Modelos Autorregresivos

Juan Isaula

Modelos Autorregresivos

Juan Isaula

Contents

1. Modelo Autorregresivo (AR)

1.1 Operador de Rezago

Un operador de serie de tiempo es un "proceso" que transforma una o más series de tiempo en nuevas series de tiempo.

Ejemplos:

- **Multiplicación Esclar:** $r_t = \beta x_t$
- **Suma:** $r_t = x_t + w_t$
- **Identidad:** $r_t = 1 r_t$

Nótese que:

$$r_t = \beta(x_t + w_t) = \beta x_t + \beta w_t$$

El operador de rezago se denota por L y se define como:

$$L x_t = x_{t-1} \tag{1.1}$$

En general

$$L^k x_t = x_{t-k} \tag{1.2}$$

Algunas transformaciones de series pueden expresarse con el operador de rezagos, por ejemplo:

- **Primera diferencia:** $\Delta r_t = r_t - r_{t-1} = r_t - L r_t = (1 - L) r_t$
- **Suavisado por media móvil:**

$$r_t = \frac{1}{4}(r_r + r_{t-1} + r_{t-2} + r_{t-3}) = \frac{1}{4}(1 + L + L^2 + L^3) r_t$$

- **Tasa de crecimiento:** $\Delta\% r_t \approx 100(\ln r_t - \ln r_{t-1}) = 100(1 - L)\ln r_t$

Propiedades del Operador de Rezagos

- $L(\beta x_t) = \beta L x_t$
- $L(x_t + w_t) = Lx_t + Lw_t$
- $L(c) = c$
- $L^{-h}x_t = x_{t+h}$
- $L^0 x_t = x_t$
- $(\alpha L^h + \beta L^k)x_t = \alpha x_{t-h} + \beta x_{t-k}$

donde α, β, c son constantes.

Demostraciones

- Por definición de rezago, sabemos que el rezago de $L(\beta x_t) = \beta x_{t-1}$, por tanto:

$$\begin{aligned} L(\beta x_t) &= \beta x_{t-1} \\ &= \beta L x_t \quad \text{Por (1.1)} \end{aligned}$$

- Para probar la segunda propiedad, hacemos:

$$L(x_t + w_t) = x_{t-1} + w_{t-1} = Lx_t + Lw_t \quad \text{Por (1.1)}$$

- El rezago de una constante es el mismo, es decir no cambia.
-

$$\begin{aligned} L^{-h}x_t &= x_{t-(-h)} \quad \text{Por (1.2)} \\ &= x_{t+h} \end{aligned}$$

- El operador L^0 es exactamente lo mismo que la identidad. Por tanto $L^0 x_t = x_t$.

Lo interesante del operador de rezago es que mantiene las leyes del álgebra, como veremos a continuación:

Polinomio de Rezagos

El operador de rezagos sigue las reglas usuales de operaciones algebraicas. Por ejemplo:

$$\begin{aligned} (a + bL)(c + dL)x_t &= (a + bL)(cx_t + dx_{t-1}) \\ &= acx_t + adx_{t-1} + bcx_{t-1} + bdx_{t-2} \\ &= [ac + (ad + bc)L + bdL^2]x_t \end{aligned}$$

Así, definimos un polinomio de rezagos de orden p:

$$(1+\phi_1 L+\phi_2 L^2+...+\phi_p L^p)x_t = x_t+\phi_1 x_{t-1}+\phi_2 x_{t-2}+...+\phi_p x_{t-p}$$

donde $\phi_1, \phi_2, ..., \phi_p$ son escalares.

Inverso de un polinomio de rezagos de grado 1

Considere la operación:

$$\begin{aligned}(1-\phi L)(1+\phi L+...+\phi^k L^k)x_t &= (1+\phi L+...+\phi^k L^k-\phi L-\phi^2L^2-...-\phi^{k+1}L^k)x_t\\ &= (1-\phi^{k+1}L^{k+1})x_t\\ &= x_t-\phi^{k+1}x_{t-k-1} \quad \text{Por (1.2)}\end{aligned}$$

Note que si $|\phi|<1$, entonces es claro que:

$$\lim_{k\to\infty}\phi^{k+1}x_{t-k-1}=0$$

Por lo que

$$(1-\phi L)(1+\phi L+\phi^2L^2+...+)x_t = x_t$$

Con base en los hechos previos, escribimos:

$$(1-\phi L)^{-1} = 1+\phi L+\phi^2 L^2+... \tag{1.3}$$

$$\tag{1.4}$$

Inverso de un polinomio de rezagos de grado p

Consideremos el polinomio:

$$\Phi(L)=1-\phi_1 L-...-\phi_p L^p$$

Si factorizamos el polinomio como:

$$\Phi(L)=(1-\lambda_1 L)(1-\lambda_2 L)\cdots(1-\lambda_p L)$$

Encontramos su inverso como:

$$\begin{aligned}\Phi^{-1}(L) &= (1-\lambda_1 L)^{-1}\cdots(1-\lambda_p L)^{-1}\\ &= (1+\lambda_1 L+\lambda_1^2L^2+..)\cdots(1+\lambda_p L+\lambda_p^2L^2+...)\end{aligned}$$

dicha operación es factible siempre y cuando $|\lambda_i|<1$, pero se tiene que tener en cuenta que nadie me garantiza que los λ no siempre serán números reales, dado que pueden ser complejos, por tanto en el caso complejo el valor absoluto se tiene que ver como el modulo.

1.2 Estacionariedad

Si la media μ_t ni las covarianzas γ_{jt} dependen de la fecha t, entonces decimos que el proceso es r_t es débilmente estacionario y por tanto:

$$\begin{aligned} \mathbb{E}(r_t) &= \mu \qquad \forall\, t \\ \mathbb{E}(r_t-\mu)(r_{t-j}-\mu) &= \gamma_j \quad \forall\, t \text{ y cualquier } \ j \end{aligned}$$

Supongamos que $r_t \sim N(\mu_t, \sigma_t^2)$, decimos que es estacionario si μ_t y σ_t^2 son constantes.

No estacionario si μ_t está cambiando con el tiempo.

No estacionario si σ_t^2 está cambiando con el tiempo.

1.3 Ruido Blanco

Sea $\{a_t\}$ una sucesión cuyos elementos tiene media cero y varianza σ^2,

$$\begin{aligned} \mathbb{E}(a_t) &= 0 \qquad\quad \text{(media cero)} \\ \mathbb{E}(a_t^2) &= \sigma^2 \qquad\quad \text{(Varianza constante)} \\ \mathbb{E}(a_t a_\tau) &= 0 \ \ para\ t \neq \tau \qquad \text{Términos no correlacionados} \end{aligned}$$

Si los términos están normalmente distribuidos,

$$a_t \sim N(0, \sigma^2)$$

entonces tenemos que a_t es **ruido blanco gaussiano**.

Cuando se estiman modelos de series de tiempo, es importante evaluar si los residuos de la estimación contienen ruido blanco.

Una forma natural de evaluar si los residuos son ruido blanco es determinar si las autocorrelaciones

$$\rho_1 = \rho_2 = \cdots = \rho_p = 0$$

para todo $p \geq 1$.

1.4 Test de Box - Pierce (1970)

¿Es esta serie un caso de ruido blanco?

$$\begin{aligned} H_0 &: \rho_1 = \rho_2 = \cdots = \rho_m = 0 \quad \text{(Si es ruido blanco)} \\ Test &: Q^* = T\sum_{j=1}^{m} \hat{\rho}_j^2 \end{aligned}$$

Si $Q^* > \chi_{m-k}(1-\alpha)$, rechazar H_0 con $100\alpha\%$ de significancia: la serie no es ruido blanco.

La mayoria de los paquetes de software proporcionarán el valor p de $Q^*(m)$. Entonces la regla de decisión es rechazar H_0 si $p < \alpha$.

1.5 Modelo Autorregresivo de primer orden: AR(1)

Sea $\{a_t\}$ una serie de ruido blanco. Entonces, definimos el modelo $AR(1)$ como:

$$r_t = \phi_0 + \phi_1 r_{t-1} + a_t \tag{1.5}$$

Note que de (1.5) podemos hacer:

$$\begin{aligned} r_t - \phi_1 r_{t-1} &= \phi_0 + a_t \\ r_t - \phi_1 L r_t &= \phi_0 + a_t \quad \text{Por (1.1)} \\ (1-\phi_1 L) r_t &= \phi_0 + a_t \end{aligned}$$

Por (1.3) obtenemos que siempre y cuando $|\phi_1| < 1$ podemos invertir el término $(1-\phi_1 L)$, entonces:

$$\begin{aligned} r_t &= (1-\phi_1 L)^{-1}(\phi_0 + a_t) \\ &= (1-\phi_1 L)^{-1}\phi_0 + (1-\phi_1 L)^{-1} a_t \\ &= \frac{\phi_0}{1-\phi_1} + (1 + \phi_1 L + \phi_1^2 L^2 + \ldots) a_t \quad \text{Por (1.3)} \\ &= \frac{\phi_0}{1-\phi_1} + a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \ldots \end{aligned}$$

Podemos ver entonces que si $|\phi_1| < 1$ entonces el modelo $AR(1)$ se puede escribir como:

$$r_t = \frac{\phi_0}{1-\phi_1} + a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + \ldots \tag{1.6}$$

Por tanto, si queremos obtener su valor esperado, tenemos que:

$$\begin{aligned}
\mathbb{E}(r_t) &= \frac{\phi_0}{1-\phi_1} + \cancelto{0}{\mathbb{E}(a_t)} + \phi_1 \cancelto{0}{\mathbb{E}(a_{t-1})} + \phi_1^2 \cancelto{0}{\mathbb{E}(a_{t-2})} + .. \quad \text{Dado que } \{a_t\} \text{ es ruido blanco} \\
\mathbb{E}(r_t) &= \frac{\phi_0}{1-\phi_1} = \mu
\end{aligned}$$

Y su varianza es:

$$\begin{aligned}
Var(r_t) = E[(r_t - \mathbb{E}(r_t)^2] &= E[(r_t-\mu)^2] \\
&= E[(a_t + \phi_1 a_{t-1} + \phi_1^2 a_{t-2} + ..)^2] \quad \text{Por (1.6)} \\
&= \sigma_a^2 + \phi_1^2\sigma_a^2 + \phi_1^4\sigma_a^2 + \ldots \\
&= \underbrace{(1+\phi_1^2+\phi_4+\ldots)}_{serie\ geometrica}\sigma_a^2 \\
&= \frac{\sigma_a^2}{1-\phi_1^2} \quad \text{Por definición de serie geométrica} \\
&= \gamma_0
\end{aligned}$$

Luego, sabiendo que un modelo $AR(1)$ es estacionario si $|\phi_1| < 1$, entonces:

$$r_t = \phi_0 + \phi_0 r_{t-1} + a_t \tag{1.7}$$
$$E(r_t) = \phi_0 + \phi_1 \mathbb{E}(r_{t-1}) + \mathbb{E}(a_t) \tag{1.8}$$
$$\mu = \phi_0 + \phi_1\mu + 0 \tag{1.9}$$
$$\mu = \frac{\phi_0}{1-\phi_1} \tag{1.10}$$

Note que si restamos (1.7) de (1.9), tenemos:

$$\underbrace{r_t - \mu}_{\tilde{r}_t} = \phi_1 \underbrace{(r_{t-1}-\mu)}_{\tilde{r}_{t-1}} + a_t \tag{1.11}$$

Si continuamos repitiendo esta relación hacia atras k veces, obtenemos:

$$\begin{aligned}
r_t - \mu &= \phi_1(r_{t-1}-\mu) + a_t \\
&= \phi_1(\phi_1(r_{t-2}-\mu) + a_{t-1}) + a_t \\
&\vdots \\
&= \phi^k r_{t-k} + \sum_{i=0}^{k-1} \phi_1^i a_{t-i}
\end{aligned}$$

si continuamos con el proceso hacia el pasado (atras), siempre que $|\phi_1| < 1$ y r_t sea estacionario, entonces:

$$r_t - \mu = \sum_{i=0}^{\infty} \phi_1^i a_{t-i}$$

Para obtener su varianza, elevamos al cuadrado la expresión anterior y tomamos su esperanza, es decir:

$$\begin{aligned}
\mathbb{E}(\widetilde{r}_t^2) &= \mathbb{E}[(\phi_1 \widetilde{r}_{t-1} + a_t)^2] \qquad \text{Por (1.11)} \\
\gamma_0 &= \mathbb{E}[\phi_1^2 \widetilde{r}_{t-1}^2 + 2\phi_1 \widetilde{r}_{t-1} a_t + a_t^2] \\
&= \phi_1^2 \gamma_0 + 2 \times 0 + \sigma_a^2 \\
\gamma_0 &= \frac{\sigma_a^2}{1 - \phi_1^2}
\end{aligned}$$

Luego, para obtener su autocovarianza γ_i, para $i \geq 1$, escribimos

$$\widetilde{r}_t = \phi_1 \widetilde{r}_{t-1} + a_t$$

Multiplicamos ambos lados por $\widetilde{r}_{t-i}$ y tomamos la esperanza,

$$\begin{aligned}
\mathbb{E}[\widetilde{r}_t \widetilde{r}_{t-i}] &= \mathbb{E}[\phi_1 \widetilde{r}_{t-1} \widetilde{r}_{t-i} + \widetilde{r}_{t-i} a_t] \\
\gamma_i &= \phi_1 \gamma_{i-1}
\end{aligned}$$

Note que la función de autocorrelación es la solución de una ecuación en diferencia homogénea de primer orden, cuya solucion es

$$\gamma_j = \phi^j \gamma_0$$

<u>Resumen de los resultados obtenido</u>

$$\mathbb{E}(r_t) = \mu = \frac{\phi_0}{1 - \phi_1} \tag{1.12}$$

$$Var(r_t) = \gamma_0 = \frac{\sigma_a^2}{1 - \phi_1^2} \tag{1.13}$$

$$Cov(r_t, r_{t-j}) = \gamma_j = \phi_j \gamma_0 \tag{1.14}$$

1.6 Modelo Autorregresivo de orden p: AR(p)

Es fácil extender el modelo $AR(1)$ para incluir más rezagos. El proceso $AR(p)$ es:

$$\begin{aligned} r_t &= \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_p r_{t-p} + a_t \\ r_t - \phi_1 r_{t-1} - \ldots - \phi_p r_{t-p} &= \phi_0 + a_t \\ (1 - \phi_1 L - \ldots - \phi_p L^p) r_t &= \phi_0 + a_t \ \text{ utilizamos polinomios de regazgos visto en seccion 1.1} \\ \Phi(L) r_t &= \phi_0 + a_t \quad \text{con } \Phi(L) = (1 - \phi_1 L - \ldots - \phi_p L^p) \end{aligned}$$

En este punto es importante, observar dos cosas:

- El proceso $AR(p)$ es una ecuación en diferencia de orden p.

- Esta ecuación es estable si y solo si las raíces del polinomio $\underbrace{1 - \phi_1 z^1 - \ldots - \phi_p z^p}_{\Phi(L)}$ están todas **fuera** del círculo unitario.

Si el proceso es estable, yo lo puedo factorizar y nos ayuda a realizar la siguiente operación:

$$\begin{aligned} r_t &= \Phi^{-1}(L)(\phi_0 + a_t) \\ &= \Phi^{-1}(1)\phi_0 + \Phi^{-1}(L) a_t \\ &= \frac{\phi_0}{1 - \phi_1 - \ldots - \phi_p} + \Phi^{-1}(L) a_t \end{aligned}$$

Note que si obtenemos el valor esperado del resultado previo, tendremos que:

$$\mathbb{E}(r_t) = \frac{\phi_0}{1 - \phi_1 - \ldots - \phi_p} = \mu$$

Ahora, similar a lo que obtuvimos para el caso $AR(1)$, podemos escribir el proceso $AR(p)$ en términos de desviación de la media $\widetilde{r} = r - \mu$

$$\widetilde{r}_t = \phi_1 \widetilde{r}_{t-1} + \ldots + \phi_p \widetilde{y}_{t-p} + a_t$$

Luego, para obtener su varianza y covarianzas multiplicamos la expresión anterior por $\widetilde{r}_{t-j}$, con $j \geq 0$, y calculamos el valor esperado

$$\begin{aligned} \mathbb{E}(\widetilde{r}_t \widetilde{r}_{t-j}) &= \mathbb{E}[\phi_1 \widetilde{r}_{t-1} \widetilde{r}_{t-j} + \ldots + \phi_p \widetilde{r}_{t-p} \widetilde{r}_{t-j} + a_t \widetilde{r}_{t-j}] \\ \underbrace{\gamma_j}_{por def.} &= \begin{cases} \phi_1 \gamma_1 + \ldots + \phi_p \gamma_p + \sigma^2 & si\ j = 0 \\ \phi_1 \gamma_{j-1} + \ldots + \phi_p \gamma_{j-p} & si\ j > 0 \end{cases} \end{aligned}$$

Vea que podemos calcular las autocorrlaciones dividiendo lo anterior por γ_0, tendríamos algo así:

$$\begin{aligned}\frac{\gamma_j}{\gamma_0} &= \phi_1\frac{\gamma_{j-1}}{\gamma_0} + \phi_2\frac{\gamma_{j-2}}{\gamma_0} + \ldots + \phi_p\frac{\gamma_{j-p}}{\gamma_0} \\ \rho_j &= \phi_1\rho_{j-1} + \phi_2\rho_{j-2} + \ldots + \phi_p\rho_{j-p}\end{aligned}$$

La ecuación previa recibe el siguiente nombre:

Ecuaciones Yule - Walker

$$\rho_j = \phi_1\rho_{j-1} + \ldots + \phi_p\rho_{j-p},\ con\ j > 0$$

- Esta es la misma ecuación en diferencia que el modelo original, por lo que en principio se puede resolver con los métodos convencionales, una vez que se tengan p valores iniciales $\rho_0, \rho_1, \ldots, \rho_{p-1}$.
- En general, no es tan sencillo obtener despejar los valores de las autocorrelaciones ρ_j.

Para obtener la función de impulso respuesta podemos, podríamos partir de la forma

$$r_t = \mu + \Phi^{-1}(L)a$$

pero esto requiere que obtengamos la forma explícita de un polinomio de rezagos

$$\Phi^{-1}(L) = 1 + \psi_1 L + \psi_2 L^2 + \ldots +$$

Luego, como un modelo $AR(p)$ es una ecuación en diferencia de orden p, encontramos que

$$\frac{\partial r_{t+j}}{\partial a_t} = c_1\lambda_1^j + c_2\lambda_2^j + \ldots + c_p\lambda_p^j$$

donde λ_j son las raíces del polinomio característico $\lambda^p - \phi_1\lambda^{p-1} - \ldots - \phi_p$

Ahora, como para un modelo $AR(p)$ estacionario, sabemos que todas las raíces λ_j están dentro del circulo unitario.

Por ello, a partir de

$$\frac{\partial r_{t+j}}{\partial a_t} = c_1\lambda_1^j + c_2\lambda_2^j + \ldots + c_p\lambda_p^j$$

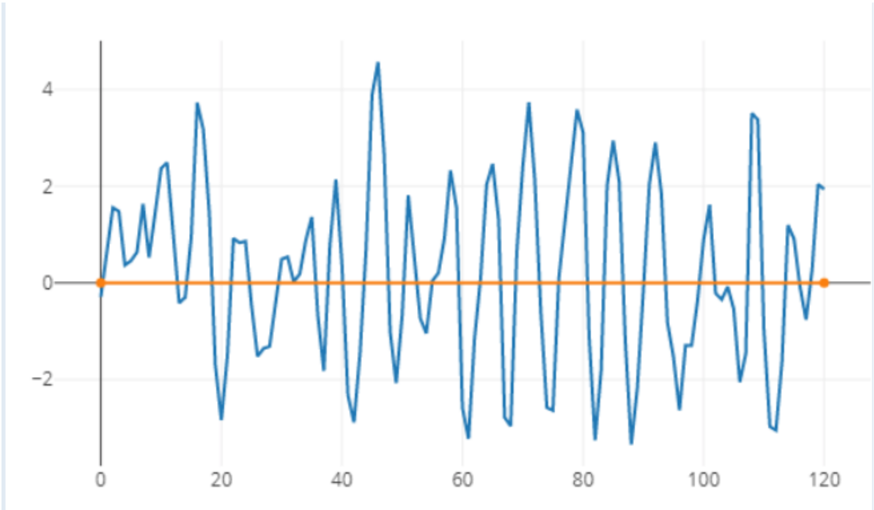
sabemos que la función de impulso respuesta converge a cero en el largo plazo.

La forma de la función depende especialmente del valor de λ_j más grande:

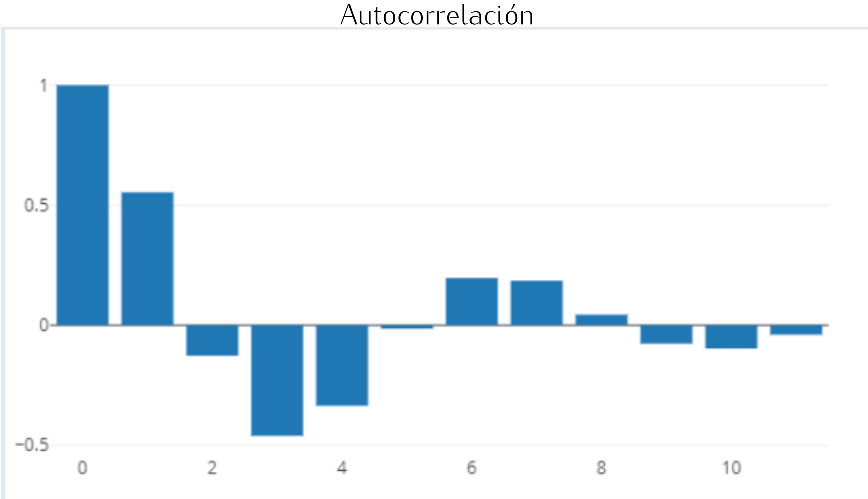
- Si es positivo, decae geométricamente.
- Si es negativo, decae alternando de signo.
- si es complejo, decae oscilando periodicamente

Ejemplo 1 AR(2)

$$r_t = 0.9r_{t-1} - 0.625r_{t-2} + a_t$$



Autocorrelación



Note que con las ecuaciones de Yule - Walker

$$\phi_1 + \phi_2\rho_1 = \rho_1$$
$$\phi_1\rho_1 + \phi_2 = \rho_2$$

encontramos que:

$$\rho_1 = \frac{\phi_1}{1-\phi_2} \approx \boxed{0.5538}$$
$$\rho_2 = \phi_1\rho_1 + \phi_2 \approx \boxed{-0.1265}$$

Las demás autocorrelaciones las encontramos con:

$$\rho_j = 0.9\rho_{j-1} - 0.625\rho_{j-2}$$

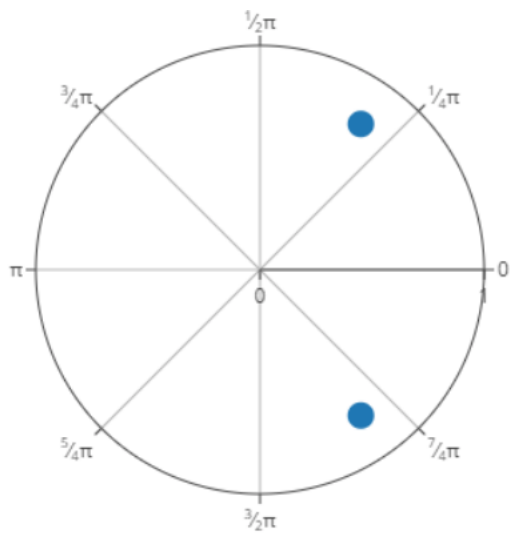
La ecuación característica es:

$$\lambda^2 - 0.9\lambda + 0.625 = 0$$

con raíces

$$\lambda = 0.45 \pm 0.65i$$

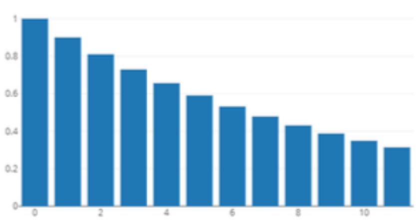
Por lo que el proceso es estacionario.

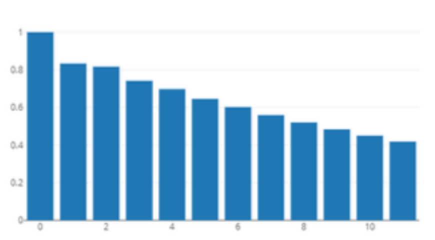


Ejemplo 2: AR(1) vs AR(2)

Considere estos dos procesos autorregresivos:

$$r_t = 0.9r_{t-1} + a_t \qquad\qquad r_t = 0.5r_{t-1} + 0.4r_{t-2} + a_t$$





- Este caso ilustra que es sumamente díficil identificar el orden p de un proceso $AR(p)$ a partir de su autocorrelograma.
- Para resolver este problema, a continuación estudiaremos la **autorrelación parcial**.

1.7 Autocorrelación Parcial

La autocorrelación parcial mide la correlación **restante** entre r_t y r_{t-k} una vez que se han eliminado la influencia de $r_{t-1}, r_{t-2}, ..., r_{t-k+1}$.

$$r_t = \underbrace{a_1^{(k)} r_{t-1} + a_2^{(k)} r - t - 2 + ... + a_{k-1}^{(k)} r_{t-k+1}}_{eliminamos\ este\ efecto} + a_k^{(k)} r_{t-k}$$

Es decir, las primeras m autocorrelaciones parciales vienen de

$$\begin{aligned}
r_t &= a_1^{(1)} r_{t-1}; \quad a_1^{(1)} \text{ es mi autocorrelación parcial de un rezago} \\
r_t &= a_1^{(2)} r_{t-1} + a_2^{(2)} r_{t-2} \quad a_2^{(2)} \text{ segunda autocorrelación parcial} \\
&\vdots \\
r_t &= a_1^{(m-1)} r_{t-1} + a_2^{(m-1)} r_{t-2} + ... + a_{m-1}^{(m-1)} r_{t-m+1}
\end{aligned}$$

Para encontrar el valor de $a_k^{(k)}$ basta con resolver

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \rho_2 & \rho_1 & 1 & \cdots & \rho_{k-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_1^{(k)} \\ a_2^{(k)} \\ a_3^{(k)} \\ \vdots \\ a_k^{(k)} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \rho_3 \\ \vdots \\ \rho_k \end{bmatrix}$$

A partir de este punto, denotaremos la k - ésima correlación parcial por

$$\phi(k) = a_k^{(k)}$$

comparando las ecuaciones del modelo $AR(p)$ y de la autocorrelación parcial k:

$$\begin{aligned} r_t &= \phi_1 r_{t-1} + \phi_2 r_{t-2} + \dots + \phi_p r_{t-p} + a_t \\ r_t &= a_1^{(k)} r_{t-1} + a_2^{(k)} r_{t-2} + \dots + a_{k-1}^{(k)} r_{t-k+1} + a_k^{(k)} r_{t-k} \end{aligned}$$

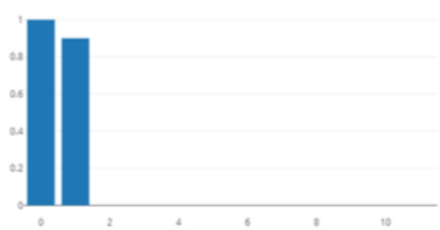
vemos que

- si $k = p$, entonces $\phi_k = \theta_p$
- si $k > p$, entonces $\phi_k = 0$
- si $k = 1$, entonces $\phi_1 = \rho_1$

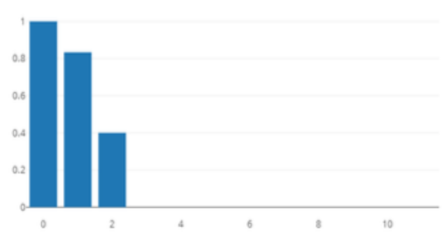
Ejemplo 2: AR(1) vs AR(2) con autocorrelación parcial

Consideremos de nuevos estos modelos autorregresivos:

$r_t = 0.9 r_{t-1} + a_t$



$r_t = 0.5 r_{t-1} + 0.4 r_{t-2} + a_t$



ahora si es sencillo identificar el orden p de los modelos AR.

Para encontrar la segunda autocorrelación parcial, resolvemos:

$$\begin{bmatrix} a_1 \\ \phi_2 \end{bmatrix} = \begin{bmatrix} 1 & \rho_1 \\ \rho_1 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix} = \frac{1}{1-\rho_1^2} \begin{bmatrix} 1 & -\rho_1 \\ -\rho_1 & 1 \end{bmatrix} \begin{bmatrix} \rho_1 \\ \rho_2 \end{bmatrix}$$

donde $\phi_2 = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$.

- Para el proceso $AR(1)$ sabemos que $\rho_k = \phi^k$, con lo que comprobamos que $\phi_2 = \frac{\theta^2 - \theta^2}{1-\theta^2} = 0$

- En un ejemplo anterior encontramos que para un modelo $AR(2)$ se cumple:

$$\rho_1 = \frac{\theta_1}{1-\theta_2} = \frac{5}{6} \qquad \rho_2 = \theta_1\rho_1 + \theta_2 = \frac{49}{60}$$

Al sustituir en la expresión para ϕ_2 comprobamos que

$$\phi_2 = \frac{\frac{49}{60} - \frac{25}{36}}{1 - \frac{25}{36}} = 0.4 = \theta_2$$

1.8 Función de verosimilitud

Consideremos la serie de rendimientos $\{r_{it}\}_{i=1}^{T}$ para el activo i con distribución conjunta:

$$F(r_{i1}, ..., r_{iT}; \theta) = F(r_{i1})F(r_{i2|r_{1t}}) \cdots F(r_{it}|r_i, T-1, ..., r_{i1}) = F(r_{i1}) \prod_{t=2}^{T} F(r_{it}|r_i, t-1, ..., r_{i1}) \quad (1.15)$$

donde por simplicidad, se omite el parámetro θ.

Es habitual tratar los rendimientos de los activos como variables aleatorias continuas, especialmente para los rendimientos de índices o rendimientos de acciones calculados con una frecuencia baja, y utilizar sus funciones de densidad de probabilidad. Podemos escribir la partición en la ecuación. (1.15) como

$$f(r_{i1}, ..., r_{iT}; \theta) = f(r_{i1}; \theta) \prod_{t=2}^{T} f(r_{it}|r_i, t-1, ..., r_{i1}, \theta) \quad (1.16)$$

La partición de Eq. (1.15) se puede utilizar para obtener la función de verosimilitud de los rendimientos $(r_1, ..., r_T)$ de un activo. Donde para facilitar la notación, el subíndice i se omite:

$$f(r_1, ..., r_T; \theta) = f(r_1; \theta) \prod_{t=2}^{T} \frac{1}{\sqrt{2\pi}\sigma_t} e^{\left(\frac{-(r_t-\mu_t)^2}{2\sigma_t^2)} \right)} \quad (1.17)$$

donde $f(r_1; \theta)$ es la función de densidad marginal de la primera observación r_1. El valor de θ que maximiza esta función de verosimilitud es la estimación de máxima verosimilitud (MLE) de θ. Dado que la función logarítmica es monótona, el MLE se puede obtener maximizando la función logarítmica de verosimilitud,

$$\ln f(r_1, ..., r_T; \theta) = \ln f(r_1; \theta) - \frac{1}{2} \sum_{t=2}^{T} \left(\ln(2\pi) + \ln(\sigma_t^2) + \frac{(r_t-\mu_t)^2}{\sigma_t^2} \right)$$

que es más fácil de manejar en la práctica. La función de probabilidad logarítmica de los datos se puede obtener de manera similar si la distribución condicional $f(r_t|r_{t-1},...,r_1;\theta)$ no es normal.

1.9 Criterios de Información

Hay varios criterios de información disponibles para determinar el orden p de un proceso de AR. Todos ellos se basan en la verosimilitud. Por ejemplo, el conocido **criterio de información de Akaike (AIC) (Akaike, 1973)** se define como

$$AIC = -\frac{2}{T}\ln(verosimilitud) + \frac{2}{T} \times (número\ de\ parámetros) \tag{1.18}$$

donde la función de verosimilitud se evalúa en las estimaciones de máxima verosimilitud y T es el tamaño de la muestra. Para un modelo $AR(\ell)$ Gaussiano, AIC se reduce a

$$AIC(\ell) = \ln(\widetilde{\sigma}_\ell^2) + \frac{2\ell}{T}$$

donde $\widetilde{\sigma}_\ell^2$ es la estimación de maxima verosimilitud de σ_a^2 que es la varianza de a_t, y T es el tamaño de la muestra.

El primer término del AIC en (1.18) mide la bondad de ajuste del modelo $AR(\ell)$ a los datos, mientras que el segundo término se denomina función de penalización del criterio porque penaliza a un modelo candidato por el número de parámetros utilizados. Diferentes funciones de penalización dan como resultado diferentes criterios de información.

Otra función de criterio comúnmente utilizada es el **criterio de información Bayesiano (Schwarz) (BIC)**. Para un modelo Gaussiano $AR(\ell)$, el criterio es

$$BIC = \ln(\widetilde{\sigma}_\ell^2) + \frac{\ell\ln(T)}{T}$$

La penalización por cada parámetro utilizado es 2 para AIC e $\ln(T)$ para BIC. Por lo tanto, BIC tiende a seleccionar un modelo de AR más bajo cuando el tamaño de la muestra es moderado o grande.

1.9.1 Regla de selección

Para usar AIC para seleccionar un modelo AR en la práctica, se calcula $AIC(\ell)$ para $\ell = 0, ..., P$, donde P es un número entero positivo preespecificado, y se selecciona el orden k que tiene el valor AIC mínimo.

1.10 Estimación de Parámetros

Para un modelo $AR(p)$, el método de mínimos cuadrados condicionales, que comienza con la observación $p+1$, se utiliza a menudo para estimar los parámetros. Específicamente, condicionando las primeras p observaciones, tenemos

$$r_t = \phi_0 + \phi_1 r_{t-1} + \ldots + \phi_p r_{t-p} + a_t, \quad t = p+1, \ldots., T$$

se parece a la forma de una regresión lineal múltiple y se puede estimar mediante el método de mínimos cuadrados. Denote la estimación de ϕ_i por $\hat{\phi}_i$. El modelo ajustado es:

$$\hat{r}_t = \hat{\phi}_0 + \hat{\phi}_1 r_{t-1} + \ldots + \hat{\phi}_p r_{t-p}$$

y el residual asociado es:

$$\hat{a}_t = r_t - \hat{r}_t$$

La serie $\{\hat{a}_t\}$ se llama serie residual, de la cual obtenemos:

$$\hat{\sigma}_a^2 = \frac{\sum_{t=p+1}^{T} \hat{a}_t^2}{T-2p-1}$$

Si se utiliza el método de verosimilitud condicional la estimación de ϕ_i permanece sin cambios, pero la estimación de σ_a^2 se convierte en $\widetilde{\sigma}_a^2 = \hat{\sigma}_a^2 \times (T-2p-1)/(T-p)$.

1.10.1 Modelo Adecuado

- Si el modelo es adecuado, entonces la serie residual debería comportarse como un ruido blanco.

- Se puede usar los ACF y Test de Ljung - Box para comprobar la proximidad de un $\hat{a}_t$ a un ruido blanco.

- Para un modelo $AR(p)$, el estadístico de Ljung - Box $Q(m)$ sigue una distribución χ^2_{m-g}, donde g es el número de coeficientes AR usados en el modelo.

1.11 Bonda de Ajuste

Un estadístico comúnmente utilizado para medir la bondad de ajuste de un modelo estacionario es el R^2 definido como

$$R^2 = 1 - \frac{\text{suma de cuadrado de los residuales}}{\text{Suma de cuadrado de total}}$$

Para un modelo $AR(p)$ estacionario con T observaciones $\{r_t|t = 1, ..., T\}$, la medida se convierte en

$$R^2 = 1 - \frac{\sum_{t=p+1}^{2} \hat{a}_t^2}{\sum_{t=p+1}^{2}(r_t - \bar{r})^2}$$

donde $\bar{r} = \frac{\sum_{t=p+1}^{T} r_t}{T-p}$. Un R^2 grande indica que el modelo proporciona un ajuste mas cercano a los datos. Sin embargo, esto solo es cierto para una serie de tiempo estacionaria. Para un conjunto de datos dados es bien sabido que R^2 es una función decreciente del número de parámetros utilizados. Para superar esa debilidad se propone un R^2 ajustado, que se define como

$$\begin{aligned} R^2_{aj} &= 1 - \frac{\text{Varianza de residuales}}{\text{Varianza de } r_t} \\ &= 1 - \frac{\hat{\sigma}_a^2}{\hat{\sigma}_r^2} \end{aligned}$$

donde $\hat{\sigma}_r^2$ es la varianza simple de r_t. Esta nueva medida tiene en cuenta el número de parámetros utilizafos en el modelo ajustado. Sin embargo, ya no está entre 0 y 1.

1.12 Forecasting

Para el modelo $AR(p)$ en la ecuación

$$r_t = \phi_0 + \phi_1 r_{t-1} + ... + \phi_p r_{t-p} + a_t$$

suponga que estamos en el indice h y estamos interesados en pronosticar $r_{h+\ell}$, donde $\ell \geq 1$.

- h: origen del pronóstico.
- ℓ: horizonte del pronóstico.

Sea $\hat{r}_h(\ell)$ el pronóstico de $r_{h+\ell}$ usando la función de perdida de error de minimos cuadrados y F_h la recopilación de información disponible en el origen del pronóstico h. Entonces $\hat{r}_h(\ell)$ se elige de manera que

$$E\{[r_{h+\ell} - \hat{r}_h(\ell)]^2|F_h\} \leq \min_g E[(r_{h+\ell} - g)^2|F_h]$$

donde g es una función de la información disponible en el tiempo h, es decir, es una función de F_h.

1. **Pronóstico de 1 paso adelante**

   De un modelo $AR(p)$ tenemos:

   $$r_{h+1} = \phi_0 + \phi_1 r_h + ... + \phi_p r_{h+1-p} + a_{h+1}$$

   Bajo la función de perdida de error cuadrado mínimo el pronóstico puntual de r_{h+1} dado $F_h = \{r_h, r_{h-1}, ...\}$ es la esperanza condicional

   $$\hat{r}_h(1) = E(r_{h+1}|F_h) = \phi_0 + \sum_{i=1}^{p} \phi_i r_{h+1-i}$$

   y el error pronóstico asociado es:

   $$e_h(1) = r_{h+1} - \hat{r}_h(1) = a_{h+1}$$

   con varianza

   $$V(e_h(1)) = Var(a_{h+1}) = \sigma_a^2$$

2. **Pronóstico de 2 pasos adelante**

   Considere el pronóstico r_{h+2} en el origen del pronóstico h. De el modelo $AR(p)$ tenemos

   $$r_{h+2} = \phi_0 + \phi_1 \hat{r}_h(1) + \phi_2 r_h + ... + \phi_p r_{h+2-p}$$

   tomando la esperanza condicional tenemos:

   $$\hat{r}_h(2) = E(r_{h+2}|F_h) = \phi_0 + \phi_1 \hat{r}(1) + \phi_2 r_h + ... + \phi_p r_{h+2-p}$$

   con error

   $$e_h(2) = r_{h+2} - \hat{r}_h(2) = \phi_1[r_{h+1} - \hat{r}_h(1)] + a_{h+2} = a_{h+2} + \phi_1 a_{h+1}$$

   Y varianza

   $$Var(e_h(2)) = (1 + \phi_1^2)\sigma_a^2$$

   Es interesante ver que $Var(e_h(2)) \geq Var(e_h(1)$ es decir, a medida que aumenta el horizonte de pronóstico, también aumenta la incertidumbre en el pronóstico.

3. **Pronóstico de multiples pasos adelante**

   En general tenemos

   $$r_{h+\ell} = \phi_0 + \phi_1 r_{h+\ell-1} + ... + \phi_p r_{h+\ell-p} + a_{h+\ell}$$

   El pronostico de ℓ pasos adelante basado en la función de perdida de error de minimos cuadrados es la esperanza condicional de $r_{h+\ell}$ dado F_h, es decir

   $$\hat{r}_h(\ell) = \phi_0 + \sum_{i=1}^{p} \phi_i \hat{r}_h(\ell - i)$$

   donde $\hat{r}_h(i) = r_{h+i}$, si $i \leq 0$. Dicho pronóstico se puede calcular de forma recursiva si usamos $\hat{r}_h(i)$, para $i = 1, ..., \ell - 1$. Con $e_h(\ell) = r_{h+\ell} - \hat{r}_h(\ell)$.

2. Modelo de media Móvil ARMA(q)

Sea $\{a_t\}$ una serie de ruido blanco. Se fine el modelo $MA(1)$ como:

$$r_t = c_0 + a_t - \theta_1 a_{t-1} \quad o \quad r_t = c_0 + (1 - \theta_1 L) a_t \tag{2.1}$$

c_0 es una constante. De manera similar, un $MA(2)$ es de la forma

$$r_t = c_0 a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} \tag{2.2}$$

y un modelo $MA(q)$ es

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \ldots - \theta_q a_{t-q} \quad o \quad r_t = c_0 + (1 - \theta_1 L - \ldots - \theta_q L^q) a_t, \quad q > 0$$

2.1 Propiedades de los modelos MA

Nos enfocamos en los modelos simples $MA(1)$ y $MA(2)$. Los resultados de los modelos $MA(q)$ pueden obtenerse fácilmente mediante las mismas técnicas.

2.1.1 Estacionariedad

Los modelos MA son siempre débilmente estacionarios porque son combinaciones lineales finitas de una secuencia de ruido blanco para la cual los dos primeros momentos son invariantes en el tiempo. Por ejemplo, considere el modelo $MA(1)$ en la ecuación. (2.1). Teniendo en cuenta las esperanzas del modelo, tenemos

$$\begin{aligned} \mathbb{E}(r_t) &= c_0 + \mathbb{E}(a_t) - \theta_1 \mathbb{E}(a_{t-1}) \\ &= c_0 + \cancelto{0}{\mathbb{E}(a_t)} - \cancelto{0}{\theta_1 \mathbb{E}(a_{t-1})}, \quad \text{ya que } a_t \text{ sn ruido blanco} \\ &= c_0 \end{aligned}$$

Ahora, si tomamos la varianza en (2.1), tenemos:

$$\begin{aligned}
Var(r_t) &= \mathbb{E}[(r_t - E(r_t))^2] \\
&= \mathbb{E}[(r_t - c_0)^2] \quad \text{Por resultado previo} \\
&= \mathbb{E}[(a_t - \theta_1 a_{t-1})^2] \\
&= \mathbb{E}[a_t^2 - 2\theta_1 a_t a_{t-1} + \theta_1^2 a_{t-1}^2] \\
&= \mathbb{E}[a_t^2] - 2\theta_1 \mathbb{E}(a_t a_{t-1}) + \theta_1^2 \mathbb{E}(a_{t-1}^2) \\
&= \sigma_a^2 + \theta_1^2 \sigma_a^2 \\
&= (1 + \theta_1^2)\sigma_a^2
\end{aligned}$$

Estó también se aplica a los modelos $MA(q)$ generales y obtenemos dos propiedades generales:

- El término constante de un modelo $MA(q)$ es la media de la serie (es decir, $E(r_t) = c_0$).
- La varianza de un modelo MA es

$$Var(r_t) = (1 + \theta_1^2 + \theta_2^2 + ... + \theta_q^2)\sigma_a^2$$

2.1.2 Función de Autocorrelación

Su autocovarianza j, para $j \geq 1$, es

$$\begin{aligned}
Cov(r_t, r_{t-1}) &= \mathbb{E}[(r_t - c_0)(r_{t-j} - c_0)] \\
&= \mathbb{E}[(a_t - \theta_1 a_{t-1})(a_{t-j} - \theta_1 a_{t-j-1}] \\
&= \mathbb{E}[a_t a_{t-j} - \theta_1 a_t a_{t-j-1} - \theta_1 a_{t-1} a_{t-j} + \theta_1^2 a_{t-1} a_{t-j-1}] \\
&= 0 + 0 - \theta_1 \mathbb{E}(a_{t-1} a_{t-j}) + 0 \\
&= -\theta_1 \mathbb{E}[a_{t-1} a_{t-j}] \\
&= \begin{cases} -\theta_1 \sigma_a^2 & si\ j = 1 \\ 0 & si\ j > 1 \end{cases}
\end{aligned}$$

Por tanto, su función de autocorrelación es

$$\rho_j = \begin{cases} -\frac{\theta_1}{1+\theta_1^2} & si\ j = 1 \\ 0 & si\ j > 1 \end{cases}$$

Por lo tanto, para un modelo $MA(1)$, ACF de rezago 1 no es cero, pero todos las ACF de orden superior son cero. En otras palabras, el ACF de un modelo $MA(1)$

se corta en el rezago 1. Para el modelo $MA(2)$ en la Ec. (2.2), los coeficientes de autocorrelación son

$$\begin{cases} -\frac{\theta_2}{1+\theta_1^2+\theta_2^2} & si\ j = 2 \\ 0 & si\ j > 2 \end{cases}$$

Aquí el ACF se corta en el retraso 2. Esta propiedad se generaliza a otros modelos MA. Para un modelo $MA(q)$, el rezago q no es cero. En consecuencia, una serie $MA(q)$ solo está relacionada linealmente con sus primeros valores q rezagados y, por lo tanto, es un modelo de **"memoria finita"**.

2.1.3 Invertibilidad de un modelo $MA(1)$

Supongamos que $c_0 = 0$, con lo que el modelo $r_t = a_t - \theta_1 a_{t-1}$ lo cual implica que

$$a_t = r_t + \theta_1 a_{t-1} \tag{2.3}$$

Vea que si repitó con $r_{t-1} = a_{t-1} - \theta_1 a_{t-2} \Longrightarrow a_{t-1} = r_{t-1} + \theta_1 r_{t-2}$, sustituyendo en (2.3), tenemos:

$$a_t = r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2}$$

Si repetimos este procedimiento, obtendremos:

$$a_t = r_t + \theta_1 r_{t-1} + \theta_1^2 r_{t-2} + \theta_1^3 r_{t-3} + \ldots$$

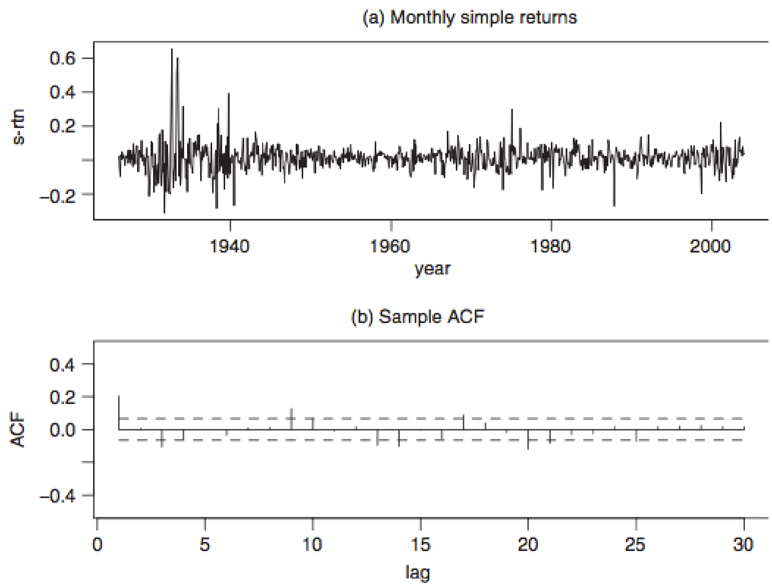
Esta ecuación expresa el impacto actual en como una combinación lineal de los rendimientos presentes y pasados. Intuitivamente, θ_1^j debería ir a cero a medida que j aumenta porque el retorno remoto r_{t-j} debería tener muy poco impacto en el choque actual, si lo hay. En consecuencia, para que un modelo $MA(1)$ sea plausible, se requiere que $|\theta_1| < 1$. Si $|\theta_1| = 1$, entonces el modelo $MA(1)$ no es reversible.

2.1.4 Identificación del orden de un modelo MA

las ACF son útiles para identificar el orden de un modelo MA. Para una serie de tiempo r_t con ACF ρ_j, si $\rho_q \neq 0$, pero $\rho_j = 0$ para $j > q$, entonces r_t sigue un modelo $MA(q)$.

La Figura 2.1 muestra la gráfica de tiempo de los retornos mensuales simples del índice CRSP ponderado equitativamente desde enero de 1926 hasta diciembre de 2003 y el ACF muestral de la serie. Las dos líneas punteadas que se muestran en el gráfico ACF denotan los dos límites de error estándar. Se ve que la serie tiene un ACF significativo en los rezagos 1, 3 y 9.

Figura 2.1



Hay algunos ACF marginalmente significativos en rezagos más altos, pero no los consideramos aquí. Basado en el ACF de muestra, el siguiente modelo *MA*(9) para la serie.

$$r_t = c_0 + a_t - \theta_1 a_{t-1} - \theta_3 a_{t-3} - \theta_9 a_{t-9}$$

Tenga en cuenta que, a diferencia del PACF de muestra, el ACF de muestra proporciona información sobre los retrasos *MA* distintos de cero del modelo.