INGENIERIA INDUSTRIAL
UNIVERSIDAD DE CHILE

Auxiliar 5

IN540

Profesor : Mattia Makovec
Auxiliar : Gonzalo Viveros A.
Semestre : Primavera 2009

Pregunta 1

Sea el modelo definido por:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \beta_3 X_{3t} + \beta_4 X_{4t} + u_t, \qquad t = 1, \ldots, T,$$

el cual cumple las hipótesis del MLG con los errores normales. Este modelo fue estimado por MCO con 26 observaciones, obteniendo los siguientes resultados:

$$Y_t = 2 + 3.5\, X_{2t} - 0.7\, X_{3t} + 2.1\, X_{4t} + \varepsilon_t,$$

obteniendo además los siguientes estadísticos:

Estad. t_2	Estad. t_3	Estad. t_4	R^2
1.9	−2.2	1.5	0.982

El mismo modelo fue estimado con la restricción $\beta_2 = \beta_4$ y se obtuvo el siguiente resultado:

$$Y_t = 1.5 + 3.1\,(X_{2t} + X_{4t}) - 0.6\, X_{3t} + \varepsilon_t,$$

dando los siguientes estadísticos:

Estad. t_2	Estad. t_3	R^2
2.7	−2.4	0.876

a) Contraste la significatividad global del modelo original.

b) Contraste la restricción $\beta_2 = \beta_4$ frente a una alternativa bilateral, utilizando un intervalo de confianza apropiado.

c) ¿Qué ocurriría con el R^2 de las dos estimaciones realizadas si se eliminara la variable X_{3t}?

Pregunta 2

Se desea estudiar la relación entre las variables "*Consumo Agregado*" (C_t) y "*Producto Nacional Bruto*" (P_t) de un determinado país, para lo cual dispondremos de 20 observaciones correspondientes al período 1963-1982. Llamaremos D_{1t} a una variable binaria que vale 1 si la observación corresponde al período 1963-1970 y 0 en el caso contrario. Conocemos los resultados de las estimaciones MCO siguientes, realizadas todas ellas con las 20 observaciones disponibles.

$$
\begin{array}{llll}
\mathcal{M}_1: & C_t = -24 + 0.64\,P_t + \varepsilon_t, & R^2 = 0.9955, & SCR_1 = 1626. \\
\mathcal{M}_2: & C_t = 33\,D_{1t} - 49\,D_{2t} + 0.58\,(D_{1t}\,P_t) + 0.66\,(D_{2t}\,P_t) + \varepsilon_t, & & SCR_2 = 1331. \\
\mathcal{M}_3: & C_t = -17 + 0.63\,(D_{1t}\,P_t) + 0.64\,(D_{2t}\,P_t) + \varepsilon_t, & & SCR_3 = 1616. \\
\mathcal{M}_4: & C_t = 835 - 0.22\,(D_{1t}\,P_t) + \varepsilon_t, & & SCR_4 = 129560. \\
\mathcal{M}_5: & C_t = 601 + 0.18\,(D_{2t}\,P_t) + \varepsilon_t, & & SCR_5 = 69418. \\
\mathcal{M}_6: & C_t = 0.64\,P_t - 23.63\,D_{1t} - 23.38\,D_{2t} + \varepsilon_t, & & SCR_6 = 1625.
\end{array}
$$

Conocemos también los resultados de otras dos estimaciones MCO, una de ellas realizadas únicamente con las observaciones del primer periodo (modelo $\mathcal{M}_A$), y la otra con las observaciones del segundo periodo (modelo $\mathcal{M}_B$):

$$
\begin{array}{lll}
\mathcal{M}_A: & C_t = 33 + 0.58\,P_t + \varepsilon_t, & SCR_A = 186. \\
\mathcal{M}_B: & C_t = -49 + 0.66\,P_t + \varepsilon_t, & SCR_B = 1145.
\end{array}
$$

a) Suponiendo que la varianza de los errores es la misma en los dos modelos, contraste si puede admitirse que el término independiente (es decir, el parámetro de la variable constante) es el mismo en estos dos modelos.

b) Suponiendo que la varianza de los errores es la misma en los dos modelos, contraste si puede admitirse que la pendiente (es decir, el parámetro de la variable P_t) es el mismo en estos dos modelos.

c) Suponiendo que la varianza de los errores es la misma en los dos modelos, contraste si puede admitirse que los parámetros de estos modelos son iguales a los del otro.

Pregunta 3 (Propuesto)

Se dispone de 17 datos de la economía de la India correspondientes a las siguientes variables:

- M_t: Agregado monetario (nominal), en millones de rupias.
- Y_t: Renta nacional (nominal), en cientos de millones de rupias
- P_t: Indicador del nivel de precios.
- R_t: Tipo de interés a largo plazo, en tanto por ciento.

Para explicar la demanda de dinero en la India se consideran los siguientes modelos:

$$
\begin{aligned}
&\text{Modelo 1:} \quad & M_t^* &= \beta_0 \, Y_t^{*\beta_1} \, R_t^{\beta_2} \, \exp\{u_t\}. \\
&\text{Modelo 2:} \quad & \tfrac{M_t^*}{Y_t^*} &= \alpha \, R_t^{\beta} \, \exp\{u_t\}. \\
&\text{Modelo 3:} \quad & M_t &= \alpha_0 \, Y_t^{*\alpha_1} \, R_t^{\alpha_2} \, P_t^{\alpha_3} \, \exp\{u_t\}. \\
&\text{Modelo 4:} \quad & M_t &= \lambda_0 \, Y_t^{\lambda_1} \, R_t^{\lambda_2} \, P_t^{\lambda_3} \, \exp\{u_t\}.
\end{aligned}
$$

Las variables sin asterisco son variables nominales y las variables con asterisco son reales, obtenidas dividiendo las variables nominales por el indicador del nivel de precios. Se sabe que las varianzas muestrales de las variables $\ln M$, $\ln M^*$ y $\ln \frac{M^*}{Y^*}$ son, respectivamente, 0.0682, 0.03976 y 0.005015, (en el calculo de estas varianzas muestrales se utilizó $T = 17$ como el denominador). Se conocen también las estimaciones MCO siguientes:

$$
\begin{aligned}
&\text{i)} \quad & \ln M_t^* &= 3.64 - 1.49 \ln Y_t^* - 0.5 \ln R_t + \varepsilon_t, & R^2 &= 0.8942. \\
&\text{ii)} \quad & \ln \tfrac{M_t^*}{Y_t^*} &= 2.89 + 0.09 \ln R_t + \varepsilon_t, & R^2 &= 0.0283. \\
&\text{iii)} \quad & \ln M_t &= 5.45 + 1.70 \ln Y_t^* - 0.60 \ln R_t + 0.64 \ln P_t + \varepsilon_t, & R^2 &= 0.9444. \\
&\text{iv)} \quad & \ln M_t &= 5.45 + 1.70 \ln Y_t - 0.60 \ln R_t + 1.06 \ln P_t + \varepsilon_t, & R^2 &= 0.9444.
\end{aligned}
$$

a) Suponiendo que el modelo estimado en la ecuación (i) satisface las hipótesis del MLG con normalidad, contraste si puede admitirse como válido el modelo 2.

b) Suponiendo que el modelo estimado en la ecuación (iii) satisface las hipótesis del MLG con normalidad, contraste:

- La validez del Modelo 2.
- La validez del Modelo 1.

c) ¿Qué relación hay entre los modelos 2 y 4? ¿Qué relación hay entre las estimaciones obtenidas en las ecuaciones (iii) y (iv)?

Solución

Pregunta 1

Se reescriben los resultados estadísticos de los modelos expuestos en esta pregunta. Supongamos que el modelo estimado viene dado por:

$$Y_t \;=\; \beta_1 + \beta_2\, X_{2t} + \beta_3\, X_{3t} + \beta_4\, X_{4t} + u_t, \qquad t = 1, \ldots, T, \tag{1}$$

donde se tiene el resumen estadístico de este modelo:

Coeficiente	Estimación	Estadístico t	R^2
β_1	2.0	-	0.982
β_2	3.5	1.9	
β_3	-0.7	-2.2	
β_4	2.1	1.5	

Modelo con restricción $\beta_2 = \beta_4$:

$$Y_t \;=\; \alpha_1 + \alpha_2\,(X_{2t} + X_{4t}) + \alpha_3\, X_{3t} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{2}$$

con el resumen estadístico de este modelo:

Coeficiente	Estimación	Estadístico t	R^2
α_1	1.5	-	0.876
α_2	3.1	2.7	
α_3	-0.6	-2.4	

a) Significatividad del Modelo Estimado (1).

Consiste en determinar si las variables que han sido consideradas en el modelo, son estadísticamente significativas para explicar Y_t, i.e., el test de hipótesis se realiza para los coeficientes del modelo, excluyendo el intercepto.

(i) Hipótesis:

$$\mathrm{H}_0 : \beta_2 \;=\; \beta_3 \;=\; \beta_4 \;=\; 0 \qquad \text{v/s} \qquad \mathrm{H}_1 : \beta_j \;\neq\; 0, \quad \text{para algún } j = 2, 3, 4.$$

(ii) Estadístico:

$$F = \frac{\dfrac{R^2 - R_r^2}{q}}{\dfrac{1 - R^2}{T - k}} \sim \mathrm{F}(q, T - k),$$

donde:

q : Número de restricciones lineales.

T : Número de observaciones.

k : Número de parámetros del modelo (incluyendo intercepto).

Bajo H_0, se tiene que $R_r^2 = 0$, entonces el estadístico observado para esta prueba es:

$$F_{obs} = \frac{\dfrac{R^2}{q}}{\dfrac{1 - R^2}{T - k}} = \frac{\dfrac{0.982}{3}}{\dfrac{1 - 0.982}{26 - 4}} = 400.07.$$

(iii) Región de Rechazo.

Se tiene que:

$$\mathrm{F}(q, T - k) = \mathrm{F}(3, 22) \Rightarrow \mathrm{F}_{0.95}(3, 22) = 3.05,$$

donde se está utilizando una significancia de $\alpha = 0.05$. Luego, la región de rechazo es:

$$\mathrm{RC}(\mathrm{H}_0) = \{F : F > 3.05\},$$

donde se consigue que, $F_{obs} \in \mathrm{RC}(\mathrm{H}_0)$.

(iv) Conclusión:

Existe evidencia estadística para rechazar H_0, i.e., el modelo posee al menos un parámetro significativo, con un 95 % de confianza.

b) Contrastar $\beta_2 = \beta_4$, frente a una alternativa bilateral.

(i) Hipótesis:

$$\mathrm{H}_0 : \beta_2 = \beta_4 \quad \text{v/s} \quad \mathrm{H}_1 : \beta_2 \neq \beta_4$$

(ii) Estadístico:

$$F \;=\; \frac{\dfrac{R^2 - R_r^2}{q}}{\dfrac{1-R^2}{T-k}} \;\sim\; \mathrm{F}(q, T-k),$$

donde bajo H_0 se tiene que $R_r^2 = 0.376$, dado por el modelo (2).

Obs: Una restricción lineal ($q = 1$) es equivalente a un test t-student de 2 colas, ya que:

$$\mathrm{F}(1,\nu) \;=\; \big(t(\nu)\big)^2 \quad\Leftrightarrow\quad \pm\sqrt{\mathrm{F}(1,\nu)} \;=\; t(\nu).$$

Entonces el estadístico observado a utilizar en esta prueba es:

$$F_{obs} \;=\; \frac{\dfrac{R^2 - R_r^2}{q}}{\dfrac{1-R^2}{T-k}} \;=\; \frac{\dfrac{0.982-0.876}{1}}{\dfrac{1-0.982}{26-4}} \;=\; 129.5 \quad\Rightarrow\quad t_{obs} \;=\; \sqrt{F_{obs}} \;=\; 11.38.$$

El resultado es el mismo si se considera la raíz positiva o negativa.

(iii) Región de Rechazo.

Ahora estamos ante una distribución t-student, definida por:

$$t(T-k) \;=\; t(22) \quad\Rightarrow\quad t_{0.975}(22) \;=\; 2.07.$$

Entonces, la región de rechazo es:

$$\mathrm{RC}(\mathrm{H}_0) \;=\; \{t:\; |t| \;>\; 2.07\},$$

consiguiéndose, $t_{obs} \in \mathrm{RC}(\mathrm{H}_0)$.

(iv) Conclusión:

Existe evidencia estadística para rechazar H_0, i.e., con un 95 % de confianza se puede concluir que $\beta_2 \neq \beta_4$.

Intervalo de Confianza

De manera general,

$$\mathrm{IC}\big(R\,\beta, 100(1-\alpha)\,\%\big) \;\equiv\; R\,\widehat{\beta} \pm t_{\alpha/2}(T-k)\,\sqrt{\widehat{\sigma}^2\, R(X'X)^{-1}R^{-1}}.$$

En este caso, se tiene que:

$$R \;=\; [0 \quad 1 \quad 0 \; -1],$$

entonces:

$$R\widehat{\beta} \;=\; [0 \quad 1 \quad 0 \; -1]\begin{bmatrix} 2.0 \\ 3.5 \\ -0.7 \\ 2.1 \end{bmatrix} \;=\; 1.4,$$

y suponiendo que se tiene la información necesaria para obtener:

$$\sqrt{\widehat{\sigma}^2\, R(X'X)^{-1}R^{-1}} \;=\; 0.1318,$$

el Intervalo de Confianza, para esta restricción queda como:

$$\begin{aligned} & \mathrm{IC}\big(\beta_2 - \beta_4, 95\,\%\big) && \equiv && 1.4 \pm 2.07 \cdot 0.1318 \\ \Rightarrow\quad & \mathrm{IC}\big(\beta_2 - \beta_4, 95\,\%\big) && = && (1.145; 1.6546). \end{aligned}$$

En el IC se tiene que el número cero no esta contenido en este intervalo, i.e.,

$$\begin{aligned} \Rightarrow &\quad 0 \;\notin\; \mathrm{IC}\big(\beta_2 - \beta_4, 95\,\%\big) \\ \Rightarrow &\quad \beta_2 - \beta_4 \;\neq\; 0 \\ \Rightarrow &\quad \beta_2 \;\neq\; \beta_4, \end{aligned}$$

con un 95 % de confianza. Luego, se concluye lo mismo que lo realizado con el test de hipótesis.

c) Si se elimina una variable en el modelo siendo significativa o no, entonces se tiene que el R^2 disminuiría. No obstante, el $\bar{R}^2$ es posible que aumente o disminuya.

Pregunta 2

En primera instancia, Sean,

C_t: Consumo Agregado.

P_t: PNB.

Se tienen 20 observaciones correspondientes a datos anuales, entre 1963-1982.

Se consideran las siguientes variables binarias:

$$D_{1t} \;=\; \begin{cases} 1, & 1963 \leq t \leq 1970; \\ 0, & 1970 < t \leq 1982. \end{cases}$$

$$D_{2t} \;=\; \begin{cases} 1, & 1970 < t \leq 1982; \\ 0, & 1963 \leq t \leq 1970. \end{cases}$$

En esta pregunta, denotaremos como β_j^i al coeficiente j del modelo $\mathcal{M}_i$, con $i = 1, \ldots, 6$. Cada coeficiente vendrá dado por el orden presentado en cada modelo. Por ejemplo, el modelo $\mathcal{M}_2$ quedaría como:

$$C_t \;=\; \beta_1^2\, D_{1t} + \beta_2^2\, D_{2t} + \beta_3^2\,(D_{1t}\, P_t) + \beta_4^2\,(D_{2t}\, P_t) + \varepsilon_t.$$

a) Contrastar si el coeficiente de intercepto del modelo $\mathcal{M}_A$ y $\mathcal{M}_B$ son iguales.
Hasta el momento, hemos visto restricciones aplicadas a un solo modelo. Aquí se tienen hipótesis que están comparando parámetros de distintos modelos. La idea es expresar esta hipótesis sobre algún modelo relevante. De los seis modelos expuestos, se puede apreciar que el modelo $\mathcal{M}_2$ es aquel más completo, considera todo el intervalo de tiempo, y hace la diferencia de sus parámetros en los dos intervalos de tiempo, además se puede observar que $\mathcal{M}_2 = \mathcal{M}_A + \mathcal{M}_B$.

(i) Hipótesis:

$$\mathrm{H}_0 : \beta_1^A \;=\; \beta_1^B \qquad \text{v/s} \qquad \mathrm{H}_1 : \beta_1^A \;\neq\; \beta_1^B$$

Equivalente a:

$$\mathrm{H}_0 : \beta_1^2 \;=\; \beta_2^2 \qquad \text{v/s} \qquad \mathrm{H}_1 : \beta_1^2 \;\neq\; \beta_2^2$$

(ii) Estadístico:

El estadístico R también puede ser escrito utilizando las SCR, como:

$$F \;=\; \frac{\dfrac{\widehat{\varepsilon}_r'\widehat{\varepsilon}_r - \widehat{\varepsilon}'\widehat{\varepsilon}}{q}}{\dfrac{\widehat{\varepsilon}'\widehat{\varepsilon}}{T-k}} \quad\sim\quad \mathrm{F}(q, T-k),$$

donde:

q : Número de restricciones lineales.

T : Número de observaciones.

k : Número de parámetros del modelo (incluyendo intercepto).

En este caso, se tiene que:

$\widehat{\varepsilon}_r'\widehat{\varepsilon}_r \;=\; SCR_3 \;=\; 1616.$

$\widehat{\varepsilon}'\widehat{\varepsilon} \;=\; SCR_7 + SCR_8 \;=\; SCR_2 \;=\; 1331.$

$q = 1; \quad T = 20; \quad k = 4.$

Así, el estadístico observado es:

$$F_{obs} \;=\; \frac{\dfrac{1616 - 1331}{1}}{\dfrac{1331}{20-4}} \;=\; 3.426.$$

(iii) Región de Rechazo.

Se tiene que:

$$\mathrm{F}(q, T-k) \;=\; \mathrm{F}(1,16) \quad \Rightarrow \quad \mathrm{F}_{0.95}(1,16) \;=\; 4.494,$$

con $\alpha = 0.05$. Luego, la región de rechazo es:

$$\mathrm{RC}(\mathrm{H}_0) \;=\; \{F : \; F \;>\; 4.494\},$$

donde se consigue que, $F_{obs} \notin \mathrm{RC}(\mathrm{H}_0)$.

(iv) Conclusión:

Existe evidencia estadística para no rechazar H_0, i.e., con un 95 % de confianza se puede concluir que los modelos $\mathcal{M}_A$ y $\mathcal{M}_B$ poseen el mismo término constante.

b) Contrastar si la pendiente es la misma en estos dos modelos.

(i) Hipótesis:

$$\mathrm{H}_0 : \beta_2^A \;=\; \beta_2^B \qquad \text{v/s} \qquad \mathrm{H}_1 : \beta_2^A \;\neq\; \beta_2^B$$

Equivalente a:

$$\mathrm{H}_0 : \beta_3^2 \;=\; \beta_4^2 \qquad \text{v/s} \qquad \mathrm{H}_1 : \beta_3^2 \;\neq\; \beta_4^2$$

(ii) Estadístico:

$$F \;=\; \frac{\dfrac{\widehat{\varepsilon}_r'\widehat{\varepsilon}_r - \widehat{\varepsilon}'\widehat{\varepsilon}}{q}}{\dfrac{\widehat{\varepsilon}'\widehat{\varepsilon}}{T-k}} \;\sim\; \mathrm{F}(q, T-k),$$

donde:

$\widehat{\varepsilon}_r'\widehat{\varepsilon}_r \;=\; SCR_6 \;=\; 1625.$

$\widehat{\varepsilon}'\widehat{\varepsilon} \;=\; SCR_2 \;=\; 1331.$

$q = 1; \quad T = 20; \quad k = 4.$

Así, el estadístico observado es:

$$F_{obs} \;=\; \frac{\dfrac{1625 - 1331}{1}}{\dfrac{1331}{20-4}} \;=\; 3.5342.$$

(iii) Región de Rechazo.

$$\mathrm{RC}(\mathrm{H}_0) \;=\; \{F : \; F \;>\; 4.494\},$$

donde se consigue que, $F_{obs} \notin \mathrm{RC}(\mathrm{H}_0)$.

(iv) Conclusión:

No hay evidencia estadística para rechazar H_0, i.e., con un 95 % de confianza se puede concluir que los modelos $\mathcal{M}_A$ y $\mathcal{M}_B$ poseen la misma pendiente.

c) Contrastar si los parámetros del modelo $\mathcal{M}_A$ son iguales a los parámetros del modelo $\mathcal{M}_B$.

(i) Hipótesis:

$$\mathrm{H}_0 : \beta_1^A = \beta_1^B \;\wedge\; \beta_2^A = \beta_2^B \qquad \text{v/s} \qquad \mathrm{H}_1 : \beta_1^A \neq \beta_1^B \;\vee\; \beta_2^A \neq \beta_2^B$$

Equivalente a:

$$\mathrm{H}_0 : \beta_1^2 = \beta_2^2 \;\wedge\; \beta_3^2 = \beta_4^2 \qquad \text{v/s} \qquad \mathrm{H}_1 : \beta_1^2 \neq \beta_2^2 \;\vee\; \beta_3^2 \neq \beta_4^2$$

(ii) Estadístico:

$$F \;=\; \frac{\dfrac{\widehat{\varepsilon}_r'\widehat{\varepsilon}_r - \widehat{\varepsilon}'\widehat{\varepsilon}}{q}}{\dfrac{\widehat{\varepsilon}'\widehat{\varepsilon}}{T-k}} \;\sim\; \mathrm{F}(q, T-k),$$

donde:

$$\widehat{\varepsilon}_r'\widehat{\varepsilon}_r \;=\; SCR_1 \;=\; 1626.$$

$$\widehat{\varepsilon}'\widehat{\varepsilon} \;=\; SCR_2 \;=\; 1331.$$

$$q = 2; \quad T = 20; \quad k = 4.$$

Así, el estadístico observado es:

$$F_{obs} \;=\; \frac{\dfrac{1626 - 1331}{2}}{\dfrac{1331}{20 - 4}} \;=\; 1.7731.$$

(iii) Región de Rechazo.

Se tiene que:

$$\mathrm{F}(q, T - k) \;=\; \mathrm{F}(2, 16) \quad \Rightarrow \quad \mathrm{F}_{0.95}(2, 16) \;=\; 3.6337,$$

con $\alpha = 0.05$. Luego, la región de rechazo es:

$$\mathrm{RC}(\mathrm{H}_0) \;=\; \{F : \; F \;>\; 3.6337\},$$

donde se consigue que, $F_{obs} \notin \mathrm{RC}(\mathrm{H}_0)$.

(iv) Conclusión:

No hay evidencia estadística para rechazar H_0, i.e., con un 95 % de confianza se puede concluir que los modelos $\mathcal{M}_A$ y $\mathcal{M}_B$ poseen los mismos parámetros.