5CAPÍTULO 8

Ejercicio 8.1

Se dispone de T =100 observaciones para una variable dependiente (y) , y seis variables explicativas. Para la variable y se sabe que: $\sum(\mathrm{y}_t - \bar{\mathrm{y}})^2 = 20$ Se consideran los modelos anidados M1 , M2, el primero con 3 variables y el segundo con las seis variables, con los siguientes coeficientes de determinación: $\mathrm{R}_1^2 = 0{,}8, \mathrm{y} \mathrm{R}_2^2 = 0{,}9$.

1).Suponiendo que la hipótesis nula es M1, escribir la región crítica que corresponde a los siguientes criterios: $\overline{\mathrm{R}}^2$, AIC , LR y SBIC e indicar la decisión que se tomaría utilizando cada uno de los cuatro criterios.

2). Interpretar cada uno de los anteriores contrastes como un contraste de la F derivando el correspondiente punto crítico.

3). Derivar cual sería el tamaño del error tipo 1que corresponde a cada uno de los cuatro criterios cuando el tamaño muestral tiende a infinito.

Solución

1). Primero tener en cuenta que,

$$\hat{u}_i'\hat{u}_i = (1-R_i^2)\sum(y_t - \bar{y})^2$$

por lo que será

$$\frac{\hat{u}_1'\hat{u}_1}{\hat{u}_2'\hat{u}_2} = \frac{4}{2} = 2$$

Para el criterio $\overline{\mathrm{R}}^2$ la Región Crítica es

$$\overline{R}_1^2 < \overline{R}_2^2 \Rightarrow \frac{\hat{u}_1'\hat{u}_1}{\hat{u}_2'\hat{u}_2} > \frac{T-k_1}{T-k_2} = 1{,}0319$$

por lo que la hipótesis nula se rechaza.

Para el contraste LR la región crítica es:

$$T\ln\frac{\tilde{\sigma}_1^2}{\tilde{\sigma}_2^2} > \frac{\chi^2(k_2-k_1)}{T} \Rightarrow \frac{\hat{u}_1'\hat{u}_1}{\hat{u}_2'\hat{u}_2} > \exp\left(\frac{\chi^2(k_2-k_1)}{T}\right) \approx 1 + \frac{\chi^2(k_2-k_1)}{T} = 1{,}07$$

ya que $\chi^2(3) = 7{,}81$. La hipótesis nula se rechaza.

La Región Crítica para el criterio AIC es la siguiente,

$$AIC_1 > AIC_2 \Rightarrow \frac{\hat{u}_1'\hat{u}_1}{\hat{u}_2'\hat{u}_2} > \exp\left(\frac{2(k_2-k_1)}{T}\right) \approx 1 + \frac{2(k_2-k_1)}{T} = 1{,}06$$

por lo que también se rechaza la hipótesis nula.

Para el criterio SBIC se tiene,

$$SBIC_1 > SBIC_2 \Rightarrow \frac{\hat{u}_1'\hat{u}_1}{\hat{u}_2'\hat{u}_2} > \exp\left(\frac{\ln T(k_2-k_1)}{T}\right) \approx 1 + \frac{\ln T(k_2-k_1)}{T} = 1{,}14$$

y también se rechaza la hipótesis nula.

CAPÍTULO 8
Ejercicio 8.1

2). Como puede verse en la Sección 8.2, la Región Crítica del contraste de la F puede escribirse como

$$\frac{\hat{u}_1'\hat{u}_1}{\hat{u}_2'\hat{u}_2} > 1 + \frac{k_2 - k_1}{T - k_2} F_{\varepsilon}$$

Para obtener el punto crítico correspondiente a cada uno de los cuatro criterios considerados en este ejercicio, basta igualar la parte de la derecha de esta región crítica con las correspondientes en las cuatro regiones críticas anteriormente derivadas. Por ejemplo, para el contraste $\overline{R}^2$ se tiene que

$$1 + \frac{k_2 - k_1}{T - k_2} F_{\overline{R}^2} = \frac{T - k_1}{T - k_2}$$

de donde se obtiene que: $F_{\overline{R}^2} = 1$.Siguiendo un proceso similar se obtienen los puntos críticos para los otros tres criterios,

$$F_{LR} = \frac{\chi^2(k_2 - k_1)}{T} \frac{T - k_2}{k_2 - k_1}$$

$$F_{AIC} \approx 2$$

$$F_{SBIC} \approx \ln T$$

3). Si el tamaño muestral tiende a infinito, entonces los puntos críticos de los tres primeros criterios tienden a constantes por lo que el tamaño del error tipo 1 no tenderá a cero. Solamente para el criterio SBIC el punto crítico tiende a infinito por lo que el tamaño del error tipo 1 tenderá a cero.

Ejercicio 8.2

Sea M1 un modelo lineal anidado en otro modelo lineal, M2. Sean $\hat{u}_1$ y $\hat{u}_2$ los respectivos vectores de residuos MCO.

1). Obtener las esperanzas y matrices de varianzas y covarianzas de ambos vectores de residuos generando los datos M2.

2). Demostrar que si genera los datos M1, se cumple que :

$$\hat{u}_2'\hat{u}_2 \le \hat{u}_1'\hat{u}_1$$

¿Se cumple esta desigualdad si genera los datos M2?.

3). Para discriminar entre M1 y M2 un investigador propone utilizar el criterio AIC de Akaike alegando que dicho criterio garantiza un tamaño del error tipo 1 igual al 5% que,en este caso, corresponde

CAPÍTULO 8
Ejercicio 8.2

a un punto crítico del contraste de la F igual a 3,5. Evaluar esta propuesta demostrando los resultados que sean necesarios.

4). Otro investigador propone utilizar conjuntamente los criterios $\bar{R}^2$ y el contraste de la F tomando un nivel de significación del 5%. Evaluar la coherencia de esta propuesta.

Solución

1). Sean los modelos,

$$M1: y = X_1\beta_1 + u_1$$

$$M2: y = X\beta + u = X_1\beta_1 + X^{\bullet}\beta^{\bullet} + u_2$$

Los vectores de residuos son:

$$\hat{u}_1 = M_1 y = M_1 X^{\bullet}\beta^{\bullet} + M_1 u_2$$

$$\hat{u}_2 = M_2 y = M_2 u_2$$

en donde, $M_1 = I - X_1(X_1^{'}X_1)^{-1}X_1^{'}$ y $M_2 = I - X(X^{'}X)^{-1}X^{'}$. Los momentos son:

$E(\hat{u}_1) = M_1 X^{\bullet}\beta^{\bullet}$; $Var(\hat{u}_1) = \sigma_2^2 M_1$ (Demostrar este resultado).

$E(\hat{u}_2) = 0$; $Var(\hat{u}_2) = \sigma_2^2 M_2$ (Demostrar este resultado).

2). Esa desigualdad se cumple siempre ya que la suma de cuadrados de los residuos de un modelo restringido anidado en otro siempre es mayor que la suma correspondiente al modelo amplio tal como se demuestra en el Resultado 1.3.

3). Tal como se ha demostrado en el Apartado 1 de este ejercicio, la utilización del criterio AIC equivale a utilizar el contraste F con un punto crítico que gira en torno a 2. Por lo tanto, las dos propuestas pueden llevar a resultados muy diferentes.

4). Como también se ha demostrado en el Apartado 1, el criterio $\bar{R}^2$ tambien admite una interpretación como un contraste F con un punto crítico igual a 1. Se ve que este valor es muy diferente del que corresponde a un nivel de significación del 5%, por lo que el uso de ambas estrategias puede conducir a resultados muy diferentes.

Ejercicio 8.3

Sea M1 un modelo con k_1 regresores que está anidado en otro modelo M2 con k_2 regresores. Tomando toda la información hasta el periodo T ,y utilizando el modelo M1 se hace una predicción MCO para el periodo siguiente; sea e_{T+1} el error de predicción correspondiente. A continuación,usando

CAPÍTULO 8
Ejercicio 8.3

la información hasta T+1 se formula la predicción para el periodo T+2; sea e_{T+2} el error de predicción correspondiente.Suponiendo que genera los datos M2 se pide:

1). Derivar $E\left(\frac{e_{T+1}^2+e_{T+2}^2}{2}\right)$

2). Derivar $Var\left(\frac{e_{T+1}^2+e_{T+2}^2}{2}\right)$

3). Utilizando los resultados de los dos puntos anteriores comentar la utilidad de utilizar el promedio de los cuadrados de dos errores de predicción sucesivos con un periodo de horizonte para discriminar entre M1 y M2.

Solución

1) Escribamos los modelos como

$$M1: y = X\beta_1 + u_1$$

$$M2: y = X_1\beta_1 + X_2\beta_2 + u = X\beta + u$$

Los errores de predicción se definen como (M2 genera los datos)

$$e_{T+1} = y_{T+1} - x'_{1T+1}\hat{\beta}_1 = h_{12T+1} + u_{T+1} - x'_{1T+1}(X'_1X_1)^{-1}X'_1u$$

$$e_{T+2} = y_{T+1} - x'_{1T+2}\hat{\beta} = h_{12T+2} + u_{T+2} - x'_{1T+2}(X'_{1T+1}X_{1T+1})^{-1}X'_{1T+1}u^*$$

A partir de estas expresiones se tiene que si genera los datos M2 entonces,

$$e_{T+1} \approx N(h_{12T+1}, \sigma_2^2(1+\frac{k_1}{T}))$$

$$e_{T+2} \approx N(h_{12T+2}, \sigma_2^2(1+\frac{k_1}{T+1}))$$

Por lo tanto, $\frac{e_{T+1}^2}{\sigma_2^2(1+\frac{k_1}{T})} \approx \chi^2(1, \frac{h_{12T+1}^2}{\sigma_2^2(1+\frac{k_1}{T})})$

Y teniendo en cuenta que,

$$E\chi^2(k,\lambda^2) = 1+\lambda^2$$

$$Var(\chi^2(k,\lambda^2) = 2(1+2\lambda^2)$$

podemos escribir

$$Ee_{T+1}^2 = \ldots\ldots\ldots\ldots\ldots\ldots = h_{12T+1}^2 + \sigma_2^2(1+\frac{k_1}{T})$$

CAPÍTULO 8
Ejercicio 8.3

2). Con respecto a la varianza se tiene que:

$$Var(e_{T+1}^2) = ... = 2\sigma_2^4(1+\frac{k_1}{T}) + 4\sigma_2^2(1+\frac{k_1}{T})h_{12T+1}^2$$

El resultado para e_{T+2} es el mismo cambiando T por T+1.
Al resultado se llega teniendo en cuenta que,
$Cov(e_{T+1}, e_{T+2}) = 0$ (Demostrado en 7.4).

3). Podemos pensar en discriminar entre dos modelos prestando atención al promedio de los cuadrados de los errores de predicción. Pero el problema es que su varianza puede ser muy grande y llevar a un procedimiento de selección poco eficiente ya que cualquier resultado es posible. La solución está en tomar muchos errores de predicción en lugar de dos; en este caso, se demuestra que la varianza del promedio de los cuadrados de los errores de predicción tiende a cero. (Demostrar este resultado).

Ejercicio 8.4

Para un modelo lineal con tres regresores se conoce que:
$\bar{y}=4$; $y'y=200$; $y'X(X'X)^{-1}X'y=190$; T=10;
Se pide:
1). Calcular los valores que tomarían los estadísticos $\bar{R}^2$, SBIC y AIC.
2). Suponiendo que el modelo anterior se va a comparar con otro modelo lineal que tiene cinco regresores derivar y calcular el valor de los factores de parsimonia y de los puntos críticos implicitos de la F correspondientes a cada uno de los tres criterios comentados.

CAPÍTULO 8
Ejercicio 8.4

3). Determinar qué valor debería tomar la suma de cuadrados de los residuos del modelo con cinco regresores para que dicho modelo fuera aceptado utilizando el criterio AIC y fuera rechazado utilizando el criterio SBIC. Determinar qué valor debería tomar dicha suma para que el modelo con cinco regresores fuera rechazado por los tres criterios.

Solución

1). Primero, tener en cuenta que,

$$\hat{u}_1^{'}\hat{u}_1 = y^{'}My = 10$$

$$\sum (y_t - \bar{y})^2 = y^{'}y - T\bar{y}^2 = 40 \quad ; \quad \tilde{\sigma}_1^2 = 1$$

A partir de estos resultados, se tiene que,

$$\bar{R}_1^2 = 1 - \frac{T-1}{T-k_1}\frac{\hat{u}_1^{'}\hat{u}_1}{\sum (y_t - \bar{y})^2} = 0{,}678$$

$$AIC_1 = \ln \tilde{\sigma}_1^2 + \frac{2k_1}{T} = 0{,}8$$

$$SBIC_1 = \ln \tilde{\sigma}_1^2 + \frac{\ln T k_1}{T} = 0{,}92$$

2). Para obtener el factor de parsimonia, la región crítica de cualquier contraste, A , debe escribirse como,

$$\hat{u}_1^{'}\hat{u}_1 > \hat{u}_2^{'}\hat{u}_2 h(A)$$

en donde h(A) es el factor de parsimonia. Retomando los resultados derivados en el Apartado 1) del Ejercicio 8.1), tenemos,

$$h(\bar{R}^2) = \frac{T-k_1}{T-k_2} = 1{,}4$$

$$h(AIC) = 1 + \frac{2(k_2 - k_1)}{T} = 1{,}4$$

$$h(SBIC) = 1 + \frac{\ln T(k_2 - k_1)}{T} = 1{,}46$$

Se ve que el criterio más parsimonioso es el SBIC.

Para derivar el punto crítico implícito, hay que tener en cuenta lo derivado en el Apartado 2) de ese mismo ejercicio. Se obtiene,

$$F_{\hat{R}^2} = 1 \; ; \quad F_{AIC} = 2 \quad \text{y} \quad F_{SBIC} = \ln 10 = 2{,}30 .$$

3). Para que el modelo con cinco regresores sea aceptado con el criterio AIC se debe de cumplir que,

CAPÍTULO 8
Ejercicio 8.4

$$\hat{u}_1'\hat{u}_1 > \hat{u}_2'\hat{u}_2 h(AIC)$$

es decir, que se cumpla que:

$$\hat{u}_2'\hat{u}_2 < \frac{\hat{u}_1'\hat{u}_1}{h(AIC)} = 7{,}14$$

Para que sea rechazado utilizando el criterio SBIC se debe de cumplir que

$$\hat{u}_1'\hat{u}_1 < \hat{u}_2'\hat{u}_2 h(SBIC)$$

es decir, que se cumpla que:

$$\hat{u}_2'\hat{u}_2 > \frac{\hat{u}_1'\hat{u}_1}{h(SBIC)} = 6{,}84$$

Por lo tanto, la suma de cuadrados de los residuos del modelo amplio tiene que ser mayor que 6,84 y menor que 7,14. Para que sea rechazado por los tres criterios basta con que sea rechazado por el criterio menos parsimonioso que es el $\overline{R}^2$. Para que sea rechazado por este criterio la suma de los cuadrados de los residuos del modelo amplio tiene que ser tal que.

$$\hat{u}_2'\hat{u}_2 > \frac{\hat{u}_1'\hat{u}_1}{h(\overline{R}^2)} = 7{,}14\,.$$

Ejercicio 8.5

Suponer que se están comparando tres modelos, M1, M2 y M3. El primero está anidado en el segundo y éste en el tercero. Con 96 observaciones se han obtenido los siguientes estadísticos:

	k	$\hat{u}'\hat{u}$	$\hat{\sigma}$	AIC	SBIC
M1	5	11.62	?	-2.00	-1.87
M2	7	?	.33	?	-1.91
M3	9	9.58	.33	-2.11	?

en k se incluye como parámetro la varianza de la perturbación del modelo.

Se pide:

1). Calcular las casillas que faltan.

2). Discriminar entre M1 y M2 utilizando los siguientes criterios: $R^2, \overline{R}^2$ y el contraste de la F con un nivel de significación del 5% sabiendo que F (0.05) = 3.09.

3). Derivar y calcular el factor de parsimonia de los criterios $R^2, \overline{R}^2$ y AIC cuando se discrimina entre M1 y M2.

CAPÍTULO 8
Ejercicio 8.5

Solución

1). Las casillas que faltan son

$$\hat{\sigma}_1^2 = 0{,}126 \;;\; \hat{u}_2^{'}\hat{u}_2 = \hat{\sigma}_2^2(T - k_2 + 1) = 9{,}8$$

$$AIC_2 = ... = -2{,}13$$

$$SBIC_3 = ... = -1{,}87$$

2). Si se utiliza el R^2 se elige siempre el modelo más amplio es decir, en este caso el M3.

El uso del $\bar{R}^2$ equivale a elegir el modelo que tiene un menor valor del estimador $\hat{\sigma}$, en nuestro caso el modelo M2. En lo que respecta al uso de la F, el valor que toma es 8,42 y por ser superior al punto crítico la hipótesis nula es rechazada.

3). La derivación de los factores de parsimonia se ha llevado a cabo en el Ejercicio 1. Aplicando los resultados alli obtenidos se tiene,

$$h(R^2) = 1$$

$$h(\bar{R}^2) = ... = 1{,}022$$

$$h(AIC) = ... = 1{,}04$$

Ejercicio 8.6

Sean $\mathrm{y_t}$ y $\mathrm{x_t}$ dos variables I(1) que están cointegradas. Demostrar que el coeficiente de determinación de la regresión de $\mathrm{y_t}$ sobre $\mathrm{x_t}$ tiende siempre a ser superior al de la regresión de $\Delta\mathrm{y_t}$ sobre $\Delta\mathrm{x_t}$. Demostrar que ese resultado no tiene porque cumplirse si las dos variables son I(0).

Solución

El R^2 de la regresión en niveles puede escribirse como,

$$R^2 = 1 - \frac{1}{T}\frac{\hat{u}^{'}\hat{u}/T}{\sum y_t^2/T^2}$$

La convergencia del numerador se explica porque los residuos son estacionarios y la convergencia del denominador por el Resultado 2.6. Se ve como el coeficiente de determinación será muy elevado, próximo a la unidad. La regresión en primeras diferencias es,

$$\Delta y_t = \beta\Delta x_t + u_t - u_{t-1}$$

Siendo $v_t = u_t - u_{t-1}$ podemos escribir:

CAPÍTULO 8
Ejercicio 8.6

$$R^{\bullet 2} = 1 - \frac{\hat{v}'\hat{v}}{\sum \Delta y_t^2} \rightarrow 1 - \frac{\sigma_v^2}{\sigma_{\Delta y}^2}$$

y esta expresión puede tomar cualquier valor entre 0 y 1. En el caso en que las dos variables sean I(0), entonces estamos en el caso similar al que teníamos antes con el modelo en primeras diferencias.

Ejercicio 8.7

El criterio AIC esta basado en la función de pérdida de la distancia de Kullback-Leibler.Utilizando este concepto resolver las dos cuestiones siguientes:

1). Dos expertos predicen que la proporción de consumidores de un determinado producto será respectivamente .7 y .5. Si la verdadera proporción es .6 ¿Qué predicción de las dos está más próxima de la verdadera?.

2). Suponer que la verdadera distribución viene dada por la distribución Normal estandar N(0, 1) ¿ Qué modelo, el N(.5, 1) o el N (0, 1.5) está más próximo de la verdadera distribución?.

Solución

1). Supongamos que $p = \{p_1, p_2, ...p_m\}$ es la verdadera distribución y que $q = \{q_1, q_2, ...q_m\}$ es cualquier otra distribución que se propone para aproximar la verdadera distribución. La expectativa de la variable $\log \frac{p_i}{q_i}$ viene dada por:

$$I(p:q) = E \log \frac{p}{q} = \sum_1^m p_i \log \frac{p_i}{q_i}$$

Esta es la llamada distancia de Kullback-Leibler y nos da una medida de distancia entre las dos distribuciones de probabilidad.

En este ejercicio, tenemos dos distribuciones

$$q_A = (.5 \quad .5)$$
$$q_B = (0.7 \quad .3)$$

Las distancias de cada distribución a la distribución verdadera son:

$$I(p:q_A) = ... = .0201$$

$$I(p:q_B) = ... = .0226$$

La conclusión es que la distribución A está más próxima a la distribución verdadera.

CAPÍTULO 8

Ejercicio 8.7

2). Suponer que partimos de dos funciones de densidad continuas, f(x), y g(x), y que la primera es la distribución verdadera. La distancia de Kullback-Leibler entre las dos viene dada por:

$$I(f:g) = \int_{-\infty}^{\infty} f(x) \log \frac{f(x)}{g(x)} dx$$

Supongamos que $f(x) \approx N(\mu, \sigma^2)$ y que $g(x) \approx N(\xi, \tau^2)$ .Entonces, teniendo en cuenta que, $E_x(x-\xi)^2 = \sigma^2 + (\mu - \xi)^2$, se llega a:

$$\int f(x) g(x) dx = -\frac{1}{2}\log(2\pi\tau^2) - \frac{\sigma^2 + (\mu-\xi)^2}{2\tau^2}$$

A partir de esta expresión se obtiene,

$$I(f:g) = \frac{1}{2}\left\{\log\frac{\tau^2}{\sigma^2} + \frac{\sigma^2 + (\mu-\xi)^2}{\tau^2} - 1\right\}$$

En este ejercicio, la distribución verdadera viene caracterizada por $\mu = 0$ y $\sigma^2 = 1$ por lo que,

$$I(f:g) = \frac{1}{2}\left\{\log\tau^2 + \frac{1+\xi^2}{\tau^2} - 1\right\}$$

por lo que las distancias de las dos distribuciones son

$$I(f:g_1) = \dots\dots\dots\dots\dots\dots = 0{,}125$$

$$I(f:g_2) = \dots\dots\dots\dots\dots\dots\dots = 0{,}036$$

La coclusión es que la segunda distribución está más próxima de la verdadera distribución que la primera.

Ejercicio 8.8

El criterio AIC de Akaike se escribe como:

AIC = $-2\mathrm{l}(\tilde{\theta}) + 2\mathrm{k}$

en donde $\mathrm{l}(\tilde{\theta})$ es el valor que toma el logaritmo de la función de verosimili tud sustituyendo los parámetros por sus estimadores máximo-verosímiles, y k es el número de parámetros que se estiman.

Utilizando el criterio AIC resolver la siguiente cuestión planteada en Sakamoto, Ishiguro y Kitagawa (1986):

Una máquina produce bolas esféricas. Se conoce que los diámetros de estas bolas, si la máquina funciona normalmente, tienen una media de un centímetro y una desviación típica de .01 cm.

CAPÍTULO 8
Ejercicio 8.8

Un dia se eligen aleatoriamente
20 bolas y se miden sus diámetros. Los resultados son los siguientes.

.999	1.013	.974	.993	.989
1.001	1.008	1.003	.989	1.009
1.001	.977	1.023	.994	.988
1.005	1.006	.995	1.003	1.027

¿ Podemos concluir, a partir de esta información, diciendo que la máquina sigue funcionando normalmente?.

Solución

Aceptamos que la máquina funciona normalmente si la distribución que corresponde a los datos proporcionados no está más próxima a la distribución verdadera que la distribución asociada con el funcionamiento normal. Primero, hay que tener en cuenta que $\theta = (\mu, \sigma^2)$. Por otra parte, hay que tener en cuenta que los estimadores máximoverosimiles que se obtienen a partir de los datos son:

$$\tilde{\mu} = 0{,}99985 \;; \quad \tilde{\sigma}^2 = 0{,}000168$$

El valor del logaritmo de la función de verosimilitud bajo el supuesto de funcionamiento normal es

$$l(\mu_0, \sigma_0^2) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\sigma_0^2 - \frac{T}{2\sigma_0^2}\left\{(\mu_0 - \bar{y})^2 + \tilde{\sigma}^2\right\}$$

Sustituyendo por los estimadores máximoverosímiles el valor del logaritmo es:

$$l(\tilde{\mu}, \tilde{\sigma}) = -\frac{T}{2}\log(2\pi) - \frac{T}{2}\log\tilde{\sigma}^2 - \frac{T}{2}$$

A partir de estas expresiones se llega a

$$AIC(\mu_0, \sigma_0^2) = -2l(\mu_0, \sigma_0^2) + 2\cdot 0 = -113{,}8$$

$$AIC(\tilde{\mu}, \tilde{\sigma}^2) = -2l(\tilde{\mu}, \tilde{\sigma}^2) + 2\cdot 2 = -113{,}1$$

El valor que toma el criterio AIC para el modelo restringido $\left\{\theta = \left(\mu_0 = 1, \sigma_0^2 = 0{,}01\right)\right\}$ es menor que el que toma para el modelo sin restringir, por lo que la evidencia de los datos indica que se puede aceptar el modelo asociado con el funcionamiento normal.