Demostraciones de Econometría

Agustín Sanhueza & Christián Gonzalez

Mínimos Cuadrados Ordinarios

1. El Estimador. $\hat{\beta}$

Derivación

Para el modelo

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

El problema a minimizar para obtener $\hat{\beta}_0$ y $\hat{\beta}_1$ es

$$\boxed{\min_{\{\beta_0,\beta_1\}} \sum_{i=1}^{T} u_i^2 = \sum_{i=1}^{T}(y_i - \beta_0 - \beta_1 x_i)^2}$$

Resolviendo para $\hat{\beta}_0$

$$\frac{\partial \sum_{i=1}^{T} u_i^2}{\partial \beta_0}|_{\beta=\hat{\beta}} = -2\sum_{i=1}^{T}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{T}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{T} y_i - \sum_{i=1}^{T} \hat{\beta}_0 - \sum_{i=1}^{T} \hat{\beta}_1 x_i = 0$$

$$\sum_{i=1}^{T} y_i - T * \hat{\beta}_0 - \hat{\beta}_1 \sum_{i=1}^{T} x_i = 0$$

$$T * \hat{\beta}_0 = \sum_{i=1}^{T} y_i - \hat{\beta}_1 \sum_{i=1}^{T} x_i$$

$$\hat{\beta}_o = \sum_{i=1}^{T} \frac{y_i}{T} - \hat{\beta}_1 \sum_{i=1}^{T} \frac{x_i}{T}$$

$$\boxed{\hat{\beta}_0 = \overline{y} - \hat{\beta}_1 \overline{x}} \tag{1}$$

Resolviendo para $\hat{\beta}_1$

$$\frac{\partial \sum_{i=1}^{T} u_i^2}{\partial \beta_1}|_{\beta=\hat{\beta}} = -2\sum_{i=1}^{T}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)x_i = 0$$

$$\sum_{i=1}^{T} x_i(y_i - (\overline{y} - \hat{\beta}_1\overline{x}) - \hat{\beta}_1 x_i) = 0$$

$$\sum_{i=1}^{T} x_i[(y_i - \overline{y}) - \hat{\beta}_1(x_i - \overline{x})] = 0$$

$$\sum_{i=1}^{T} x_i(y_i - \overline{y}) - \hat{\beta}_1 \sum_{i=1}^{T} x_i(x_i - \overline{x})] = 0$$

$$\sum_{i=1}^{T}(y_i - \overline{y})(x_i - \overline{x}) = \hat{\beta}_1 \sum_{i=1}^{T}(x_i - \overline{x})^2$$

$$\boxed{\hat{\beta}_1 = \frac{\sum_{i=1}^{T}(y_i - \overline{y})(x_i - \overline{x})}{\sum_{i=1}^{T}(x_i - \overline{x})^2}} \tag{2}$$

Para un modelo con k variables y constante tendremos

$$y_i = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + ... + \beta_k x_{ki} + u_i$$

Matricialmente

$$Y_{T\times 1} = X_{T\times(k+1)}\beta_{(k+1)\times 1} + u_{T\times 1}$$

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}_{T\times 1} = \begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \cdots & x_{1,k} \\ 1 & x_{2,1} & x_{2,2} & \cdots & x_{2,k} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{T,1} & x_{T,2} & \cdots & x_{T,k} \end{bmatrix}_{T\times(k+1)} * \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}_{(k+1)\times 1} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_T \end{bmatrix}_{T\times 1}$$

El problema a minimizar para obtener el vector $\hat{\beta}$ es

$$\min_{\{\beta\}} \; u'u = (Y - X\beta)'(Y - X\beta)$$

$$u'u = Y'Y - \beta' X'Y - Y'X\beta + \beta X'X\beta$$

$$u'u = Y'Y - 2Y'X\beta + \beta X'X\beta$$

$$\frac{\partial u'u}{\partial \beta}|_{\beta=\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

$$X'X\hat{\beta} = X'Y$$

$$\boxed{\hat{\beta} = (X'X)^{-1}(X'Y)}$$

Si queremos evaluar que el óptimo encontrado es efectivamente un mínimo, debemos calcular las condiciones de segundo orden. Teníamos que

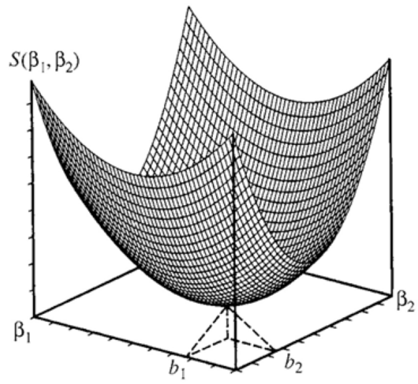
$$\frac{\partial u'u}{\partial \beta}|_{\beta=\hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0$$

Derivando nuevamente (respecto a β' dado que estamos en un contexto de matrices), llegamos a que:

$$\frac{\partial^2 u'u}{\partial\beta\partial\beta'}|_{\beta=\hat{\beta}} = 2X'X$$

Dado que la matriz X'X es una matriz semidefinida positva (valores propios mayores o iguales a 0), se demuestra que $\hat{\beta}$ es el vector que minimiza la suma de los residuos al cuadrados del problema.

Gráficamente se podría ver algo así (en un problema con dos parámetros a estimar):



2. Insesgamiento. $\mathbb{E}[\hat{\beta}]$

Derivación

Para el modelo

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Tendremos

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{T}(x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{T}(x_i - \overline{x})^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{T}(x_i - \overline{x})[(\beta_0 + \beta_1 x_i + u_i) - (\beta_0 + \beta_1\overline{x} + \overline{u})]}{\sum_{i=1}^{T}(x_i - \overline{x})^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{T}(x_i - \overline{x})(\beta_1(x_i - \overline{x}) + u_i)}{\sum_{i=1}^{T}(x_i - \overline{x})^2}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{T}\beta_1(x_i - \overline{x})^2 + (x_i - \overline{x})u_i}{\sum_{i=1}^{T}(x_i - \overline{x})^2}$$

$$\hat{\beta}_1 = \beta_1 + \frac{\sum_{i=1}^{T}(x_i - \overline{x})u_i}{\sum_{i=1}^{T}(x_i - \overline{x})^2}$$

$$\mathbb{E}[\hat{\beta}_1] = \beta_1 + \mathbb{E}\left[\frac{\sum_{i=1}^{T}(x_i - \overline{x})u_i}{\sum_{i=1}^{T}(x_i - \overline{x})^2}\right]$$

$$\mathbb{E}[\hat{\beta}_1] = \beta_1 + \frac{\sum_{i=1}^{T}(x_i - \overline{x})\mathbb{E}(u_i)}{\sum_{i=1}^{T}(x_i - \overline{x})^2}$$

$$\boxed{\mathbb{E}[\hat{\beta}_1] = \beta_1} \tag{3}$$

$$\hat{\beta}_0 = \overline{y} - \hat{\beta}_1\overline{x}$$

$$\hat{\beta}_0 = \beta_0 + \beta_1\overline{x} + \overline{u} - \hat{\beta}_1\overline{x}$$

$$\mathbb{E}[\hat{\beta}_0] = \beta_0 + \beta_1\overline{x} - \mathbb{E}[\hat{\beta}_1]\overline{x}$$

$$\mathbb{E}[\hat{\beta}_0] = \beta_0 + \beta_1\overline{x} - \beta_1\overline{x}$$

$$\boxed{\mathbb{E}[\hat{\beta}_0] = \beta_0} \tag{4}$$

Matricialmente para el modelo

$$Y = X\beta + u$$

Tendremos que

$$\hat{\beta} = (X'X)^{-1}X'Y$$
$$\hat{\beta} = (X'X)^{-1}X'(X\beta + u)$$
$$\hat{\beta} = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u$$
$$\hat{\beta} = \beta + (X'X)^{-1}X'u$$
$$\mathbb{E}[\hat{\beta}] = \beta + (X'X)^{-1}X'\mathbb{E}[u]$$
$$\boxed{\mathbb{E}[\hat{\beta}] = \beta}$$

3. La Eficiencia. $Var(\hat{\beta})$

Derivación

Para el modelo

$$y_i = \beta_0 + \beta_1 x_i + u_i$$

Tendremos

$$Var(\hat{\beta}_1) = \mathbb{E}[(\hat{\beta}_1 - \mathbb{E}(\hat{\beta}_1))^2]$$

$$Var(\hat{\beta}_1) = \mathbb{E}\left[\frac{\sum_{i=1}^{T}(x_i - \overline{x})u_i}{\sum_{i=1}^{T}(x_i - \overline{x})^2}\right]^2$$

$$Var(\hat{\beta}_1) = \mathbb{E}\left[\frac{\sum_{i=1}^{T}(x_i - \overline{x})^2 u_i^2}{\left(\sum_{i=1}^{T}(x_i - \overline{x})^2\right)^2}\right]$$

$$Var(\hat{\beta}_1) = \frac{\sum_{i=1}^{T}(x_i - \overline{x})^2 \mathbb{E}(u_i^2)}{\left(\sum_{i=1}^{T}(x_i - \overline{x})^2\right)^2}$$

$$\boxed{Var(\hat{\beta}_1) = \frac{\sigma^2}{\sum_{i=1}^{T}(x_i - \overline{x})^2}} \quad (5)$$

Matricialmente para el modelo

$$Y = X\beta + u$$

Tendremos que

$$V(\hat{\beta}) = \mathbb{E}\left[(\hat{\beta} - \mathbb{E}[\hat{\beta}])(\hat{\beta} - \mathbb{E}[\hat{\beta}])'\right]$$

$$V(\hat{\beta}) = \mathbb{E}\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right]$$

$$V(\hat{\beta}) = \mathbb{E}\left[((X'X)^{-1}X'u)((X'X)^{-1}X'u)'\right]$$

$$V(\hat{\beta}) = \mathbb{E}\left[(X'X)^{-1}X'uu'X(X'X)^{-1}\right]$$

$$V(\hat{\beta}) = (X'X)^{-1}X'\mathbb{E}[uu']X(X'X)^{-1}$$

$$V(\hat{\beta}) = \sigma^2(X'X)^{-1}X'X(X'X)^{-1}$$

$$\boxed{V(\hat{\beta}) = \sigma^2(X'X)^{-1}}$$

4. Logaritmos como una aproximación a cambios porcentuales.

Derivación

Podemos escribir la variación porcentual de una variable Y de la siguiente manera

$$\Delta \%Y = \frac{y_t - y_{t-1}}{y_{t-1}}$$

Paralelamente tenemos que

$$\Delta log(Y) = log(y_t) - log(y_{t-1})$$

$$\Delta log(Y) = log\left(\frac{y_t}{y_{t-1}}\right)$$

$$\Delta log(Y) = log\left(\frac{y_t}{y_{t-1}} + 1 - 1\right)$$

$$\Delta log(Y) = log\left(\frac{y_t}{y_{t-1}} + 1 - \frac{y_{t-1}}{y_{t-1}}\right)$$

$$\Delta log(Y) = log\left(1 + \frac{y_t - y_{t-1}}{y_{t-1}}\right)$$

Si tenemos que $\Delta \%Y = \frac{y_t - y_{t-1}}{y_{t-1}}$ es cercano a 0 entonces.

$$\boxed{\Delta log(Y) = log\left(1 + \frac{y_t - y_{t-1}}{y_{t-1}}\right) \approx \frac{y_t - y_{t-1}}{y_{t-1}}}$$

Por ejemplo, si hay un cambio porcentual de 3 %

$$ln(1 + 0{,}03) \approx 0{,}03$$

$$ln(1{,}03) \approx 0{,}03$$

$$0{,}0295588 \approx 0{,}03$$

Una aproximación bastante razonable

Por lo tanto, si queremos estimar el cambio porcentual de la variable Y ante cambios porcentuales de la variable X, podemos estimar lo siguiente

$$log(y_t) = \beta_0 + \beta_1 log(x_t) + u_t$$

¿Por qué?

$$\beta_1 = \frac{\Delta log(y_t)}{\Delta log(x_t)} \approx \frac{\frac{y_t - y_{t-1}}{y_{t-1}}}{\frac{x_t - x_{t-1}}{x_{t-1}}}$$

Nota: Típicamente se usa el logaritmo natural (logaritmo en base e) para estimar regresiones dado que este tiene propiedades que facilitan mucho la vida al hacer operaciones.

5. Sesgo por variable omitida

Derivación

Si el modelo real es:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i \tag{6}$$

Pero nosotros estimamos

$$y_i = \gamma_0 + \gamma_1 x_{1,i} + e_i \tag{7}$$

Entonces llegaremos a que:

$$\hat{\gamma_1} = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x})^2}$$

Reemplazamos y_i y el $\overline{y}$ del modelo real.

$$\hat{\gamma_1} = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})[(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + u_i) - (\beta_0 + \beta_1 \overline{x_1} + \beta_2 \overline{x_2} + \overline{u})]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma_1} = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})[\beta_1 x_{1,i} - \beta_1 \overline{x_1} + \beta_2 x_{2,i} - \beta_2 \overline{x_2} + u_i - \overline{u}]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma_1} = \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})[\beta_1(x_{1,i} - \overline{x_1}) + \beta_2(x_{2,i} - \overline{x_2}) + u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma_1} = \frac{\beta_1 \sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2 + \beta_2 \sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2}) + \sum_{i=1}^{T}(x_{1,i} - \overline{x_1})u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\hat{\gamma_1} = \beta_1 + \frac{\beta_2 \sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2} + \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

Aplicamos esperanza

$$\mathbb{E}[\hat{\gamma_1}] = \mathbb{E}[\beta_1] + \mathbb{E}\left[\frac{\beta_2 \sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}\right] + \mathbb{E}\left[\frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}\right]$$

$$\mathbb{E}[\hat{\gamma_1}] = \beta_1 + \frac{\beta_2 \sum_{i=1}^{T}(x_{1,i} - \overline{x_1})(x_{2,i} - \overline{x_2})}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2} + \frac{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})\mathbb{E}[u_i]}{\sum_{i=1}^{T}(x_{1,i} - \overline{x_1})^2}$$

$$\boxed{\mathbb{E}[\hat{\gamma_1}] = \beta_1 + \underbrace{\beta_2 \frac{Cov(X_1, X_2)}{Var(X_1)}}_{Sesgo}}$$

La dirección del sesgo depende del signo que tomen β_2 y $Cov(X_1, X_2)$. Se dice que el estimador $\hat{\gamma_1}$ está "sobrevalorado" si el sesgo es positivo, y está "subvalorado" si el sesgo es negativo. Esto se resume en la siguiente tabla:

	$Cov(X_1, X_2) > 0$	$Cov(X_1, X_2) < 0$
$\beta_2 > 0$	Sobrevalorado	Subvalorado
$\beta_2 < 0$	Subvalorado	Sobrevalorado

Modelo AR

1. Descomposición de Wald

Derivación

Permite expresar un proceso AR(P) estacionario como un MA(∞)

Sea $|\phi_1| < 1$

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$$

$$y_t = \phi_0 + \phi_1 L y_t + \varepsilon_t$$

$$(1 - \phi_1 L) y_t = \phi_0 + \varepsilon_t$$

$$y_t = \frac{\phi_0}{(1 - \phi_1 L)} + \frac{\varepsilon_t}{(1 - \phi_1 L)}$$

Usamos la propiedad de sumatoria $\sum_{i=0}^{\infty} a^i = \frac{1}{1-a}$ $\forall a < 1$. También usamos la propiedad del operador de rezagos $L^i c = c$

$$y_t = \frac{\phi_0}{(1 - \phi_1)} + \sum_{i=0}^{\infty} (\phi_1 L)^i \varepsilon_t$$

$$y_t = \frac{\phi_0}{(1 - \phi_1)} + \sum_{i=0}^{\infty} \phi_1^i L^i \varepsilon_t$$

$$\boxed{y_t = \frac{\phi_0}{(1 - \phi_1)} + \sum_{i=0}^{\infty} \phi_1^i \varepsilon_{t-i}}$$

$$y_t = \underbrace{\frac{\phi_0}{(1 - \phi_1)}}_{Constante} + \underbrace{(\varepsilon_t + \phi_1 \varepsilon_{t-1} + \phi_1^2 \varepsilon_{t-2} + ... + \phi_1^{\infty} \varepsilon_{t-\infty})}_{MA(\infty)}$$

2. Estacionariedad

Derivación

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$$

$$y_t = \phi_0 + \phi_1 L y_t + \phi_2 L^2 y_t + ... + \phi_p L^p y_t + \varepsilon_t$$

$$y_t \underbrace{(1 - \phi_1 L - \phi_2^2 L^2 - ... - \phi_p^p L^p)}_{Polinomio\ autoregresivo} = \phi_0 + \varepsilon_t$$

$$y_t \Phi(L) = \phi_0 + \varepsilon_t$$

Para que el proceso sea estacionario, es necesario que todas las raíces del polinomio autoregresivo $\Phi(L)$ deben estar fuera del circulo unitario (deben ser mayores a 1 en valor absoluto).

Alternativamente podemos reescribir el polinomio autoregresivo $\Phi(L)$ de la siguiente manera

$$\underbrace{\lambda^p - \phi_1\lambda^{p-1} - \phi_2\lambda^{p-2} - ... - \phi_{p-1}\lambda - \phi_p}_{\textit{Ecuación Característica}}$$

En paquetes estadísticos como Stata se trabaja más con esta expresión. Implícitamente está la relación $\lambda = \frac{1}{L}$. Por lo tanto, para que el proceso sea estacionario, es necesario que todas las raíces de la ecuación característica deben estar dentro del circulo unitario (deben ser menores a 1 en valor absoluto).

Del item anterior teníamos un proceso AR(1) donde

$$\underbrace{(1 - \phi_1 L)}_{\textit{Polinomio autoregresivo}}$$

Por lo tanto

$$\underbrace{(\lambda - \phi_1)}_{\textit{Ecuación Característica}}$$

Saquemos la raiz de la ecuación característica (solución de incógnita igualando a 0)

$$\lambda - \phi_1 = 0$$

$$\lambda = \phi_1$$

Para que exista estacionariedad en este proceso AR(1) es necesario que

$$|\lambda| = \boxed{|\phi_1| < 1}$$

Análogamente con el polinomio autoregresivo

$$|L| = \frac{1}{|\phi_1|} > 1 \Rightarrow \boxed{|\phi_1| < 1}$$

3. Esperanza

Derivación

Suponiendo estacionariedad, definimos $E[y_t] = \mu$ y que $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t$$

$$\mathbb{E}[y_t] = \phi_0 + \phi_1\mathbb{E}[y_{t-1}] + \phi_2\mathbb{E}[y_{t-2}] + ... + \phi_p\mathbb{E}[y_{t-p}] + \mathbb{E}[\varepsilon_t]$$
$$\mathbb{E}[y_t] = \phi_0 + \phi_1\mathbb{E}[y_t] + \phi_2\mathbb{E}[y_t] + ... + \phi_p\mathbb{E}[y_t] + 0$$

$$\mu = \phi_0 + \phi_1\mu + \phi_2\mu + ... + \phi_p\mu$$
$$\mu(1 - \phi_1 - \phi_2 - ... - \phi_p) = \phi_0$$
$$\mu(1 - \sum_{i=1}^{p}\phi_i) = \phi_0$$
$$\boxed{\mu = \frac{\phi_0}{(1 - \sum_{i=1}^{p}\phi_i)}}$$

4. Varianza

Derivación

Suponiendo estacionariedad, definimos $\gamma_0 = Var[y_t]$. Por simplicidad primero usaremos AR(1) y luego extrapolamos

$$y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t$$
$$Var[y_t] = Var[\phi_0 + \phi_1 y_{t-1} + \varepsilon_t]$$
$$Var[y_t] = Var[\phi_0] + \phi_1^2 Var[y_{t-1}] + Var[\varepsilon_t] + 2\phi_1 Cov(\phi_0, y_{t-1}) + 2\phi_1 Cov(y_{t-1}, \varepsilon_t) + 2Cov(\phi_0, \varepsilon_t)$$
$$\gamma_0 = \phi_1^2\gamma_0 + \sigma_\varepsilon^2$$
$$(1 - \phi_1^2)\gamma_0 = \sigma_\varepsilon^2$$
$$\boxed{\gamma_0 = \frac{\sigma_\varepsilon^2}{(1 - \phi_1^2)}}$$

Extrapolando para un AR(p) tendremos que

$$\boxed{\gamma_0 = \sigma_\varepsilon^2 + \sum_{i=1}^{p}\phi_i\gamma_i} \Rightarrow \ Ecuación\ de\ Yule - Walker$$

(Queda pendiente demostrar esto último)

5. Autocovarianza

Derivación

Suponiendo estacionariedad, definimos $\gamma_s = Cov[y_t, y_{t-s}]$

Primero veremos el caso para s=1 y luego extrapolamos.

$$Cov(y_t, y_{t-1}) = E[(y_t - E[y_t])(y_{t-1} - E[y_{t-1}])]$$
$$Cov(y_t, y_{t-1}) = E[(y_t - \mu)(y_{t-1} - \mu)]$$

Usamos la descomposición de Wald para expresar y_t e y_{t-1}

$$\gamma_1 = E\left[\left(\frac{\phi_0}{1-\phi_1} + \varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_1^2\varepsilon_{t-2} + ... - \frac{\phi_0}{1-\phi_1}\right)\left(\frac{\phi_0}{1-\phi_1} + \varepsilon_{t-1} + \phi_1\varepsilon_{t-2} + \phi_1^2\varepsilon_{t-3} + ... - \frac{\phi_0}{1-}\right.\right.$$

$$\gamma_1 = E[(\varepsilon_t + \phi_1\varepsilon_{t-1} + \phi_1^2\varepsilon_{t-2} + \phi_1^3\varepsilon_{t-3} + ...)(\varepsilon_{t-1} + \phi_1\varepsilon_{t-2} + \phi_1^2\varepsilon_{t-3} + \phi_1^3\varepsilon_{t-4} + ...)]$$

$$\gamma_1 = E[\varepsilon_t\varepsilon_{t-1} + \varepsilon_t\phi_1\varepsilon_{t-2} + ... + \phi_1\varepsilon_{t-1}^2 + \phi_1^2\varepsilon_{t-1}\varepsilon_{t-2} + ... + \phi_1^2\varepsilon_{t-2}\varepsilon_{t-1} + \phi_1^3\varepsilon_{t-2}^2 + ...]$$

$$\gamma_1 = E[\varepsilon_t\varepsilon_{t-1}] + \phi_1E[\varepsilon_t\varepsilon_{t-2}] + ... + \phi_1E[\varepsilon_{t-1}^2] + \phi_1^2E[\varepsilon_{t-1}\varepsilon_{t-2}] + ... + \phi_1^2E[\varepsilon_{t-2}\varepsilon_{t-1}] + \phi_1^3E[\varepsilon_{t-2}^2] + ...$$

$$\gamma_1 = 0 + \phi_1\cdot 0 + ... + \phi_1\cdot\sigma_\varepsilon^2 + \phi_1^2\cdot 0 + ... + \phi_1 0 + \phi_1^3\cdot\sigma_\varepsilon^2 + ...$$

$$\gamma_1 = \phi_1\cdot\sigma_\varepsilon^2 + \phi_1^3\cdot\sigma_\varepsilon^2 + \phi_1^5\cdot\sigma_\varepsilon^2 + ...$$

$$\gamma_1 = \phi_1(\sigma_\varepsilon^2 + \phi_1^2\sigma_\varepsilon^2 + \phi_1^4\sigma_\varepsilon^2 + ...)$$

$$\gamma_1 = \phi_1(\sigma_\varepsilon^2\cdot\sum_{i=0}^{\infty}\phi_1^{2i})$$

$$\gamma_1 = \phi_1(\sigma_\varepsilon^2\cdot\frac{1}{1-\phi_1^2})$$

$$\gamma_1 = \phi_1\cdot\frac{\sigma_\varepsilon^2}{1-\phi_1^2}$$

$$\boxed{\gamma_1 = \phi_1\cdot\gamma_0}$$

Generalizando el resultado anterior tendremos que la autocovarianza de orden S para un AR(1) es

$$\boxed{\gamma_s = \phi_1^s\cdot\gamma_0}$$

Extrapolando la autocovarianza de orden 1 para un AR(P) tendremos

$$\boxed{\gamma_j = \sum_{i=1}^{P}\phi_i\gamma_{j-i} \quad \forall j\geq 1} \Rightarrow \ Ecuación\ de\ Yule-Walker$$

(queda pendiente demostrar esto último)

6. Autocorrelación

Derivación

$$\rho_1 = \frac{\gamma_1}{\gamma_0} = \frac{\phi_1\cdot\gamma_0}{\gamma_0}$$

$$\rho_2 = \frac{\gamma_2}{\gamma_0} = \frac{\phi_1^2\cdot\gamma_0}{\gamma_0}$$

$$\rho_3 = \frac{\gamma_3}{\gamma_0} = \frac{\phi_1^3\cdot\gamma_0}{\gamma_0}$$

Por lo tanto

$$\rho_s = \frac{\gamma_s}{\gamma_0} = \frac{\phi_1^s \cdot \gamma_0}{\gamma_0}$$

$$\boxed{\rho_s = \phi_1^s}$$

Modelos MA

1. Invertibilidad

Derivación

Consideramos un modelo MA(1)

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

Despejamos ε_t y vamos hacia atrás

$$\varepsilon_t = y_t - \theta_1 \varepsilon_{t-1}$$

$$\varepsilon_{t-1} = y_{t-1} - \theta_1 \varepsilon_{t-2}$$

Reemplazamos este último en nuestro proceso

$$y_t = \varepsilon_t + \theta_1 (y_{t-1} - \theta_1 \varepsilon_{t-2})$$

$$y_t = \varepsilon_t + \theta_1 y_{t-1} - \theta_1^2 \varepsilon_{t-2}$$

Reemplazamos lo que sería ε_{t-2} y luego así con los demás términos

$$y_t = \varepsilon_t + \theta_1 y_{t-1} - \theta_1^2 (y_{t-2} - \theta_1 \varepsilon_{t-3})$$

$$y_t = \varepsilon_t + \theta_1 y_{t-1} - \theta_1^2 y_{t-2} - \theta_1^3 \varepsilon_{t-3}$$

$$y_t = \varepsilon_t + \theta_1 y_{t-1} - \theta_1^2 y_{t-2} - \theta_1^3 (y_{t-3} - \theta_1 \varepsilon_{t-4})$$

$$y_t = \varepsilon_t + \theta_1 y_{t-1} - \theta_1^2 y_{t-2} - \theta_1^3 y_{t-3} - \theta_1^4 \varepsilon_{t-4}$$

Identificando el patrón dejamos expresado este término como

$$\boxed{y_t = \varepsilon_t + \underbrace{\sum_{i=1}^{\infty} \theta_1^i y_{t-i}}_{AR(\infty)}}$$

A partir de un proceso MA(1) logramos llegar a un proceso AR(∞)

Para que esto sea posible, es necesario que se cumpla algunas condiciones. Analicemos un MA(q)

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$
$$y_t = \underbrace{(1 + \theta_1 L + \theta_2 L + ... + \theta_q L)}_{\textit{Polinomio autoregresivo}} \varepsilon_t$$
$$y_t = \Theta(L)\varepsilon_t$$

Para que el proceso sea invertible es necesario que todas las raíces del polinomio autoregresivo $\Theta(L)$ deben estar fuera del circulo unitario (deben ser mayores a 1 en valor absoluto)

Alternativamente podemos reescribir el polinomio autoregresivo $\Theta(L)$ de la siguiente manera

$$\underbrace{\lambda^q + \theta_1 \lambda^{q-1} + \theta_2 \lambda^{q-2} + ... + \theta_{q-1}\lambda - \theta_q}_{\textit{Ecuación Característica}}$$

En paquetes estadísticos como Stata se trabaja más con esta expresión. Implícitamente está la relación $\lambda = \frac{1}{L}$. Por lo tanto, para que el proceso sea invertible, es necesario que todas las raíces de la ecuación característica deben estar dentro del circulo unitario (deben ser menores a 1 en valor absoluto).

En un proceso MA(1) tenemos que

$$\underbrace{(1 + \theta_1 L)}_{\textit{Polinomio autoregresivo}}$$

Por lo tanto

$$\underbrace{(\lambda + \theta_1)}_{\textit{Ecuación Característica}}$$

Saquemos la raiz de la ecuación característica (solución de incógnita igualando a 0)

$$\lambda + \theta_1 = 0$$

$$\lambda = -\theta_1$$

Para que exista invertibilidad en este proceso MA(1) es necesario que

$$|\lambda| = \boxed{|-\theta_1| < 1}$$

Análogamente con el polinomio autoregresivo

$$|L| = \frac{1}{|-\theta_1|} > 1 \Rightarrow \boxed{|-\theta_1| < 1}$$

2. Esperanza

Derivación

Sea $\varepsilon_t \sim N(0, \sigma_\varepsilon^2)$ Definimos $E[y_t] = \mu$

$$y_t = \theta_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$

$$E[y_t] = \theta_0 + E[\varepsilon_t] + \theta_1[\varepsilon_{t-1}] + \theta_2[\varepsilon_{t-2}] + ... + \theta_q[\varepsilon_{t-q}]$$

$$\boxed{\mu = \theta_0}$$

3. Varianza

Derivación

Por simplicidad usaremos un MA(1) y luego extrapolamos

$$y_t = \theta_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$Var[y_t] = Var[\theta_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1}]$$

$$Var[y_t] = Var[\theta_0] + Var[\varepsilon_t] + \theta_1^2 Var[\varepsilon_{t-1}] + 2Cov[\theta_0, \varepsilon_t] + 2\theta_1^2 Cov[\theta_0, \varepsilon_{t-1}] + 2\theta_1^2 Cov[\varepsilon_t, \varepsilon_{t-1}]$$

$$Var[y_t] = 0 + \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 + 0 + 0 + 0$$

$$\boxed{\gamma_0 = (1 + \theta_1^2)\sigma_\varepsilon^2}$$

Si extrapolamos a un proceso MA(q) tendremos

$$\boxed{\gamma_0 = \left(1 + \sum_{i=1}^{q} \theta_i^2\right) \sigma_\varepsilon^2}$$

4. Autocovarianza

Derivación

Definimos $\gamma_s = Cov[y_t, y_{t-s}]$. Por simplicidad usamos un MA(1) y luego extrapolamos

$$y_t = \theta_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

$$Cov(y_t, y_{t-1}) = E[(y_t - E[y_t])(y_{t-1} - E[y_{t-1}])$$

$$\gamma_1 = E[(y_t - \mu)(y_{t-1} - \mu])$$

$$\gamma_1 = E[(\theta_0 + \varepsilon_t + \theta_1 \varepsilon_{t-1} - \theta_0)(\theta_0 + \varepsilon_{t-1} + \theta_1 \varepsilon_{t-2} - \theta_0)$$

$$\gamma_1 = E[(\varepsilon_t + \theta_1 \varepsilon_{t-1})(\varepsilon_{t-1} + \theta_1 \varepsilon_{t-2})]$$

$$\gamma_1 = E[\varepsilon_t \varepsilon_{t-1} + \theta_1 \varepsilon_t \varepsilon_{t-2} + \theta_1 \varepsilon_{t-1}^2 + \theta_1^2 \varepsilon_{t-1} \varepsilon_{t-2}]$$

$$\gamma_1 = E[\varepsilon_t\varepsilon_{t-1}] + \theta_1 E[\varepsilon_t\varepsilon_{t-2}] + \theta_1 E[\varepsilon_{t-1}^2] + \theta_1^2 E[\varepsilon_{t-1}\varepsilon_{t-2}]$$
$$\gamma_1 = 0 + \theta_1 \cdot 0 + \theta_1\sigma_\varepsilon^2 + \theta_1^2 \cdot 0$$
$$\boxed{\gamma_1 = \theta_1\sigma_\varepsilon^2}$$

Notar que

$$\gamma_2 = E[(\varepsilon_t + \theta_1\varepsilon_{t-1})(\varepsilon_{t-2} + \theta_1\varepsilon_{t-3})]$$
$$\gamma_2 = E[(\varepsilon_t\varepsilon_{t-2} + \theta_1\varepsilon_t\varepsilon_{t-3} + \theta_1\varepsilon_{t-1}\varepsilon_{t-2} + \theta_1^2\varepsilon_{t-1}\varepsilon_{t-3}]$$
$$\gamma_2 = E[(\varepsilon_t\varepsilon_{t-2}] + \theta_1 E[\varepsilon_t\varepsilon_{t-3}] + \theta_1 E[\varepsilon_{t-1}\varepsilon_{t-2}] + \theta_1^2 E[\varepsilon_{t-1}\varepsilon_{t-3}]$$
$$\gamma_2 = 0 + \theta_1 \cdot 0 + \theta_1 \cdot 0 + \theta_1^2 \cdot 0$$
$$\boxed{\therefore \gamma_2 = 0}$$

Esto mismo ocurre para covarianzas de ordenes superiores. Por lo tanto, la covarianza de orden S de un modelo MA(1) es

$$\gamma_s = \begin{cases} \theta_1\sigma_\varepsilon^2 & \text{si } s \leq 1 \\ 0 & \text{si } s > 1 \end{cases}$$

Extrapolando tenemos que la covarianza de orden S de un modelo MA(q) es

$$y_t = \theta_0 + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + ... + \theta_q\varepsilon_{t-q}$$

$$\boxed{\gamma_s = \begin{cases} \sigma_\varepsilon^2 \sum_{j=0}^{q-s} \theta_j\theta_{j+s} & \text{si } s \leq q \quad \textit{donde} \ \ \theta_0 = 1 \\ 0 & \text{si } s > q \end{cases}}$$

De aquí viene el que las series MA(q) tienen una "memoria finita" de orden q

5. Autocorrelación

Derivación

$$\boxed{\rho_s = \frac{\gamma_s}{\gamma_0} = \begin{cases} \frac{\sum_{j=0}^{q-s} \theta_j\theta_{j+s}}{(1+\sum_{i=1}^{q} \theta_i^2)} & \text{si } s \leq q \quad \textit{donde} \ \ \theta_0 = 1 \\ 0 & \text{si } s > q \end{cases}}$$

Modelos ARMA

1. Estacionariedad e Invertibilidad

Derivación

Consideramos un proceso ARMA(p,q)

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + ... + \phi_p y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$$

$$(1 - \phi_1 L - \phi_2 L^2 - ... - \phi_p L^p) y_t = (1 + \theta_1 L + \theta_2 L^2 + ... + \theta_q L^q) \varepsilon_t$$

$$\boxed{\Phi(L) y_t = \Theta(L) \varepsilon_t}$$

Para que este proceso sea estacionario, todas las raíces del polinomio autoregresivo $\Phi(L)$ deben estar afuera del círculo unitario.
Para que este proceso sea invertible, todas las raíces del polinomio autoregresivo $\Theta(L)$ deben estar afuera del círculo unitario.

Apéndice

Apéndice N°1: Sea $\{x_i, y_i : i = 1, \ldots, n\}$ un conjunto de n par de observaciones, tendremos de que:

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0$$

Demostración

$$\sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x} = \sum_{i=1}^{n} x_i - \bar{x} \underbrace{\sum_{i=1}^{n} 1}_{=n}$$

$$\boxed{\therefore \quad \sum_{i=1}^{n}(x_i - \bar{x}) = \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i = 0}$$

Apéndice N°2: Sea $\{x_i, y_i : i = 1, \ldots, n\}$ un conjunto de n par de observaciones, tendremos de que:

$$\sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i \cdot (x_i - \bar{x})$$

Demostración

$$\begin{aligned}
\sum_{i=1}^{n} (x_i - \bar{x})^2 &= \sum_{i=1}^{n} \left(x_i^2 - 2 \cdot x_i \cdot \bar{x} + \bar{x}^2\right) \\
&= \sum_{i=1}^{n} x_i^2 - 2 \cdot \bar{x} \cdot \sum_{i=1}^{n} x_i + \bar{x}^2 \cdot \sum_{i=1}^{n} 1 \quad / \quad \sum_{i=1}^{n} x_i = n \cdot \bar{x} \\
&= \sum_{i=1}^{n} x_i^2 - 2 \cdot n \cdot \bar{x}^2 + n \cdot \bar{x}^2 \\
&= \sum_{i=1}^{n} x_i^2 - n \cdot \bar{x}^2 = \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \cdot \bar{x} = \sum_{i=1}^{n} \left(x_i^2 - x_i \cdot \bar{x}\right)
\end{aligned}$$

$$\boxed{\therefore \quad \sum_{i=1}^{n} (x_i - \bar{x})^2 = \sum_{i=1}^{n} x_i \cdot (x_i - \bar{x})}$$

Apéndice N°3: Sea $\{x_i, y_i : i = 1, \ldots, n\}$ un conjunto de n par de observaciones, tendremos de que:

$$\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})=\sum_{i=1}^{n}y_i\cdot(x_i-\bar{x})=\sum_{i=1}^{n}x_i\cdot(y_i-\bar{y})$$

Demostración

$$\begin{aligned}
\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y}) &= \sum_{i=1}^{n}(x_i\cdot y_i-x_i\cdot\bar{y}-y_i\cdot\bar{x}+\bar{x}\cdot\bar{y}) \quad / \quad \sum_{i=1}^{n}z_i=n\cdot\bar{z}\\
&= \sum_{i=1}^{n}x_i\cdot y_i-n\cdot\bar{x}\cdot\bar{y}-n\cdot\bar{y}\cdot\bar{x}+\underbrace{\sum_{i=1}^{n}\bar{x}\cdot\bar{y}}_{=n\cdot\bar{x}\cdot\bar{y}}\\
&= \sum_{i=1}^{n}x_i\cdot y_i-n\cdot\bar{x}\cdot\bar{y}=\sum_{i=1}^{n}x_i\cdot y_i-\sum_{i=1}^{n}\bar{x}\cdot y_i=\sum_{i=1}^{n}(x_i\cdot y_i-\bar{x}\cdot y_i)
\end{aligned}$$

$$\boxed{\therefore \quad \sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})=\sum_{i=1}^{n}y_i\cdot(x_i-\bar{x})=\sum_{i=1}^{n}x_i\cdot(y_i-\bar{y})}$$

Apéndice N°4:

- Introduciremos el concepto de esperanza iterada, la cual dice que si $\mathbb{E}|y| < \infty$ entonces para cualquier vector aleatorio de x, se cumple que:

$$\mathbb{E}(\mathbb{E}(y|x)) = \mathbb{E}(y) \tag{8}$$

A esta se le conoce como la Ley simple de las expectativas iteradas, la cual nos dice, en resumidas cuentas, que el promedio de los promedios condicionales es el promedio incondicional.

Por ejemplo, consideremos la siguiete tabla:

Y	X
1	1
2	1
3	2
4	2

En este caso tendremos que:

$$E[Y] = \frac{1+2+3+4}{4} = 2{,}5$$

$$E[Y/X=1] = \frac{1+2}{2} = 1{,}5$$

$$E[Y/X=2] = \frac{3+4}{2} = 3{,}5$$

$$E[E[Y/X]] = \frac{1{,}5 + 3{,}5}{2} = 2{,}5$$

Por lo que es fácil de ver que

$$\boxed{E[E[Y/X]] = E[Y]}$$

Cuando x es discreta:

$$\mathbb{E}(\mathbb{E}(y|x)) = \sum_{j=1}^{\infty} \mathbb{E}(y|x_j)Pr(x = x_j)$$

Cuando x es continua:

$$\mathbb{E}(\mathbb{E}(y|x)) = \int_{\mathbb{R}^k} \mathbb{E}(y|x)f_x(x)dx$$

La Ley general de las esperanzas iteradas permite dos condiciones para las variables establecidas, la Ley de la experanza iterada nos dice que:
Si $\mathbb{E}|y| < \infty$ entonces para cualquier vectores x_1 y x_2 aleatorios, se cumple de que:

$$\mathbb{E}(\mathbb{E}(y|x_1, x_2)|x_1) = \mathbb{E}(y|x_1)$$

Esta nos dice, en resumidas cuentas que la variable que contenga la mayor cantidad de información es la que termina ganando.

Apéndice N°5: Teniendo de que $\mu_x = \mathbb{E}(x)$, la varianza se puede expresar como:

$$V(x) = \mathbb{E}\left((x - \mathbb{E}(x))^2\right) = \mathbb{E}(x^2) - \left(\mathbb{E}(x^2)\right)^2$$

Apéndice N°6: Operador de Rezagos. $L^i y_t = y_{t-i}$

Demostración

Propiedades:
Lc=c
$\beta(Ly_t) = L(\beta y_t) = \beta y_{t-1}$
$L^i L^j y_t = y_{t-i-j}$
$(L^i + L^j)y_t = y_{t-i} + y_{t-j}$
$L^{-i} = y_{t+i}$
$\Delta y_t = y_t - y_{t-1} = (1 - L)y_t$