

Capítulo 8

PREDICCIÓN CON MODELOS ARIMA

Andrei Nikolaevich Kolmogorov (1903-1987)

Matemático ruso, fundador del cálculo de probabilidades moderno. Desde muy joven mostró una extraordinaria aptitud para las matemáticas. Pionero en muchas ramas de la matemática, estableció las bases rigurosas para el estudio de los procesos estocásticos y desarrolló la teoría de predicción para procesos estocásticos estacionarios. Dedicó mucho esfuerzo en su madurez a la educación de niños superdotados.

8.1 Introducción

En este capítulo vamos a estudiar la predicción mediante modelos ARIMA cuando los parámetros son conocidos. Definiremos los predictores óptimos como aquellos que minimizan en promedio los errores de predicción al cuadrado, y que estos predictores se obtienen calculando los valores esperados, condicionados a los valores observados, de las observaciones futuras. Veremos que la función de predicción de un modelo ARIMA tienen una estructura simple: los operadores no estacionarios, es decir, las diferencias y la constante si existe, determinan la predicción a largo plazo, mientras que los operadores estacionarios, AR y MA, determinan la predicción a corto plazo.

Las predicciones tienen poca utilidad si no damos también una medida de su precisión, y veremos como calcular intervalos de confianza de las predicciones. Finalmente, estudiaremos como revisar las predicciones cuando recibimos nueva información.

8.2 La ecuación de Predicción de un modelo ARIMA

8.2.1 La esperanza condicionada como predictor óptimo

Supongamos que se ha observado una realización de longitud T , de una serie temporal $\mathbf{z}_T = (z_1, \dots, z_T)$, y se desea obtener una predicción de un valor futuro $k > 0$ periodos adelante, z_{T+k} . Llamemos $\hat{z}_{T+k} = f(\mathbf{z}_T)$ a un predictor de z_{T+k} obtenido como función de los valores observados. En particular, en este capítulo vamos a estudiar predictores lineales, que son aquellos que se construyen como una combinación lineal de los datos disponibles:

$$\hat{z}_{T+k} = \alpha_1 z_T + \alpha_2 z_{T-1} + \dots + \alpha_T z_1$$

Definir un predictor supone indicar las constantes $\alpha_1, \dots, \alpha_T$ utilizadas para construirlo. Naturalmente queremos que el predictor esté lo más cerca posible del verdadero valor que queremos prever, de manera que el error de predicción

$$C = z_{T+k} - \hat{z}_{T+k}$$

sea lo más pequeño posible. Como la variable z_{T+k} es desconocida, es imposible conocer a priori el error que tendremos al predecirla. Sin embargo, si conocemos su distribución de probabilidad, es decir, los valores

posibles y sus probabilidades, para un predictor definido podemos calcular la probabilidad de que el error esté entre unos valores dados. Para decidir entre posibles predictores podemos especificar un criterio de predicción y seleccionar el predictor que se comporte mejor de acuerdo con este criterio.

El criterio de predicción dependerá de las consecuencias de los errores de predicción. Si nos resulta indiferente cometer errores de predicción negativos que positivos, es decir, predecir por defecto o por exceso, podemos prescindir del signo del error y minimizar el valor promedio de una función del error simétrica, como $l(e_{T+k}) = e_{T+k}^2$ o $l(e_{T+k}) = |e_{T+k}|$, para ciertos valores de las constantes. Hay situaciones donde los errores por defecto son más costosos que los errores por exceso, y entonces podemos minimizar el valor promedio de una función asimétrica del error. Por ejemplo,

$$l(e_{T+k}) = \begin{cases} c_1 |e_{T+k}|, & \text{si } e_{T+k} \geq 0 \\ c_2 |e_{T+k}|, & \text{si } e_{T+k} \leq 0 \end{cases} \quad (8.1)$$

para ciertos c_1 y c_2 positivos. Cuanto mayor sea la relación c_2/c_1 más penalizaremos los errores por defecto. La figura 8.1 presenta distintas funciones del error cuyo valor promedio podemos minimizar. En esta figura hemos representado las dos funciones simétricas $l_1(e_{T+k}) = e_{T+k}^2$, $l_2(e_{T+k}) = 2 |e_{T+k}|$ y (8.1) con $c_1 = 2/3$ y $c_2 = 9/2$. Si comparamos las dos funciones simétricas vemos que la cuadrática da menos importancia a errores de predicción menores de 2 que la función de valores absolutos considerada, mientras que penaliza más los errores mayores que 2. La función asimétrica da muy poco peso a los errores positivos, pero penaliza mucho los negativos.

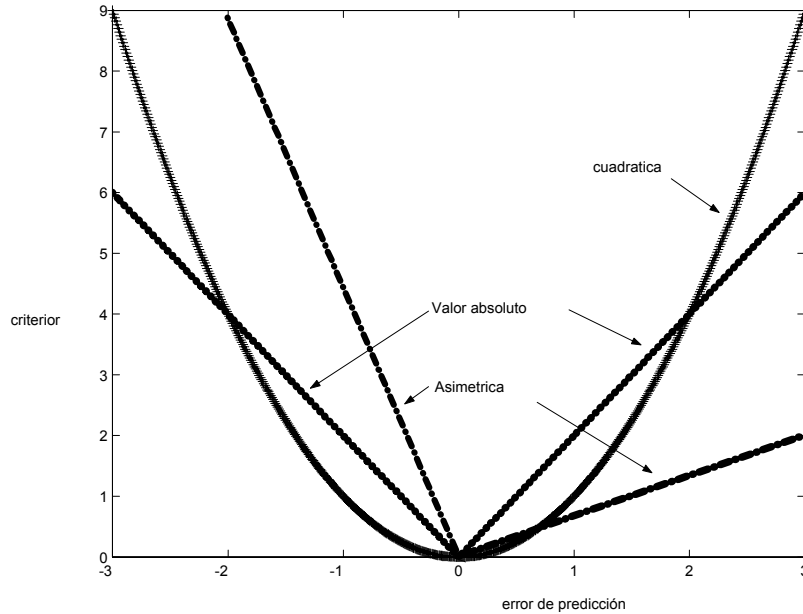


Figura 8.1: Tres funciones del error de predicción que pueden considerarse. La función cuadrática, la de valor absoluto y una asimétrica.

Supongamos para simplificar que queremos minimizar el error cuadrático medio de predicción (ECMP) de z_{T+k} cuando disponemos de la información dada por \mathbf{z}_T . Entonces, tendremos que minimizar :

$$ECMP(z_{T+k}|\mathbf{z}_T) = E[(z_{T+k} - \hat{z}_{T+k})^2|\mathbf{z}_T] = E[e_{T+k}^2|\mathbf{z}_T]$$

donde la esperanza se toma respecto a la distribución de la variable z_{T+k} condicionada a los valores observados \mathbf{z}_T . Vamos a demostrar que el predictor que minimiza este error cuadrático medio es la esperanza z_{T+k} condicionada a la información disponible. Por tanto, si podemos calcular esta esperanza podemos disponer del predictor óptimo sin necesidad de conocer la distribución condicionada completa.

Para demostrarlo, llamemos

$$\mu_{T+k|T} = E[z_{T+k}|\mathbf{z}_T]$$

a la media de esta distribución condicionada. Restando y sumando $\mu_{T+k|T}$ en la expresión de $ECMP(z_{T+k})$ y desarrollando el cuadrado tenemos que

$$ECMP(z_{T+k}|\mathbf{z}_T) = E[(z_{T+k} - \mu_{T+k|T})^2|\mathbf{z}_T] + E[(\mu_{T+k|T} - \hat{z}_{T+k})^2|\mathbf{z}_T] \quad (8.2)$$

ya que el doble producto se anula, por ser $\mu_{T+k|T} - \hat{z}_{T+k}$ una constante (como \hat{z}_{T+k} es función de los valores pasados y estamos condicionando a ellos será constante) que saldrá fuera de la esperanza, y :

$$E[(\mu_{T+k|T} - \hat{z}_{T+k})(z_{T+k} - \mu_{T+k|T})|\mathbf{z}_T] = (\mu_{T+k|T} - \hat{z}_{T+k})E[(z_{T+k} - \mu_{T+k|T})|\mathbf{z}_T] = 0$$

con lo que obtenemos (8.2), que podemos escribir:

$$ECMP(z_{T+k}|\mathbf{z}_T) = var(z_{T+k}|\mathbf{z}_T) + E[(\mu_{T+k|T} - \hat{z}_{T+k})^2|\mathbf{z}_T].$$

Como el primer término de esta expresión es constante, minimizaremos $ECMP$ haciendo cero el segundo término. Esto ocurrirá si tomamos:

$$\hat{z}_{T+k} = E[z_{T+k}|\mathbf{z}_T] = \mu_{T+k|T}$$

Hemos demostrado que el predictor que minimiza el error cuadrático medio de predicción es la esperanza de la observación que queremos prever condicionada a los datos observados. Podríamos preguntarnos por el predictor que minimiza este error en sentido absoluto, es decir, no sólo promediando para los posibles valores de z_{T+k} , sino también promediando para todas las posibles series temporales \mathbf{z}_T que pueden observarse. En el apéndice 8.1 se demuestra que este predictor es también $\mu_{T+k|T}$, la media condicionada a la información disponible.

Ejemplo 8.1

Supongamos que disponemos de 50 datos generados por un proceso AR(1) $z_t = 10 + 0.5z_{t-1} + a_t$. El último valor observado corresponde al instante $t=50$, y es $z_{50} = 18$, mientras que el valor anterior ha sido $z_{49} = 15$. Queremos calcular predicciones para los dos periodos siguientes, $t=51$ y $t=52$.

La primera observación que queremos prever es la z_{51} . Su expresión es

$$z_{51} = 10 + 0.5z_{50} + a_{51}$$

y su esperanza, condicionada a los valores observados es

$$\hat{z}_T(1) = 10 + 0.5(18) = 19.$$

Para dos periodos adelante, la expresión de la variable es:

$$z_{52} = 10 + 0.5z_{51} + a_{52}$$

y tomando esperanzas condicionadas a los datos

$$\hat{z}_T(2) = 10 + 0.5\hat{z}_T(1) = 10 + 0.5(19) = 19.5.$$

8.2.2 Cálculo de las predicciones

Supongamos que se dispone de una realización de tamaño T , $\mathbf{z}_T = (z_1, \dots, z_T)$, de un proceso ARIMA(p, d, q) con parámetros conocidos $\phi(B)\nabla^d z_t = c + \theta(B)a_t$. Vamos a ver como aplicar el resultado anterior para calcular las predicciones. Como los parámetros son conocidos podemos obtener todas las innovaciones a_t fijando unos valores iniciales. Por ejemplo, si el proceso es ARMA(1,1) tomando $a_1 = 0$ las innovaciones, a_2, \dots, a_T , se calculan recursivamente mediante la ecuación:

$$a_t = z_t - c - \phi z_{t-1} + \theta a_{t-1}.$$

En consecuencia, supondremos en adelante que tanto las observaciones como las innovaciones son conocidas hasta el instante T .

De acuerdo con el resultado de la sección anterior, la predicción que minimiza el error cuadrático medio de z_{T+k} , que llamaremos para simplificar en adelante la predicción óptima de z_{T+k} , es la esperanza de la variable condicionada a los valores observados. Definamos

$$\begin{aligned}\widehat{z}_T(j) &= E[z_{T+j}|\mathbf{z}_T] \quad j = 1, 2, \dots \\ \widehat{a}_T(j) &= E[a_{t+j}|\mathbf{z}_T] \quad j = 1, 2, \dots\end{aligned}$$

donde el subíndice T representa el origen de la predicción, que suponemos fijo, y j el horizonte de la misma, que irá cambiando para generar predicciones de distintas variables futuras desde el origen T . Llamando $\varphi_h(B) = \phi_p(B)\nabla^d$ al operador de orden $h = p + d$ que se obtiene multiplicando el operador AR(p) estacionario y el operador de diferencia (si existe), tendremos que la expresión de la variable z_{T+k} es:

$$z_{T+k} = c + \varphi_1 z_{t+k-1} + \dots + \varphi_h z_{T+k-h} + a_{T+k} - \theta_1 a_{T+k-1} - \dots - \theta_q a_{T+k-q}. \quad (8.3)$$

Tomando esperanzas condicionadas a \mathbf{z}_T en todos los términos de esta expresión, se obtiene:

$$\widehat{z}_T(k) = c + \varphi_1 \widehat{z}_T(k-1) + \dots + \varphi_h \widehat{z}_T(k-h) - \theta_1 \widehat{a}_T(k-1) - \dots - \widehat{a}_T(k-q) \quad (8.4)$$

En esta ecuación algunas esperanzas se calculan para variables observadas y otras para variables no observadas. Cuando $i > 0$ la expresión $\widehat{z}_T(i)$ es la esperanza condicionada de la variable z_{T+i} que no se ha observado todavía. Sin embargo, cuando $i \leq 0$, $\widehat{z}_T(i)$ es la esperanza condicionada de la variable $z_{T-|i|}$, que ya se ha observado y es conocida, por lo que esta esperanza coincidirá con la observación y $\widehat{z}_T(-|i|) = z_{T-|i|}$. Con relación a las innovaciones, la esperanza de las innovaciones futuras condicionada a las historia de la serie es igual a su esperanza absoluta, ya que la variable a_{T+i} es independiente de \mathbf{z}_T . Como las innovaciones tienen esperanza nula, concluimos que para $i > 0$, las innovaciones futuras $\widehat{a}_T(i)$ serán cero. Cuando $i \leq 0$ las innovaciones $a_{T-|i|}$ son conocidas, por lo que sus esperanzas coinciden con los valores observados y $\widehat{a}_T(-|i|) = a_{T-|i|}$. De esta manera, podemos calcular recursivamente las predicciones, comenzando con $j = 1$ y continuando con $j = 2, 3, \dots$, hasta el horizonte deseado.

Observemos que restando (8.3) de (8.4) para $k = 1$, se obtiene que:

$$a_{T+1} = z_{T+1} - \widehat{z}_T(1)$$

que indica que las innovaciones pueden interpretarse como los errores de predicción a un período por delante cuando conocemos los parámetros del modelo.

La ecuación (8.4) indica que después de q valores iniciales los términos de media móvil desaparecerán, y la predicción quedará determinada exclusivamente por la parte autorregresiva del modelo. En efecto, para $j > q$ todas las innovaciones que aparecen en la ecuación de predicción son no observadas, sus esperanzas serán cero y desaparecerán de la ecuación de predicción. Entonces, las predicciones de (8.4) para $j > q$ satisfacen la ecuación

$$(1 - \varphi_1 B - \dots - \varphi_h B^h) \widehat{z}_T(j) = \phi(B) \nabla^d \widehat{z}_T(j) = c \quad j > q \quad (8.5)$$

donde en esta ecuación $B\widehat{z}_T(j) = \widehat{z}_T(j-1)$ ya que el operador B actúa ahora sobre el horizonte de la previsión j siendo T , el origen de la predicción, un valor fijo. Esta ecuación se denomina la *ecuación de predicción final* porque establece como están ligadas las predicciones para horizontes altos cuando ha desaparecido la parte de media móvil. Observemos que la relación entre las predicciones es similar a la existente entre las autocorrelaciones de un proceso estacionario ARMA, aunque en las predicciones además del operador estacionario $\phi(B)$ aparece también el operador no estacionario ∇^d , y la ecuación no está igualada a cero sino a la constante c .

Ejemplo 8.2

Supongamos el proceso AR(1) $z_t = c + \phi_1 z_{t-1} + a_t$. La ecuación de predicción un periodo adelante es

$$\widehat{z}_T(1) = c + \phi_1 z_T,$$

y para dos periodos:

$$\hat{z}_T(2) = c + \phi_1 \hat{z}_T(1) = c(1 + \phi_1) + \phi_1^2 z_T,$$

Estas ecuaciones se obtuvieron para unos valores particulares de los parámetros en el ejemplo 8.1. Generalizando, para cualquier periodo $k > 0$ tenemos:

$$\hat{z}_T(k) = c + \phi_1 \hat{z}_T(k-1) = c(1 + \phi_1 + \dots + \phi_1^k) + \phi_1^k z_T.$$

Para k grande como $|\phi_1| < 1$ el término $\phi_1^k z_T$ será próximo a cero y la predicción será el término $c(1 + \phi_1 + \dots + \phi_1^k) = c/(1 - \phi_1)$, que es la media del proceso. Veremos en la sección siguiente que para cualquier proceso estacionario ARMA (p, q) la predicción para un horizonte k grande es la media del proceso, $\mu = c/(1 - \phi_1 - \dots - \phi_p)$. Como caso particular si $c = 0$ la predicción a largo plazo será cero, la media del proceso. La información actual deja de ser útil para la predicción a largo plazo.

Ejemplo 8.3

Supongamos el paseo aleatorio $\nabla z_t = c + a_t$. La predicción a un paso con este modelo es

$$\hat{z}_T(1) = c + z_T$$

y para dos:

$$\hat{z}_T(2) = c + \hat{z}_T(1) = 2c + z_T$$

y para horizonte k cualquiera:

$$\hat{z}_T(k) = c + \hat{z}_T(k-1) = kc + z_T.$$

Como la predicción del siguiente periodo se obtiene sumando siempre la misma constante, c , concluimos que las predicciones seguirán una línea recta con pendiente c . Si $c = 0$ las predicciones de todos los periodos son iguales al último valor observado, y seguirán una recta horizontal. Vemos que la constante determina la pendiente de la ecuación de predicción. Sin embargo, el origen de la predicción es siempre la última información disponible.

Ejemplo 8.4

Consideremos el modelo estacional para datos mensuales $\nabla_{12} z_t = (1 - 0.8B^{12})a_t$. La predicción para cualquier mes del año siguiente, j , donde $j = 1, 2, \dots, 12$, será

$$\hat{z}_T(j) = z_{T+j-12} - 0.8a_{T+j-12-1}, \quad j = 1, \dots, 12$$

y para el segundo año, como las perturbaciones a_t ya no serán observadas:

$$\hat{z}_T(12+j) = \hat{z}_T(j)$$

Esta ecuación indica que las predicciones para todos los meses del segundo año serán idénticas a las del primer año. Lo mismo ocurrirá para los años posteriores. Por tanto, la ecuación de predicción está formada por 12 coeficientes que se repiten año tras año. Para interpretar estos coeficientes, sea

$$\bar{z}_T(12) = \frac{1}{12} \sum_{j=1}^{12} \hat{z}_T(j) = \frac{1}{12} \sum_{j=1}^{12} z_{T+j-12} - 0.8 \frac{1}{12} \sum_{j=1}^{12} a_{T+j-12-1}$$

la media de las predicciones para los doce meses del primer año. (En la práctica esta media no será muy distinta de la media de los últimos doce meses observados, ya que como las innovaciones tienen esperanza nula $\sum_{j=1}^{12} a_{T+j-12-1}$ será próximo a cero). Entonces, las predicciones pueden escribirse:

$$\hat{z}_T(12k+j) = \bar{z}_T(12) + S_T(j), \quad j = 1, \dots, 12; k = 1, 2, \dots$$

donde $\bar{z}_T(12)$ puede interpretarse como una media general y los coeficientes $S_T(j) = \hat{z}_T(j) - \bar{z}_T(12)$, que suman cero para los doce meses por construcción, pueden interpretarse como los coeficiente estacionales. La predicción con este modelo es pues la suma de un nivel constante más el coeficiente estacional del mes. Una forma simple de obtener estos coeficientes estacionales es generar 12 predicciones y restar a cada mes la media de las 12 predicciones obtenidas. La diferencia, $S_T(j) = \hat{z}_T(j) - \bar{z}_T(12)$, son los coeficientes estacionales.

8.3 Interpretación de las Predicciones

8.3.1 Procesos no estacionales

Para modelos ARIMA no estacionales hemos visto en la sección anterior que las predicciones efectuadas desde el origen T con horizonte k satisfacen para $k > q$ la llamada *ecuación de predicción final* :

$$\phi(B) \nabla^d \hat{z}_T(k) = c \quad (8.6)$$

donde el operador B actúa sobre k y ahora T es fijo.

Supongamos inicialmente el caso más simple de $c = 0$. Como sabemos la constante es muy importante para procesos no estacionarios y esto va a aparecer en las predicciones. La ecuación (8.6) requiere $p+d$ valores iniciales que se calculan teniendo en cuenta la parte MA. Hemos visto en el apéndice 4.1 que la solución de una ecuación en diferencias que puede expresarse como producto de operadores primos (sin raíces en común) es la suma de las soluciones correspondientes a cada uno de estos operadores. Como los polinomios $\phi(B)$ y ∇^d son primos (no tienen ninguna raíz en común), la solución de esta ecuación será la suma de las soluciones correspondientes al operador estacionario, $\phi(B)$, y al no estacionario, ∇^d . Podemos escribir:

$$\hat{z}_T(k) = P_T(k) + t_T(k) \quad (8.7)$$

donde estos componentes deben satisfacer:

$$\nabla^d P_T(k) = 0 \quad (8.8)$$

$$\phi(B) t_T(k) = 0 \quad (8.9)$$

Es inmediato comprobar que (8.7) con las condiciones (8.8) y (8.9) satisface (8.6). Además, la ecuación (8.7) será válida para $k > q - (p + d)$ ya que la solución de (8.6) requiere $p + d$ valores iniciales. La solución del operador no estacionario, $P_T(k)$, representa la *componente permanente* de la predicción y la solución del operador estacionario, $t_T(k)$, es la *componente transitoria*.

La solución de (8.8) es la tendencia prevista:

$$P_T(k) = \beta_0^{(T)} + \beta_1^{(T)} k + \dots + \beta_{d-1}^{(T)} k^{(d-1)} \quad (8.10)$$

donde los parámetros $\beta_j^{(T)}$ son constantes a determinar que dependen del origen T , y se calculan utilizando las últimas observaciones disponibles. La solución de (8.9) es el componente transitorio previsto que, como obtuvimos en el apéndice 4.1, viene dado por:

$$t_T(k) = \sum_{i=1}^P A_i G_i^k$$

siendo G_i^{-1} las raíces del operador AR estacionario. Este componente tiende a cero con k , ya que al ser $|G_i| \leq 1$, G_i^k disminuirá con k , lo que justifica su nombre de componente transitorio.

En consecuencia, la ecuación de predicción final es la suma de dos componentes:

- (1) un componente permanente o de tendencia prevista, que es un polinomio de orden $d-1$ con coeficientes que dependen del origen de la predicción;
- (2) un componente transitorio, que es una mezcla de términos cíclicos y exponenciales (según las raíces de $\phi(B) = 0$) y que tiende a cero con k .

En la mayoría de los casos $d = 1$ ó 2 , con lo que el componente de tendencia es una constante (si $d = 1$) o una recta (si $d = 2$). Para determinar los coeficientes $\beta_j^{(k)}$ de la ecuación de la tendencia prevista, el método más simple es generar predicciones con (8.4) para un horizonte k grande hasta que:

- (a) si $d = 1$, sean constantes. Entonces esa constante es $\beta_0^{(T)}$,

- (b) si $d = 2$, se verifique que la diferencia entre predicciones, $\widehat{z}_T(k) - \widehat{z}_T(k-1)$ sea constante. Entonces las predicciones están en una recta, y su pendiente es $\beta_1^{(T)} = \widehat{z}_T(k) - \widehat{z}_T(k-1)$, que va adaptándose en función de las observaciones disponibles.

Consideremos ahora el caso donde la constante del modelo no es cero. En lo anterior hemos supuesto que la serie estacionaria tiene media cero. Si no es así, la ecuación final de predicción final es:

$$\phi(B) \nabla^d \widehat{z}_T(k) = c \quad (8.11)$$

que también puede escribirse como

$$\phi(B) (\nabla^d \widehat{z}_T(k) - \mu) = 0$$

donde $\mu = c/\phi(1)$ es la media de la serie estacionaria $\nabla^d \widehat{z}_T(k)$. Hemos visto que la solución de una ecuación en diferencias que no está igualada a cero, es decir no es homogénea, es la solución de la homogénea (la ecuación igualada a cero), más una solución particular. La solución de (8.11) igualada a cero es (8.10). La solución particular debe ser tal que al diferenciar d veces se obtenga una constante. Esto ocurre con funciones del tipo $\beta_d k^d$, donde el valor β_d se obtiene imponiendo la condición de que esta expresión verifique la ecuación, es decir:

$$\phi(B) \nabla^d (\beta_d k^d) = c.$$

Vamos a demostrar que al aplicar el operador ∇^d a $\beta_d k^d$ se obtiene $d! \beta_d$, donde $d! = (d-1)(d-2)\dots 2$. Esto es cierto para $d = 1$, ya que

$$(1-B)\beta_d k = \beta_d k - \beta_d(k-1) = \beta_d$$

y para $d = 2$, ya que

$$(1-B)^2 \beta_d k^2 = (1-2B+B^2)\beta_d k^2 = \beta_d k^2 - 2\beta_d(k-1)^2 + \beta_d(k-2)^2 = 2\beta_d$$

y es fácil comprobar que si es cierto para $d-1$ lo es para d . Además, al aplicar el operador $\phi(B)$ a una constante se obtiene $\phi(1)$ por esa constante, con lo debe verificarse

$$\phi(1) d! \beta_d = c$$

y β_d viene dado por :

$$\beta_d = \frac{c}{\phi(1) d!} = \frac{\mu}{d!}. \quad (8.12)$$

En consecuencia, la solución de la ecuación (8.11) será

$$\widehat{z}_T(k) = P_T(k) + t_T(k) + \beta_d k^d.$$

En esta ecuación existe una diferencia básica entre los coeficientes del polinomio $P_T(k)$, que dependen del origen de las predicciones y van cambiando en el tiempo, y el coeficiente β_d , dado por (8.12), que es constante para cualquier origen de predicción porque sólo depende de la media de la serie estacionaria. Como el comportamiento a largo plazo dependerá del término de orden más alto vemos que las predicciones a largo plazo de un modelo con constante quedan determinadas por esta constante.

Si incluimos el término fijo $\beta_d k^d$ dentro del componente permanente, podemos escribir la solución general como suma de un componente permanente, obtenido como solución de

$$\nabla^d P_T^*(k) = \mu$$

que será un polinomio de grado d con coeficientes hasta $d-1$ que dependen del origen de la predicción y con el último coeficiente constante, dado por (8.12):

$$P_T^*(k) = \beta_0^{(T)} + \beta_1^{(T)} k + \dots + \beta_d k^d$$

El componente transitorio se define como en el caso anterior.

Ejemplo 8.5

Comparar la predicción a largo plazo de un paseo aleatorio con constante, $\nabla z_t = c + a_t$, con la del modelo $\nabla^2 z_t = (1 - \theta B) a_t$.

La predicción del paseo aleatorio es, de acuerdo a lo anterior, una recta con pendiente c (véase también el ejemplo 8.3), que es la media de la serie estacionaria $w_t = \nabla z_t$ y se estima con

$$\hat{c} = \frac{1}{T-1} \sum_{t=2}^{t=T} \nabla z_t = \frac{z_T - z_1}{T-1}$$

Como

$$\hat{z}_T(k) = \hat{c} + \hat{z}_T(k-1) = k\hat{c} + z_T$$

La predicción del crecimiento futuro en cualquier periodo $\hat{z}_T(k) - \hat{z}_T(k-1)$ será igual el crecimiento medio observado en toda la serie \hat{c} . Observemos que esta pendiente es constante.

El modelo con dos diferencias si $\theta \rightarrow 1$ será muy parecido al $\nabla z_t = c + a_t$, ya que la solución de

$$\nabla^2 z_t = \nabla a_t$$

es

$$\nabla z_t = c + a_t$$

ya que al cancelar las diferencias queda una constante indeterminada. En efecto al tomar diferencias en la última ecuación se obtiene la anterior. Sin embargo, cuando θ no es uno aunque este próxima a uno la predicción de ambos modelos puede ser muy diferente a largo plazo. En el modelo con dos diferencias la ecuación de predicción final será también una línea recta, pero con pendiente que va cambiando con el retardo, mientras que en el modelo con una diferencia y constante la pendiente es siempre la constantes. Por ejemplo, la predicción para un periodo adelante del modelo con dos diferencias será

$$\hat{z}_T(1) = 2z_T - z_{T-1} - \theta a_T = z_T + \hat{\beta}_T$$

donde hemos llamado

$$\hat{\beta}_T = \nabla z_T - \theta a_T.$$

a cantidad que sumamos a la última observación que podemos considerar como la pendiente aparente en el instante T . Para dos periodos

$$\hat{z}_T(2) = 2\hat{z}_T(1) - z_T = z_T + 2\hat{\beta}_T$$

y vemos que sumamos dos veces la pendiente aparente, confirmando que las predicciones seguirán una línea recta. De la misma forma, es fácil comprobar que:

$$\hat{z}_T(k) = z_T + k\hat{\beta}_T$$

que muestra que las predicciones siguen una línea recta con pendiente $\hat{\beta}_T$. Observemos que la pendiente va cambiando en el tiempo, ya que depende del último crecimiento observado, ∇z_T , y del último error de previsión cometido a_T . Sustituyendo $a_T = (1 - \theta B)^{-1} \nabla^2 z_T$ en la definición de $\hat{\beta}_T$, podemos escribir

$$\hat{\beta}_T = \nabla z_T - \theta(1 - \theta B)^{-1} \nabla^2 z_T = (1 - \theta(1 - \theta B)^{-1} \nabla) \nabla z_T,$$

y el operador resultante es

$$1 - \theta(1 - \theta B)^{-1} \nabla = 1 - \theta(1 - (1 - \theta)B - \theta(1 - \theta)B^2 - \dots)$$

con lo que $\hat{\beta}_T$ puede escribirse

$$\hat{\beta}_T = (1 - \theta) \sum_{i=0}^{T-1} \theta^i \nabla z_{T-i}$$

Esta expresión muestra que la pendiente de la ecuación de predicción final es una media ponderada de todos los crecimientos observados con pesos que decrecen geométricamente. Observemos que si θ es próximo a uno

esta media será próxima a la media aritmética, mientras que si θ es próximo a cero la pendiente se estima únicamente con lo últimos valores observados. Este ejemplo muestra la mayor flexibilidad del modelo con dos diferencias frente al modelo con una diferencia y constante.

Es interesante comparar estos modelos con uno determinista que genere también predicciones siguiendo una línea recta. Consideremos el modelo lineal determinista:

$$\hat{z}_T(k) = a + b_R(T + k)$$

Supongamos para simplificar el análisis que $T = 5$, y escribamos $t = (-2, -1, 0, 1, 2)$ para que $\bar{t} = 0$. Sean entonces las cinco observaciones $(z_{-2}, z_{-1}, z_0, z_1, z_2)$. La pendiente de la recta se estima por mínimos cuadrados por

$$b_R = \frac{-2z_{-2} - z_{-1} + z_1 + 2z_2}{10}$$

y esta expresión puede escribirse como:

$$b_R = .2(z_{-1} - z_{-2}) + .3(z_0 - z_{-1}) + .3(z_1 - z_0) + .2(z_2 - z_1)$$

que es una media ponderada de los crecimientos observados pero dando peso mínimo al último. Puede demostrarse (Peña, 1995) que en el caso general la pendiente se escribe

$$b_R = \sum_{t=2}^T w_t \nabla z_t$$

donde $\sum w_t = 1$ y los pesos son simétricos y su valor mínimo corresponde a los crecimientos al principio y al fin de la muestra. Este ejemplo muestra las limitaciones de los modelos deterministas para la predicción, que ya comentamos en el capítulo 2, y la mayor flexibilidad que puede obtenerse tomando diferencias.

8.3.2 Procesos estacionales

Si el proceso es estacional la descomposición anterior sigue siendo válida: habrá un componente permanente que dependerá de los operadores no estacionarios, ∇^d , ∇_s y la media μ del proceso estacionario si es no nula, y un componente transitorio, que englobará el efecto de los operadores AR estacionarios (regular y estacional).

Para separar la tendencia de la estacionalidad en el componente permanente los operadores asociados a estos componentes no pueden tener raíces en común. Si la serie tiene una diferencia estacional como:

$$\nabla_s = (1 - B^s) = (1 + B + B^2 + \dots + B^{s-1})(1 - B)$$

el operador estacional $(1 - B^s)$ puede interpretarse como la mezcla de dos operadores: el operador diferencia y un operador estacional puro, dado por

$$S_s(B) = 1 + B + \dots + B^{s-1} \quad (8.13)$$

que produce la suma de s términos consecutivos. Por ejemplo, en una serie mensual, $s = 12$, y esta operación suma los doce meses del año, por lo que se obtiene un valor libre del efecto estacional. La descomposición del operador ∇_s indica que la diferencia entre dos observaciones del mismo mes en dos años consecutivos equivale a calcular los totales anuales y hacer la diferencia entre dos totales consecutivos. Separando la raíz $B = 1$ del operador ∇_s , el modelo estacional, suponiendo una constante distinta de cero, puede escribirse:

$$\Phi(B^s) \phi(B) S_s(B) \nabla^{d+1} z_t = c + \theta(B) \Theta(B^s) a_t$$

donde ahora los cuatro operadores $\Phi(B^s)$, $\phi(B)$, $S_s(B)$ y ∇^{d+1} no tienen raíces en común. La ecuación de predicción final será:

$$\Phi(B^s) \phi(B) S_s(B) \nabla^{d+1} \hat{z}_t(k) = c \quad (8.14)$$

ecuación que será válida para $k > q + sQ$ y que requiere $d + s + p + sP$ valores iniciales (orden máximo de B en el operador de la derecha). Como ahora los operadores no estacionarios, $S_s(B)$ y ∇^{d+1} , no tienen raíces en común, la componente permanente puede escribirse como suma de las soluciones de cada operador. La solución de la ecuación homogénea será:

$$\hat{z}_t(k) = T_T(k) + S_T(k) + t_T(k)$$

donde $T_T(k)$ es el componente de tendencia que se obtiene de la ecuación :

$$\nabla^{d+1}T_T(k) = 0$$

y será un polinomio de grado d con coeficientes que se van adaptando en el tiempo. El componente estacional, $S_T(k)$, es la solución de:

$$S_s(B) S_T(k) = 0 \quad (8.15)$$

que tiene como solución cualquier función de período s con coeficientes que sumen cero. En efecto, una secuencia $S_T(k)$ es solución de (8.15) si verifica:

$$\sum_{j=1}^s S_T(j) = \sum_{j=s+1}^{2s} S_T(j) = 0$$

los coeficientes $S_T(1), \dots, S_T(s)$ que se obtienen al resolver esta ecuación se denominan coeficientes estacionales. Finalmente, el componente transitorio incluirá las raíces de los polinomios $\Phi(B^s)$ y $\phi(B)$ y tendrá la expresión

$$t_T(k) = \sum_{i=1}^{p+P} A_i G_i^k$$

donde los G_i^{-1} son las soluciones de las ecuaciones $\Phi(B^s) = 0$ y $\phi(B) = 0$.

Una solución particular de la ecuación (8.14) será del tipo $\beta_{d+1}k^{d+1}$, donde la constante β_{d+1} viene determinada por la condición

$$\Phi(1)\phi(1)s(d+1)!\beta_{d+1} = c$$

ya que el resultado de aplicar el operador ∇^{d+1} a $\beta_{d+1}k^{d+1}$ es $(d+1)!\beta_{d+1}$ y si aplicamos a una constante el operador $S_s(B)$ obtenemos s veces esa constante. Como la media de la serie estacionaria es $\mu = c/\Phi(1)\phi(1)$ la constante β_{d+1} es

$$\beta_{d+1} = \frac{c}{\Phi(1)\phi(1)s(d+1)!} = \frac{\mu}{s(d+1)!}.$$

que generaliza los resultados del modelo con constante pero sin estacionalidad.

En resumen, la solución general de la ecuación de predicción final de un proceso estacional tendrá tres componentes: (1) un término de tendencia, que será un polinomio de grado d con coeficientes que se van adaptando en el tiempo si no hay constante en el modelo y un polinomio de grado $d+1$ con el coeficiente del término de mayor orden, β_{d+1} determinista y dado por $\mu/s(d+1)!$, siendo μ la media de la serie estacionaria; (2) un componente estacional que cambiará con las condiciones iniciales; (3) Un término transitorio determinado por las raíces de los operadores AR regulares y estacionales. A continuación presentamos el cálculo de estos coeficientes en un modelo muy frecuente en la práctica.

El modelo de pasajeros de avión

El modelo ARIMA estacional más utilizado es el llamado modelo de pasajeros de avión:

$$\nabla \nabla_s z_t = (1 - \theta B)(1 - \Theta B^{12})a_t$$

cuya ecuación para calcular las predicciones es, para $k > 0$

$$\hat{z}_t(k) = \hat{z}_t(k-1) + \hat{z}_t(k-12) - \hat{z}_t(k-13) - \theta \hat{a}_t(k-1) - \Theta \hat{a}_t(k-12) + \theta \Theta \hat{a}_t(k-13).$$

Por otro lado, según los resultados anteriores sabemos que la predicción de este modelo puede escribirse como:

$$\widehat{z}_t(k) = \beta_0^{(t)} + \beta_1^{(t)}k + S_t^{(k)} \quad (8.16)$$

Esta ecuación de predicción contiene 13 parámetros. (Recuérdese que $\sum_{k=1}^{12} S_t^{(k)} = 0$). Esta predicción es la suma de una tendencia lineal –que va cambiando con el origen de la predicción– y de doce coeficientes estacionales, que también se adaptan con el origen de la predicción, y que suman cero. La ecuación (8.16) será válida para $k > q + Qs = 13$, que es el momento futuro en que desaparecen los términos de media móvil, pero como son necesarios $d + s = 13$ valores iniciales para determinarla, la ecuación será válida para $k > q + Qs - d - s = 0$.

Para determinar los coeficientes $\beta_0^{(t)}$ y $\beta_1^{(t)}$ correspondientes a un origen dado y los coeficientes estacionales, podemos resolver el sistema de ecuaciones obtenido al igualar las predicciones para trece períodos, a la estructura (8.16), resultando:

$$\begin{aligned} \widehat{z}_t(1) &= \widehat{\beta}_0^{(t)} + \widehat{\beta}_1^{(t)} + S_1^{(t)} \\ \vdots &\dots \dots\dots \\ \widehat{z}_t(12) &= \widehat{\beta}_0^{(t)} + 12\widehat{\beta}_1^{(t)} + S_{12}^{(t)} \\ \widehat{z}_t(13) &= \widehat{\beta}_0^{(t)} + 13\widehat{\beta}_1^{(t)} + S_1^{(t)} \end{aligned}$$

y con estas ecuaciones podemos obtener $\beta_0^{(t)}$ y $\beta_1^{(t)}$ con la restricción $\sum S_j^{(t)} = 0$. Restando la primera ecuación de la última y dividiendo por 12:

$$\widehat{\beta}_1^{(t)} = \frac{\widehat{z}_t(13) - \widehat{z}_t(1)}{12} \quad (8.17)$$

y la pendiente mensual se obtiene dividiendo por 12 el crecimiento anual, medido a su vez por la diferencia entre las predicciones de un mes cualquiera en dos años consecutivos.

Sumando las 12 primeras ecuaciones se anulan los coeficientes estacionales, y se obtiene:

$$\bar{z}_t = \frac{1}{12} \sum_{j=1}^{12} \widehat{z}_t(j) = \widehat{\beta}_0^{(t)} + \widehat{\beta}_1^{(t)} \left(\frac{1 + \dots + 12}{12} \right)$$

que resulta en:

$$\widehat{\beta}_0^{(t)} = \bar{z}_t - \frac{13}{2} \widehat{\beta}_1^{(t)}$$

finalmente, los coeficientes estacionales se obtienen por diferencia

$$S_j^{(t)} = \widehat{z}_t(j) - \widehat{\beta}_0^{(t)} - \widehat{\beta}_1^{(t)}j$$

y es inmediato comprobar que suman cero dentro del año.

8.4 Varianza de las predicciones

La varianza de las predicciones se calcula fácilmente para procesos MA(q). En efecto, si $z_t = \theta_q(B)a_t$, tenemos que:

$$z_{T+k} = a_{T+k} - \theta_1 a_{T+k-1} - \dots - \theta_q a_{T+k-q}$$

y tomando esperanzas condicionadas a las observaciones hasta el instante T y suponiendo que $q > k$, las innovaciones no observadas se anulan por tener esperanza cero y resulta:

$$\widehat{z}_T(k) = -\theta_k a_T - \theta_{k+1} a_{T-1} - \dots - \theta_q a_{T+k-q}$$

Restando estas dos últimas ecuaciones se obtiene el error de predicción:

$$e_T(k) = z_{T+k} - \hat{z}_T(k) = a_{T+k} - \theta_1 a_{T+k-1} - \dots - \theta_{k-1} a_{T+1}$$

elevando al cuadrado y tomando esperanzas se obtiene el valor esperado del error cuadrático de predicción de z_{T+k} , que será igual a la varianza de su distribución condicionada a las observaciones hasta el instante T :

$$ECMP(z_{T+k} | \mathbf{z}_T) = ECMP(e_T(k)) = E[(z_{T+k} - \hat{z}_T(k))^2 | \mathbf{z}_T] = \sigma^2 (1 + \theta_1^2 + \dots + \theta_{k-1}^2) \quad (8.18)$$

Esta idea puede extenderse a cualquier proceso ARIMA como sigue. Sea $z_t = \psi(B) a_t$ la representación MA(∞) del proceso. Entonces:

$$z_{T+k} = \sum_{i=0}^{\infty} \psi_i a_{T+k-i} \quad (\psi_0 = 1) \quad (8.19)$$

La predicción óptima será, tomando esperanzas condicionadas a las primeras T observaciones

$$\hat{z}_T(k) = \sum_{j=0}^{\infty} \psi_{k+j} a_{T-j} \quad (8.20)$$

y restando (8.20) de (8.19) se obtiene el error de predicción.

$$e_T(k) = z_{T+k} - \hat{z}_T(k) = a_{T+k} + \psi_1 a_{T+k-1} + \dots + \psi_{k-1} a_{T+1}$$

cuya varianza será:

$$Var(e_T(k)) = \sigma^2 (1 + \psi_1^2 + \dots + \psi_{k-1}^2) \quad (8.21)$$

Esta ecuación muestra que la incertidumbre de la predicción es muy diferente para modelos estacionarios y no estacionarios. En un modelo estacionario $\psi_k \rightarrow 0$ si $k \rightarrow \infty$, y la varianza de la predicción a largo plazo converge a un valor constante, la varianza marginal del proceso. Esto es consecuencia de que la predicción a largo plazo es la media del proceso. Por ejemplo, para un AR(1), $\psi_k = \phi^k$ y, cuando k es grande, tenemos que $Var(e_T(k)) = \sigma^2 / (1 - \phi^2)$, que es la varianza marginal del proceso. Si ϕ es grande, la varianza marginal será mucho mayor que la condicionada, σ^2 , pero la incertidumbre es finita y, como máximo, tendremos un error cuadrático de predicción esperado igual a la varianza marginal del proceso. Esto ocurrirá cuando la observación que queremos prever este incorrelada con los valores observados y, por lo tanto, su predicción sea la media del proceso, y su error cuadrático de predicción la varianza del proceso..

Para modelos no estacionarios la serie $\sum \psi_i^2$ no es convergente y la incertidumbre de la predicción a largo plazo crece hasta infinito. A largo plazo no podemos prever el comportamiento de un proceso no estacionario.

Ejemplo 8.6

Dado el modelo ARIMA:

$$(1 - 0.2B) \nabla z_t = (1 - 0.8B) a_t$$

con $\sigma^2 = 4$, y suponiendo que $z_{49} = 30$, $z_{48} = 25$, $a_{49} = z_{49} - \hat{z}_{49/48} = -2$, obtener predicciones desde el último valor observado para $t = 49$ y construir intervalos de confianza aproximados.

La parte AR desarrollada es :

$$(1 - 0.2B)(1 - B) = 1 - 1.2B + 0.2B^2$$

y el modelo puede escribirse:

$$z_t = 1.2z_{t-1} - 0.2z_{t-2} + a_t - 0.8a_{t-1}$$

Entonces:

$$\begin{aligned}
\hat{z}_{49}(1) &= 1.2(30) - 0.2(25) - 0.8(-2) = 32.6 \\
\hat{z}_{49}(2) &= 1.2(32.6) - 0.2(30) = 33.12 \\
\hat{z}_{49}(3) &= 1.2(33.12) - 0.2(32.6) = 33.25 \\
\hat{z}_{49}(4) &= 1.2(33.25) - 0.2(33.12) = 33.28
\end{aligned}$$

Los intervalos de confianza de estas previsiones requieren calcular los coeficientes $\psi(B)$. Igualando los operadores puestos en la forma $MA(\infty)$ tenemos $(1 - 1.2B + 0.2B^2)^{-1}(1 - 0.8B) = \psi(B)$, que implica:

$$(1 - 1.2B + 0.2B^2)(1 + \psi_1 B + \psi_2 B^2 + \dots) = (1 - 0.8B)$$

y operando en el primer miembro:

$$1 - B(1.2 - \psi_1) - B^2(1.2\psi_1 - 0.2 - \psi_2) - B^3(1.2\psi_2 - 0.2\psi_1 - \psi_3) - \dots = (1 - 0.8B).$$

e igualando potencias de B , se obtiene $\psi_1 = 0.4$, $\psi_2 = 0.28$, $\psi_3 = 0.33$. Las varianzas de los errores de predicción serán:

$$\begin{aligned}
Var(e_{49}(1)) &= \sigma^2 = 4 \\
Var(e_{49}(2)) &= \sigma^2(1 + \psi_1^2) = 4 \times 1.16 = 4.64 \\
Var(e_{49}(3)) &= \sigma^2(1 + \psi_1^2 + \psi_2^2) = 4.95 \\
Var(e_{49}(4)) &= \sigma^2(1 + \psi_1^2 + \psi_2^2 + \psi_3^2) = 5.38
\end{aligned}$$

y asumiendo normalidad, los intervalos aproximados del 95% para las cuatro predicciones serán $(32.6 \pm 1.96 \times 2)$, $(33.12 \pm 1.96 \times \sqrt{4.64})$, $(33.25 \pm 1.96 \times \sqrt{4.95})$, y $(33.28 \pm 1.96 \times \sqrt{5.38})$.

8.5 Adaptación de las predicciones

Supongamos que se generan predicciones desde un instante T para períodos futuros $T+1, \dots, T+j$. Cuando se observa el valor z_{T+1} , queremos adaptar las previsiones $\hat{z}_{T+2}, \dots, \hat{z}_{T+j}$ a la vista de esta nueva información. Según (8.20) la predicción de z_{T+k} con información hasta T es:

$$\hat{z}_T(k) = \psi_k a_T + \psi_{k+1} a_{T-1} + \dots$$

mientras que al observar el valor z_{T+1} y obtener el error de predicción, $a_{T+1} = z_{T+1} - \hat{z}_T(1)$, la nueva predicción para z_{T+k} , ahora desde el instante $T+1$, será:

$$\hat{z}_{T+1}(k-1) = \psi_{k-1} a_{T+1} + \psi_k a_T + \dots$$

Restando estas dos expresiones para la predicción de z_{T+k} tenemos que:

$$\hat{z}_{T+1}(k-1) - \hat{z}_T(k) = \psi_{k-1} a_{T+1}$$

Por tanto, cuando observemos z_{T+1} y calculemos a_{T+1} podremos adaptar todas las predicciones mediante:

$$\hat{z}_{T+1}(k-1) = \hat{z}_T(k) + \psi_{k-1} a_{T+1} \quad (8.22)$$

que indica que las predicciones se adaptan añadiendo a las predicciones anteriores una fracción del último error de predicción obtenido. Si $a_{T+1} = 0$ las predicciones no se modificarán.

La ecuación (8.22) tiene una interpretación interesante. Las dos variables z_{T+1} y z_{T+k} tienen, dada la información hasta T , una distribución normal con esperanzas, $\hat{z}_T(1)$ y $\hat{z}_T(k)$ y varianzas σ^2 y $\sigma^2(1 + \psi_1^2 + \dots + \psi_{k-1}^2)$. Su covarianza es

$$cov(z_{T+1}, z_{T+k}) = E[(z_{T+1} - \hat{z}_T(1))(z_{T+k} - \hat{z}_T(k))] = E[a_{T+1}(a_{T+k} + \psi_1 a_{T+k-1} + \dots)] = \sigma^2 \psi_{k-1}$$

La mejor predicción de z_{T+k} dada z_{T+1} vendrá dada por la regresión de la segunda sobre la primera, según la expresión

$$E(z_{T+k}|z_{T+1}) = E(z_{T+k}) + \text{cov}(z_{T+1}, z_{T+k})\text{var}^{-1}(z_{T+1})(z_{T+1} - E(z_{T+1}))$$

donde toda la expresión esta condicionada a la información hasta T . Sustituyendo, se obtiene:

$$\hat{z}_{T+1}(k-1) = \hat{z}_T(k) + (\sigma^2\psi_{k-1})\sigma^{-2}a_{T+1}$$

que es la ecuación (8.22).

Ejemplo 8.7

Adaptar las predicciones del ejercicio 8.5 suponiendo que se observa el valor z_{50} igual a 34.

Entonces:

$$a_{50} = z_{50} - \hat{z}_{49}(1) = 34 - 32.6 = 1.4$$

y las nuevas predicciones serán,

$$\hat{z}_{50}(1) = \hat{z}_{49}(2) + \psi_1 a_{50} = 33.12 + 0.4 \times 1.4 = 33.68$$

$$\hat{z}_{50}(2) = \hat{z}_{49}(3) + \psi_2 a_{50} = 33.25 + 0.28 \times 1.4 = 33.64$$

$$\hat{z}_{50}(3) = \hat{z}_{49}(4) + \psi_3 a_{50} = 33.28 + 0.33 \times 1.4 = 33.74$$

con nuevos intervalos de confianza $(33.68 \pm 1.96 \times 2)$, $(33.64 \pm 1.96 \times \sqrt{4.64})$ y $(33.74 \pm 1.96 \times \sqrt{4.95})$. Observemos que al someterse un error por defecto en la predicción de z_{50} , las predicciones siguientes se revisan al alza.

Ejercicios 8

8.1 Dado el proceso $z_t = 2 + 0.8z_{t-1} - 0.1z_{t-2} + a_t$ y las cuatro observaciones (4,3,1,2.5) generar predicciones para 4 periodos adelante.

8.2 Indicar cual será la predicción a largo plazo con el proceso anterior.

8.3 Suponiendo que la varianza de las innovaciones del proceso anterior es 2, calcular intervalos de confianza para las predicciones a uno y dos pasos.

8.4 Calcular predicciones para $t=100, 101$ y 102 y la ecuación de predicción final del proceso MA(2) $z_t = 5 + a_t - 0.5a_{t-1}$ sabiendo que las predicciones realizadas con información hasta $t = 97$ han sido $\hat{z}_{97}(1) = 5.1$, y $\hat{z}_{97}(2) = 5.3$ y que después hemos observado $z_{98} = 4.9$ y $z_{99} = 5.5$.

8.5 Explicar la estructura de las predicciones en el proceso $\nabla z_t = 3 + (1 - 0.7B)a_t$

8.6 Explicar la estructura de las predicciones a largo plazo en el proceso $\nabla \nabla_{12} z_t = (1 - 0.7B)a_t$

8.6 Apéndice 8.1. Predicción y esperanza condicionada

Vamos a demostrar que la esperanza condicionada minimiza el error cuadrático medio de predicción en un sentido absoluto, es decir, no sólo para la realización observada, sino también para las otras realizaciones posibles que no hemos observado. Es decir, hemos demostrado que dada una serie temporal la predicción con la esperanza condicionada minimiza el error cuadrático de las observaciones futuras, es decir, pero no es inmediato que si tenemos muchas realizaciones del mismo proceso, muchas series temporales, y prevemos cada una de ellas con este criterio minimizamos el error promedio calculado ahora para todas las series. Vamos a ver que esto es cierto.

Para demostrar este resultado partiremos de la importante propiedad siguiente: La esperanza de una función $g(x, y)$ respecto a la distribución conjunta de las dos variables puede obtenerse en dos etapas (a veces se conoce esta regla como de las esperanzas iteradas). Primero, se calcula la esperanza de la función respecto a la distribución de una de ellas condicionada a la otra, por ejemplo tomamos la esperanza respecto a la distribución condicionada de y dada la x , y, segundo, calculamos la esperanza del resultado anterior respecto a la distribución la segunda, en el ejemplo de x . Es decir:

$$E_{xy}[g(x, y)] = E_x[E_{y|x}[g(x, y)]] \quad (8.23)$$

donde E_{xy} quiere decir la esperanza respecto a la distribución conjunta de ambas variables, $E_{y|x}$ la esperanza respecto a la condicionada de y dada x y E_x respecto a la marginal de x .

Demostración

$$\begin{aligned} E_{xy}[g(x, y)] &= \int \int g(x, y) f(x, y) dx dy = \int \left[\int g(x, y) f(y|x) dy \right] f(x) dx = \\ &= \int E_{y|x}[g(x, y)] f(x) dx = E_x[E_{y|x}[g(x, y)]] \end{aligned}$$

y este resultado es válido para dos variables escalares o vectoriales.

Sea ahora z_{t+k} una variable aleatoria parte del proceso estocástico $\{z_t\}$. Cuando se carece de información sobre los valores previos del proceso, la predicción óptima de z_{t+k} es su media μ_{t+k} , ya que la constante que minimiza $E[(z_{t+k} - c)^2]$ es la esperanza de la variable. Sin embargo, dados t valores previos de la realización, $\mathbf{z} = (z_1, \dots, z_t)$, la predicción de z_{t+k} que minimiza el error cuadrático medio condicionada a los datos observados, \mathbf{z} , es la esperanza de la variable z_{t+k} condicionada a lo observado. En consecuencia, el predictor óptimo será el que minimiza:

$$E[(z_{t+k} - \hat{z}_{t+k})^2 | \mathbf{z}_t]$$

que resulta en

$$\hat{z}_{t+k} = E(z_{t+k} | \mathbf{z}_t) \quad (8.24)$$

y el predictor óptimo es la esperanza de la variable desconocida, z_{t+k} , condicionada a la información disponible, \mathbf{z} .

Vamos a demostrar que esta predicción minimiza el error cuadrático medio respecto a la distribución conjunta de (z_{t+k}, \mathbf{z}) , es decir, promediando no sólo respecto a los valores futuros de z_{t+k} , sino también respecto a las posibles secuencias, \mathbf{z} . Consideremos otro predictor cualquiera z_{t+k}^* . Entonces:

$$\begin{aligned} E[(z_{t+k} - z_{t+k}^*)^2] &= E[(z_{t+k} - \hat{z}_{t+k} + \hat{z}_{t+k} - z_{t+k}^*)^2] = \\ &= E[(z_{t+k} - \hat{z}_{t+k})^2] + E[(\hat{z}_{t+k} - z_{t+k}^*)^2] + 2E[(z_{t+k} - \hat{z}_{t+k})(\hat{z}_{t+k} - z_{t+k}^*)] \end{aligned}$$

donde las esperanzas se toman respecto a la distribución conjunta de z_{t+k} y \mathbf{z} . Demostraremos que el doble producto es cero tomando esperanzas respecto a la distribución de $z_{t+k} | \mathbf{z}$ y luego respecto a \mathbf{z} en virtud de (8.23). Como:

$$E[(\hat{z}_{t+k} - z_{t+k}^*)(z_{t+k} - \hat{z}_{t+k}) | \mathbf{z}_t] = (\hat{z}_{t+k} - z_{t+k}^*) E[(z_{t+k} - \hat{z}_{t+k}) | \mathbf{z}_t] = 0$$

ya que tanto \hat{z}_{t+k} como z_{t+k}^* serán funciones de \mathbf{z} que es fijo y \hat{z}_{t+k} está dado por (8.24). En consecuencia, resulta que:

$$E[(z_{t+k} - z_{t+k}^*)^2] = E[(z_{t+k} - \hat{z}_{t+k})^2] + E[(\hat{z}_{t+k} - z_{t+k}^*)^2]$$

que será mínimo cuando $\hat{z}_{t+k} = z_{t+k}^*$. Por tanto, la esperanza condicionada es el predictor óptimo no sólo en un sentido condicional sino también incondicional.

Este resultado, aunque de gran interés teórico, sólo permite obtener el predictor cuando conozcamos las distribuciones condicionadas. En general, esto no ocurre, y buscaremos predictores que siendo funciones lineales de las observaciones minimicen el error de predicción. Si:

$$\hat{z}_{T+k} = b_{k,0} z_T + \dots + b_{k,T-1} z_1 = \mathbf{b}_k' \mathbf{z}_T \quad (8.25)$$

vamos a determinar la expresión del vector $\mathbf{b}_k = (b_{k,0}, \dots, b_{k,T-1})'$. El error cuadrático medio de este predictor es:

$$\begin{aligned}
E[(z_{t+k} - \hat{z}_{t+k})^2] &= E[(z_{t+k} - \mathbf{b}'_k \mathbf{z}_t)(z_{t+k} - \mathbf{b}'_k \mathbf{z}_t)'] \\
&= E[z_{t+k}^2 - 2z_{t+k} \mathbf{b}'_k \mathbf{z}_t + \mathbf{b}'_k \mathbf{z}_t \mathbf{z}'_t \mathbf{b}_k] \\
&= \text{Var}(z_{t+k}) - 2\mathbf{b}'_k \boldsymbol{\gamma}(t, t+k) + \mathbf{b}'_k \boldsymbol{\Gamma}(t) \mathbf{b}_k
\end{aligned}$$

siendo $\boldsymbol{\gamma}(t, t+k)$ el vector de covarianzas entre la variable a prever, z_{t+k} , y los predictores, z_T, \dots, z_1 , y $\boldsymbol{\Gamma}(t)$ la matriz de varianzas y covarianzas de estos predictores. Para obtener el valor que minimiza esta expresión, derivando respecto a \mathbf{b}_k e igualando a cero:

$$-2\boldsymbol{\gamma}(t, t+k) + 2\boldsymbol{\Gamma}(t) \mathbf{b}_k = 0$$

que proporciona el predictor definido por:

$$\mathbf{b}_k = \boldsymbol{\Gamma}(t)^{-1} \boldsymbol{\gamma}(t, t+k).$$

Este resultado es previsible ya que si consideramos (8.25) como una ecuación de regresión, la ecuación de \mathbf{b}_k es la expresión usual de estimación de los parámetros en regresión múltiple. Para procesos normales la esperanza condicionada es siempre función lineal de las observaciones y, por tanto, el mejor predictor lineal coincide con el mejor predictor global.