Tema 4

Contrastes de restricciones lineales y predicción

Contenido

4.1. Contrastes de restricciones lineales 78

4.2. Contrastes utilizando Gretl 80

4.3. Estimación bajo restricciones lineales 87

4.4. Estadísticos equivalentes 89

4.5. Predicción 91

4.1. Contrastes de restricciones lineales

En el Tema 3 hemos estudiado la forma más común de realizar los contrastes de significatividad individual y el contraste de significatividad conjunta sobre los coeficientes que acompañan a las variables explicativas en un modelo de regresión lineal general. Estos contrastes son los más habituales y en general cualquier programa econométrico, como también es el caso de Gretl, muestra por defecto los valores de los estadísticos correspondientes para contrastar estas restricciones en el mismo output de estimación.

En ocasiones, además de éstas, también podemos estar interesados en contrastar hipótesis que implican otro tipo de restricciones lineales en los coeficientes poblacionales del modelo. En general, podemos denotar la hipótesis nula y la alternativa como:

$$H_0 : \underset{(q \times K)}{R} \cdot \underset{(K \times 1)}{\beta} = \underset{(q \times 1)}{r}$$

$$H_a : R\beta \neq r$$

siendo q el número de restricciones bajo la hipótesis nula y K el número de parámetros en el modelo no restringido. La hipótesis alternativa implicaría que **al menos una** de las igualdades no se satisface[1].

Por ejemplo en el modelo sobre el precio de la vivienda que hemos visto ya en temas anteriores,

$$P_i = \beta_1 + \beta_2 F2_i + \beta_3 BEDRMS_i + \beta_4 BATHS_i + u_i \tag{4.1}$$

podemos expresar de esta forma los siguientes contrastes:

1. Contraste de significación individual de la variable $BEDRMS$: $H_0 : \beta_3 = 0$

$$H_0 : R\beta = r \quad \Rightarrow \quad \begin{bmatrix} 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = 0$$

2. Contraste de significación conjunta: $H_0 : \beta_2 = \beta_3 = \beta_4 = 0$

$$H_0 : R\beta = r \quad \Rightarrow \quad \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

3. Contraste de un subconjunto de coeficientes igual a cero, por ejemplo los que acompañan a las variables BEDRMS y BATHS: $H_0 : \beta_3 = \beta_4 = 0$

$$H_0 : R\beta = r \quad \Rightarrow \quad \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

[1] Cuidado que esto no es lo mismo que **todas** las igualdades **no** se satisfagan.

Podemos ilustrar el interés de contrastar otro tipo de restricciones lineales en el siguiente modelo para la inversión agregada de un país,

$$INVERR_t = \beta_1 + \beta_2\, t + \beta_3 PNBR_t + \beta_4 INTERES_t + \beta_5 INFLACION_t + u_t \qquad (4.2)$$

donde las variables implicadas son:

INVERR:	Inversión agregada,, en términos reales.
t :	Tiempo $t = 1, 2, \ldots, T$
PNBR:	Producto Nacional Bruto, en términos reales.
INTERES:	Tipo de Interés nominal.
INFLACION:	Tasa de Inflación.

Además de realizar los contrastes de significatividad individual y conjunta, podríamos estar interesados en contrastar las siguientes restricciones lineales:

1. $H_0 : \beta_3 = 1$, la propensión marginal a invertir es igual a 1, esto es, si aumenta el PNB real en una unidad, la inversión aumentará en la misma proporción, manteniendo el valor del resto de variables constante.

$$H_0 : R\beta = r \quad \Rightarrow \quad \begin{bmatrix} 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = 1$$

2. $H_0 : \beta_4 + \beta_5 = 0$, los inversores tienen en cuenta el tipo de interés real. Esto es, la inversión no variará si un aumento del tipo de interés nominal viene acompañado por un aumento de la misma magnitud de la tasa de inflación, manteniendo el resto de factores constantes.

$$H_0 : R\beta = r \quad \Rightarrow \quad \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} 0 \end{bmatrix}$$

3. $H_0 : \beta_2 = 0,\ \beta_3 = 1,\ \beta_4 + \beta_5 = 0$. Contraste conjunto de las dos restricciones anteriores además de la restricción de que la inversión en media no presenta una tendencia lineal.

$$H_0 : R\beta = r \quad \Rightarrow \quad \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \\ \beta_5 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$$

El siguiente estadístico, conocido como estadístico F de Wald, se puede utilizar para contrastar una o más restricciones lineales en el contexto de un MRLG. Esta forma de realizar el contraste solamente requiere estimar el modelo sin restringir.

Como ya hemos visto en el Tema 3, bajo las hipótesis básicas la distribución del estimador MCO del modelo sin restringir es: $\hat{\beta} \sim \mathcal{N}(\beta, \sigma^2(X'X)^{-1})$. Por lo tanto, dado que R es una matriz de constantes de rango q, se tiene que **bajo la hipótesis nula**:

$$\underset{(q\times 1)}{R\hat{\beta}} \sim \mathcal{N}(\underset{(q\times 1)}{r}, \underbrace{\sigma^2 R(X'X)^{-1}R'}_{(q\times q)}) \tag{4.3}$$

Utilizando este resultado y el estimador $\hat{\sigma}^2 = \frac{\hat{u}'\hat{u}}{T-K}$ del parámetro σ^2, tenemos que el estadístico de contraste y su distribución bajo la hipótesis nula es el siguiente:

$$F = \frac{(R\hat{\beta}-r)'[R(X'X)^{-1}R']^{-1}(R\hat{\beta}-r)/q}{\hat{u}'\hat{u}/(T-K)} \overset{H_0}{\sim} \mathcal{F}(q, T-K) \tag{4.4}$$

Si no es cierta la H_0, la diferencia $(R\hat{\beta}-r)$ será grande por lo que el estadístico F tomará valores grandes en ese caso. Rechazaremos la H_0 con un nivel de significatividad α si el valor muestral del estadístico es mayor que el valor crítico, $F > \mathcal{F}(q, T-K)_\alpha$, no rechazando H_0 en caso contrario.

4.2. Contrastes utilizando Gretl

En esta sección vamos a utilizar Gretl para contrastar las restricciones vistas en los ejemplos anteriores utilizando ese estadístico. En general, una vez que hemos leído los datos de las variables de interés la forma de proceder es la siguiente:

- Especificar y estimar por MCO el modelo sin imponer las restricciones o **el modelo no restringido** en *Modelo ⇒ Mínimos cuadrados ordinarios*

- En la ventana donde se muestran los resultados de la estimación del modelo no restringido, **gretl: modelo1** elegir *Contrastes ⇒ Restricciones lineales*

- Dentro de la ventana que aparece *gretl: restricciones lineales* podemos escribir las restricciones a contrastar.

  Cada restricción del conjunto de restricciones tiene que ir en una línea como una ecuación, donde a la izquierda del signo igual tiene que ir la combinación lineal de los parámetros y a la derecha el valor numérico correspondiente. Los parámetros en la restricción se denotan de la forma bJ donde J representa la posición en la lista de regresores comenzando por J= 1. Lo que nosotros hemos denotado en el MRLG como β_1, coeficiente que normalmente, aunque no necesariamente, acompaña a la constante, en Gretl se denomina b1, nuestro β_2 es b2, β_3 es b3 y así sucesivamente con todos los coeficientes del modelo.

En el **ejemplo del modelo para el precio de la vivienda**, que hemos utilizado en el Tema 3, vamos a contrastar la hipótesis de que conjuntamente variaciones en el número de habitaciones y el número de baños, manteniendo el tamaño de la vivienda constante, no influyen en el precio de la vivienda. Vamos a denotar los coeficientes como Gretl lo haría, suponiendo que al especificar el modelo mantenemos el mismo orden en el listado de variables explicativas

$$P_i = \beta_1 + \beta_2 F2_i + \beta_3 BEDRMS_i + \beta_4 BATHS_i + u_i \tag{4.5}$$

Una vez estimado el modelo con *Modelo ⇒ Mínimos cuadrados ordinarios*, en la ventana de resultados de la estimación *gretl:modelo1* seleccionamos con el cursor

Contrastes ⇒ Restricciones lineales

Aparecerá la ventana *gretl: restricciones lineales*. Dentro de la ventana escribimos

```
b3=0
b4=0
```

Al seleccionar *Aceptar* en esta ventana obtenemos los siguientes resultados:

```
Conjunto de restricciones
 1: b[BEDRMS] = 0
 2: b[BATHS] = 0

Estadístico de contraste:

F(2, 10) = 0,471106, con valor p = 0,637492

Estimaciones restringidas:

 VARIABLE        COEFICIENTE       DESV.TÍP.         ESTAD T   VALOR P
  const             52,3509           37,2855          1,404   0,18565
  F2                 0,138750          0,0187329       7,407  <0,00001 ***
  BEDRMS             0,000000          0,000000         indefinido
  BATHS              0,000000          0,000000         indefinido

  Desviación típica de los residuos = 39,023
```

No rechazamos la hipótesis nula al nivel de significación por ejemplo del 5 % ya que el valor $p = 0,637492 > 0,05$. Si miramos a las tablas de la distribución F con 2 y 10 grados de libertad, eligiendo en la ventana principal de Gretl

Herramientas → Tablas estadísticas → F con gln 2 y gld 10

obtenemos la siguiente información,

```
Valores críticos aproximados de F(2, 10)
 10% en la cola derecha 2,92
  5%              4,10
  1%              7,56
```

De igual forma vemos que, para los tres niveles de significación del 1, 5 y 10 % no se rechaza la hipótesis nula, ya que el valor muestral del estadístico es menor que el valor crítico correspondiente. Además también se muestran las estimaciones del modelo restringido bajo esas dos restricciones. Notar que los coeficientes que acompañan a BEDRMS y BATHS son igual a cero y sus desviaciones típicas también. La razón es que esos coeficientes no son estimaciones ya que toman un valor dado conocido.

Cuando las restricciones a contrastar son simplemente de exclusión de uno o más regresores del modelo de partida, otra forma de llevar a cabo este contraste en Gretl es elegir en el menú de la ventana de estimación del modelo de partida,

Contrastes ⇒ *Omitir variables*

Seguidamente en la ventana que surge, **gretl: contrastes del modelo**, se seleccionan las variables que acompañan a los coeficientes que bajo la hipótesis nula son cero. En el ejemplo en concreto que estamos viendo, sería elegir las variables BEDRMS y BATHS. Al pulsar *Aceptar* se muestra una nueva ventana con la estimación del modelo restringido bajo esas dos restricciones

$$P_i = \beta_1 + \beta_2 F2_i + u_i \tag{4.6}$$

que implican excluir de la regresión a BEDRMS y BATHS,

```
 Modelo Restringido: estimaciones MCO utilizando las 14 observaciones 1-14
 Variable dependiente: P

   VARIABLE       COEFICIENTE       DESV.TÍP.      ESTAD T     VALOR P
   const           52,3509          37,2855        1,404      0,18565
   F2               0,138750         0,0187329     7,407     <0,00001 ***

  Media de la var. dependiente = 317,493
  Desviación típica de la var. dependiente. = 88,4982
  Suma de cuadrados de los residuos = 18273,6
  Desviación típica de los residuos = 39,023
  R-cuadrado = 0,820522
  R-cuadrado corregido = 0,805565
  Grados de libertad = 12
  Log-verosimilitud = -70,0842
  Criterio de información de Akaike (AIC) = 144,168
  Criterio de información Bayesiano de Schwarz (BIC) = 145,447
  Criterio de Hannan-Quinn (HQC) = 144,05

Comparación entre el modelo restringido y no restringido:

  Hipótesis nula: los parámetros de regresión son cero para las variables
    BEDRMS
    BATHS

Estadístico de contraste:  F(2, 10) = 0,471106, con valor p = 0,637492
```

La ventaja de realizar de esta forma el contraste es que, además de tener la estimación del modelo restringido (4.6), en esta nueva ventana tenemos otra vez todos los menús que Gretl ofrece para el análisis de esta nueva especificación[2].

En esta ventana también se muestra el resultado del contraste, esto es, el valor muestral del estadístico F que contrasta esas dos restricciones de exclusión, y el *valor-p*. Como se puede observar, el resultado que se obtiene es exactamente el mismo que el que se ofrece en la ventana **gretl: restricciones lineales**.

Seguidamente vamos a utilizar el ejemplo del modelo de la Función de Inversión, para ilustrar otro tipo de restricciones lineales que no sean simplemente de exclusión.

Escribimos el modelo no restringido

$$INVERR_t = \beta_1 + \beta_2\, t + \beta_3\ PNBR_t + \beta_4 INTERES_t + \beta_5 INFLACION_t + u_t \tag{4.7}$$

y para el análisis usamos los datos de la siguiente Tabla[3]:

Año	PNB nominal	Inversión nominal	IPC	Tipo de Interés
1968	73,4	133,3	82,54	5,16
1969	944,0	149,3	86,79	5,87
1970	992,7	144,2	91,45	5,95
1971	1077,6	166,4	96,01	4,88
1972	1185,9	195,0	100,00	4,50
1973	1326,4	229,8	105,75	6,44
1974	1434,2	228,7	115,08	7,83
1975	1549,2	206,1	125,79	6,25
1976	1718,0	257,9	132,34	5,50
1977	1918,3	324,1	140,05	5,46
1978	2163,9	386,6	150,42	7,46
1979	2417,8	423,0	163,42	10,28
1980	2633,1	402,3	178,64	11,77
1981	2937,7	471,5	195,51	13,42
1982	3057,5	421,9	207,23	11,02

Tabla 4.1: Datos para el estudio de la Función de Inversión

Las series de Inversión y Producto Nacional Bruto en términos reales, INVERR y PNBR, se han obtenido de dividir las series nominales por el IPC con año base en 1972 y multiplicar por 10^{-1}, tal que están medidas en trillones de dólares. La tasa de inflación se ha calculado como el porcentaje de variación del IPC. Por lo tanto, los datos utilizados para estimar el modelo, son los de la siguiente tabla:

[2] El estimador restringido será $\hat{\beta}_R = \left[\ \hat{\beta}_{R,1}\ \ \hat{\beta}_{R,2}\ \ 0\ \ 0\ \right]'$ donde $\hat{\beta}_{R,1}$ y $\hat{\beta}_{R,2}$ son los obtenidos de la regresión excluyendo BEDRMS y BATHS.

[3] Corresponden a la Tabla F3.1 publicada en Greene (2008), p.1082 y disponible en: http://pages.stern.nyu.edu/~wgreene/Text/econometricanalysis.htm. Fuente: Economic Report of the President, U.S. Government Printing Office, Washington, D.C., 1983. El IPC de 1967 es 79,06. El tipo de interés es el promedio anual de la tasa de descuento del Banco de la Reserva Federal de Nueva York.

Año	INVERR	PNBR	INFLACION	INTERES
1968	0,161	1,058	4,40	5,16
1969	0,172	1,088	5,15	5,87
1970	0,158	1,086	5,37	5,95
1971	0,173	1,122	4,99	4,88
1972	0,195	1,186	4,16	4,50
1973	0,217	1,254	5,75	6,44
1974	0,199	1,246	8,82	7,83
1975	0,163	1,232	9,31	6,25
1976	0,195	1,298	5,21	5,50
1977	0,231	1,370	5,83	5,46
1978	0,257	1,439	7,40	7,46
1979	0,259	1,479	8,64	10,28
1980	0,225	1,474	9,31	11,77
1981	0,241	1,503	9,44	13,42
1982	0,204	1,475	5,99	11,02

Tabla 4.2: Datos en términos reales

Primeramente creamos el fichero de datos a partir de la tabla anterior incluyendo la variable $t = 1, \dots, 15$, con la opción de Gretl

Archivo → Nuevo conjunto de datos

Seguidamente estimamos por MCO el modelo no restringido arriba especificado, eligiendo en el menú *Modelo → Mínimos Cuadrados ordinarios* y obtenemos los siguientes resultados

Modelo 1: estimaciones MCO utilizando las 15 observaciones 1968–1982
Variable dependiente: INVERR

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−0,509071	0,0551277	−9,2344	0,0000
t	−0,0165804	0,00197176	−8,4089	0,0000
PNBR	0,670383	0,0549972	12,1894	0,0000
INTERES	−0,00232593	0,00121887	−1,9083	0,0854
INFLACION	−9,40107e-05	0,00134748	−0,0698	0,9458

Media de la var. dependiente	0,203333
D.T. de la variable dependiente	0,0341774
Suma de cuadrados de los residuos	0,000450812
Desviación típica de los residuos ($\hat{\sigma}$)	0,00671425
R^2	0,972433
$\bar{R}^2$ corregido	0,961406
$F(4, 10)$	88,1883
Estadístico de Durbin–Watson	1,96364
Coef. de autocorr. de primer orden	−0,0981367
Criterio de información de Akaike	−103,62
Criterio de información Bayesiano de Schwarz	−100,07

Contrastes de restricciones lineales:

a) Contraste de que la propensión marginal a invertir es la unidad, $H_0 : \beta_3 = 1$, frente a la hipótesis alternativa de que es distinto de la unidad. En la ventana **gretl: modelo1** seleccionamos *Contrastes* → *Restricciones lineales* y en la ventana que surge escribimos `b3` $= 1$. Al aceptar se obtiene el siguiente resultado,

```
Restricción:
 b[PNBR] = 1

Estadístico de contraste:
F(1, 10) = 35,92, con valor p = 0,000133289

Estimaciones restringidas:
  VARIABLE      COEFICIENTE       DESV.TÍP.      ESTAD T      VALOR P
  const          -0,837112         0,0134320   -62,322    <0,00001 ***
  t              -0,0276707        0,00139136  -19,888    <0,00001 ***
  PNBR            1,00000          0,000000     indefinido
  INTERES        -0,00311914       0,00247563   -1,260     0,23377
  INFLACION      -0,000342359      0,00275183   -0,124     0,90323

Desviación típica de los residuos = 0,0137184
```

Se muestran también las estimaciones de los coeficientes del modelo restringido, donde se ha impuesto que el coeficiente que acompaña a PNBR es igual a la unidad. Como damos ese valor a β_3, no estamos estimando ese coeficiente, por lo tanto su desviación típica es cero y el estadístico t no está definido.

Dado que el *valor-p*, asociado al valor muestral del estadístico de contraste, es más pequeño que $0,01$ se rechaza la hipótesis nula al $1\,\%$ de significación.

b) Contraste de que la inversión real responde al tipo de interés real, $H_0 : \beta_4 + \beta_5 = 0$, frente a $H_a : \beta_4 + \beta_5 \neq 0$. De la misma forma que antes, en la ventana **gretl: modelo1** seleccionamos *Contrastes* → *Restricciones lineales*. En la nueva ventana que aparece escribimos `b4`+`b5`$= 0$. Al aceptar se obtiene el siguiente resultado

```
Restricción:
 b[INTERES] + b[INFLACION] = 0

Estadístico de contraste:
  F(1, 10) = 3,25354, con valor p = 0,10143

Estimaciones restringidas:
  VARIABLE     COEFICIENTE       DESV.TÍP.      ESTAD T      VALOR P
  const         -0,505855        0,0604801      -8,364     <0,00001 ***
  t             -0,0170255       0,00214732     -7,929     <0,00001 ***
  PNBR           0,657533        0,0598599      10,985     <0,00001 ***
  INTERES       -0,00133784      0,00119517     -1,119      0,28683
  INFLACION      0,00133784      0,00119517      1,119      0,28683
```

De nuevo se muestran las estimaciones del modelo restringido. En este caso se estiman todos los coeficientes bajo la restricción de que $\beta_4 = -\beta_5$. El coeficiente estimado que acompaña a INTERES es el mismo valor pero con signo contrario que el obtenido para el coeficiente de INFLACION. Este resultado surge de la restricción impuesta ($\beta_4 = -\beta_5$). De igual forma coinciden las varianzas estimadas y las desviaciones típicas.

Dado que el *valor-p*, asociado al valor muestral del estadístico de contraste, es mayor que $0,1$ no se rechaza la hipótesis nula al 10 % (ni al 5 % o 1 %) de significación.

c) Por último, realizamos el contraste conjunto de estas dos restricciones lineales, la propensión marginal a invertir es la unidad y la inversión real responde al tipo de interés real. Esto es $H_0 : \beta_3 = 1,\ \beta_4 + \beta_5 = 0$ frente a la alternativa de que al menos una de ellas no se satisface, $H_a : \beta_3 \neq 1, \text{y} \setminus \text{o}\, \beta_4 + \beta_5 \neq 0$.

De nuevo, en la ventana **gretl: modelo1** seleccionamos

Contrastes $\rightarrow$ *Restricciones lineales*

y escribimos

```
b3=1
b4+b5=0
```

Al aceptar se obtiene el siguiente resultado:

```
Conjunto de restricciones
 1: b[PNBR] = 1
 2: b[INTERES] + b[INFLACION] = 0

Estadístico de contraste:
   F(2, 10) = 21,3453, con valor p = 0,000246226

Estimaciones restringidas:
 VARIABLE       COEFICIENTE      DESV.TÍP.      ESTAD T     VALOR P
  const           -0,851039       0,00799803    -106,406   <0,00001 ***
  t            -0,0289471      0,000989688   -29,249    <0,00001 ***
  PNBR              1,00000       0,000000       indefinido
  INTERES         -0,00172664     0,00227790     -0,758    0,46308
  INFLACION        0,00172664     0,00227790      0,758    0,46308

  Desviación típica de los residuos = 0,0140693
```

Se rechaza la hipótesis nula al 1 % de significación, ya que el *valor-p* es menor que 0,01. Por lo tanto, al menos una de las restricciones parece no satisfacerse. Viendo los resultados de los contrastes individuales, parece que la evidencia es contra la primera restricción.

4.3. Estimación bajo restricciones lineales

El estimador resultante de minimizar la suma de los residuos al cuadrado sujeto a restricciones lineales del tipo $R\beta = r$, esto es

$$\min_{\hat{\beta}_R} \sum_{i=1}^{N} (Y_i - \hat{\beta}_{R,1} - \hat{\beta}_{R,2} X_{2i} - \hat{\beta}_{R,3} X_{3i} - \cdots - \hat{\beta}_{R,K} X_{Ki})^2$$

$$\texttt{sujeto a} \qquad R\hat{\beta}_R = r$$

se puede expresar como:

$$\hat{\beta}_R = \hat{\beta} - (X'X)^{-1} R' \big[R(X'X)^{-1} R' \big]^{-1} (R\hat{\beta} - r) \tag{4.8}$$

donde $\hat{\beta} = (X'X)^{-1} X'Y$ es el estimador de los parámetros β sin imponer las restricciones. Dado que el estimador no restringido $\hat{\beta}$ se ha obtenido sin imponer que éste satisfaga tales restricciones, en general $(R\hat{\beta} - r) \neq 0$. La solución restringida, $\hat{\beta}_R$, es igual a la solución no restringida, $\hat{\beta}$, menos un término de ajuste que tiene en cuenta en qué medida la solución no restringida no satisface las restricciones. Si hemos obtenido ya $\hat{\beta}$ podemos utilizar directamente la expresión (4.8) para obtener el estimador de β restringido, es decir $\hat{\beta}_R$.

Hemos visto en la sección anterior que el programa Gretl muestra las estimaciones del modelo restringido cuando se selecciona la opción de contrastar restricciones lineales, a la vez que el valor muestral del estadístico de contraste.

Otra posibilidad es la de estimar el modelo imponiendo la o las restricciones. Cuando las restricciones implican solamente la exclusión de variables explicativas del modelo de partida, no hay mayor problema en llevar a cabo la estimación del modelo restringido. Bien se realiza la regresión eliminando del listado de regresores esas variables o, como hemos visto antes en Gretl, se puede utilizar la opción *Contrastes* $\Rightarrow$ *Omitir variables* a la vez que se contrasta.

Si las restricciones no son simplemente de exclusión, entonces se pueden sustituir en el modelo de partida y reorganizarlo en función del conjunto de $(K - q)$ parámetros que quedan sin determinar. Una ventaja de proceder así es que se dispone de las mismas opciones que en la ventana de estimación de un modelo por mínimos cuadrados ordinarios. Por ejemplo, se pueden hacer otro tipo de contrastes en el modelo restringido, guardar sus residuos, etc.

Por ejemplo, si queremos obtener **el estimador de los parámetros bajo la restricción de que la propensión marginal a invertir sea la unidad**, podemos hacerlo sustituyendo en el modelo

$$INVERR_t = \beta_1 + \beta_2\, t + \beta_3\, PNBR_t + \beta_4\, INTERES_t + \beta_5\, INFLACION_t + u_t \tag{4.9}$$

la restricción $\beta_3 = 1$ y reorganizar tal que nos quedaría la siguiente regresión:

$$INVERR_t - PNBR_t = \beta_1 + \beta_2\, t + \beta_4 INTERES_t + \beta_5 INFLACION_t + u_t \tag{4.10}$$

en función de $K - q = 5 - 1 = 4$ parámetros a estimar. El quinto ya está determinado por la restricción. Definimos una nueva variable llamémosla R, calculada como $R_t = INVERR_t - PNBR_t$, utilizando la opción en Gretl de

Variable → *Definir nueva variable*

y en la ventana que aparece escribimos R = INVERR-PNBR. De esta forma se añade la variable R al conjunto de variables disponibles que aparecen en la ventana principal o de inicio. Seguidamente, se realiza la regresión de esta variable sobre la constante, t, INTERES e INFLACION con *Modelo* → *Mínimos cuadrados ordinarios* y se obtienen los siguientes resultados:

Modelo Restringido (4.10): estimaciones MCO utilizando las 15 observaciones 1968–1982
Variable dependiente: R

Variable	Coeficiente	Desv. típica	Estadístico t	valor p
const	−0,837112	0,0134320	−62,3223	0,0000
t	−0,0276707	0,00139136	−19,8875	0,0000
INTERES	−0,00311914	0,00247563	−1,2599	0,2338
INFLACION	−0,000342359	0,00275183	−0,1244	0,9032

Media de la var. dependiente	−1,0840
D.T. de la variable dependiente	0,131901
Suma de cuadrados de los residuos	0,00207013
Desviación típica de los residuos ($\hat{\sigma}$)	0,0137184
R^2	0,991501
$\bar{R}^2$ corregido	0,989183
$F(3, 11)$	427,751
Estadístico de Durbin–Watson	0,995558
Coef. de autocorr. de primer orden.	0,441936
Log-verosimilitud	45,3774
Criterio de información de Akaike	−82,754
Criterio de información Bayesiano de Schwarz	−79,922
Criterio de Hannan–Quinn	−82,784

Recordamos lo que se obtenía al realizar el contraste de esa restricción en la ventana de estimación por MCO del modelo no restringido mediante *Contrastes* → *Restricciones Lineales*:

```
Restricción: b[PNBR] = 1

Estadístico de contraste: F(1, 10) = 35,92, con valor p = 0,000133289

Estimaciones restringidas:
   VARIABLE       COEFICIENTE       DESV.TÍP.    ESTAD T    VALOR P
    const          -0,837112         0,0134320   -62,322    <0,00001 ***
    t           -0,0276707        0,00139136  -19,888    <0,00001 ***
    PNBR            1,00000         0,000000        indefinido
    INTERES        -0,00311914       0,00247563   -1,260     0,23377
    INFLACION      -0,000342359      0,00275183   -0,124     0,90323
  Desviación típica de los residuos = 0,0137184
```

Los coeficientes estimados corresponden a las realizaciones del estimador de Mínimos Cuadrados Restringidos para los cuatro coeficientes que quedaban sin determinar por la restricción[4]. El valor para el coeficiente de PNBR viene dado por la restricción y es igual a la unidad. Su varianza por lo tanto es igual a cero ya que su valor está dado.

Hay que notar que el R^2, y por lo tanto el corregido, obtenidos en este ajuste no son comparables con los resultantes de estimar el modelo no restringido, ya que en este caso la Suma de Cuadrados Total corresponde a la variable $R = INVERR - PNBR$ que es el regresando de esta regresión y no a $INVERR$ que es realmente la variable endógena de interés a explicar. Para que los R^2 sean comparables entre el modelo no restringido y el restringido la Suma de Cuadrados Total tiene que ser la misma. Veremos en la sección siguiente los que sí son comparables y un estadístico de contraste basado en ellos.

4.4. Estadísticos equivalentes

Partimos del modelo $Y = X\beta + u$ donde se quiere contrastar las restricciones lineales $H_0 : R\beta = r$. Podemos obtener la suma de los residuos al cuadrado y el coeficiente de determinación correspondientes a la estimación del modelo sin restringir y al modelo restringido, de la siguiente forma:

$$SCR_{NR} = \hat{u}'\hat{u} = (Y - X\hat{\beta})'(Y - X\hat{\beta}) \qquad R^2_{NR} = 1 - \frac{\hat{u}'\hat{u}}{\sum(Y_t - \overline{Y})^2}$$

$$SCR_R = \hat{u}'_R\hat{u}_R = (Y - X\hat{\beta}_R)'(Y - X\hat{\beta}_R) \qquad R^2_R = 1 - \frac{\hat{u}'_R\hat{u}_R}{\sum(Y_t - \overline{Y})^2}$$

Por otra parte, utilizando las sumas de cuadrados de los residuos correspondientes a la estimación del modelo restringido y no restringido, SCR_R y SCR_{NR} respectivamente y sus grados de libertad, gl_R y gl_{NR}, es posible realizar el contraste de las restricciones lineales con el siguiente estadístico:

$$F = \frac{(SCR_R - SCR_{NR})/q}{SCR_{NR}/(T-K)} \overset{H_0}{\sim} \mathcal{F}(q, T-K) \tag{4.11}$$

Nótese que los grados de libertad de la distribución del estadístico bajo la hipótesis nula son en el numerador $gl_R - gl_{NR} = (T - (K - q)) - (T - K) = q$, el número de restricciones, y en el denominador $gl_{NR} = T - K$. Se puede demostrar que este estadístico es el mismo que el estadístico anterior (4.4). La diferencia radica en que calcularlo de esta forma requiere estimar tanto el modelo sin restringir como el restringido.

Su interpretación puede ser más intuitiva. Imponer restricciones en la estimación siempre empeora el ajuste tal que la diferencia de las sumas de cuadrados residuales del modelo restringido y no restringido, $(SCR_R - SCR_{NR})$, es mayor o igual a cero. Ahora bien, cuanto más grande sea esta diferencia más evidencia habrá de que las restricciones no sean ciertas, es decir contra la hipótesis nula. Se rechazará esta hipótesis nula si el valor muestral del estadístico es suficientemente grande como para caer en una región crítica establecida.

[4]El estimador restringido será $\hat{\beta}_R = \begin{bmatrix} \hat{\beta}_{R,1} & \hat{\beta}_{R,2} & 1 & \hat{\beta}_{R,4} & \hat{\beta}_{R,5} \end{bmatrix}'$ donde $\hat{\beta}_{R,1}$, $\hat{\beta}_{R,2}$, $\hat{\beta}_{R,4}$ y $\hat{\beta}_{R,5}$, son los obtenidos de la regresión bajo la restricción de que el coeficiente que acompaña al PNBR en el modelo para la Inversión real es igual a 1.

Si dividimos numerador y denominador por la suma de cuadrados total $SCT = \sum_t (Y_t - \overline{Y})^2$ podemos expresar el estadístico en términos de los coeficientes de determinación[5] :

$$F = \frac{(R^2_{NR} - R^2_R)/q}{(1 - R^2_{NR})/(T-K)} \overset{H_0}{\sim} \mathcal{F}_{(q,T-K)} \tag{4.12}$$

El contraste se realizará del mismo modo que con los otros estadísticos equivalentes.

Vamos a ilustrar esta forma de realizar el contraste en el ejemplo del modelo para la inversión agregada. Para realizar el contraste de la restricción de que la propensión marginal a invertir es igual a la unidad, utilizamos las sumas de cuadrados residuales de la estimación del modelo restringido (4.10) y el modelo no restringido (4.9). Esto ya lo obtuvimos en la secciones anteriores. En la ventana donde hemos realizado la regresión en cada caso podemos guardar las sumas de cuadrados residuales y añadirlo a las variables ya definidas con *Guardar* → *Suma de cuadrados de lo residuos*. En concreto se obtienen las siguientes sumas de cuadrados residuales:

$$SCR_R = 0,00207013 \qquad\qquad SCR_{NR} = 0,000450812$$

Sustituyendo en el estadístico (4.11) obtenemos el siguiente valor muestral[6]:

$$F = \frac{(0,00207013 - 0,000450812)/(15-4) - (15-5)}{0,000450812/(15-5)} = 35,92$$

siendo este el mismo valor que obtuvimos anteriormente con el estadístico utilizando *Contrastes* → *Restricciones lineales*, y por lo tanto obtenemos la misma conclusión del contraste, se rechaza la hipótesis nula de que la propensión marginal a invertir sea la unidad.

A su vez, utilizando el dato que nos da Gretl de la Desviación típica para la variable dependiente $INVERR$, podemos obtener la Suma de Cuadrados Total como,

$$SCT = \sum (INVERR_t - \overline{INVERR})^2 = (15-1)(D.T.\,INVERR)^2 = 14(0,0341774)^2$$

obteniendo el valor $SCT = 0,016353325$. Por lo tanto la realización de R^2_R es en este caso,

$$R^2_R = 1 - \frac{\hat{u}'_R \hat{u}_R}{SCT} = 1 - (0,00207013/0,016353325) = 0,87341$$

que no coincide con el que muestra la regresión del modelo (4.10). Esta vez este valor sí es comparable con el valor obtenido para el coeficiente de determinación de estimar el modelo no restringido, $R^2_{NR} = 0,972433$. Se puede apreciar, como era de esperar, que el valor obtenido del R^2_R es menor que el del R^2_{NR}, el ajuste empeora al imponer la restricción. La cuestión es si esto es aceptable, con un nivel de confianza elegido, para aceptar la hipótesis nula como cierta o no.

[5] Este es el estadístico que se introdujo en el Tema 3. En ese tema se vió como caso particular el estadístico de significación conjunta

$$F = \frac{R^2/(K-1)}{(1-R^2)/(T-K)} = \frac{R^2}{(1-R^2)} \frac{(T-K)}{(K-1)} \overset{H_0}{\sim} \mathcal{F}(K-1, T-K)$$

En ese caso $R^2_R = 0$

[6] Se puede hacer el cálculo con Gretl utilizando *Datos* → *Definir nueva variable* y escribiendo la fórmula del estadístico en términos de los nombres asignados a las variables sumas de cuadrados residuales.

El valor del estadístico (4.12) para este caso es,

$$F = \frac{(R^2_{NR} - R^2_R)/q}{(1 - R^2_{NR})/(T-K)} = F = \frac{(0,972433 - 0,87341)/1}{(1 - 0,972433)/(15-5)} = 35,92$$

obteniendo de nuevo el mismo valor para el estadístico y la misma conclusión del contraste.

4.5. Predicción

Uno de los objetivos de la econometría consiste en predecir. Una vez estimado un modelo que se considera que recoge bien el comportamiento de una variable en función de otros factores o variables explicativas, se quiere determinar con cierta confianza el valor o intervalo de valores que puede tomar la variable dependiente, supuestos unos valores para esos factores.

Supongamos que se ha estimado el siguiente modelo[7]:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \ldots + \beta_K X_{Kt} + u_t$$

con una muestra de tamaño T, obteniendo la siguiente función de regresión muestral (FRM):

$$\hat{Y}_t = \hat{\beta}_1 + \hat{\beta}_2 X_{2t} + \ldots + \hat{\beta}_K X_{Kt}$$

Entonces, disponiendo de nuevas observaciones de las variables explicativas,

$$X'_p = \left[\begin{array}{cccc} 1 & X_{2p} & \ldots & X_{Kp} \end{array}\right] \qquad p \notin \{1, 2, \ldots, T\}$$

podemos utilizar el modelo estimado por MCO para predecir el valor que tomará la variable endógena en el periodo de predicción p. A este proceso se le llama predicción por punto, donde el valor estimado para la variable endógena Y en el periodo de predicción se obtiene sustituyendo estos valores de las variables exógenas en la FRM.

$$\hat{Y}_p = X'_p \hat{\beta}_{MCO}$$

Equivalentemente:

$$\hat{Y}_p = \hat{\beta}_1 + \hat{\beta}_2 X_{2p} + \ldots + \hat{\beta}_K X_{Kp}.$$

El error de predicción se define como $e_p = Y_p - \hat{Y}_p = -X'_p(\hat{\beta} - \beta) + u_p$. Para obtener la predicción por intervalo, nos basaremos en la distribución del error de predicción, ya que si u_p y $\hat{\beta}$ son variables aleatorias normales, el error de predicción también lo será:

$$e_p \sim \mathcal{N}(0, \sigma^2 (\, 1 + X'_p \left(X'X \right)^{-1} X_p \,))$$

Sin embargo, en general, σ^2 es desconocido por lo que utilizaremos su estimador insesgado propuesto en temas anteriores obteniendo el siguiente resultado:

$$\frac{e_p}{\hat{\sigma}\sqrt{1 + X'_p\,(\, X'X\,)^{-1}\, X_p}} \sim t_{(T-K)}$$

[7]En lo que sigue, como siempre, se satisfacen las hipótesis básicas tanto en el periodo de estimación como de predicción

A partir de este estadístico podemos obtener un intervalo con un nivel de confianza del $1-\alpha$ alrededor de la predicción por punto para la variable endógena en el momento p.

$$IC_{1-\alpha}(Y_p) = \left(\hat{Y}_p - t_{\frac{\alpha}{2}(T-K)}\,\hat{\sigma}_{e_p}\ ,\ \hat{Y}_p + t_{\frac{\alpha}{2}(T-K)}\,\hat{\sigma}_{e_p}\right)$$

donde $\hat{\sigma}^2_{e_p} = \hat{\sigma}^2(\,1 + X_p'\,(\,X'X\,)^{-1}\,X_p\,)$.

¿Cómo utilizar Gretl para predecir por punto y por intervalo?

Utilizaremos el ejemplo de los precios de las viviendas para analizar los pasos a seguir en el programa Gretl.

Uno de los modelos propuestos era

$$P_i = \beta_1 + \beta_2 F2_i + \beta_3 BEDRMS_i + \beta_4 BATHS_i + u_i$$

Supongamos que tenemos información de una nueva vivienda, por ejemplo, $F2 = 3200$, $BEDRMS = 5$ y $BATHS = 3$ y nos piden $P = 500$, en miles de euros, por ella. Mediante este modelo, podemos obtener una predicción del precio que tendría una vivienda con estas características y analizar si el precio solicitado es razonable o no.

Para ello, incorporamos los nuevos datos (X_p) a la base de datos mediante

Datos → Seleccionar todos

A continuación, pincharemos la opción

Datos → Añadir Observaciones

indicando el número de observaciones que queremos añadir, en este caso 1. En la fila correspondiente incluimos los valores de las variables explicativas en el periodo de predicción, en este caso la observación 15, incorporando cada observación en la casilla correspondiente. Si no incorporamos el valor para la variable P que es la que vamos a predecir, gretl nos mostrará un aviso (Atención: había observaciones perdidas). Podemos simplemente ignorarlo y darle a aceptar.

Posteriormente, estimaremos el modelo sin considerar esta nueva observación (recordar que inicialmente teníamos 14 observaciones en la muestra). Para ello, tenemos que especificar el rango muestral, es decir, en la opción

Muestra → Establecer rango

especificaremos del rango de observaciones de la muestra para estimar el modelo, en nuestro caso de la 1 a la 14 y elegimos *Aceptar*.

Tal y como explicamos en los temas anteriores, estimaremos el modelo por MCO y en la ventana de los resultados elegimos

Análisis → Predicciones

En la nueva ventana podemos determinar el dominio de predicción, es decir el *Inicio* y *Fin* que en este caso es en ambos la observación número 15, y también cuantas observaciones se quieren representar antes de la prediccion[8].

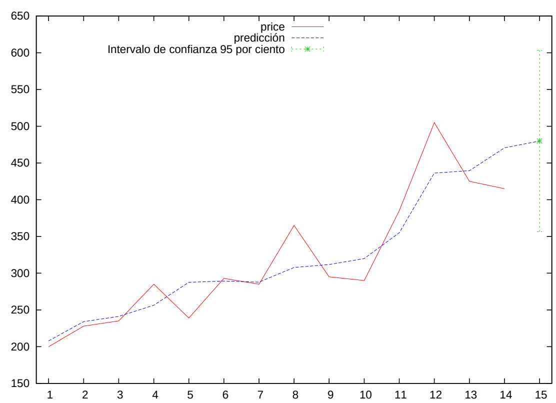
[8]En este caso hemos elegido todas pero esto es opcional.

Los resultados que muestra Gretl son los siguientes:

```
 Para intervalos de confianza 95%, t(10, .025) = 2,228

     Obs        price    predicción  desv. típica   Intervalo de confianza 95%

       1         199,9        207,8
       2         228,0        234,0
       3         235,0        241,2
       4         285,0        256,3
       5         239,0        287,6
       6         293,0        289,2
       7         285,0        287,8
       8         365,0        307,8
       9         295,0        311,8
      10         290,0        319,9
      11         385,0        355,1
      12         505,0        436,3
      13         425,0        439,6
      14         415,0        470,5
      15                      479,9        55,39        356,5 -    603,3
```



El gráfico que se obtiene junto a los resultados muestra la serie de precios (P) observada en color rojo y estimada con el modelo para las 14 observaciones anteriores a la predicción y la predicción en color azul, junto con su intervalo de confianza en color verde.

La predicción por punto del precio de una vivienda con estas características es de $479,905$ miles de euros, mientras que la predicción por intervalo con un nivel de confianza del $95\,\%$ es $(356,5; 603,3)$ en miles de euros, por lo que el precio que nos piden, que era de 500 miles de euros por la vivienda, está dentro del intervalo. Este precio para una vivienda de esas características se aceptaría como razonable dado nuestro modelo y la información muestral utilizada para su estimación, con un nivel de confianza del $95\,\%$.

Bibliografía

Greene, W. (2008), *Econometric Analysis*, 6^{a} edn., Prentice-Hall.