hat{\mu}_{T+1} &= \hat{\mu}_T(1) + (1 - \theta)(z_{T+1} - \hat{\mu}_T(1)) = \\ &= \hat{\mu}_T + \hat{\beta}_T + (1 - \theta)(z_{T+1} - \hat{\mu}_T - \hat{\beta}_T)\end{aligned}$$

donde  $\theta < 1$  es un factor de descuento. Observemos que la estructura de esta ecuación es análoga a la (2.17). La nueva estimación del crecimiento futuro con información hasta  $T+1$ ,  $\hat{\beta}_{T+1}$  se realiza modificando la última estimación por una fracción del último error cometido:

$$\hat{\beta}_{T+1} = \hat{\beta}_T + (1 - \gamma)(\hat{\mu}_{T+1} - \hat{\mu}_T - \hat{\beta}_T)$$

donde  $\gamma < 1$  es otro factor de descuento sobre el error anterior en la estimación del crecimiento. Estas ecuaciones permiten obtener recursivamente las estimaciones a partir de unos valores iniciales. Los parámetros  $\theta$  y  $\gamma$  se determinan como en el caso anterior, probando con una rejilla de valores y escogiendo aquellos valores que minimizan la suma cuadrática de los errores de predicción.

#### Ejemplo 2.4

Vamos a determinar el mejor valor de los parámetros  $\theta$  y  $\gamma$  para la serie de población de mayores de 16 años en España. Estimando los parámetros con el programa SPSS, se obtienen los siguientes mejores valores para los parámetros, en el sentido de conducir a menor error cuadrático de predicción dentro de la muestra:

$\theta$	1.00	.90	1.00	.90	.80	.80	.80	1.00
$\gamma$	.40	.40	.20	.60	.60	.80	.40	.60
SSE	22499	22690	23056	23133	23609	23956	23963	24057

La figura 2.5 muestra los residuos del mejor modelo que corresponde a  $\theta = 1$  y  $\gamma = .4$ . Si comparamos estos resultados con el modelo lineal vemos que los errores de predicción son mucho menores y que ahora los residuos no muestran una tendencia destacada y las predicciones son bastante buenas en muchos periodos.

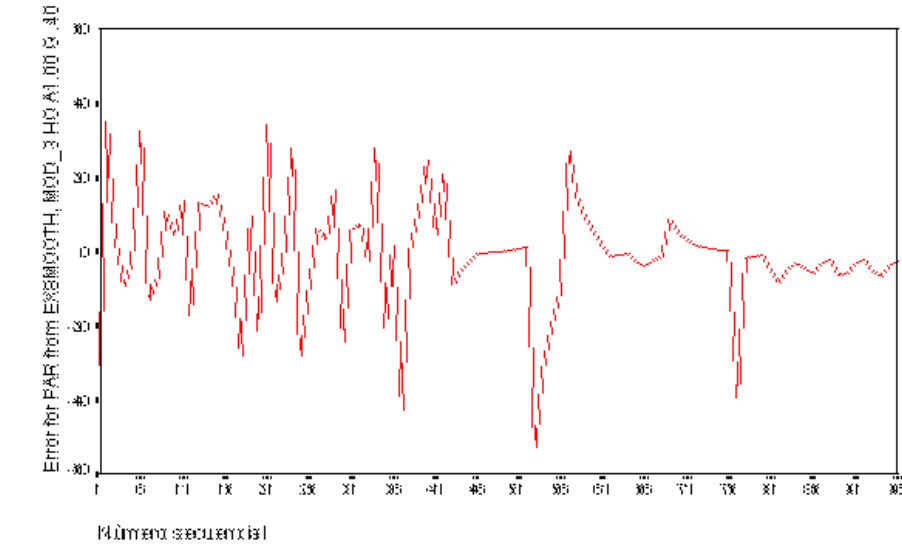


Figura 2.5: Residuos de la serie de mayores de 16 años con el método de Holt

## 2.4 Métodos de descomposición para series estacionales

Cuando la serie además de tendencia y componente aleatorio tiene estacionalidad, los métodos de descomposición suponen que los datos se generan como suma de esos tres efectos:

$$z_t = \mu_t + S_t + a_t$$

donde  $\mu_t$  es el nivel de la serie,  $S_t$  es el componente estacional y  $a_t$  es el componente aleatorio que, como en modelos anteriores, es una secuencia de variables incorreladas de media cero y varianza constante. Los métodos clásicos de descomposición suponen que tanto el nivel como la estacionalidad son deterministas. El nivel se modela normalmente como en el modelo de tendencias deterministas mediante un polinomio en  $t$ , de orden menor o igual a dos, y la estacionalidad como una función periódica, que verifica la condición :

$$S_t = S_{t-s}$$

donde  $s$  es el periodo de la función que depende de la clase de estacionalidad de los datos. Una serie mensual con estacionalidad anual tiene periodo  $s = 12$  meses, ya que suponemos que los coeficientes estacionales,  $S_t$ , se repetirán cada 12 observaciones, una serie trimestral con estacionalidad anual tendrá periodo  $s = 4$  y una serie diaria con estacionalidad semanal tendrá  $s = 7$ . El procedimiento de construcción del modelo para la serie se realiza en las tres etapas siguientes:

1. Se estima el nivel de la serie observada mediante el modelo de tendencias deterministas. La forma de la tendencia se elige con el gráfico de la serie y se realiza como hemos visto en la sección 2.2. A continuación se resta a la serie el nivel estimado,  $\hat{\mu}_t$ , para obtener una serie residual,  $E_t = z_t - \hat{\mu}_t$ , que contendrá la estacionalidad y el componente aleatorio. Esta serie se denomina serie sin tendencia.

2. Se obtiene la media de la serie  $E_t$ ,  $\bar{E} = \sum E_t / T$  y la media de las observaciones que tienen el mismo coeficiente estacional. Supongamos, por ejemplo, que tenemos  $n$  años exactos de datos de una serie mensual. Entonces,  $t = 1, \dots, 12n$ , y podemos etiquetar las observaciones para poner de manifiesto el mes y el año al que pertenecen. Definamos  $t = 12i + j$ , donde la variable  $i$  toma los valores  $0, 1, \dots, n-1$  y representa los años transcurridos y la  $j$  los meses dentro del año y toma valores  $1, \dots, 12$ . Con esta notación, las observaciones de la serie sin tendencia pueden escribirse como  $E_{12i+j}$ . Las medias de cada mes en los distintos años serán

$$\bar{E}_j = \sum_{i=1}^n E_{12i+j} / n$$

Se definen los factores estacionales como un conjunto de coeficientes,  $S_1, \dots, S_{12}$  que suman cero y que se repiten cada año. Los estimaremos como la diferencia entre la media del mes y la media general

$$\hat{S}_j = \overline{E}_j - \overline{E}.$$

Por construcción, es fácil comprobar que la suma de los doce factores estacionales es cero.

3. Se obtiene la serie aleatoria restando a la serie sin tendencia el factor estacionales de cada observación

$$\hat{a}_{12i+j} = E_{12i+j} - \hat{S}_j$$

La predicción de la serie se realiza sumando las estimaciones de la tendencia y del factor estacional que corresponde a cada observación este mes. Por otro lado, si restamos a la serie original el coeficiente estacional del mes se obtiene la *serie desestacionalizada*.

Como se ha visto en las secciones anteriores existen series que claramente no tienen una tendencia constante y para las que ajustar una tendencia determinista no es adecuado. Una posibilidad alternativa es estimar el nivel de la serie mediante una media móvil de doce meses como sigue: el valor del nivel en el centro del año se obtiene mediante la operación

$$\hat{\mu}_t = \frac{z_{t-5} + \dots + z_{t+5} + z_{t+6}}{12}$$

es decir, se construye una media de doce observaciones. Esta operación se conoce como media móvil, ya que las observaciones que entran en el cálculo de  $\hat{\mu}_t$  van variando en el tiempo. Aplicando este método se obtiene una estimación del nivel de la serie en los instantes  $t = 6, \dots, T - 6$ . Para obtener los valores del nivel en los extremos se ajusta una recta a los últimos valores y de esta manera se completa la serie de niveles. A continuación se realiza la descomposición de la serie de la manera ya explicada mediante los pasos 2 y 3.

### Ejemplo 2.5

Vamos a analizar la serie de paro en España. La figura 2.6 presenta el resultado de la descomposición con el programa SPSS. La primera serie es la serie original a la que se ha ajustado una tendencia lineal para obtener la serie sin tendencia (detrended) que se presenta a continuación. Haciendo las medias de los cuatrimestres y restando la media general se obtienen los coeficientes estacionales indicados en la figura 2.7. En esta figura se observa que el primer cuatrimestre, que corresponde al invierno, es el de más alto número de paro. Estos disminuyen en primavera y verano y vuelven a aumentar en otoño. En los otros paneles de esta figura se observa que la variabilidad es similar en los cuatro periodos estacionales.

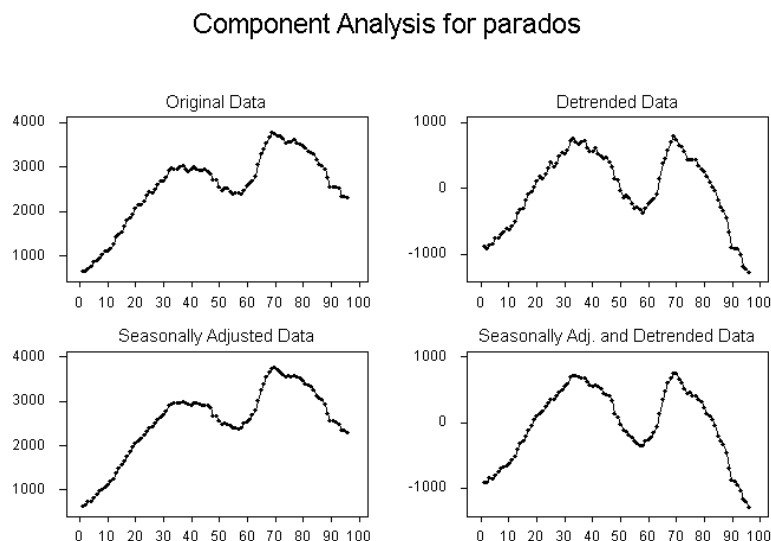


Figura 2.6: Descomposición de la serie del paro en España en el periodo primer cuatrimestre de 1977 al cuarto del 2000

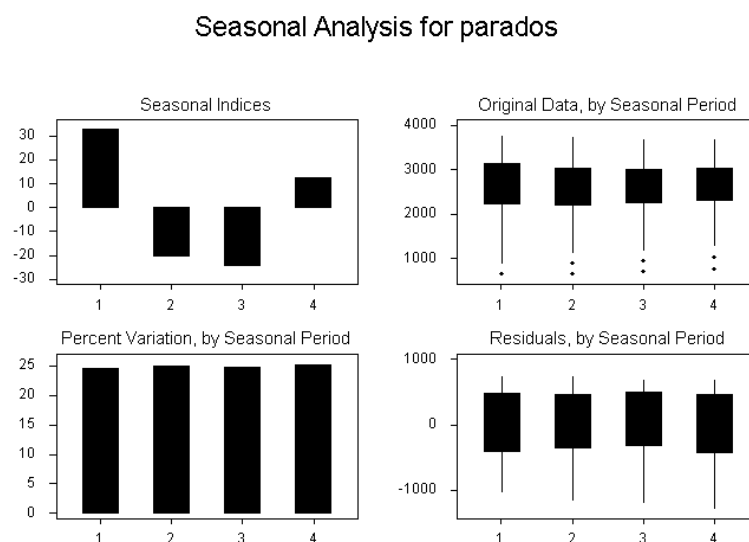


Figura 2.7: Coeficientes estacionales para la serie del paro en España

Para ilustrar los resultados de obtener la tendencia mediante medias móviles las figuras 2.8, 2.9 y 2.10 presentan al tendencia obtenida por este procedimiento y los factores estacionales y residuos resultantes. Se observa que este segundo procedimiento parece dar mejores resultados en esta serie que el de tendencias deterministas, ya que la estimación del nivel de la serie mediante medias móviles de la figura 2.8 es mucho más fiel a la serie que la estimación del nivel mediante una línea recta.

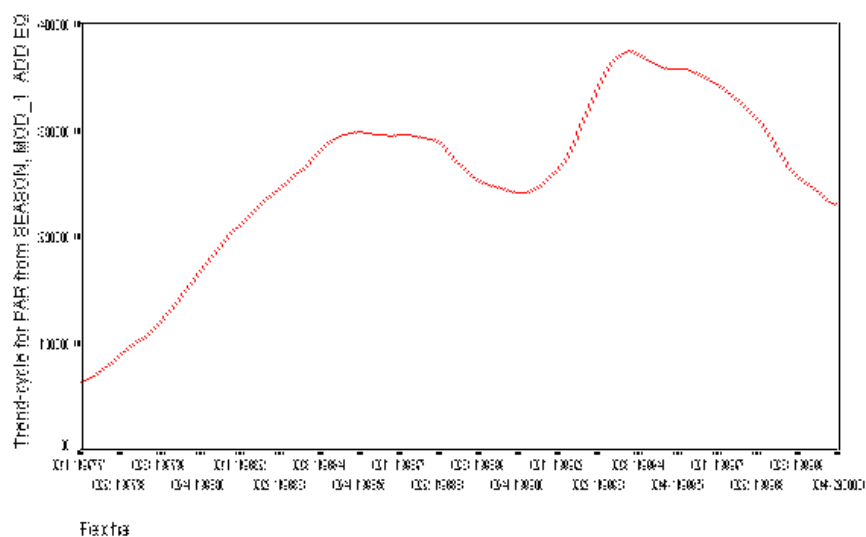


Figura 2.8: Tendencia estimada por medias móviles para la serie de paro en España

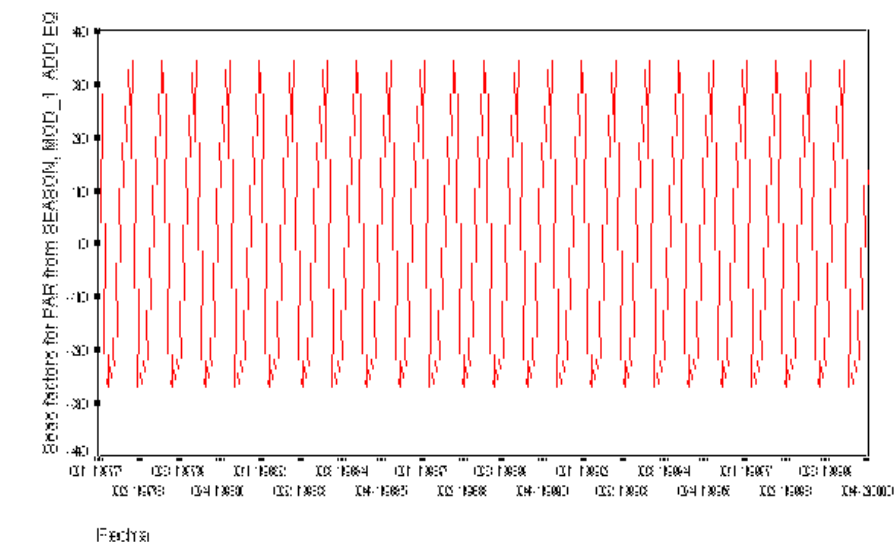


Figura 2.9: Coeficientes estacionales para la series del paro en España

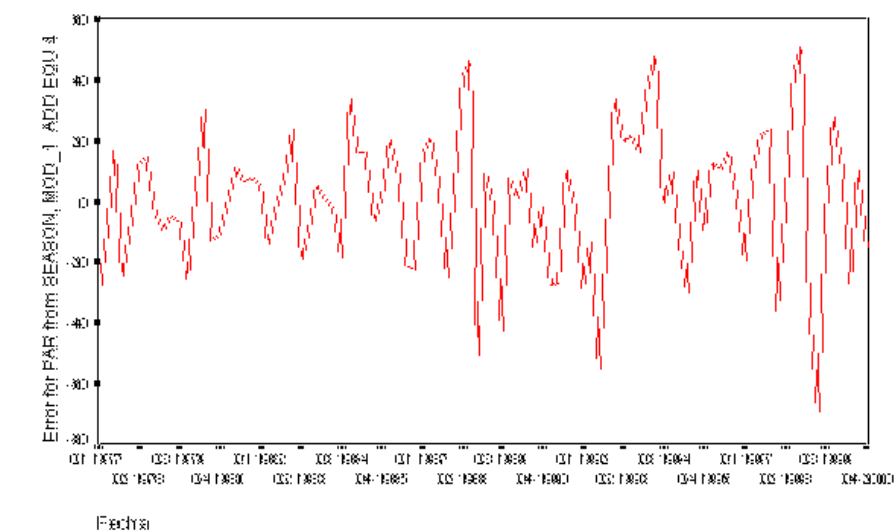


Figura 2.10: Residuos para la serie del paro en España

## 2.5 Estacionalidad y el ajuste de Ciclos.

### 2.5.1 Definiciones básicas

Un procedimiento alternativo para modelar la estacionalidad de una serie es representarla por una función armónica de periodo  $s$ . Suponiendo que hemos eliminado la tendencia, si la tuviese, consideremos series que tienen sólo componente estacional, con estructura:

$$z_t = S_t + a_t$$

La alternativa más simple para representar  $S_t$  como una función periódica, con  $S_t = S_{t-s}$ , es suponer una función armónica, como el seno o el coseno. La figura 2.11 representa las funciones  $\text{sen}(2\pi t/s)$  y  $\text{cos}(2\pi t/s)$  para  $t = 1, \dots, T$ , donde se ha tomado el tamaño muestral,  $T$ , exactamente igual al periodo de la función,  $s = 100$ . Es fácil comprobar que estas funciones sinusoidales se repiten exactamente cada  $s$  observaciones. Por ejemplo,  $\text{sen}(2\pi(t+s)/s) = \text{sen}(2\pi t/s + 2\pi) = \text{sen}(2\pi t/s)$ . La cantidad inversa del periodo,

$$f = \frac{1}{s}$$

se denomina *frecuencia*, e indica el número de ciclos completos por unidad de tiempo. Por ejemplo, en una serie trimestral ( $s = 4$ )  $f = 1/4 = .25$ , indicando que la unidad de tiempo considerada, el trimestre, ocupa .25 del periodo de la función o un 25% de un ciclo completo. En una serie mensual ( $s = 12$ ) la frecuencia es  $1/12 = .0833$ , indicando que la unidad de tiempo, mes, es el 8.33% del ciclo estacional. La cantidad

$$w = 2\pi f = 2\pi/s$$

se denomina *frecuencia angular*, e indica en radianes el ángulo recorrido en una unidad de tiempo teniendo en cuenta que el ciclo completo son  $2\pi$  radianes. Por ejemplo, una serie trimestral tiene una frecuencia angular de  $w = 2\pi/4 = \pi/2$ , indicando que en un trimestre se recorre un ángulo de  $\pi/2$  respecto al ciclo total de  $2\pi$ .

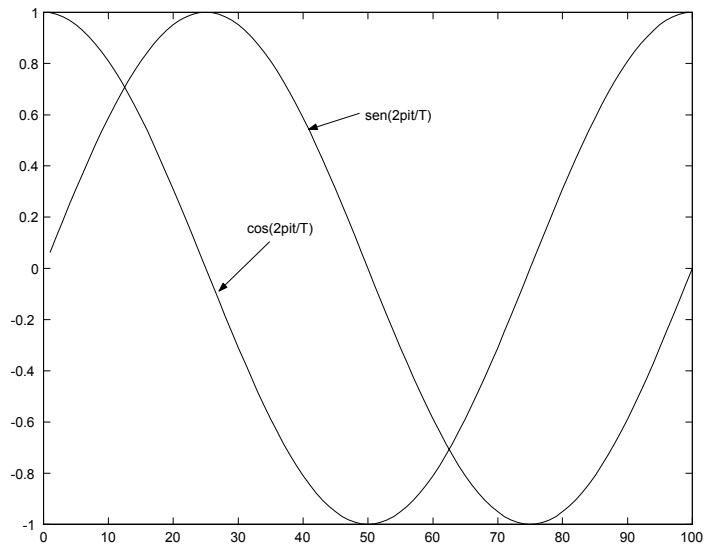


Figura 2.11: Representación de las funciones seno y coseno con periodo igual a los datos observados.

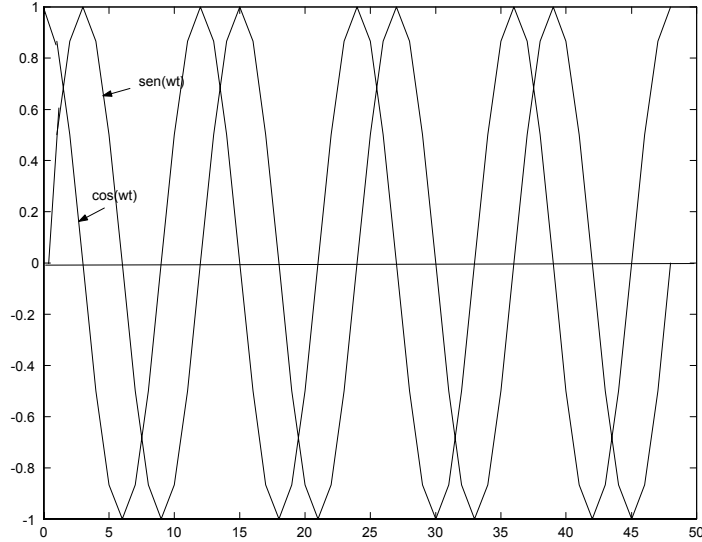


Figura 2.12: Tres años de datos mensuales para una serie con estacionalidad sinusoidal

En una serie real el tamaño muestral es típicamente mayor que el periodo de la función. Si el tamaño muestral fuese menor que el periodo, sólo observaremos una fracción del ciclo. Por ejemplo, si tuviésemos ocho datos mensuales de una serie con estacionalidad anual. Para simplificar la exposición vamos a suponer que el tamaño muestral es múltiplo exacto de  $s$ , de manera que observamos  $j = T/s$  ciclos completos. Además, vamos a suponer que el tamaño muestral es un número par. Entonces,  $w = 2\pi/s = 2\pi j/T$ , con  $j$  entero. Por ejemplo, la figura 2.12 presenta una muestra de tamaño  $T = 48$  de las dos series deterministas  $x_{1t} = \text{sen}(wt)$  y  $x_{2t} = \cos(wt)$  mensuales,  $s = 12$ , con frecuencia angular  $w = 2\pi/12 = \pi/6$ . Es intuitivo en el gráfico, y se demuestra rigurosamente en el apéndice 2.1, que al tener ciclos completos:

$$\sum_{t=1}^T \text{sen}(wt) = \sum_{t=1}^T \cos(wt) = 0$$

y además

$$\sum_{t=1}^T \text{sen}(wt) \cos(wt) = 0$$

La primera condición indica que las variables  $x_{1t} = \text{sen}(wt)$  y  $x_{2t} = \cos(wt)$  observadas en  $t = 1, \dots, T$  tienen media nula. Esto es consecuencia de que el tamaño muestral es igual a uno o más ciclos completos. La segunda que, en este caso, estas variables tienen covarianzas nulas y están por tanto incorreladas en la muestra.

### 2.5.2 Representación de la estacionalidad por un ciclo

Supongamos una serie  $(z_1, \dots, z_T)$  que tiene una estacional cíclica de periodo  $s$ , y en la que observamos  $j$  ciclos completos, es decir  $T = js$ , con  $j$  entero. Vamos a modelar la estacionalidad mediante una función seno con frecuencia angular  $w = 2\pi/s$ . Por ejemplo, tenemos  $j = 10$  años de datos de una serie mensual, con  $T = 120$ , y queremos ajustar un ciclo anual, de manera que  $s = 12$  y la onda sinusoidal tendrá una frecuencia angular  $w = 2\pi/12$ . Por ejemplo, la figura 2.13 presenta una serie de estas características. La primera observación de la serie no será, en general, el valor medio del ciclo, como corresponde a la función seno, sino que la onda sinusoidal que describe la estacionalidad comenzará en la primera observación con un cierto ángulo  $\theta$  de desfase con relación al comienzo del ciclo. Además, el ciclo tendrá una amplitud desconocida, que llamaremos  $R$ . Finalmente, supondremos que existe un error aleatorio superpuesto a la estacionalidad,



que representamos como en modelos anteriores por una secuencia de variables independientes,  $a_t$ , de media cero, varianza constante y distribución normal. El modelo para la serie será:

$$z_t = \mu + R \sin(wt + \theta) + a_t. \quad (2.18)$$

Para ajustar este modelo a los datos vamos a escribirlo de una forma más conveniente. Utilizaremos que el seno de la suma de dos ángulos es la suma de los productos de los senos más la suma del producto de los cosenos, con lo que podemos escribir la expresión anterior como

$$z_t = \mu + R \sin(wt) \sin(\theta) + R \cos(wt) \cos(\theta) + a_t$$

y llamando  $A = R \sin \theta$  y  $B = R \cos \theta$  tenemos que:

$$z_t = \mu + A \sin(wt) + B \cos(wt) + a_t. \quad (2.19)$$

Esta expresión es más simple que la (2.18) ya que representa la serie como suma de dos funciones sinusoidales de amplitudes desconocidas que tenemos que estimar a partir de los datos. El modelo (2.19) es lineal en los dos parámetros, A y B, ya que  $w$  es conocida, y podemos estimarlo por mínimos cuadrados. Las ecuaciones resultantes son (véase el apéndice 2.1)

$$\begin{bmatrix} \hat{\mu} \\ \hat{A} \\ \hat{B} \end{bmatrix} = \begin{bmatrix} T & \sum_{t=1}^T \sin(wt) & \sum_{t=1}^T \cos(wt) \\ \sum_{t=1}^T \sin(wt) & \sum_{t=1}^T \sin^2(wt) & \sum_{t=1}^T \sin(wt) \cos(wt) \\ \sum_{t=1}^T \cos(wt) & \sum_{t=1}^T \sin(wt) \cos(wt) & \sum_{t=1}^T \cos^2(wt) \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^T z_t \\ \sum_{t=1}^T z_t \sin(wt) \\ \sum_{t=1}^T z_t \cos(wt) \end{bmatrix}$$

Se demuestra en el apéndice 2.1 que cuando el tamaño muestral es un múltiplo del periodo, como estamos suponiendo aquí, las variables  $\sin(wt)$  y  $\cos(wt)$  son ortogonales y tienen media cero varianza 1/2. Por tanto la matriz de productos cruzados es diagonal, y los estimadores son

$$\hat{\mu} = \frac{1}{T} \sum_{t=1}^T z_t$$

y

$$\hat{A} = \frac{2}{T} \sum_{t=1}^T z_t \sin(wt) \quad (2.20)$$

$$\hat{B} = \frac{2}{T} \sum_{t=1}^T z_t \cos(wt) \quad (2.21)$$

Podemos obtener un estimador de la amplitud  $R$  mediante:

$$\hat{R}^2 = \hat{A}^2 + \hat{B}^2 \quad (2.22)$$

La amplitud de la onda determina la varianza de la variable observada. Elevando al cuadrado en (2.19) y sumando para todos los datos resulta:

$$\frac{1}{T} \sum_{t=1}^T (z_t - \hat{\mu})^2 = \frac{\hat{A}^2}{2} + \frac{\hat{B}^2}{2} + \hat{\sigma}^2 = \frac{\hat{R}^2}{2} + \hat{\sigma}^2, \quad (2.23)$$

por la ortogonalidad de las variables y llamando  $\hat{\sigma}^2$  a la varianza de los residuos estimados.

### Ejemplo 2.6

Como ejemplo vamos a ajustar una función sinusoidal a la temperatura media mensual en Santiago de Compostela. La figura 2.13 ilustra la serie y su ajuste sinusoidal, que aplicando las fórmulas anteriores es:

$$\hat{z}_t = 13.2417 - 3.3351 \sin(2\pi t/12) - 4.2035 \cos(2\pi t/12)$$

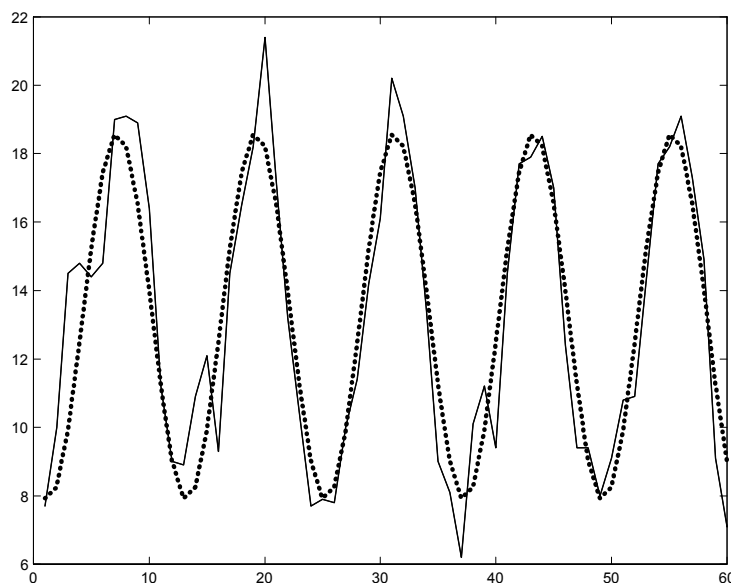


Figura 2.13: Temperatura media mensual en Santiago en los años 1997 a 2001 (línea gruesa) y función sinusoidal ajustada (línea de trazos)

Se observa que el ajuste de la serie es bastante bueno. La amplitud de la onda se calcula como  $\hat{R} = \sqrt{\hat{A}^2 + \hat{B}^2} = 5.36$ . El ángulo inicial puede también calcularse con  $-3.3351 = 5.36\sin(\theta)$ , con lo que se obtiene que  $\theta = -0.6727$  radianes. En resumen, la temperatura en Santiago puede representarse bien por una onda seno de amplitud 5.36 grados centígrados y que comienza en enero con un desfase angular de -0.6727 radianes. Dos veces la amplitud de la onda indica la diferencia promedio máxima entre el mes más frío y el más caliente, 10.7 grados en este caso.

### Ejemplo 2.7

Se realiza el mismo ajuste sinusoidal con una función de periodo 12 a la serie de lluvias en Santiago. El resultado se presenta en la figura 2.14. El modelo ajustado es ahora

$$\hat{z}_t = 145.35 + 2.0858\sin(2\pi t/12) + 88.2449\cos(2\pi t/12)$$

Se observa que, aunque el modelo sinusoidal explica parte de la variabilidad de la serie, el ajuste ahora no es tan bueno como en la serie de temperaturas en el ejemplo anterior, ya que la función sinusoidal no es capaz de captar ni las asimetrías ni los picos de la serie observada.

## 2.6 Exploración de múltiples ciclos: el periodograma

La representación de una serie estacional mediante la ecuación (2.19) es adecuada cuando la estacionalidad es exactamente sinusoidal de periodo  $s$ , pero no nos sirve para describir funciones periódicas generales. Fourier, un matemático francés, demostró a principios del siglo XIX que toda función periódica puede representarse como suma de funciones sinusoidales de distinta amplitud y frecuencia. Este resultado sugiere generalizar el análisis realizado para un ciclo permitiendo que la función periódica sea la suma de varias funciones armónicas con distintas frecuencias.

Dada una serie de longitud  $T$  se denominan *periodos básicos* o de *Fourier* a los que son fracciones exactas completas del tamaño muestral. Es decir, los periodos básicos están definidos por:

$$s_j = \frac{T}{j}, \text{ para } j = 1, 2, \dots, T/2$$

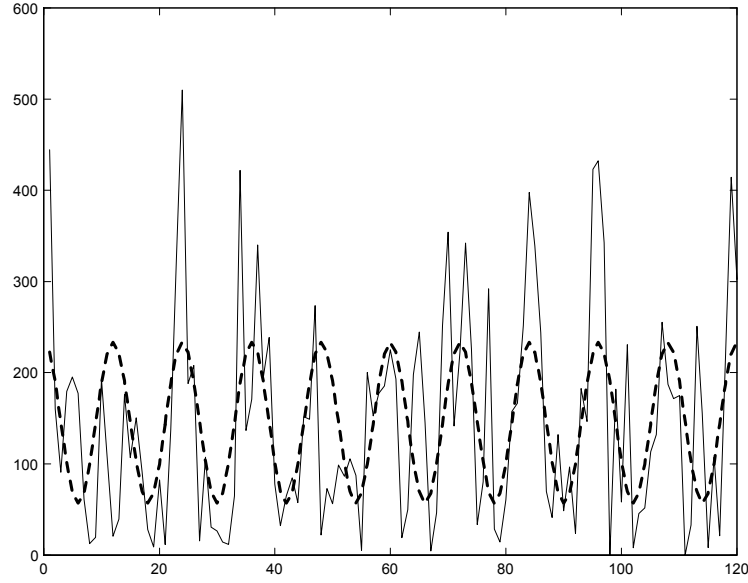


Figura 2.14: Lluvia mensual en Santiago de Compostela y función sinusoidal ajustada.

El valor máximo del periodo básico se obtiene para  $j = 1$  y es  $T$ , el tamaño muestral. Entonces observamos la onda sólo una vez, como en la figura 2.11. El valor mínimo del periodo básico se obtiene para  $j = T/2$  y es 2, ya que periodos que duran menos de dos observaciones no podremos observarlos. Por ejemplo, si tenemos datos diarios será imposible observar ciclos de periodo doce horas o que ocurren dentro del día. Sólo podremos observar ciclos que tienen un periodo mínimo de dos días. En el ajuste de ciclos suele trabajarse con las frecuencias, en lugar de con los periodos, y se definen las frecuencias básicas o de Fourier como las inversas de los periodos básicos:

$$f_j = \frac{j}{T}, \text{ para } j = 1, 2, \dots, T/2$$

con lo que tenemos  $1/2 \geq f_j \geq 1/T$ , y el valor máximo de la frecuencia que podemos observar es  $f = .5$ . Este valor de las frecuencias básicas se conoce como la *frecuencia de Nyquist*.

Utilizando estas definiciones, podemos obtener una representación general de una función periódica como suma de ondas asociadas a todas las frecuencias básicas, como

$$z_t = \mu + \sum_{j=1}^{T/2} A_j \sin(w_j t) + \sum_{j=1}^{T/2} B_j \cos(w_j t). \quad (2.24)$$

Esta ecuación contiene tantos parámetros como observaciones por lo que siempre ajustará exactamente a cualquier serie observada. En efecto, para  $j = T/2$ ,  $s_j = 2$ , y  $w_j = 2\pi/s_j = \pi$ , y la variable  $\sin \pi t$  es siempre igual a cero para  $t = 1, \dots, T$ , con lo que el coeficiente  $B_{T/2}$  no puede estimarse y tenemos  $T$  parámetros:  $\mu$  más los  $T/2$  de  $A_j$  y los  $T/2 - 1$  de  $B_j$ . En consecuencia tenemos que encontrar un procedimiento para seleccionar las frecuencias que debemos incluir para explicar la evolución de la serie. Este es el objetivo del periodograma.

Observemos que la ecuación (2.24) permite descomponer exactamente una serie temporal observada como una suma de componentes armónicos. Según la ecuación (2.23), la contribución de una onda a la varianza de una serie es la amplitud de la onda al cuadrado dividido por dos. Por tanto, ondas con amplitud estimada alta serán importantes para explicar la serie, mientras aquellas ondas de amplitud baja contribuyen poco a la explicación de la serie. Para seleccionar las frecuencias importantes podemos calcular los parámetros  $A_j$  y  $B_j$  para todas las frecuencias básicas y representar la contribución a la varianza de la serie, es decir la

amplitud de la onda al cuadrado dividido por dos, como una función de la frecuencia. Este ajuste se realiza fácilmente, ya que las variables explicativas  $\sin(w_j t)$  y  $\cos(w_j t)$  son ortogonales (véase el apéndice 2.1), por lo que los coeficientes de regresión  $A_j$  y  $B_j$  serán iguales calculados en el modelo (2.24), que incluye todas las frecuencias básicas, que en un modelo como el (2.18) que incluye sólo la frecuencia  $w_j$ . Por tanto, podemos estimar los coeficientes  $A_j$  y  $B_j$  por (2.20) y (2.21), con  $w_j = 2\pi f_j$  y  $f = j/T$ .

Dados los coeficientes estimados  $\hat{A}_j$  y  $\hat{B}_j$  para la frecuencia  $w_j$  calculamos  $\hat{R}_j = \hat{A}_j^2 + \hat{B}_j^2$ , y utilizando la ecuación (2.24), descomponemos la varianza de la serie en los componentes asociados a cada una de las funciones armónicas. Llamando  $s_z^2$  a la varianza muestral de la serie, podemos escribir:

$$Ts_z^2 = \sum_{t=1}^T (z_t - \mu)^2 = \sum_{j=1}^{T/2} \frac{T}{2} \hat{R}_j^2 \quad (2.25)$$

Se conoce como *periodograma* la representación de la contribución de cada frecuencia,  $T\hat{R}_j^2/2$ , en función de la frecuencia  $w_j$  o  $f_j$ . En este segundo caso se define como:

$$I(f_j) = \frac{T\hat{R}_j^2}{2}, \quad \text{con } 1/T \leq f_j \leq .5 \quad (2.26)$$

Observemos que, con esta representación, el valor medio de las alturas de las ordenadas del periodograma es el doble de la varianza del proceso, ya que por (2.25) y (2.26), como el número de frecuencias básicas es  $T/2$

$$\bar{I} = \frac{2}{T} \sum_{j=1}^{T/2} \frac{T}{2} \hat{R}_j^2 = 2s_z^2$$

El análisis anterior supone que estamos interesados únicamente en las frecuencias básicas. Esta hipótesis es poco restrictiva si el tamaño muestral es grande, ya que entonces el número de frecuencias analizadas es muy grande y siempre existirá alguna frecuencia básica muy próxima a la que puede interesarnos. Por ejemplo, si tenemos 140 datos mensuales y queremos estimar la amplitud del periodo 12, el periodo básico para  $j = 11$  es  $T_{11} = 12.7$ , y para  $j = 12$ ,  $T_{12} = 11.66$ . Por tanto, podemos tener una idea aproximada de la amplitud de la onda de periodo 12 por la amplitud calculada con  $j = 11$  y 12. Naturalmente podemos ajustar un modelo para cualquier periodo  $T_j$ ,  $2 \leq T_j \leq T/2$ , pero entonces, al perder la simetría, las formulas presentadas para la estimación son sólo aproximadas, aunque siempre podemos calcular los coeficientes exactamente mediante un programa de ordenador que estime regresiones múltiples.

El periodograma puede verse como una herramienta para la detección de posibles ciclos deterministas en una serie temporal. Por ejemplo, en una serie mensual estacional esperamos encontrar un valor alto del periodograma para  $f = 1/12$ , pero también podemos encontrar valores altos para  $f = j/12$ , es decir,  $1/6$ ,  $1/4$ ,  $1/3$ , que son armónicos del periodo estacional. Por otro lado la serie puede tener otros ciclos no necesariamente ligados al periodo estacional, y el periodograma es una buena herramienta para detectar estos posibles componentes. La utilidad del periodograma aumenta si notamos que cuando estimamos la amplitud de una onda para una frecuencia determinada estamos de hecho calculando una amplitud media de todos los posibles ciclos con frecuencias próximas a la que estimamos. Como las distancias entre las frecuencias básicas son  $(j+1)/T - j/T = 1/T$ , podemos considerar la amplitud calculada para la frecuencia  $f_j$  como un promedio de las amplitudes existentes en las frecuencias situadas en el intervalo  $f_j \pm 1/2T$ . En consecuencia, en lugar de representar barras en las frecuencias básicas con alturas  $I(f_j)$ , podemos construir una *periodograma suavizado* construyendo rectángulos con centro en  $f_j$ , base igual a  $1/T$  y con alturas  $I(f_j) = T\hat{R}_j^2/2$ , de manera que el área del rectángulo sea igual a  $\hat{R}_j^2/2$ . De esta manera, distribuimos la variabilidad estimada para la frecuencia  $f_j$  en todo el rango de frecuencias. Este periodograma suavizado estará definido ahora para todas las frecuencias en el intervalo  $.5 \geq f_j \geq 0$  y el área total encerrada bajo esta función es, por (2.25), la varianza del proceso. Podemos dar un paso adicional y en lugar de representar una función en escalones, como la que acabamos de describir, podemos pensar en una función continua, y suavizar la estimación anterior ponderando las estimaciones adyacentes de forma similar a como se suaviza un histograma para obtener una estimación de la función de densidad. En este periodograma suavizado la

ordenada en cada punto se calcula por

$$I(f) = \sum_{f'_i - a}^{f_i + a} p_i I(f_i), \quad \text{con } .5 \geq f \geq 0$$

donde  $a$  representa la ventana utilizada y los  $p_i$  son pesos simétricos que suman la unidad.

### Ejemplo 2.8

Vamos a calcular el periodograma para la serie de lluvias en Santiago. Los programas de ordenador presentan siempre el periodograma suavizado como una curva continua. Para ello se estiman por mínimos cuadrados las amplitudes para las frecuencias básicas y luego se suavizan estas estimaciones para obtener una curva aplicando medias ponderadas sobre las ordenadas adyacentes. El gráfico resultante se observa en la figura 2.15 y se observa un pico para la frecuencia estacional mensual,  $f = 1/12 = .083$ . El gráfico muestra también picos en las frecuencias  $1/6 = .16$  y  $1/3 = .33$ , que están asociadas a las frecuencias estacionales. Este resultado indica que la estacionalidad se modela mejor como una suma de estos tres componentes armónicos que con una onda determinista de periodo 12.

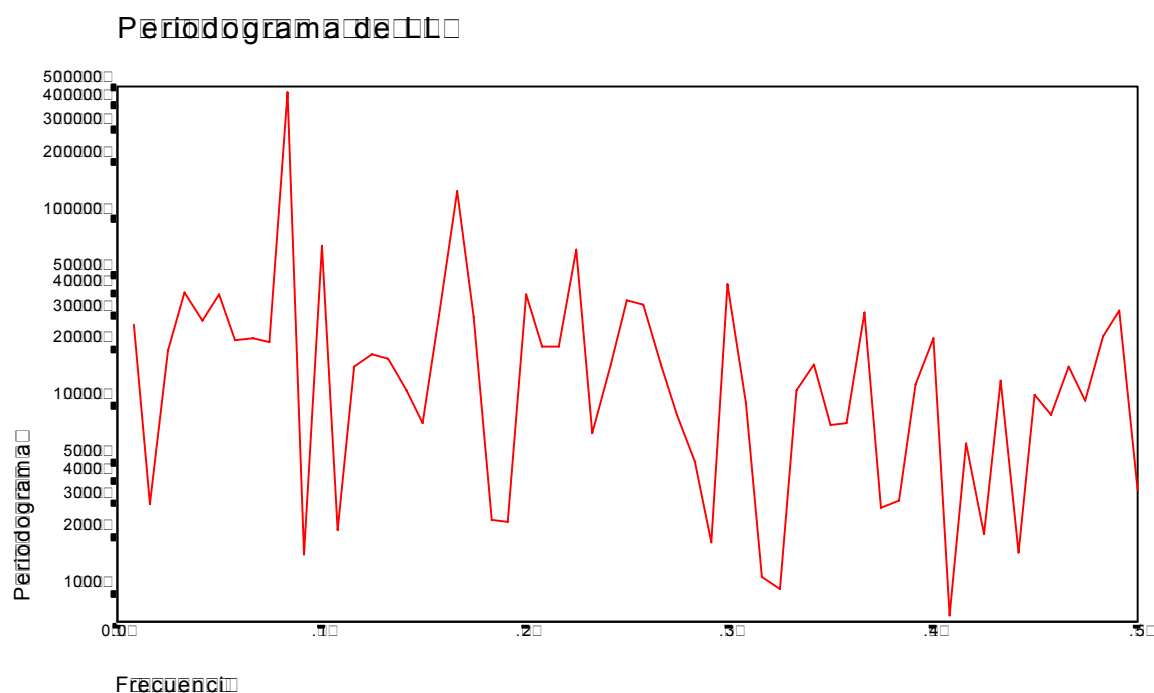


Figura 2.15: Periodograma de la serie de lluvias en Santiago

## 2.7 Conclusiones y lecturas complementarias

Hemos visto en este capítulo que los modelos de tendencias deterministas tienen muchas limitaciones para representar series reales. Los métodos adaptativos basados en dar peso decreciente al pasado funcionan mejor, pero suponen una estructura de dependencia que, aunque flexible, no es esperable que se aplique a todas las series reales. Los métodos de descomposición son útiles, pero es necesario disponer de modelos más flexibles para los componentes. El periodograma es una herramienta muy valiosa para detectar componentes sinusoidales deterministas en la serie, como efectos estacionales cíclicos.

Abraham y Ledolter (1983) es una excelente referencia sobre los métodos de alisado. Las propiedades de las funciones sinusoidales y del periodograma se presenta de forma clara en Broomfield (2000) y Wei (1990) y a un nivel más matemático en Brockwell and Davies (1996). Tres libros clásicos sobre el tema que recomendamos consultar al lector interesado son Granger and Hatanaka (1964), Jenkins and Watts (1968) y Brillinger (1981)..

### Ejercicios 2

2.1 Estimar la tendencia de la serie de consumo de gasolina en España mediante una recta en el periodo 1945 a 1995 y generar predicciones para 24 meses. Comparar el resultado con el método de Holt.

2.2 Aplicar a la serie de consumo de gasolina un método de descomposición y estimar los coeficientes estacionales. Comparar los resultados al estimar la tendencia mediante una línea recta, por el método de Holt y mediante promedios móviles.

2.3 Calcular el periodograma para la serie del viaje de Colón e interpretarlo.

2.4 Calcular el periodograma para la serie de consumo de gasolina e interpretarlo.

2.5 Demostrar que las predicciones realizadas con el método de Holt verifican la ecuación recursiva-  
 $\hat{z}_{T+1}(1) = \hat{z}_T(1) + \hat{\beta}_T + \gamma(1 - \theta)(z_{T+1} - \hat{\mu}_T - \hat{\beta}_T)$

### Apéndice 2.1

En este apéndice resumimos la estimación de una ecuación de regresión múltiple por mínimos cuadrados y su aplicación a funciones sinusoidales. El lector puede encontrar un tratamiento más amplio de la regresión múltiple en Peña (2001).

Suponemos que se desea construir una ecuación para prever una variable  $y$ , en función de un conjunto de  $p$  variables explicativas,  $x_1, \dots, x_p$ . La ecuación a estimar es

$$\hat{y}_i = \bar{y} + \hat{\beta}_1(x_{i1} - \bar{x}_1) + \dots + \hat{\beta}_p(x_{ip} - \bar{x}_p), \quad i = 1, \dots, n \quad (2.27)$$

y es inmediato comprobar que en esta ecuación si las variables explicativas toman un valor igual a su media,  $\bar{x}_j$ , la predicción  $\hat{y}_i$  de la variable respuesta es también su media,  $\bar{y}$ . Los  $p$  coeficientes  $\hat{\beta}_j$ , para  $j = 1, \dots, p$  se determinan de manera que la ecuación proporcione, en promedio, la mejor predicción posible de los valores de  $y_i$ . Llamando  $a_i = y_i - \hat{y}_i$  a los residuos o errores de predicción, el criterio de mínimos cuadrados minimiza

$$M = \sum_{i=1}^n a_i^2,$$

y al derivar esta expresión respecto a los parámetros  $\hat{\beta}_j$ , se obtiene el sistema de  $p$  ecuaciones:

$$2 \sum_{i=1}^n \left[ y_i - \bar{y} + \hat{\beta}_1(x_{i1} - \bar{x}_1) + \dots + \hat{\beta}_p(x_{ip} - \bar{x}_p) \right] (x_{ij} - \bar{x}_j),$$

que puede escribirse:

$$\sum e_i x_{ij} = 0 \quad j = 1, \dots, p;$$

que tiene una clara interpretación intuitiva. Indica que los residuos, o errores de predicción, deben de estar incorrelados con las variables explicativas, de manera que la covarianza entre ambas variables sea cero. En efecto, si existiese relación entre ambas variables podría utilizarse para prever los errores de predicción y reducirlos, con lo que la ecuación de predicción no podría ser óptima. Geométricamente, este sistema establece que el vector de residuos debe ser ortogonal al espacio generado por las variables explicativas. Llamando

$$\tilde{\mathbf{X}} = \begin{bmatrix} x_{11} - \bar{x}_1 & \dots & x_{1p} - \bar{x}_p \\ \dots & \dots & \dots \\ x_{n1} - \bar{x}_1 & \dots & x_{np} - \bar{x}_p \end{bmatrix}$$

a la matriz de variable explicativas centradas, llamando  $\tilde{\mathbf{y}} = (y_1 - \bar{y}, \dots, y_n - \bar{y})'$  al vector de los valores de la respuesta centrados, y  $\tilde{\boldsymbol{\beta}} = (\hat{\beta}_1, \dots, \hat{\beta}_p)'$  al vector de coeficientes, el sistema de ecuaciones para obtener los parámetros es:

$$\tilde{\mathbf{X}}' \tilde{\mathbf{y}} = \tilde{\mathbf{X}}' \tilde{\mathbf{X}} \tilde{\boldsymbol{\beta}}$$

que conduce a :

$$\hat{\beta} = (\tilde{\mathbf{X}}' \tilde{\mathbf{X}})^{-1} \tilde{\mathbf{X}}' \tilde{\mathbf{y}},$$

Vamos a demostrar que aplicar esta formula a componentes sinusoidales con frecuencias básicas equivale a utilizar variables de media cero e incorreladas, con lo que la matriz  $\tilde{\mathbf{X}}' \tilde{\mathbf{X}}$  es diagonal y la estimación de cada coeficiente es simplemente el cociente entre la covarianza de la variable y la respuesta dividida por la varianza del regresor. Es decir, si las variables explicativas son del tipo  $x_{1t} = \text{sen}(wt)$  y  $x_{2t} = \cos(wt)$ , siendo  $w = 2\pi j/T$ , donde  $T$  es el tamaño muestral y  $j$  un entero positivo menor de  $T/2$ , tenemos  $\bar{x}_1 = \bar{x}_2 = 0$ ,  $\text{var}(x_1) = \text{var}(x_2) = T/2$  y  $\text{cov}(x_1 x_2) = 0$ .

Comencemos recordando la fórmula para la suma de una progresión geométrica. Escribiendo

$$S_T = a + \dots + a^T$$

entonces

$$aS_T = a^2 + \dots + a^{T+1}$$

y restando ambas ecuaciones y despejando  $S_T$

$$S_T = \frac{a - a^{T+1}}{1 - a}$$

En segundo lugar, recordemos que si  $i = \sqrt{-1}$ , la exponencial de un número complejo  $i\omega$ , se define como

$$e^{i\omega} = \cos(\omega) + i\text{sen}(\omega),$$

y que aplicando la fórmula de la suma a la progresión geométrica:

$$S_T = \sum_{t=1}^T e^{i\omega t} = \frac{e^{i\omega} - e^{i\omega(T+1)}}{1 - e^{i\omega}} = \frac{e^{i\omega}(1 - e^{i\omega T})}{1 - e^{i\omega}}$$

Supongamos ahora que  $w = 2\pi j/T$ , donde  $j$  es un entero positivo. Entonces  $e^{i\omega T} = e^{i2\pi j} = \cos(2\pi j) + i\text{sen}(2\pi j) = \cos(2\pi) + i\text{sen}(2\pi) = 1$ . Sustituyendo este resultado en la fórmula anterior para  $S_T$  concluimos que  $S_T = 0$  y tenemos que

$$S_T = \sum_{t=1}^T e^{i\omega t} = \sum_{t=1}^T \cos(\omega t) + i \sum_{t=1}^T \text{sen}(\omega t) = 0.$$

lo que exige que  $\sum_{t=1}^T \cos(\omega t) = \sum_{t=1}^T \text{sen}(\omega t) = 0$ . Por otro lado, como  $\text{sen}(a+b) = \text{sen}(a)\cos(b) + \cos(a)\text{sen}(b)$  tenemos que tomando  $a = b = \omega t$  y sumando para todos los datos muestrales:

$$\sum_{t=1}^T \cos(\omega t)\text{sen}(\omega t) = \frac{1}{2} \sum_{t=1}^T \text{sen}(2\omega t) = 0$$

Esta ecuación indica que la covarianza entre las variables seno y coseno para las frecuencias básicas es cero. Para calcular las varianzas utilizamos que  $\cos(a+b) = \cos(a)\cos(b) - \text{sen}(a)\text{sen}(b)$  y tomando  $a = b = \omega t$

$$\cos(2\omega t) = \cos^2 \omega t - \text{sen}^2 \omega t = \cos^2 \omega t - (1 - \cos^2 \omega t)$$

y despejando  $\cos^2(\omega t)$  y sumando para todos los datos:

$$\sum_{t=1}^T \cos^2(\omega t) = \frac{1}{2} \sum_{t=1}^T (1 + \cos(2\omega t)) = \frac{T}{2}.$$

De la misma forma se obtiene:

$$\sum_{t=1}^T \text{sen}^2(\omega t) = \frac{1}{2} \sum_{t=1}^T (1 - \cos(2\omega t)) = \frac{T}{2}$$

con lo que concluimos que la varianza de las variables  $x_{1t} = \text{sen}(wt)$  y  $x_{2t} = \cos(wt)$  es  $1/2$ .

Observemos que la prueba es válida para cualquier  $w_j = 2\pi j/T$ , donde  $j$  es un entero positivo. Vamos a comprobar que si consideramos dos frecuencias de Fourier distintas y definimos las cuatro variables  $x_{1t} = \text{sen}(w_1t)$ ,  $x_{2t} = \cos(w_1t)$ ,  $x_{3t} = \text{sen}(w_2t)$  y  $x_{4t} = \cos(w_2t)$  ligadas a las dos frecuencias, las cuatro variables están incorreladas. En efecto, consideremos, por ejemplo, las variables  $x_{1t}$  y  $x_{4t}$ . Entonces:

$$x_{1t}x_{4t} = \text{sen}(w_1t) \cos(w_2t) = \frac{1}{2} [\text{sen}(w_1 + w_2)t - \text{sen}(w_1 - w_2)t]$$

y  $\text{sen}(w_1 + w_2)t = \text{sen}(w_3t)$  donde  $w_3$  es también de la forma  $w_3 = 2\pi j/T$ , siendo  $j$  un entero positivo, esta variable tiene media cero. La prueba es similar en los otros casos.

En resumen, hemos comprobado que si definimos variables explicativas  $x_{jt} = \text{sen}(w_jt)$  y  $x_{ht} = \cos(w_ht)$  para las frecuencias de Fourier estas variables tienen media cero, varianza  $1/2$  y están incorreladas. Si el tamaño muestral no es un múltiplo exacto de la frecuencia de interés, es decir, no se verifica que  $w_3 = 2\pi j/T$ , los resultados son sólo aproximados pero si el tamaño muestral es grande serán bastante exactos. Por ejemplo, si nos interesa  $f = 1/12$  porque tenemos datos mensuales y  $T = 121$  no podemos expresar  $2\pi/12$  como  $2\pi j/121$ . Sin embargo, como  $2\pi/12 = 2\pi 10/120$ , tomando  $j=10$ , que corresponde a la frecuencia  $f = 10/121$  esta aproximación a la frecuencia básica será buena.