

FACULTAD DE CIENCIAS ECONÓMICAS UNC

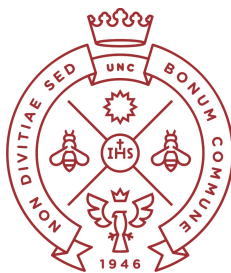
Resolución de la Guía de Ejercicios de Econometría II

Autores:

Joaquín Aguirre¹

Alan Jonás Alvarado²

Marzo 2022



FACULTAD
DE CIENCIAS
ECONÓMICAS

¹Legajo: 40.522.631

²Legajo: 41.886.044

Índice

Ejercicio I	3
Inciso a)	5
Inciso b)	5
Inciso c)	9
Inciso d)	10
Ejercicio II	20
Gráfico de la SCR	20
Inciso a)	24
Inciso b)	25
Inciso c)	28
Ejercicio III	30
Inciso a)	30
Inciso b)	32
Inciso c)	36
Inciso d)	42
Inciso e)	43
Inciso f)	45
Ejercicio IV	53
Inciso a)	53
Inciso b)	62
Inciso c)	70
Método de Newton-Raphson	71
Quadratic Hill-Climbing (Goldfeld-Quandt)	73
Berndt, Hall, Hall, and Hausman (BHHH)	74
Inciso d)	74
Inciso e)	76

Ejercicio V	82
Coincidencias y diferencias con Croissant y Millo (2008)	84
Ejercicio VI	86
Ejercicio VII	89
Inciso a)	89
Inciso b)	99
Inciso c)	102
Ejercicio VIII	109
Inciso a)	109
Inciso b)	111
Ítem I	111
Ítem II	112
Ítem III	113
Anexo	115
Demostración Slutsky	115
Estimaciones Ejercicio 1 inciso b)	115
Referencias	119

Ejercicio I

Resolver el ejercicio 11.3.7 de las páginas 467 y 468 del libro: Judge, George G., Hill, R. Carter, Griffiths, William E., Luetkepohl, Helmut and Lee, Tsoung Chao (1988), *Introduction to the Theory and Practice of Econometrics*, Second Edition, Wiley, New York, explicando breve, pero cuidadosamente, el sentido económico de las diferentes restricciones que se plantean en dicho libro de texto.

Respuesta:

El ejercicio plantea un sistema de 3 ecuaciones de demanda como el que sigue:

$$\ln q_{it} = \beta_{i0} + \beta_{i1} \ln p_{1t} + \beta_{i2} \ln p_{2t} + \beta_{i3} \ln p_{3t} + \beta_{i4} \ln y_t + e_{it}$$

$$i = 1, 2, 3$$

Donde los coeficientes están sujetos a las siguientes 3 restricciones:

Condición de Engel: $w_1\beta_{14} + w_2\beta_{24} + w_3\beta_{34} = 1$

Condiciones de Homogeneidad: $\beta_{i1} + \beta_{i2} + \beta_{i3} + \beta_{i4} = 0 \quad i = 1, 2, 3$

Condiciones de Simetría: $\frac{\beta_{ij}}{w_j} + \beta_{i4} = \frac{\beta_{ji}}{w_i} + \beta_{j4} \quad i, j = 1, 2, 3 \quad (i \neq j)$

Donde $w_i = \frac{p_i q_i}{y}$ ($i = 1, 2, 3$) es la cuota de presupuesto del bien i -ésimo. Empleando las medias muestrales de las cuotas de presupuesto:

$$\bar{w}_i = T^{-1} \sum_{t=1}^T \left(\frac{p_{it} q_{it}}{y_t} \right)$$

Se nos pide resolver 4 incisos. Antes de eso pasamos a explicar el sentido económico de cada una de las 3 restricciones que la teoría económica sugiere que los coeficientes deben satisfacer para que las ecuaciones de demanda estimadas sean consistentes con el problema de maximización de utilidad del agente. Recordemos que los coeficientes β_{i1} , β_{i2} y β_{i3} representan elasticidades precio de la demanda, mientras que β_{i4} representa la elasticidad ingreso de la demanda.

Condición de Engel:

$$w_1\beta_{14} + w_2\beta_{24} + w_3\beta_{34} = 1$$

Nos dice cómo debe responder la cantidad demanda ante una variación en el ingreso, *ceteris paribus*. Plantea que el promedio ponderado de elasticidades ingreso de la demanda debe ser igual a la unidad, siendo los ponderadores las cuotas de presupuesto de cada bien:

$\sum_i w_i \beta_{i4} = 1$. De esto se deriva que algunas elasticidades deben ser mayores a uno (bienes de lujo) y otras menores a 1 (bienes necesarios). En el primer caso, ante un aumento en el ingreso, el gasto en esos bienes aumenta más rápido en términos relativos, por lo que su cuota en el presupuesto total aumenta. En el segundo caso, el gasto en esos bienes aumenta más lentamente que el ingreso (y en el caso de los bienes con elasticidades negativas, llamados bienes inferiores, su cantidad demandada disminuye), por lo que su participación en el presupuesto disminuye.

Condiciones de Homogeneidad:

$$\beta_{i1} + \beta_{i2} + \beta_{i3} + \beta_{i4} = 0 \quad i = 1, 2, 3$$

Nos dice cómo debe responder la cantidad demandada ante un cambio de igual proporción tanto en el ingreso como en los precios (por ejemplo, ambos se multiplican por 2). El resultado sería que la demanda no cambia (de ahí que la suma de las elasticidades sea igual a cero), porque la restricción presupuestaria no se ve alterada. Simplemente se produce un cambio en los valores nominales que no modifica las decisiones de los agentes, es decir, no hay ilusión monetaria. Dicho de otra forma, las cantidades demandadas de los bienes dependen sólo de los precios relativos.

Condiciones de Simetría:

$$\frac{\beta_{ij}}{w_j} + \beta_{i4} = \frac{\beta_{ji}}{w_i} + \beta_{j4} \quad i, j = 1, 2, 3 \quad (i \neq j) \quad (1)$$

Esta ecuación representa las condiciones de simetría de Slutsky en forma de elasticidades. Expresa que el cambio en la demanda del bien i ante una modificación en el precio del bien j debe ser equivalente al cambio en la demanda del bien j ante un cambio en el precio del bien i , lo cual es posible gracias a que la matriz de Slutsky o de efectos sustitución V es simétrica:

$$V = \begin{bmatrix} \frac{\partial q_1}{\partial p_1} & \frac{\partial q_1}{\partial p_2} & \cdots & \frac{\partial q_1}{\partial p_n} \\ \frac{\partial q_2}{\partial p_1} & \ddots & & \vdots \\ \vdots & & & \\ \frac{\partial q_n}{\partial p_1} & \frac{\partial q_n}{\partial p_2} & \cdots & \frac{\partial q_n}{\partial p_n} \end{bmatrix}$$

Entonces:

$$\frac{\partial q_i}{\partial p_j} = \frac{\partial q_j}{\partial p_i} \quad (2)$$

Por otro lado, recordemos que la ecuación de Slutsky nos dice que las derivadas de las ecuaciones de demanda respecto al precio, o dicho de otra forma, que la respuesta de la cantidad demandada a cambios en los precios, se puede descomponer en un efecto sustitución (variación de la cantidad demandada ante un cambio en el precio relativo) y un efecto ingreso (variación en la cantidad demandada ante un cambio en el ingreso real del consumidor, como consecuencia del cambio en los precios). Entonces, aplicando la ecuación de Slutsky sobre (2) se llega a:

$$\frac{\partial q_i}{\partial p_j} + q_j \frac{\partial q_i}{\partial y} = \frac{\partial q_j}{\partial p_i} + q_i \frac{\partial q_j}{\partial y}$$

Que operando algebraicamente es equivalente a la ecuación (1). ([Demostración en Anexo](#)).

Inciso a)

Escribir las 7 restricciones en la forma $R\beta = r$

Las 7 restricciones son:

$$\begin{aligned} w_1\beta_{14} + w_2\beta_{24} + w_3\beta_{34} &= 1 \\ \beta_{11} + \beta_{12} + \beta_{13} + \beta_{14} &= 0 \\ \beta_{21} + \beta_{22} + \beta_{23} + \beta_{24} &= 0 \\ \beta_{31} + \beta_{32} + \beta_{33} + \beta_{34} &= 0 \\ \frac{\beta_{12}}{w_2} + \beta_{14} &= \frac{\beta_{21}}{w_1} + \beta_{24} \\ \frac{\beta_{13}}{w_3} + \beta_{14} &= \frac{\beta_{31}}{w_1} + \beta_{34} \\ \frac{\beta_{23}}{w_3} + \beta_{24} &= \frac{\beta_{32}}{w_2} + \beta_{34} \end{aligned}$$

Las últimas 3 ecuaciones se pueden reescribir como:

$$\begin{aligned} \frac{\beta_{12}}{w_2} + \beta_{14} - \frac{\beta_{21}}{w_1} - \beta_{24} &= 0 \\ \frac{\beta_{13}}{w_3} + \beta_{14} - \frac{\beta_{31}}{w_1} - \beta_{34} &= 0 \\ \frac{\beta_{23}}{w_3} + \beta_{24} - \frac{\beta_{32}}{w_2} - \beta_{34} &= 0 \end{aligned}$$

Entonces, en la forma $R\beta = r$:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & w_1 & 0 & 0 & 0 & 0 & w_2 & 0 & 0 & 0 & 0 & w_3 \\ 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & \frac{1}{w_2} & 0 & 1 & 0 & \frac{-1}{w_1} & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{w_3} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{w_1} & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{w_3} & 1 & 0 & 0 & \frac{-1}{w_2} & 0 & -1 \end{bmatrix} \begin{bmatrix} \beta_{10} \\ \beta_{11} \\ \beta_{12} \\ \beta_{13} \\ \beta_{14} \\ \beta_{20} \\ \beta_{21} \\ \beta_{22} \\ \beta_{23} \\ \beta_{24} \\ \beta_{30} \\ \beta_{31} \\ \beta_{32} \\ \beta_{33} \\ \beta_{34} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Inciso b)

Estimar las tres ecuaciones usando tanto mínimos cuadrados como el estimador SUR restringido.

Para hacer ambas estimaciones, primero cargamos los datos y definimos las ecuaciones:

```
library(readxl)
judge_11_3 <- read_excel(judge_path)
head(judge_11_3)
```

```
## # A tibble: 6 x 7
##       p1      p2      p3      y      q1      q2      q3
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1 10.8    4.47  6.63  488.  11.6   13.2   45.8
## 2 13.0    10.8  13.8  365.  12.0    2.18  13.4
## 3  9.24    5.86  4.06  541.   8.92   5.59  105.
## 4  4.61   14.0   3.87  760.  33.9    5.23  137.
## 5 13.0    11.4  14.9  422.   4.56  10.9   15.9
## 6  7.71    8.76  14.3  578.  17.6   11.9   23.7
```

```
demanda_1 <- log(q1) ~ log(p1) + log(p2) + log(p3) + log(y)
demanda_2 <- log(q2) ~ log(p1) + log(p2) + log(p3) + log(y)
demanda_3 <- log(q3) ~ log(p1) + log(p2) + log(p3) + log(y)
system <- list(d1 = demanda_1, d2 = demanda_2, d3 = demanda_3)
```

Luego definimos las restricciones (las matrices R y r de $R\beta = r$), para lo cual debemos calcular w_1 , w_2 y w_3 con anterioridad.

```
p1 <- judge_11_3[,1]
p2 <- judge_11_3[,2]
p3 <- judge_11_3[,3]
q1 <- judge_11_3[,5]
q2 <- judge_11_3[,6]
q3 <- judge_11_3[,7]
y <- judge_11_3[,4]
w1 <- (p1*q1)/y
w2 <- (p2*q2)/y
w3 <- (p3*q3)/y
```

El ejercicio pide que trabajemos con las medias muestrales de las cuotas de presupuesto, que se calculan a continuación:

```
w1_prom <- sum(w1)/30
w2_prom <- sum(w2)/30
w3_prom <- sum(w3)/30
```

Finalmente definimos las restricciones

```
Rdata <- c(rep(0,4),w1_prom,rep(0,4),w2_prom,rep(0,4),w3_prom,0,
           rep(1,4),rep(0,10),rep(0,6),rep(1,4),rep(0,5),rep(0,11),
           rep(1,4),0,0,1/w2_prom,0,1,0,-1/w1_prom,0,0,-1,rep(0,5),
           0,0,0,1/w3_prom,1,rep(0,6),-1/w1_prom,0,0,-1,rep(0,8),
           1/w3_prom,1,0,0,-1/w2_prom,0,-1)
R <- matrix(Rdata,nrow = 7,ncol = 15, byrow = T)
r <- matrix(c(1,rep(0,6)),nrow = 7, ncol = 1)
```

Ahora sí podemos hacer las estimaciones con el paquete `systemfit`. Primero por MCO, luego por SUR (sin las restricciones):

```
library(systemfit)

fitols <- systemfit(system ,
                    method = "OLS", data = judge_11_3)
summolS <- summary(fitols, residCov = FALSE, equations = FALSE)

fitsur <- systemfit(system ,
                    method = "SUR", data = judge_11_3)
summsur <- summary(fitsur, residCov = FALSE, equations = FALSE)
```

Finalmente por los métodos MCO y SUR pero aplicando la restricción $R\beta = r$:

```
fitolsR <- systemfit(system ,
                    method = "OLS", data = judge_11_3,
                    restrict.matrix = R,
                    restrict.rhs = r)
summolSR <- summary(fitolsR, residCov = FALSE, equations = FALSE)

fitsurR <- systemfit(system ,
                    method = "SUR", data = judge_11_3,
                    restrict.matrix = R,
                    restrict.rhs = r)
summsurR <- summary(fitsurR, residCov = FALSE, equations = FALSE)
```

En el [anexo](#) se encuentran los `summary` de cada una de estas estimaciones. En el Cuadro 1 se resumen los resultados.

Entre paréntesis se tienen los errores estándar asociados a los coeficientes estimados. Como se puede observar, los resultados de aplicar MCO son los mismos que los que se obtienen de emplear Mínimos Cuadrados Generalizados (MCG), que es el método efectuado cuando sospechamos de correlación entre los errores, siendo el vector $\hat{\beta}$ estimado resultante el siguiente:

$$\hat{\beta} = [X'(\Sigma^{-1} \otimes I)X]^{-1}X'(\Sigma^{-1} \otimes I)Y$$

Donde Σ es la matriz de varianzas y covarianzas. Siguiendo a Judge et al.(1988), esto se debe a que las variables explicativas son idénticas en todas las ecuaciones. En ese caso (y también cuando las covarianzas son nulas) es indistinto aplicar MCO o MCG.

Cuadro 1: Resultados MCO y SUR, restringido y sin restringir

Parámetro	MCO/SUR sin restringir	MCO restringido	SUR restringido
β_{10}	-4.885 (1.502)	-2.468 (0.666)	-2.783 (0.764)
β_{11}	-0.568 (0.219)	-0.804 (0.149)	-0.865 (0.169)
β_{12}	0.073 (0.168)	-0.094 (0.079)	-0.125 (0.098)
β_{13}	-0.203 (0.138)	-0.289 (0.109)	-0.278 (0.126)
β_{14}	1.429 (0.227)	1.187 (0.167)	1.267 (0.192)
β_{20}	-4.311 (1.757)	-1.528 (0.673)	-1.756 (0.879)
β_{21}	0.191 (0.257)	-0.076 (0.128)	-0.118 (0.158)
β_{22}	-0.632 (0.196)	-0.809 (0.132)	-0.836 (0.171)
β_{23}	0.135 (0.161)	0.027 (0.117)	0.037 (0.153)
β_{24}	1.141 (0.265)	0.858 (0.168)	0.917 (0.219)
β_{30}	1.017 (0.635)	-0.306 (0.252)	-0.164 (0.304)
β_{31}	-0.250 (0.093)	-0.045 (0.048)	-0.018 (0.056)
β_{32}	-0.144 (0.071)	-0.010 (0.031)	0.005 (0.041)
β_{33}	-0.965 (0.058)	-0.917 (0.041)	-0.923 (0.052)
β_{33}	0.879 (0.096)	0.972 (0.061)	0.935 (0.076)

Cuando comparamos los resultados de los modelos restringidos, los mismos derivan en estimaciones muy similares, sin lograrse percibir mayor eficiencia de alguno de ellos en base a los errores estándar. Aun así, en base a estos últimos sí queda claro que los estimadores resultantes de los modelos restringidos son más eficientes que los que surgen de los modelos sin restringir. Por otro lado, los dos modelos sin restringir (MCO y SUR) presentan un R^2 de McElroy más alto (0.982) que el de los modelos restringidos con los tres conjuntos de restricciones (0.978 y 0.979 respectivamente), pero sólo marginalmente:

```

mcelroys_r <- c(summols$mcelroy.r.squared,
               summsur$mcelroy.r.squared,
               summolsR$mcelroy.r.squared,
               summsurR$mcelroy.r.squared)
models_nom <- c('MCO sin restringir', 'SUR sin restringir',
               'MCO restringido', 'SUR restringido')
df_mcelroys <- data.frame('Modelo estimado'=models_nom,
                          'McElroy'=mcelroys_r)

library(knitr)
kable(df_mcelroys)

```

Modelo.estimado	McElroy
MCO sin restringir	0.9822944
SUR sin restringir	0.9822944
MCO restringido	0.9779708
SUR restringido	0.9785361

Inciso c)

Testear correlación entre los residuos de las ecuaciones (test for contemporaneous correlation)

En Judge et al.(1988) se sugiere aplicar el test propuesto por Breusch y Pagan para contrastar la existencia de correlación entre los residuos. La hipótesis nula sería que las covarianzas son todas nulas (inexistencia de correlación), mientras que la alternativa que al menos una covarianza es distinta de 0. El estadístico (el multiplicador de Lagrange) para 3 ecuaciones es el siguiente.

$$\lambda = T(r_{21}^2 + r_{31}^2 + r_{23}^2)$$

Donde $r_{ij}^2 = \hat{\sigma}_{ij}^2 / \hat{\sigma}_{ii} \hat{\sigma}_{jj}$ es la correlación al cuadrado y T el número de observaciones (30 en este ejemplo). Bajo hipótesis nula, λ se distribuye asintóticamente Chi-cuadrado con 3 grados de libertad ($\lambda \sim \chi_{(3)}^2$).

A partir de la estimación por MCO, se obtuvo la siguiente matriz de varianzas y covarianzas de los residuos:

```
summols$residCov
```

```

##           d1           d2           d3
## d1  0.15318243 -0.01563740 -0.04649085
## d2 -0.01563740  0.20945563 -0.03887287
## d3 -0.04649085 -0.03887287  0.02737039

```

A partir de la misma se obtuvieron los siguientes resultados:

$$r_{12}^2 = \frac{\hat{\sigma}_{12}^2}{\hat{\sigma}_{11}\hat{\sigma}_{22}} = \frac{(-0.016)^2}{0.153 \times 0.209} = 0.008$$

$$r_{13}^2 = \frac{\hat{\sigma}_{13}^2}{\hat{\sigma}_{11}\hat{\sigma}_{33}} = \frac{(-0.046)^2}{0.153 \times 0.027} = 0.512$$

$$r_{23}^2 = \frac{\hat{\sigma}_{23}^2}{\hat{\sigma}_{22}\hat{\sigma}_{33}} = \frac{(-0.039)^2}{0.209 \times 0.027} = 0.269$$

Por ende el estadístico queda:

$$\lambda = 30(0.008 + 0.512 + 0.269) = 23.67$$

Y el valor crítico:

```
qchisq(0.95,3)
```

```
## [1] 7.814728
```

El valor crítico para un nivel de confianza del 95 % con 3 grados de libertad es 7.81, por lo que el valor observado es mucho mayor (18.201). En consecuencia, se rechaza la hipótesis nula de inexistencia de correlación entre los residuos. Si este no fuera el caso, la estimación por MCO sería eficiente, pero como sí existe correlación contemporánea, corresponde emplear el método SUR.

Además, al observar la matriz de correlaciones entre los residuos de las ecuaciones que surge de la estimación SUR restringida, queda clara la existencia de una correlación positiva significativa entre los residuos de las ecuaciones 1 y 3 (-0.75) y entre los residuos de las ecuaciones 2 y 3 (-0.58).

```
summsurR$residCor
```

```
##           d1           d2           d3
## d1  1.00000000  0.04022589 -0.7516158
## d2  0.04022589  1.00000000 -0.5785523
## d3 -0.75161581 -0.57855231  1.0000000
```

Inciso d)

Testear la validez de las restricciones propuestas por la teoría de la demanda: (i) todas al mismo tiempo y (ii) cada grupo por separado

(i)

Se pasa a aplicar una d cima de k restricciones lineales:

$$H_0) R\beta = r$$

$$H_1) R\beta \neq r$$

Para ello empleamos el estad stico siguiente:

$$\frac{(SCR^R - SCR^{NR})/k}{SCR^{NR}/(n-p)} \sim F_{k,n-p}$$

Donde SCR^{NR} es la Suma de Cuadrados de los Residuos del modelo No Restringido (el SUR sin restringir), SCR^R es la Suma de Cuadrados de los Residuos del modelo Restringido (el SUR restringido), k es el n mero de restricciones (7), n el n mero de observaciones (90) y p el n mero de par metros del modelo (15). Las SCR surgen de las estimaciones efectuadas en el inciso **b)** (SSR en los **summary**). Entonces:

$$F_{obs} = \frac{(11.225 - 9.75)/7}{9.75/(90 - 15)} = 1.62$$

Mientras que el valor cr tico para una distribuci n F de *Snedecor* con 7 y 75 grados de libertad y un nivel de confianza del 95 % es:

```
qf(0.95, df1=7, df2=75)
```

```
## [1] 2.134314
```

Por lo que no se rechaza la hip tesis nula de que todas las restricciones juntas son v lidas al mismo tiempo. Tambi n se podr a haber usado el paquete **car**, que incluye la funci n **linearHypothesis()**, la cual nos permite hacer la d cima de restricciones lineales de la siguiente manera:

```
library(car)
lh1 <- linearHypothesis(fitsur, hypothesis.matrix = R, rhs = r)
lh1$F
```

```
## [1] NA 1.96989
```

```
lh1$`Pr(>F)`
```

```
## [1] NA 0.0704158
```

El p-value de la prueba F (**Pr>F**) es mayor al nivel de significaci n del 5 %, por lo que se concluye lo mismo, pero el F observado (**F**) es distinto. Esto puede deberse a que el estad stico usado por la funci n **linearHypothesis()** no sea exactamente el mismo al usado por nosotros previamente.

(ii)

Condición de Engel: $w_1\beta_{14} + w_2\beta_{24} + w_3\beta_{34} = 1$

En este caso hacemos una d cima para una  nica restricci n lineal, $r\beta = 1$, o m s precisamente:

$$\begin{bmatrix} 0 & 0 & 0 & 0 & \bar{w}_1 & 0 & 0 & 0 & 0 & \bar{w}_2 & 0 & 0 & 0 & 0 & \bar{w}_3 \end{bmatrix} \begin{bmatrix} \beta_{10} \\ \vdots \\ \beta_{34} \end{bmatrix} = 1$$

El estad stico a emplear es el mismo que en (i), pero ahora $k = 1$. Antes debemos estimar el modelo SUR restringido s lo con la Condici n de Engel, para contrastar la SCR resultante con la del modelo SUR sin restringir, que a lo largo de todo el inciso va a servir como nuestro modelo No Restringido.

```
fitsurEngel <- systemfit(system ,
                          method = "SUR", data = judge_11_3,
                          restrict.matrix = R[1,],
                          restrict.rhs = r[1,])

summary(fitsurEngel, residCov = F, equations = F)

##
## systemfit results
## method: SUR
##
##          N DF      SSR  detRCov   OLS-R2 McElroy-R2
## system 90 76 9.7596 0.000136 0.714059   0.981573
##
##      N DF      SSR      MSE      RMSE      R2  Adj R2
## d1 30 25 3.831448 0.153258 0.391482 0.655487 0.600365
## d2 30 25 5.236736 0.209469 0.457678 0.561784 0.491670
## d3 30 25 0.691415 0.027657 0.166303 0.937485 0.927483
##
##
## Coefficients:
##              Estimate Std. Error  t value  Pr(>|t|)
## d1_(Intercept) -5.0342606  1.4954632  -3.36636  0.0011972 **
## d1_log(p1)      -0.5687032  0.2195026  -2.59087  0.0114722 *
## d1_log(p2)       0.0731326  0.1679534   0.43543  0.6644819
## d1_log(p3)      -0.2027823  0.1379404  -1.47007  0.1456701
## d1_log(y)        1.4543094  0.2258430   6.43947 9.7876e-09 ***
## d2_(Intercept) -4.3745317  1.7558378  -2.49142  0.0149035 *
## d2_log(p1)       0.1902408  0.2566688   0.74119  0.4608613
## d2_log(p2)      -0.6323199  0.1963893  -3.21973  0.0018884 **
## d2_log(p3)       0.1353190  0.1612952   0.83895  0.4041272
```

```
## d2_log(y)      1.1514176  0.2654366   4.33782  4.3690e-05 ***
## d3_(Intercept) 1.3071902  0.5699809   2.29339  0.0245908 *
## d3_log(p1)     -0.2482203  0.0929983  -2.66908  0.0092947 **
## d3_log(p2)     -0.1446080  0.0711743  -2.03174  0.0456751 *
## d3_log(p3)     -0.9659309  0.0584523 -16.52511 < 2.22e-16 ***
## d3_log(y)      0.8305950  0.0834802   9.94960  2.2204e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La SCR de este modelo es 9,76. Entonces el estadístico queda:

$$F_{obs} = \frac{(SCR^R - SCR^{NR})/k}{SCR^{NR}/(n-p)} = \frac{(9.76 - 9.75)/1}{9.75/(90 - 15)} = 0.011$$

El valor crítico para una distribución F con 1 y 75 grados de libertad y un nivel de confianza del 95 % es:

```
qf(0.95, df1=1, df2=75)
```

```
## [1] 3.968471
```

Por lo que no rechazamos la validez de esta restricción.

Condiciones de Homogeneidad: $\beta_{i1} + \beta_{i2} + \beta_{i3} + \beta_{i4} = 0 \quad i = 1, 2, 3$

Requerimos hacer una dócima de $k = 3$ restricciones lineales. En este caso $R\beta = r$ queda:

$$\begin{bmatrix} 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \beta_{10} \\ \vdots \\ \beta_{34} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Ahora estimamos un nuevo modelo SUR restringido únicamente con las Condiciones de Homogeneidad:

```
fitsurHomog <- systemfit(system ,
                          method = "SUR", data = judge_11_3,
                          restrict.matrix = R[2:4,],
                          restrict.rhs = r[2:4,])

summary(fitsurHomog, residCov = F, equations=F)
```

```
##
## systemfit results
## method: SUR
##
##          N DF      SSR  detRCov  OLS-R2 McElroy-R2
## system  90  78 11.1422 0.000154 0.673551  0.981932
```

```
##
##      N DF      SSR      MSE      RMSE      R2      Adj R2
## d1 30 25 4.339067 0.166887 0.408518 0.609844 0.564826
## d2 30 25 5.899243 0.226894 0.476334 0.506345 0.449385
## d3 30 25 0.903888 0.034765 0.186454 0.918274 0.908844
##
##
## Coefficients:
##              Estimate Std. Error   t value   Pr(>|t|)
## d1_(Intercept) -2.5143096  0.7864658  -3.19697  0.0020065 **
## d1_log(p1)      -0.8225402  0.1766372  -4.65666  1.2978e-05 ***
## d1_log(p2)      -0.0842728  0.1504434  -0.56016  0.5769737
## d1_log(p3)      -0.2926415  0.1345545  -2.17489  0.0326683 *
## d1_log(y)        1.1994546  0.1973605   6.07748  4.2086e-08 ***
## d2_(Intercept) -1.6066718  0.9170220  -1.75205  0.0836948 .
## d2_log(p1)      -0.1000693  0.2059597  -0.48587  0.6284226
## d2_log(p2)      -0.8117235  0.1754176  -4.62738  1.4501e-05 ***
## d2_log(p3)       0.0330868  0.1568911   0.21089  0.8335231
## d2_log(y)        0.8787060  0.2301231   3.81842  0.0002679 ***
## d3_(Intercept) -0.5399140  0.3589541  -1.50413  0.1365862
## d3_log(p1)      -0.0829816  0.0806197  -1.02930  0.3065215
## d3_log(p2)      -0.0410225  0.0686645  -0.59743  0.5519483
## d3_log(p3)      -0.9064608  0.0614126 -14.76018 < 2.22e-16 ***
## d3_log(y)        1.0304649  0.0900781  11.43968 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La SCR obtenida aquí fue de 11,14. El estadístico queda:

$$F_{obs} = \frac{(SCR^R - SCR^{NR})/k}{SCR^{NR}/(n-p)} = \frac{(11.14 - 9.75)/3}{9.75/(90 - 15)} = 1.53$$

El valor crítico para una distribución F con 3 y 75 grados de libertad y un nivel de confianza del 95 % es:

```
qf(0.95, df1=3, df2=75)
```

```
## [1] 2.726589
```

Por lo que no rechazamos la hipótesis nula ($R\beta = r$) y validamos este conjunto de restricciones.

Condiciones de Simetría: $\frac{\beta_{ij}}{w_j} + \beta_{i4} = \frac{\beta_{ji}}{w_i} + \beta_{j4} \quad i, j = 1, 2, 3 \ (i \neq j)$

Nuevamente, hacemos una dócima de $k = 3$ restricciones lineales. Esta vez $R\beta = r$ se ve de la siguiente manera:

$$\begin{bmatrix} 0 & 0 & \frac{1}{w_2} & 0 & 1 & 0 & \frac{-1}{w_1} & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \frac{1}{w_3} & 1 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{-1}{w_1} & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{w_3} & 1 & 0 & 0 & \frac{-1}{w_2} & 0 & -1 \end{bmatrix} \begin{bmatrix} \beta_{10} \\ \vdots \\ \beta_{34} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Finalmente, estimamos un nuevo modelo SUR restringido únicamente con las Condiciones de Simetría:

```
fitsurSim <- systemfit(system ,
                        method = "SUR", data = judge_11_3,
                        restrict.matrix = R[5:7,],
                        restrict.rhs = r[5:7,])

summary(fitsurSim, residCov = F, equations=F)

##
## systemfit results
## method: SUR
##
##          N DF      SSR  detRCov   OLS-R2 McElroy-R2
## system 90 78 10.8318 0.000216 0.682644   0.976689
##
##      N DF      SSR      MSE      RMSE      R2  Adj R2
## d1 30 25 4.206166 0.168247 0.410179 0.621794 0.561281
## d2 30 25 5.660670 0.226427 0.475843 0.526309 0.450518
## d3 30 25 0.964993 0.038600 0.196468 0.912749 0.898789
##
##
## Coefficients:
##              Estimate   Std. Error   t value   Pr(>|t|)
## d1_(Intercept) -2.869430792  1.016063225  -2.82407 0.00601661 **
## d1_log(p1)      -0.789924888  0.181974507  -4.34085 4.2152e-05 ***
## d1_log(p2)      -0.068559814  0.104015596  -0.65913 0.51175346
## d1_log(p3)      -0.275727184  0.127780518  -2.15782 0.03401909 *
## d1_log(y)        1.237144890  0.200335223   6.17537 2.7792e-08 ***
## d2_(Intercept) -2.153298520  1.184569819  -1.81779 0.07293608 .
## d2_log(p1)      -0.041717492  0.164270874  -0.25396 0.80019819
## d2_log(p2)      -0.778834712  0.175624067  -4.43467 2.9837e-05 ***
## d2_log(p3)       0.054886318  0.152530534   0.35984 0.71994057
## d2_log(y)        0.932498013  0.231553364   4.02714 0.00013001 ***
## d3_(Intercept) -0.569954616  0.517637627  -1.10107 0.27425301
## d3_log(p1)      -0.038863769  0.063017248  -0.61672 0.53921776
## d3_log(p2)       0.000291214  0.045838974   0.00635 0.99494731
## d3_log(p3)      -0.925174458  0.069871998 -13.24099 < 2.22e-16 ***
## d3_log(y)        1.013108798  0.107831697   9.39528 1.8652e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La SCR obtenida fue de 10.83. El estadístico queda:

$$F_{obs} = \frac{(SCR^R - SCR^{NR})/k}{SCR^{NR}/(n-p)} = \frac{(10.83 - 9.75)/3}{9.75/(90 - 15)} = 1.18$$

Al igual que para el conjunto de restricciones anterior, el valor crítico para una distribución F con 3 y 75 grados de libertad y un nivel de confianza del 95 % es 2.73, por lo que las Condiciones de Simetría también son válidas consideradas en conjunto.

Podríamos haber resuelto este inciso (ii) recurriendo a la función `linearHypothesis()` nuevamente. En ese caso, para la **Condición de Engel** tendríamos:

```
lh2 <- linearHypothesis(fitsur, hypothesis.matrix = R[1,], rhs = r[1,],
                        test = "F")
lh2$F
```

```
## [1]      NA 1.094031
```

```
lh2$`Pr(>F)`
```

```
## [1]      NA 0.2989393
```

Por lo que tampoco se rechaza la hipótesis nula. Para las **Condiciones de Homogeneidad**:

```
lh3 <- linearHypothesis(fitsur, hypothesis.matrix = R[2:4,], rhs = r[2:4,])
lh3$F
```

```
## [1]      NA 2.722602
```

```
lh3$`Pr(>F)`
```

```
## [1]      NA 0.05024383
```

Aquí tampoco se rechaza la hipótesis nula, aunque por poco. Finalmente, para las **Condiciones de Simetría**:

```
lh4 <- linearHypothesis(fitsur, hypothesis.matrix = R[5:7,], rhs = r[5:7,])
lh4$F
```

```
## [1]      NA 3.29813
```

```
lh4$`Pr(>F)`
```

```
## [1]      NA 0.02492535
```

En este caso, y contrariamente a lo que nosotros habíamos obtenido, sí se rechaza la hipótesis nula, por lo que estas restricciones no son válidas individualmente consideradas.

Teniendo en cuenta las conclusiones de este inciso, y como un agregado, pasamos a estimar un modelo SUR restringido únicamente con la Condición de Engel y las Condiciones de Homogeneidad, obviando las Condiciones de Simetría, cuya invalidez ya fue comprobada, con el objetivo de contrastar lo obtenido con los modelos previamente estimados en el inciso **b)**.

```
fitsur_sinSim <- systemfit(system ,
                           method = "SUR", data = judge_11_3,
                           restrict.matrix = R[1:4,],
                           restrict.rhs = r[1:4,])
summsur_sinSim <- summary(fitsur_sinSim, residCov = F, equations = F)
summsur_sinSim
```

```
##
## systemfit results
## method: SUR
##
##          N DF      SSR  detRCov  OLS-R2  McElroy-R2
## system 90 79 11.2108 0.000181 0.67154   0.979161
##
##          N DF      SSR      MSE      RMSE      R2  Adj R2
## d1 30 25 4.357152 0.167583 0.409369 0.608218 0.563012
## d2 30 25 5.909771 0.227299 0.476759 0.505464 0.448402
## d3 30 25 0.943913 0.036304 0.190537 0.914655 0.904808
##
##
## Coefficients:
##          Estimate Std. Error  t value  Pr(>|t|)
## d1_(Intercept) -2.7719436  0.7754039  -3.57484 0.00060144 ***
## d1_log(p1)      -0.8586124  0.1756932  -4.88700 5.2647e-06 ***
## d1_log(p2)      -0.1033040  0.1501502  -0.68800 0.49346541
## d1_log(p3)      -0.3025068  0.1344808  -2.24944 0.02726794 *
## d1_log(y)        1.2644232  0.1945568   6.49899 6.6665e-09 ***
## d2_(Intercept) -1.8032517  0.9115056  -1.97832 0.05137832 .
## d2_log(p1)      -0.1275931  0.2054856  -0.62093 0.53643013
## d2_log(p2)      -0.8262447  0.1752683  -4.71417 1.0264e-05 ***
## d2_log(p3)       0.0255594  0.1568516   0.16295 0.87097187
## d2_log(y)        0.9282784  0.2287250   4.05849 0.00011530 ***
## d3_(Intercept) -0.1566377  0.3037904  -0.51561 0.60756602
## d3_log(p1)      -0.0293178  0.0764727  -0.38338 0.70247035
## d3_log(p2)      -0.0127103  0.0676844  -0.18779 0.85152513
## d3_log(p3)      -0.8917845  0.0614550 -14.51118 < 2.22e-16 ***
## d3_log(y)        0.9338126  0.0760781  12.27439 < 2.22e-16 ***
## ---
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

En base a esta estimación, se vuelve a armar una tabla comparativa (Cuadro 3). Si nos centramos en los errores estándar asociados a los coeficientes, el modelo restringido sin las Condiciones de Simetría es claramente menos eficiente, ya que los mismos son sistemáticamente mayores a los del modelo estimado con las 7 restricciones.

Cuadro 3: Resultados SUR con las 7 restricciones y sin las Condiciones de Simetría

Parámetro	SUR c/ 7 restricciones	SUR sin C. de Simetría
β_{10}	-2.783 (0.764)	-2.772 (0.775)
β_{11}	-0.865 (0.169)	-0.859 (0.176)
β_{12}	-0.125 (0.098)	-0.103 (0.150)
β_{13}	-0.278 (0.126)	-0.303 (0.134)
β_{14}	1.267 (0.192)	1.264 (0.195)
β_{20}	-1.756 (0.879)	-1.803 (0.912)
β_{21}	-0.118 (0.158)	-0.127 (0.205)
β_{22}	-0.836 (0.171)	-0.826 (0.175)
β_{23}	0.037 (0.153)	0.026 (0.157)
β_{24}	0.917 (0.219)	0.928 (0.229)
β_{30}	-0.164 (0.304)	-0.157 (0.304)
β_{31}	-0.018 (0.056)	-0.029 (0.076)
β_{32}	0.005 (0.041)	-0.013 (0.068)
β_{33}	-0.923 (0.052)	-0.892 (0.061)
β_{33}	0.935 (0.076)	0.934 (0.076)

Si además se tiene en cuenta que los R^2 de McElroy son marginalmente diferentes (como se ve a continuación), el mejor modelo sigue siendo el SUR con todas las restricciones. Podemos terminar de corroborar esto con el criterio de Akaike (AIC). Si el modelo con las 7 restricciones es el que tiene el menor AIC concluimos que es, efectivamente, el mejor de los dos modelos. Y eso es lo que sucede:

```
aics <- c(AIC(fitsurR)/nobs(fitsurR),  
          AIC(fitsur_sinSim)/nobs(fitsur_sinSim))  
mcelroy <- c(summsurR$mcelroy.r.squared,summsur_sinSim$mcelroy.r.squared)  
nombres_f <- c('SUR 7 restricc.','SUR sin Simet.')  
df_sur <- data.frame('Modelo'=nombres_f,'AIC'=aics,'McElroy'=mcelroy)  
  
library(knitr)  
kable(df_sur)
```

Modelo	AIC	McElroy
SUR 7 restricc.	0.1439821	0.9785361
SUR sin Simet.	0.1996069	0.9791614

Ejercicio II

Utilizando los datos de la tabla 12.1, pág. 500 del libro de Judge y otros (1988) antes citado, graficar la suma de cuadrados de residuos del modelo planteado en el texto. Esta función es un polinomio de cuarto grado en el parámetro beta. Se le pide:

a) Calcular las raíces del polinomio de tercer grado que se obtiene al calcular la primera derivada de esa función con respecto a beta. Para esto explore y utilice alguna de las alternativas que tiene R para obtener estas raíces.

b) Calcular los puntos de inflexión del polinomio de cuarto grado e indicar qué tipo de peligros pueden presentarse cuando se emplea el algoritmo de Newton-Raphson para obtener el estimador beta sombrero que minimiza la suma de los cuadrados de los residuos.

c) *Opcional*: Explique detalladamente el procedimiento que proponen los autores para computar el tamaño del paso (denominado “step length” en inglés).

Respuesta:

Gráfico de la SCR

Judge y los demás autores comentan que la no linealidad puede darse en los modelos económicos de diversas maneras, ya sea porque los parámetros son no lineales o porque las variables y parámetros no lo son, pero en general, es posible encontrar una transformación tal que se puedan linealizar los modelos. A pesar de ello, existen casos en los que no se pueden aplicar transformaciones para encontrar especificaciones lineales, son modelos intrínsecamente no lineales.

La base de su análisis teórico parte del siguiente modelo:

$$y_t = f(x_t, \beta) + e_t$$

En donde:

- x_t es un vector de dimensión $(N \times 1)$ de variables independientes
- β es el vector de parámetros, es de dimensión $(K \times 1)$
- y_t es la variable dependiente, su media es función de x_t y β
- e_t es el término de error aleatorio

Un ejemplo puede ser la función de producción de tipo Cobb-Douglas:

$$Q_t = \alpha L_t^{\beta_1} K_t^{\beta_2} + e_t$$

Ellos analizan cómo realizar la estimación de los parámetros en los modelos intrínsecamente no lineales, y se basan en considerar dos tipos de función objetivo:

1. Suma de cuadrados de los residuos
2. Función de máxima verosimilitud

Bajo los supuestos estándar de los modelos lineales se minimiza o maximiza resolviendo un sistema de ecuaciones normales lineales, pero en el caso de los modelos no lineales, las ecuaciones normales (en general) son no lineales en los parámetros, por lo que se necesitan utilizar algoritmos para poder realizar las estimaciones, tales como Gauss-Newton o Newton-Raphson. Para una estimación de máxima verosimilitud siguiendo este último algoritmo, describen el método de Scoring y el algoritmo BHHH³ (1974). Por lo que los autores parten del modelo antes mencionado y a partir de ahí realizan la estimación de Mínimos Cuadrados No Lineales:

$$y_t = f(x_t, \beta) + e_t$$

$$y_t = \beta x_{t1} + \beta^2 x_{t2} + e_t$$

En donde e_t son los términos de error aleatorio, independientes e idénticamente distribuidos, con media 0 y varianza σ^2 , es decir, son ruido blanco.

Entonces, la estimación de β se basa en los valores de β que minimizan la suma de cuadrados residuales:

$$S(\beta) = \sum_{t=1}^T e_t^2 = \sum_{t=1}^T [y_t - f(x_t, \beta)]^2$$

$$S(\beta) = \sum_{t=1}^T [y_t - \beta x_{t1} - \beta^2 x_{t2}]^2$$

Las condiciones de primer orden de la ecuación normal para minimizar la función están dada por:

$$\frac{dS}{d\beta} = 2 \sum_{t=1}^T [y_t - f(x_t, \beta)] \left(-\frac{df(x_t, \beta)}{d\beta} \right)$$

$$\frac{dS}{d\beta} = 2 \sum_{t=1}^T [y_t - \beta x_{t1} - \beta^2 x_{t2}] (-x_{t1} - 2\beta x_{t2})$$

Reordenando:

$$2\beta^3 \sum_{t=1}^T x_{t2}^2 + 3\beta^2 \sum_{t=1}^T x_{t1}x_{t2} + \beta \left(\sum_{t=1}^T x_{t1}^2 - 2 \sum_{t=1}^T x_{t2}y_t \right) - \sum_{t=1}^T x_{t1}y_t$$

Podemos ver que es un polinomio de tercer grado en β , por lo tanto, existen 3 posibles soluciones. La estimación no lineal de mínimos cuadrados, $\hat{\beta}$, nos dará la solución que

³Berndt, Hall, Hall, Hausman

minimice la suma de cuadrados de los residuos, siendo el mínimo global, mientras que las demás soluciones pueden ser mínimos o máximos (locales).

Ahora bien, considerando los datos que nos ofrecen los autores para tomar como ejemplo, se procederá a graficar la suma de cuadrados residuales. Mediante el paquete `readxl` se leerá el archivo Excel que contiene la tabla:

```
library(readxl)

table <- read_xlsx(table_12_1_path,
                  col_names = TRUE)

head(table) # Podemos ver las primeras 6 observaciones de la base de datos

## # A tibble: 6 x 3
##       y    x_1    x_2
##   <dbl> <dbl> <dbl>
## 1  3.28  0.286 0.645
## 2  3.15  0.973 0.585
## 3  2.88  0.384 0.31
## 4 -0.467 0.276 0.058
## 5  1.21  0.973 0.455
## 6  1.39  0.543 0.779

tail(table) # Podemos ver las últimas 6 observaciones de la base de datos

## # A tibble: 6 x 3
##       y    x_1    x_2
##   <dbl> <dbl> <dbl>
## 1  1.86  0.399 0.842
## 2  0.388 0.617 0.039
## 3  0.651 0.939 0.103
## 4  0.593 0.784 0.62
## 5  0.046 0.072 0.158
## 6  1.15  0.889 0.704

# Estableceremos a las variables como vectores

y <- as.vector(table$y)
x_1 <- as.vector(table$x_1)
x_2 <- as.vector(table$x_2)

# Comprobamos su clase

class(y)
```

```
## [1] "numeric"
```

```
class(x_1)
```

```
## [1] "numeric"
```

```
class(x_2)
```

```
## [1] "numeric"
```

Ahora bien, se crea una secuencia de datos que β pueda asumir para los cuales luego se graficará la función $S(\beta)$, y se toman como referencia los valores presentados por Judge et al. (1988) en su ejemplo:

```
beta <- seq(-3,3,by=0.01)

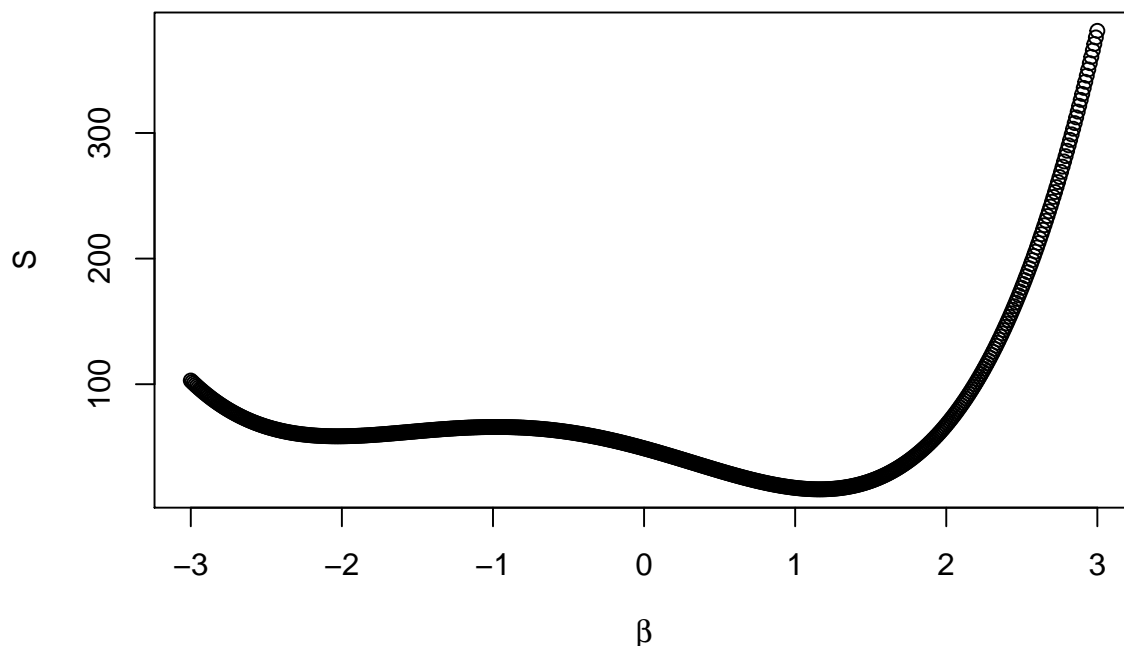
s <- function(beta){
  sum((y-beta*x_1-(beta^{2})*x_2)^{2})
}
```

Se aplica la función a la secuencia de valores creada de β mediante la función de R `sapply` (permite aplicar una función a un vector e itera con cada elemento de este) y se grafica con la función `plot`:

```
scr <- sapply(beta, s)

plot(x=beta,y=scr,
     main = "Suma de Cuadrados Residuales",
     xlab = expression(beta),
     ylab = "S")
```


Suma de Cuadrados Residuales



La gráfica nos muestra lo que dicen los autores, el mínimo de la función se encuentra cercano al valor de $\beta = 1,2$, mientras que otras soluciones a las cuales se llega por aproximación son $\beta = -2$, siendo esta un mínimo local, y otra es $\beta = -1$, un máximo.

Sin embargo, en el siguiente inciso se buscará con mayor detalle, y se realizarán las aproximaciones para encontrar las raíces buscadas.

Inciso a)

Recordando que se obtuvo un polinomio de 3° grado al buscar la derivada de la función objetivo, el cual era:

$$2\beta^3 \sum_{t=1}^T x_{t2}^2 + 3\beta^2 \sum_{t=1}^T x_{t1}x_{t2} + \beta \left(\sum_{t=1}^T x_{t1}^2 - 2 \sum_{t=1}^T x_{t2}y_t \right) - \sum_{t=1}^T x_{t1}y_t$$

Se busca calcular sus raíces, es decir, los valores para lo cual se anula la función. Con R se procede a buscar estos valores, mediante la función `polyroot`:

```
# Primero se declaran los coeficientes del polinomio

z0 <- sum(x_1*y)*(-1)
z1 <- sum(x_1^{2})*-2*sum(x_2*y)
z2 <- 3*sum(x_1*x_2)
z3 <- 2*sum(x_2^{2})
```

```
# Luego se los ingresa como un vector (en orden creciente) en la función
raices <- polyroot(c(z0,z1,z2,z3)); raices; # Observamos el resultado
```

```
## [1] 1.1612066+0i -0.9776215+0i -2.0294944-0i
```

Por lo que se obtuvieron esas 3 raíces:

- $\beta_1 = 1.16120662846748+0i$
- $\beta_2 = -0.977621493622509+0i$
- $\beta_3 = -2.0294944070943-0i$

En los 3 casos son raíces reales, pero, ¿cuál es la que minimiza la SCR? Esto nos lo preguntamos porque algunas son extremos locales, no globales, por lo tanto, se continúa:

```
# Utilizamos la función "lapply", la cual se diferencia de "sapply"
# Ya que esta última brinda un vector como salida, mientras que
# "lapply" presenta la salida como una lista

rdo_scr <- lapply(raices, s); rdo_scr
```

```
## [[1]]
## [1] 16.30797+0i
##
## [[2]]
## [1] 65.93961+0i
##
## [[3]]
## [1] 58.52343-0i
```

Los resultados nos muestran que la raíz que minimiza la suma de cuadrados residuales es β_1 , con la cual se obtiene una suma de cuadrados de los residuos igual a 16.3079700516694+0i.

Inciso b)

Los puntos de inflexión de una función se corresponden a aquellos valores que anulan la derivada segunda de ella, entonces, en base al polinomio de tercer orden de β que era la primera derivada de la función en relación a β , procedemos a derivarlo una vez más e igualarlo a cero, por lo tanto:

$$\frac{\partial^2 S}{\partial \beta^2} = 6\beta^2 \sum_{t=1}^T x_{t2}^2 + 6\beta \sum_{t=1}^T x_{t1}x_{t2} + \left(\sum_{t=1}^T x_{t1}^2 - 2 \sum_{t=1}^T x_{t2}y_t \right) = 0$$

De esta forma, ahora repetimos lo realizado en el inciso anterior, pero para esta función:

```

z0_2 <- sum(x_1^{2})-2*sum(x_2*y)

z1_2 <- 6*sum(x_1*x_2)

z2_2 <- 6*sum(x_2^{2})

inflexion <- polyroot(c(z0_2,z1_2,z2_2)); inflexion

## [1] 0.3234193+0i -1.5540255-0i

```

Por lo que los valores de β_i para los puntos de inflexión son:

- $\beta_1 = 0.323419311131089+0i$
- $\beta_2 = -1.55402549263064-0i$

Ahora bien, el ejercicio pregunta cuáles son los peligros que pueden aparecer al utilizar el algoritmo de Newton-Raphson para obtener el estimador $\hat{\beta}$ que minimiza la SCR. Nosotros conocemos que se pueden aproximar raíces mediante la utilización de este algoritmo iterativo, a su vez, los autores mencionan que también puede utilizarse el algoritmo de Gauss-Newton. Procediendo con el algoritmo de Newton-Raphson, y siguiendo a Moneta Pizarro (2021), esta es su fórmula para el caso de una sola variable:

$$x_{i+1} = x_i - \frac{f(X_i)}{f'(X_i)}$$

En donde:

- x_i es la estimación de la raíz en el paso i
- $f(X_i)$ es la función objetivo valuada en $X = X_i$
- $f'(X_i)$ es la primera derivada de la función objetivo valuada en $X = X_i$
- x_{i+1} es la estimación de la raíz que propone el algoritmo.

Mientras que, para el caso de más variables, la fórmula es:

$$\vec{x}_{i+1} = \vec{x}_i - \nabla^{-1} f(\vec{x}_i) f'(\vec{x}_i)$$

En donde:

- \vec{x}_i es el vector solución propuesto en el paso i
- $\nabla^{-1} f(\vec{x}_i)$ es la inversa del vector gradiente (sus elementos son las derivadas primeras de la función respecto de cada una de las variables independientes contenidas en el vector \vec{x}) valuada en $\vec{x} = \vec{x}_i$

- \vec{x}_{i+1} es el vector solución propuesto para el próximo paso $i + 1$.
- \vec{x}_{i+1} es la estimación del vector de raíces que el algoritmo propone

Aplicado a nuestro caso, utilizaremos la siguiente expresión, partiendo de una serie de Taylor de segundo orden alrededor de un valor inicial β_1 :

$$S(\beta) = S(\beta_1) + S'(\beta_1)(\beta - \beta_1) + \frac{1}{2}S''(\beta_1)(\beta - \beta_1)^2$$

Se deriva respecto a β y se obtiene:

$$S'(\beta) = S'(\beta_1) + S''(\beta_1)(\beta - \beta_1)$$

Al igualar a cero y despejar para beta, se obtiene que:

$$\beta_2 = \beta_1 - \frac{S'(\beta_1)}{S''(\beta_1)}$$

Y continuando de forma iterativa:

$$\beta_{i+1} = \beta_i - \frac{S'(\beta_i)}{S''(\beta_i)}$$

El proceso converge cuando $\beta_{i+1} = \beta_i$

Es importante destacar que, la derivada primera de la SCR debe ser igual a cero, que será la condición necesaria para la existencia de un máximo o un mínimo.

Entonces, en resumen, se debe saber si:

1. Se convergió o no a un mínimo
2. Dicho mínimo es local o es global

Para estar seguros de que la iteración está siendo realizada en la dirección correcta (en este caso un mínimo), se requiere que el valor de β elegido inicialmente esté lo suficientemente cerca del extremo, tal que la derivada segunda valuada en ese punto sea positiva (en el caso del mínimo) o negativa (para un máximo).

Esto podemos verlo con el ejemplo presentado por los autores. Ellos demuestran la importancia del valor elegido, en base a los datos, expresan que si comienzan con el valor inicial de $\beta_1 = 1$, la aproximación converge al mínimo global en un pequeño número de iteraciones, mientras que si comienzan con el valor de $\beta_1 = -2$, el algoritmo converge a un mínimo local. Por su parte, con un $\beta_1 = 0,1$, el resultado de la aproximación al principio se dirige hacia el máximo, se excede y luego converge al mínimo local. Finalmente si $\beta_1 = -0,9$, el algoritmo converge hacia el máximo.

Para los mismos valores elegidos, el algoritmo de Gauss-Newton, en los dos primeros casos lleva al mismo resultado que Newton-Raphson (primero converge al mínimo global con pocas iteraciones y en el segundo, convergen a un mínimo local), mientras que para

el 3° y 4° valor elegido, Gauss-Newton se aproxima al mínimo global en ambos casos. Los autores afirman que el algoritmo de Newton-Raphson es más eficiente porque converge más rápido que Gauss-Newton, pero este último se vuelve más confiable al tener una mayor probabilidad de converger al mínimo global.

Los peligros y/o limitaciones de este algoritmo surgen, por un lado, porque la función debe ser como mínimo 2 veces diferenciable, y por otro lado, si el valor que se eligió inicialmente para iterar lleva a que la derivada segunda valuada en ese punto sea igual a cero, es decir, se arriba a uno de los puntos de inflexión y no al valor óptimo de la función buscado.

Por su parte, debe tenerse en cuenta que el algoritmo puede llevar a infinitas iteraciones, por lo que es necesario saber cuándo detener el proceso. Por ejemplo, se puede poner un límite en la cantidad de iteraciones, o sino, definir un número ε tal que si se verifica que:

$$\frac{S'(\beta_i)}{S''(\beta_i)} < \varepsilon$$

se termina el proceso de iteraciones. Sino, como última opción, se pueden definir ambos criterios y finalizar el proceso cuando uno de los dos ocurra antes.

Inciso c)

Se mencionó que el algoritmo de Newton-Raphson converge cuando $\beta_{n+1} = \beta_n$, por lo que en ese caso, debiera ser cierto que $\partial S / \partial \beta|_{\beta_n} = 0$, la cual es la condición necesaria para la existencia de un mínimo (o máximo). Además, también nos preguntamos acerca de cómo saber si aplicar el algoritmo nos lleva a la dirección correcta (encontrar un mínimo) y cómo saber si el mínimo o máximo que se encuentra es global o local.

Matemáticamente se sabe que, el algoritmo encontró un mínimo si la derivada segunda es positiva $h(\beta_1) = (\partial^2 S / \partial \beta^2|_{\beta_n}) > 0$, por lo que, como $h(\beta_1)$ siempre será positivo al estar cerca de un mínimo, se está en la dirección correcta si β_1 está lo suficientemente cerca del valor de minimización.

A pesar de ello, los autores destacan que surge otro problema, uno puede sobrepasarse en la aplicación del algoritmo (lo que nos llevaría hacia arriba de nuevo), entonces, ¿qué se puede hacer?

Ante el problema de sobrepasarse, los autores expresan que puede introducirse en el algoritmo una nueva variable, conocida como “step length” (tamaño del paso), t_n , por lo que el algoritmo queda expresado de la siguiente manera:

$$\beta_{n+1} = \beta_n - t_n \frac{S'(\beta_n)}{S''(\beta_n)}$$

Esta variable t_n funciona de tal manera que $S(\beta_{n+1}) < S(\beta_n)$. Es un número positivo, y la mayoría de los algoritmos incluyen un procedimiento para determinar el “step length” óptimo en cada iteración. Flanagan, Vitale y Mendelsohn (1969) describen algunos de ellos. Siguiendo a Judge et al. (1985), un “step length” demasiado alto nos llevaría a sobrepasarnos (como ya se explicó) mientras que uno demasiado bajo sería ineficiente

Además, si se comienza en un punto cercano a un máximo, en el sentido de que $h(\beta_1)$ sea negativo, entonces el algoritmo nos estará llevando en la dirección equivocada, hacia un máximo. Por lo que los autores recomiendan que para establecer si un mínimo es global o local, se deben probar varios valores iniciales diferentes para poder encontrar los posibles mínimos de la función objetivo. En nuestro caso, buscamos encontrar los del valor crítico de β que minimiza la suma de cuadrados residuales, de tal forma que una vez que obtenemos los distintos valores de convergencia provistos por el algoritmo, se valorará la función objetivo en dichos valores y se elegirá aquella que minimice la suma de los cuadrados de los residuos.

Ejercicio III

Utilizando los datos de Kmenta, Jan (1986), *Elements of Econometrics*, Second Edition, Macmillan, New York, págs. 686 y 687, se le pide:

a) Verificar las condiciones de orden y rango para la identificación de cada una de las ecuaciones del modelo de oferta y de demanda planteado.

b) Explicar en forma esquemática en qué consisten cada uno de los métodos de estimación siguientes: Mínimos cuadrados en 2 etapas, Mínimos cuadrados en 3 etapas y Máxima verosimilitud con información completa.

c) Realizar las estimaciones pertinentes con R. Se sugiere que utilice la biblioteca “systemfit”.

d) Explicar cómo pueden utilizarse las restricciones sobre la matriz de varianzas y covarianzas de los errores de la forma estructural para identificar un modelo.

e) Resumir los principales resultados de derivación matricial planteados por: Phillips, P.C.B. y Wickens, M.R. (1978), *Exercises in Econometrics – Volume two*, Philip Allan, Oxford, págs. 331 a 332.

f) Opcional de simulaciones de Monte Carlo: generar 1000 observaciones con remuestreo de las variables del ejemplo de Kmenta. Seguidamente tomar 20 muestras de 50 observaciones cada una. Para cada muestra estimar el modelo por MCO y por MC2E y realizar el contraste de Hausman. Tabule los resultados de manera tal que sea más sencilla la interpretación de los resultados.

Respuesta:

Inciso a)

Kmenta (1986) presenta el siguiente modelo simplificado para explicar variaciones en el consumo y en los precios de la comida:

$$\begin{aligned}Q_t &= \alpha_1 + \alpha_2 P_t + \alpha_3 D_t + \mu_{1t} \\Q_t &= \beta_1 + \beta_2 P_t + \beta_3 F_t + \beta_4 A_t + \mu_{2t}\end{aligned}$$

Donde la primera es una ecuación de demanda y la segunda, de oferta. Las variables endógenas son Q_t y P_t , mientras que D_t , F_t y A_t son variables predeterminadas. Comenzamos con la verificación de la **Condición de Orden**. Siguiendo a Moneta Pizarro (2021), la misma estipula que en un modelo de M ecuaciones simultaneas, donde K es el

número de variables predeterminadas en el modelo, k el número de variables predeterminadas incluidas en la ecuación y m el número de variables endógenas que participan en la ecuación, una ecuación está identificada si se cumple $K - k \geq m - 1$. En el caso de que $K - k > m - 1$, la ecuación está sobreidentificada, mientras que si $K - k = m - 1$, la misma está exactamente identificada. Bajo estas consideraciones, pasamos a analizar cada ecuación por separado:

Ecuación de Demanda:

$$K - k \geq m - 1$$

$$3 - 1 \geq 2 - 1$$

$$2 \geq 1$$

Por ende, esta ecuación está sobreidentificada.

Ecuación de Oferta:

$$K - k \geq m - 1$$

$$2 - 1 \geq 2 - 1$$

$$1 = 1$$

Entonces, esta ecuación está exactamente identificada.

Siguiendo con la **Condición de Rango**, la misma nos dice que:

En un modelo que contiene M ecuaciones y M variables endógenas, una ecuación está identificada si y sólo si puede construirse por lo menos un determinante diferente de cero, de orden $(M - 1)(M - 1)$, a partir de los coeficientes de las variables endógenas y predeterminadas excluidas de esa ecuación particular, pero incluidas en las otras ecuaciones del modelo. (Moneta Pizarro, 2021, p. 140)

Para verificar esto, se sigue una serie de pasos:

Ecuación de Demanda:

- 1) Se despeja el término de perturbación (μ_{it}) de cada ecuación:

$$-\alpha_1 + Q_t - \alpha_2 P_t - \alpha_3 D_t = \mu_{1t}$$

$$-\beta_1 + Q_t - \beta_2 P_t - \beta_3 F_t - \beta_4 A_t = \mu_{2t}$$

- 2) Se arma una tabla en base a los coeficientes estructurales del sistema:

1	Q_t	P_t	D_t	F_t	A_t
$-\alpha_1$	1	$-\alpha_2$	$-\alpha_3$	0	0
$-\beta_1$	1	$-\beta_2$	0	$-\beta_3$	$-\beta_4$

- 3) Se eliminan los coeficientes del renglón en que aparece la ecuación bajo análisis (en este caso se elimina el primer renglón):

$$\begin{array}{cccccc} 1 & Q_t & P_t & D_t & F_t & A_t \\ -\beta_1 & 1 & -\beta_2 & 0 & -\beta_3 & -\beta_4 \end{array}$$

- 4) Se eliminan las columnas correspondientes a los coeficientes que son distintos de cero en la ecuación de interés (la de demanda):

$$\begin{array}{cc} F_t & A_t \\ -\beta_3 & -\beta_4 \end{array}$$

- 5) Se arman todas las posibles matrices de orden $M - 1 = 2 - 1 = 1$ con los coeficientes que no fueron eliminados en el paso anterior. En este caso sólo se pueden armar 2 matrices de orden 1 (un escalar) con β_3 y β_4 . Luego se calcula el determinante de cada una de esas matrices. Dado que el determinante de un escalar es el mismo escalar, y que tanto β_3 como β_4 se suponen distintos de 0, la ecuación está identificada.

Ecuación de Oferta:

Los pasos 1) y 2) son idénticos.

- 3) Se elimina el segundo renglón, en este caso:

$$\begin{array}{cccccc} 1 & Q_t & P_t & D_t & F_t & A_t \\ -\alpha_1 & 1 & -\alpha_2 & -\alpha_3 & 0 & 0 \end{array}$$

- 4) Se eliminan las columnas correspondientes a los coeficientes que son distintos de cero en la ecuación de oferta:

$$\begin{array}{c} D_t \\ -\alpha_3 \end{array}$$

- 5) Al ser $-\alpha_3$ distinto de 0, la ecuación de oferta también está identificada.

Inciso b)

En la Figura 1 se muestra un esquema para contextualizar los 3 métodos de estimación en cuestión. A continuación se los pasa a explicar uno por uno.

Mínimos Cuadrados en 2 Etapas (MC2E):

Para ecuaciones exactamente identificadas (en cuyo caso se obtienen los mismos estimadores que si se aplicara Mínimos Cuadrados Indirectos) o sobreidentificadas. El estimador MC2E es equivalente al Estimador Eficiente de Variables Instrumentales.

Pasos (2 etapas):

- Estimar cada ecuación de la forma reducida (transformación sobre la forma estructural de tal forma que las variables endógenas queden en función de exógenas y del término de perturbación) por MCO, luego obtener los valores ajustados de las variables endógenas.

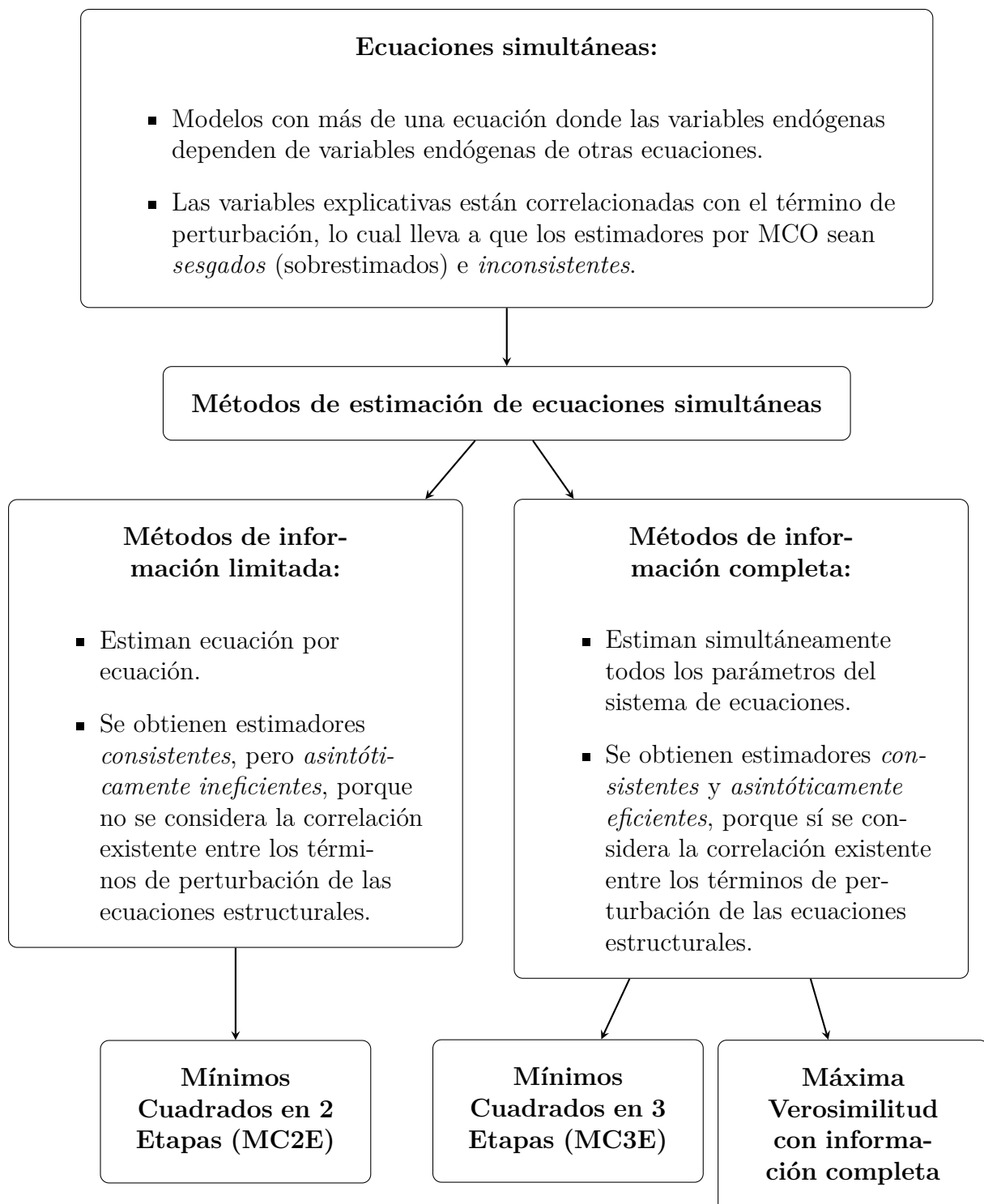


Figura 1: Esquema de métodos de estimación de ecuaciones simultáneas.

- Reemplazar las variables endógenas que aparecen como explicativas en la forma estructural (modelo original provisto por la teoría económica) por los valores ajustados obtenidos en el paso anterior, luego estimar cada ecuación por MCO.

Siguiendo a Kmenta (1986) ejemplificamos el método a partir de la primera ecuación estructural del sistema:

$$y_1 = Y_1\beta + X_1\gamma + u_1$$

Donde y_1 es el vector de la variable endógena que es igual a 1 en la primera ecuación, Y_1 es la matriz de las restantes variables endógenas en la primera ecuación, X_1 es la matriz de variables exógenas en la primera ecuación, y u_1 es el vector de perturbaciones de esta ecuación. En base a estas definiciones se concluye que $Y_1 = [y_2 \ y_3 \ \dots \ y_m]$, siendo las formas reducidas de estas variables y_i :

$$\begin{aligned} y_2 &= X\pi_2 + v_2 \\ y_3 &= X\pi_3 + v_3 \\ &\vdots \\ y_m &= X\pi_m + v_m \end{aligned} \tag{3}$$

Donde X es la matriz de *todas* las variables exógenas del sistema. Además, puedo definir $V_1 = [v_2 \ v_3 \ \dots \ v_m]$. Luego, los valores ajustados de las variables endógenas en la primera ecuación son los siguientes:

$$Y_1 - V_1 = [X\pi_2 \ X\pi_3 \ \dots \ X\pi_m]$$

El primer paso del método MC2E, entonces, es estimar (3) para así obtener los residuos de la forma reducida y luego poder calcular los valores ajustados:

$$Y_1 - \hat{V}_1 = [X\hat{\pi}_2 \ X\hat{\pi}_3 \ \dots \ X\hat{\pi}_m] = \hat{Y}_1$$

En el segundo paso, se reemplazan los valores ajustados recién obtenidos por las variables endógenas que aparecen como explicativas en la ecuación estructural:

$$\begin{aligned} y_1 &= \hat{Y}_1\beta + X_1\gamma + u_1^* \\ u_1^* &= u_1 + \hat{V}_1\beta \end{aligned}$$

Finalmente, estimando β y γ por MCO se obtienen los estimadores MC2E, de la siguiente forma:

$$\begin{bmatrix} \hat{\beta}_{MC2E} \\ \hat{\gamma}_{MC2E} \end{bmatrix} = \begin{bmatrix} \hat{Y}_1' \hat{Y}_1 & \hat{Y}_1' X_1 \\ X_1' \hat{Y}_1 & X_1' X_1 \end{bmatrix}^{-1} \begin{bmatrix} \hat{Y}_1' y_1 \\ X_1' y_1 \end{bmatrix}$$

El estimador MC2E se puede presentar como un estimador de variables instrumentales (consistente), siendo $Z = [\hat{Y}_1 X_1]$ el vector de instrumentos del vector $[Y_1 X_1]$.

Mínimos Cuadrados en 3 Etapas (MC3E):

El estimador MC3E es igual al MC2E si los términos de perturbación de las ecuaciones estructurales no están correlacionados entre sí.

La idea es tratar el sistema de ecuaciones estructurales con los valores ajustados de las variables endógenas como explicativas como un sistema de ecuaciones aparentemente no relacionadas (SUR).

“Las ecuaciones estructurales identificadas se estiman en primer lugar por MC2E y los residuos resultantes se utilizan para estimar la matriz de covarianzas de las perturbaciones, que se utiliza posteriormente para estimar conjuntamente todos los parámetros estructurales identificados” (Pérez López, 2006, p. 276). La estimación conjunta se hace a través del estimador de Aitken o de Mínimos Cuadrados Generalizados (MCG), como en el método SUR.

Siguiendo a Kmenta (1986), ahora partimos de todo el sistema de ecuaciones para la ejemplificación, con los valores ajustados de las variables endógenas ya reemplazados:

$$\begin{aligned} y_1 &= \hat{Y}_1\beta_1 + X_1\gamma_1 + u_1^* \\ y_2 &= \hat{Y}_2\beta_2 + X_2\gamma_2 + u_2^* \\ &\vdots \\ y_M &= \hat{Y}_M\beta_M + X_M\gamma_M + u_M^* \end{aligned} \quad (4)$$

Que se puede reexpresar como:

$$\begin{aligned} y_1 &= \hat{Z}_1\alpha_1 + u_1^* \\ y_2 &= \hat{Z}_2\alpha_2 + u_2^* \\ &\vdots \\ y_M &= \hat{Z}_M\alpha_M + u_M^* \end{aligned} \quad (5)$$

Donde:

$$\begin{aligned} \hat{Z}_i &= [\hat{Y}_i \quad X_i] \\ \alpha_i &= \begin{bmatrix} \beta_i \\ \gamma_i \end{bmatrix} \end{aligned}$$

El sistema (5) se puede expresar en forma compacta:

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} = \begin{bmatrix} \hat{Z}_1 & 0 & \dots & 0 \\ 0 & \hat{Z}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \hat{Z}_M \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_M \end{bmatrix} + \begin{bmatrix} u_1^* \\ u_2^* \\ \vdots \\ u_M^* \end{bmatrix}$$

O alternativamente:

$$y = \hat{Z}\alpha + u^*$$

Como ya vimos, estimando (4) ecuación por ecuación por MCO se obtienen los estimadores y las perturbaciones correspondientes al método MC2E. En base a estos se puede estimar la matriz de varianzas y covarianzas:

$$\hat{\Omega} = \begin{bmatrix} \hat{\sigma}_{11}I_T & \hat{\sigma}_{12}I_T & \dots & \hat{\sigma}_{1M}I_T \\ \hat{\sigma}_{21}I_T & \hat{\sigma}_{22}I_T & \dots & \hat{\sigma}_{2M}I_T \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{M1}I_T & \hat{\sigma}_{M2}I_T & \dots & \hat{\sigma}_{MM}I_T \end{bmatrix} = \Sigma \otimes I_T$$

Donde:

$$\hat{\sigma}_{ij} = \frac{(y_i - \hat{Y}_i \hat{\beta}_{iMC2E} - X_i \hat{\gamma}_{iMC2E})'(y_j - \hat{Y}_j \hat{\beta}_{jMC2E} - X_j \hat{\gamma}_{jMC2E})}{T - M_i + 1 - K} \quad : \quad i, j = 1, 2, \dots, M$$

Finalmente se lleva a cabo la estimación conjunta de todos los parámetros de la forma estructural, contenidos en α , mediante la fórmula de Aitken o de MCG:

$$\hat{\alpha}_{MC3E} = (\hat{Z}'\hat{\Omega}^{-1}\hat{Z})^{-1}(\hat{Z}'\hat{\Omega}^{-1}y)$$

Máxima Verosimilitud con información completa (MVIC):

Aplicación del principio de máxima verosimilitud a todas las ecuaciones estocásticas del sistema simultáneamente.

Partiendo del sistema estructural $By_t + \Gamma x_t = u_t$, donde la matriz de varianzas y covarianzas correspondiente es $\Phi = E(u_t u_t')$, los estimadores MVIC (FIML en inglés) de B , Γ y Φ se obtienen maximizando la función de log-verosimilitud para las T observaciones de y_t condicional a los valores de x_t respecto a B , Γ y Φ .

Inciso c)

Primero recordemos que el sistema planteado por Kmenta (1986) es:

$$\begin{aligned} Q_t &= \alpha_1 + \alpha_2 P_t + \alpha_3 D_t + \mu_{1t} \\ Q_t &= \beta_1 + \beta_2 P_t + \beta_3 F_t + \beta_4 A_t + \mu_{2t} \end{aligned}$$

Donde las variables endógenas son Q_t y P_t , mientras que D_t , F_t y A_t son variables exógenas. Q_t es el consumo de comida por persona, P_t es el ratio entre precios de comida y los precios generales al consumidor, D_t es el ingreso disponible en precios constantes, F_t es el ratio entre los precios percibidos por los agricultores en el año anterior por sus productos y los precios generales al consumidor, y A_t es el tiempo en años.

Antes de llevar a cabo las estimaciones, definimos las funciones de demanda y oferta en R, además del sistema compuesto por ambas. Estas variables serán usadas como argumentos en la función `systemfit`. Otro argumento a definir es el de los instrumentos, que serán todas las variables exógenas del sistema (como se mencionó en el inciso [b](#)).

```
demanda <- Q_t ~ P_t + D_t
oferta <- Q_t ~ P_t + F_t + A_t
sistema <- list(dem = demanda, of = oferta)
instrum <- ~D_t + F_t + A_t
```

Recordemos que Kmenta (1986) provee el verdadero modelo estructural:

$$\begin{aligned} Q_t &= 96.5 - 0.25P_t + 0.30D_t + \mu_{1t} \\ Q_t &= 62.5 + 0.15P_t + 0.20F_t + 0.36A_t + \mu_{2t} \end{aligned}$$

Este último podrá ser comparado con los modelos estimados a continuación. Comenzando por MC2E:

```
mc2e <- systemfit(sistema, method = "2SLS", inst = instrum,
                  data = kmenta_13_1)
summary(mc2e, residCov = F, equations = FALSE)

##
## systemfit results
## method: 2SLS
##
##          N DF      SSR detRCov   OLS-R2 McElroy-R2
## system 40 33 162.362 4.36424 0.697214   0.548127
##
##          N DF      SSR      MSE      RMSE          R2   Adj R2
## dem 20 17 65.7291 3.86642 1.96632 0.754847 0.726005
## of 20 16 96.6332 6.03958 2.45756 0.639582 0.572004
##
##
## Coefficients:
##              Estimate Std. Error  t value  Pr(>|t|)
## dem_(Intercept) 94.6333039   7.9208383 11.94738 1.0762e-09 ***
## dem_P_t         -0.2435565   0.0964843 -2.52431 0.02183240 *
## dem_D_t          0.3139918   0.0469437  6.68869 3.8109e-06 ***
## of_(Intercept)  49.5324417 12.0105264  4.12409 0.00079536 ***
## of_P_t           0.2400758   0.0999339  2.40235 0.02878451 *
## of_F_t           0.2556057   0.0472501  5.40964 5.7854e-05 ***
## of_A_t           0.2529242   0.0996551  2.53800 0.02192877 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Es decir, el modelo estimado con MC2E es el siguiente:

$$Q_t = 94.6333 - 0.2436P_t + 0.3140D_t + \tilde{\mu}_{1t}$$

$$Q_t = 49.5324 + 0.2401P_t + 0.2556F_t + 0.2529A_t + \tilde{\mu}_{2t}$$

Los coeficientes son todos significativos al 5% por lo menos, y el R^2 de McElroy es 0,55. Aun así, se percibe fuerte correlación entre los residuos de las ecuaciones (0.9):

```
summary(mc2e)$residCor
```

```
##          dem          of
## dem 1.0000000 0.9017244
## of  0.9017244 1.0000000
```

Por lo que resultaría apropiado, como ya se explicó, aplicar el método SUR. Seguimos con la estimación por MC3E:

```
mc3e <- systemfit(sistema, method = "3SLS", inst = instrum,
                  data = kmenta_13_1)
summary(mc3e, residCov = FALSE, equations = FALSE)

##
## systemfit results
## method: 3SLS
##
##          N DF      SSR detRCov   OLS-R2 McElroy-R2
## system 40 33 173.643  1.0332 0.676177   0.786468
##
##          N DF      SSR      MSE      RMSE          R2   Adj R2
## dem 20 17  65.7291 3.86642 1.96632 0.754847 0.726005
## of  20 16 107.9138 6.74461 2.59704 0.597508 0.522041
##
##
## Coefficients:
##              Estimate Std. Error  t value  Pr(>|t|)
## dem_(Intercept) 94.6333039   7.9208383 11.94738 1.0762e-09 ***
## dem_P_t         -0.2435565   0.0964843 -2.52431 0.02183240 *
## dem_D_t          0.3139918   0.0469437  6.68869 3.8109e-06 ***
## of_(Intercept)  52.1972042 11.8933720  4.38876 0.00045780 ***
## of_P_t           0.2285892   0.0996732  2.29339 0.03570648 *
## of_F_t           0.2281580   0.0439938  5.18614 9.0087e-05 ***
## of_A_t           0.3611384   0.0728894  4.95461 0.00014343 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Entonces, el modelo estimado con MC3E es:

$$Q_t = 94.6333 - 0.2436P_t + 0.3140D_t + \tilde{\mu}_{1t}$$

$$Q_t = 52.1972 + 0.2286P_t + 0.2282F_t + 0.3611A_t + \tilde{\mu}_{2t}$$

Los coeficientes estimados son todos significativos nuevamente (al 5 % por lo menos), pero el R^2 de McElroy es considerablemente más alto (0,79). Esto tiene sentido, ya que el estimador de MCG interviene en el método, lo cual corresponde dada la correlación ya mencionada entre los errores.

Finalmente, si quisiéramos estimar los coeficientes del sistema estructural por Máxima Verosimilitud con información completa (MVIC), no podríamos hacerlo con `systemfit` porque este paquete no soporta el método mencionado. Aun así, Henningsen y Hamann (2008) mencionan que, teóricamente, los resultados de emplear el método SUR iterado deberían converger a los resultados de estimar por MVIC. Siguiendo el ejemplo presentado por los autores en el Apéndice C de su trabajo:

```

mVICSUR <- systemfit(sistema, method = "SUR", data = kmenta_13_1,
                     control = systemfit.control(methodResidCov = "noDfCor",
                                                  maxiter = 100))
summary(mVICSUR, residCov = FALSE, equations = FALSE)

##
## systemfit results
## method: iterated SUR
##
## warning: convergence not achieved after 100 iterations
##
##           N DF      SSR detRCov OLS-R2 McElroy-R2
## system 40 33 531.992      0 0.0079  0.797488
##
##           N DF      SSR      MSE      RMSE      R2      Adj R2
## dem 20 17 265.995 15.6468 3.95560 0.007904 -0.108813
## of 20 16 265.997 16.6248 4.07735 0.007896 -0.178123
##
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## dem_(Intercept) 9.20332e+01 1.35421e+01 6.79606 3.1176e-06 ***
## dem_P_t         8.85773e-02 1.35171e-01 0.65530 0.52104580
## dem_D_t         4.04016e-06 6.14199e-07 6.57793 4.6957e-06 ***
## of_(Intercept) 9.20328e+01 1.35422e+01 6.79600 4.2963e-06 ***
## of_P_t         8.85827e-02 1.35172e-01 0.65533 0.52156396
## of_F_t         1.82789e-06 4.59643e-07 3.97675 0.00108373 **
## of_A_t         3.35430e-06 7.15335e-07 4.68913 0.00024627 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Como se puede ver por el **warning**, no se alcanza convergencia usando el método SUR iterado con estos datos.

Por otro lado, Henningsen y Hamann (2008) recomiendan emplear el paquete **sem** para estimar por MVIC. Basándonos en los ejemplos expuestos en Henningsen y Hamann (2008) y en Fox (2006) llegamos a lo siguiente (luego de especificar **library(sem)**):

```

modeq <- specifyEquations(covs="Q_t,P_t",
                          text="Q_t=a1*P_t+a2*D_t
                               Q_t=b1*P_t+b2*F_t+b3*A_t")

exogVar <- c("D_t","F_t","A_t")
endogVar <- c("Q_t","P_t")
allVar <- c(exogVar,endogVar)

```



```

mvicSEM <- sem(model = modeq, S = cov(kmenta_13_1[-1,allVar]),
               N = (nrow(kmenta_13_1) - 1), fixed.x = exogVar)
print(mvicSEM)

##
## Model Chisquare = 51.98657 Df = 2
##
##          a1          a2          b2          b3      V[Q_t] C[Q_t,P_t]      V[P_t]
## -1.1285032  0.9082317 -0.4121939 -0.7465698  1.4901299  0.0000010 37.0641251
##
## Iterations = 0

```

Lamentablemente, se verifica que “`sem` cannot be used to consistently estimate systems of simultaneous equations with some endogenous regressors, it can be used to estimate systems of equations where all regressors are exogenous” [`sem` no se puede usar para estimar consistentemente sistemas de ecuaciones simultáneas con variables regresoras endógenas, sí se puede usar para estimar sistemas de ecuaciones donde todas las variables regresoras son exógenas] (Henningsen y Hamann, 2008, p. 16). Dado que P_t es una variable regresora endógena, la estimación no se puede llevar a cabo. De hecho, en los resultados ni siquiera aparecen todos los coeficientes solicitados.

Una última opción es obtener los estimadores MVIC estimando por MC3E de forma iterativa. Por un lado, Schmidt (1976) menciona y demuestra (p. 224) el siguiente teorema: cuando Σ no está restringida, la distribución asintótica de los estimadores MC3E y MVIC es la misma. Pero por el otro Dhrymes (1973) dice, comparando ambos métodos de estimación, que “the two sets of equations are similar, the difference being that the two estimators purge the jointly dependent variables differently. Hence, even if three stage least squares is iterated, it will not give an estimator which is the same as the maximum likelihood one” [los dos sets de ecuaciones son similares, siendo la diferencia que ambos estimadores purgan las variables mutuamente dependientes de forma diferente. Por lo tanto, incluso si MC3E es iterado, no va a otorgar un estimador que sea el mismo al MVIC] (Dhrymes, 1973, p. 357). De cualquier forma, probamos la estimación con el paquete `systemfit`:

```

mvic3SLS <- systemfit(sistema, method = "3SLS", data = kmenta_13_1,
                     inst = instrum,
                     control = systemfit.control(maxiter = 100))
summary(mvic3SLS, residCov = FALSE, equations = FALSE)

##
## systemfit results
## method: iterated 3SLS
##
## convergence achieved after 6 iterations
##
##          N DF      SSR  detRCov   OLS-R2 McElroy-R2

```

```
## system 40 33 178.16 0.983079 0.667753 0.814118
##
##      N DF      SSR      MSE      RMSE      R2      Adj R2
## dem 20 17  65.7291 3.86642 1.96632 0.754847 0.726005
## of  20 16 112.4313 7.02696 2.65084 0.580659 0.502033
##
##
## Coefficients:
##              Estimate Std. Error  t value  Pr(>|t|)
## dem_(Intercept) 94.6333039  7.9208383 11.94738 1.0762e-09 ***
## dem_P_t         -0.2435565  0.0964843 -2.52431 0.02183240 *
## dem_D_t          0.3139918  0.0469437  6.68869 3.8109e-06 ***
## of_(Intercept)  52.6618215 12.8051061  4.11256 0.00081480 ***
## of_P_t           0.2265865  0.1074593  2.10858 0.05109577 .
## of_F_t           0.2233723  0.0467735  4.77562 0.00020634 ***
## of_A_t           0.3800062  0.0720098  5.27715 7.5166e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Finalmente, el modelo estimado con MC3E iterado (6 veces) es:

$$Q_t = 94.6333 - 0.2436P_t + 0.3140D_t + \tilde{\mu}_{1t}$$

$$Q_t = 52.6618 + 0.2266P_t + 0.2234F_t + 0.3800A_t + \tilde{\mu}_{2t}$$

Nuevamente, los coeficientes estimados son todos significativos, pero esta vez P_t en la ecuación de oferta (of_P_t) lo es al 10 %. El R^2 de McElroy es el más alto de todos los modelos estimados (0,81), lo cual nos llevaría a elegir este modelo por sobre los otros dos. Para una comparación adicional que nos ayude a elegir, podemos recurrir al criterio de información de Akaike corregido, ya que el número de observaciones (40) es más bien pequeño:

```
library(qpcR)
aiccs <- c(AICc(mc2e), AICc(mc3e), AICc(mvic3SLS))
modelos_names <- c('MC2E', 'MC3E', 'MVIC')
df_aics <- data.frame('Modelo'=modelos_names, 'AICc'=aiccs)

library(knitr)
kable(df_aics)
```

Modelo	AICc
MC2E	132.6041
MC3E	107.7884
MVIC	106.7938

Como el modelo con el AICc más bajo es el MVIC (mvic3SLS), este es el mejor.

Inciso d)

En el inciso **a)** vimos que una ecuación estructural puede ser identificada omitiendo algunas de las variables del sistema de la ecuación bajo análisis. Omitir una variable de una ecuación lineal es equivalente a especificar que el coeficiente correspondiente sea igual a 0 (restricción lineal). Teniendo eso en cuenta, debería ser posible identificar un modelo con ecuaciones no identificadas aplicando previamente restricciones en algunos de los elementos de la matriz de varianzas y covarianzas de los errores de la forma estructural. Para ejemplificar cómo funcionaría este procedimiento, nos guiamos por Kmenta (1986). Partimos de un sencillo modelo estructural de demanda y oferta:

$$\begin{aligned}Q_t &= \alpha_1 + \alpha_2 P_t + \alpha_3 Y_t + \mu_{1t} \\Q_t &= \beta_1 + \beta_2 P_t + \mu_{2t}\end{aligned}$$

La ecuación de demanda (la primera) no está identificada, mientras que la ecuación de oferta (la segunda) está exactamente identificada. La forma reducida es:

$$\begin{aligned}Q_t &= \pi_{11} + \pi_{12} Y_t + \nu_{1t} \\P_t &= \pi_{21} + \pi_{22} Y_t + \nu_{2t}\end{aligned}$$

De manera que se establecen las siguientes relaciones entre los parámetros de ambas formas:

$$\begin{aligned}\pi_{11} &= \alpha_1 + \alpha_2 \pi_{21} & \pi_{12} &= \alpha_2 \pi_{22} + \alpha_3 \\ \pi_{11} &= \beta_1 + \beta_2 \pi_{21} & \frac{\pi_{12}}{\pi_{22}} &= \beta_2\end{aligned}$$

Estas ecuaciones se pueden resolver para los coeficientes β_i , luego de estimar por MCO los coeficientes π_{ij} de la forma reducida, pero no para los α_i , justamente porque corresponden a una ecuación no identificada. Pero veamos qué sucede si añadimos una restricción adicional sobre la matriz de varianzas y covarianzas de los errores de la forma estructural. Específicamente:

$$E(\mu_{1t}\mu_{2t}) = 0$$

Para todo t . Como de la relación entre la forma reducida y la forma estructural surge que:

$$\begin{aligned}\mu_{1t} &= \nu_{1t} - \alpha_2 \nu_{2t} \\ \mu_{2t} &= \nu_{1t} - \beta_2 \nu_{2t}\end{aligned}$$

Entonces $E(\mu_{1t}\mu_{2t}) = 0$ es equivalente a:

$$E(\nu_{1t} - \alpha_2 \nu_{2t})(\nu_{1t} - \beta_2 \nu_{2t}) = 0$$

Que si hacemos $E(\nu_{it}\nu_{jt}) = \psi_{ij} \quad : \quad i, j = 1, 2$ (siendo los ψ_{ij} elementos de la matriz de varianzas y covarianzas de los errores de la forma reducida) se puede expresar:

$$\psi_{11} - \alpha_2 \psi_{12} - \beta_2 \psi_{12} + \alpha_2 \beta_2 \psi_{22} = 0$$

Como ya sabemos que $\frac{\pi_{12}}{\pi_{22}} = \beta_2$, podemos obtener α_2 :

$$\alpha_2 = \frac{\pi_{12}\psi_{12} - \pi_{22}\psi_{11}}{\pi_{12}\psi_{22} - \pi_{22}\psi_{12}}$$

Luego, como:

$$\begin{aligned}\alpha_1 &= \pi_{11} - \alpha_2 \pi_{21} \\ \alpha_3 &= \pi_{12} - \alpha_2 \pi_{22}\end{aligned}$$

Podemos obtener los valores de α_1 y α_2 también. De esta forma, la ecuación de demanda, previamente no identificada, pasa a estar identificada, gracias a la restricción aplicada sobre la matriz de varianzas y covarianzas de los errores.

Inciso e)

Primero exponemos un breve resumen de la notación y de la derivación matricial planteadas por Phillips y Wickens (1978). Luego presentamos las principales conclusiones. Los autores parten del siguiente sistema de n ecuaciones lineales simultáneas:

$$By_t + Cx_t + u_t \quad (t = 1, \dots, T) \quad (6)$$

Donde B es la matriz de coeficientes b_i de las variables endógenas, y_i , y C es la matriz de coeficientes c_i de las variables predeterminadas, x_i . Además, $u \sim N(0, \Sigma)$. La ecuación (6) se puede reescribir de forma más compacta como:

$$BY' + CX' = U'$$

Donde $Y' = [y_1, \dots, y_T]$, $X' = [x_1, \dots, x_T]$ y $U' = [u_1, \dots, u_T]$. O también:

$$AW' = U'$$

Donde $A = (B : C)$ y $W = (Y : X)$.

También es importante mencionar la notación de la forma reducida genérica:

$$Y = X\hat{P}' + \hat{V} = \bar{Y} + \hat{V}$$

Donde $\hat{P} = -\hat{B}^{-1}\hat{C}$ es la matriz de coeficientes estimados de la forma reducida y $\hat{V} = U\hat{B}^{-1}$. Por otro lado, la ecuación i -ésima de (6) es:

$$y_i = Y_i b_i + X_i c_i + u_i \quad (i = 1, \dots, n)$$

O en forma más compacta:

$$y_i = Z_i \delta_i + u_i$$

Donde $\delta_i = (b'_i, c'_i)'$ y $Z_i = (Y_i : X_i)$. Entonces, el sistema completo se puede escribir como:

$$y = Z\delta + u$$

Donde $y = (y'_1, \dots, y'_n)'$, $\delta = (\delta'_1, \dots, \delta'_n)'$, $u = (u'_1, \dots, u'_n)'$ y:

$$Z = \begin{bmatrix} Z_1 & 0 & \dots & 0 \\ 0 & Z_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Z_n \end{bmatrix}$$

A su vez se define $\bar{Z}_i = [\bar{Y}_i : X_i]$, y por ende:

$$\bar{Z} = \begin{bmatrix} \bar{Z}_1 & 0 & \dots & 0 \\ 0 & \bar{Z}_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \bar{Z}_n \end{bmatrix}$$

Phillips y Wickens (1978) buscan derivar los estimadores MVIC de los elementos sin restringir ($\neq 0$) de A y Σ . Para ello construyen la función de log-verosimilitud de $AW' = U'$, denotada $\ln L$, a partir de la función de densidad conjunta de las perturbaciones estructurales U . Luego diferencian $\ln L$ respecto a los elementos desconocidos de A (B y C) y de Σ^{-1} (σ_{ij}). Se obtienen así 3 Condiciones de Primer Orden (CPO):

$$\frac{\partial \ln L}{\partial B} = 0, \quad \frac{\partial \ln L}{\partial C} = 0, \quad \frac{\partial \ln L}{\partial \sigma_{ij}} = 0$$

A partir de estas CPO se llega a que los estimadores MVIC de δ satisfacen:

$$\hat{\delta} = [\bar{Z}' (\hat{\Sigma}^{-1} \otimes I) Z]^{-1} \bar{Z}' (\hat{\Sigma}^{-1} \otimes I) y \quad (7)$$

Además, partiendo de $\frac{\partial \ln L}{\partial \sigma_{ij}} = 0$ se arriba a:

$$\hat{\Sigma} = T^{-1} \hat{A} W' W \hat{A}' \quad (8)$$

Conclusión 1: Los estimadores MVIC de δ y de Σ se pueden obtener usando una iteración basada en (7) y (8):

A partir de *cualquier* par de estimadores consistentes de δ y Σ se puede construir \hat{P} , por ende \bar{Y} , y en consecuencia \bar{Z} . Con esta estimación de \bar{Z} y la estimación inicial de Σ es posible resolver (7) para obtener una nueva estimación de δ . Una nueva estimación de Σ se obtiene sustituyendo esta última estimación de δ en (8). Las iteraciones continúan hasta que se logre convergencia (lo cual no está garantizado). Las estimaciones resultantes serán las estimaciones MVIC de δ y Σ .

Conclusión 2: El estimador MVIC se puede interpretar como un estimador de variable instrumental iterada en la que se revisa la matriz de variables instrumentales en cada iteración:

Si definimos el set de variables instrumentales siguiente:

$$H = (\hat{\Sigma}^{-1} \otimes I_T) \bar{Z}$$

Podemos reescribir (7) como:

$$\hat{\delta} = (H' Z)^{-1} H' y$$

Que tiene la familiar forma de una variable instrumental.

Conclusión 3: El estimador MC3E es la primera iteración del estimador MVIC:

Suponiendo que nuestra estimación inicial de Σ se deriva de (8) usando estimaciones por MC2E de A y que nuestro valor inicial de \bar{Y} es:

$$\bar{Y} = X(X'X)^{-1}X'Y$$

Entonces:

$$\bar{Z}_i = [\bar{Y}_i : X_i] = [X(X'X)^{-1}X'Y_i : X_i] = X(X'X)^{-1}X'Z_i$$

Y por ende:

$$\bar{Z} = [I \otimes X(X'X)^{-1}X'] Z \quad (9)$$

Sustituyendo (9) en (7) se obtiene el estimador MC3E de δ :

$$\hat{\delta}_{MC3E} = \left\{ Z' [\hat{\Sigma}^{-1} \otimes X(X'X)^{-1}X'] Z \right\}^{-1} Z' [\hat{\Sigma}^{-1} \otimes X(X'X)^{-1}X'] y$$

Conclusión 4: MC3E es, asintóticamente, tan eficiente como MVIC:

Sustituyendo $y = Z\delta + u$ en (7) y reordenando se obtiene:

$$T^{1/2}(\hat{\delta} - \delta) = [T^{-1}\bar{Z}' (\hat{\Sigma}^{-1} \otimes I) Z]^{-1} T^{-1/2}\bar{Z}' (\hat{\Sigma}^{-1} \otimes I) u$$

$T^{1/2}(\hat{\delta} - \delta)$ tiene una distribución asintótica $N(0, V_{MVIC})$, donde:

$$V_{MVIC} = \left\{ \text{plim } T^{-1} Z' [\Sigma^{-1} \otimes X(X'X^{-1})X'] Z \right\}^{-1}$$

Se puede demostrar que la distribución asintótica de $T^{-1/2}(\hat{\delta}_{MC3E} - \delta)$ es, a su vez, $N(0, V_{MVIC})$.

Posteriormente, Phillips y Wickens (1978) arriban a dos conclusiones adicionales: (i) si Σ es diagonal, MC3E es equivalente a aplicar MC2E a cada ecuación por separado, y (ii) si cada ecuación está exactamente identificada, el estimador MC3E es equivalente a aplicar MC2E o Mínimos Cuadrados Indirectos a cada ecuación por separado.

Inciso f)

Primero generamos las 1000 observaciones de las variables del modelo, teniendo en cuenta que la verdadera forma reducida del mismo es:

$$\begin{aligned} Q_t &= 75.25 + 0.1125D_t + 0.125F_t + 0.225A_t + \nu_{1t} \\ P_t &= 85.00 + 0.75D_t - 0.50F_t - 0.90A_t + \nu_{2t} \end{aligned} \quad (10)$$

Y que:

$$\begin{aligned} \nu_{1t} &= 2\varepsilon_{1t} \\ \nu_{2t} &= -0.5\nu_{1t} + \varepsilon_{2t} \end{aligned}$$

Donde ε_{1t} y ε_{2t} son dos conjuntos de 1000 valores cada uno obtenidos aleatoriamente a partir de una distribución $N(0, 1)$. Comenzamos obteniendo los dos conjuntos mencionados y luego calculando con ellos ν_{1t} y ν_{2t} :

```
set.seed(99)

epsilon_1t <- rnorm(1000)
epsilon_2t <- rnorm(1000)

nu_1t <- 2*epsilon_1t
nu_2t <- -0.5*nu_1t+epsilon_2t
```

Las observaciones de D_t , F_t y A_t son las mismas que se encuentran en la Tabla 13-1 en Kmenta (1986), pero replicadas 50 veces para obtener las 1000 requeridas:

```
kmentaDFA <- do.call("rbind", replicate(50,
  as.vector(kmenta_13_1[,3:5]),simplify = F))

head(kmentaDFA)
```

```
## # A tibble: 6 x 3
##   D_t    F_t    A_t
##   <dbl> <dbl> <dbl>
## 1  87.4   98      1
## 2  97.6  99.1     2
## 3  96.7  99.1     3
## 4  98.2  98.1     4
## 5  99.8 111.     5
## 6 100.  108.     6
```

```
nrow(kmentaDFA)
```

```
## [1] 1000
```

Posteriormente calculamos las observaciones de Q_t y P_t en base a (10):

```
Q_random <- 75.25+0.1125*kmentaDFA[,1]+
  0.125*kmentaDFA[,2]+0.225*kmentaDFA[,3]+nu_1t
P_random <- 85.00+0.75*kmentaDFA[,1]-0.50*kmentaDFA[,2]-
  0.90*kmentaDFA[,3]+nu_2t
```

Seguidamente armamos el data frame con el que vamos a trabajar uniendo las 1000 observaciones de cada una de las 2 variables endógenas (Q_t y P_t , representadas por Q_random y P_random respectivamente) y las 3 predeterminadas (D_t , F_t y A_t , incluidas en kmentaDFA):

```
data_random <- as.data.frame(cbind(Q_random,P_random,
                                   kmentaDFA[,1:3]))

names(data_random)[1:2] <- c("Q_t","P_t")
head(data_random)
```

```
##           Q_t           P_t    D_t    F_t A_t
## 1  97.98543 100.69961  87.4  98.0   1
## 2 100.02682 107.35294  97.6  99.1   2
## 3  99.36691 105.10183  96.7  99.1   3
## 4 100.34772 105.29073  98.2  98.1   4
## 5 100.72682 100.05183  99.8 110.8   5
## 6 101.67660  99.87366 100.5 108.2   6
```

Continuamos con el ejercicio extrayendo las 20 muestras de 50 observaciones (filas) cada una:

```
muestras <- replicate(20, data_random[sample(nrow(data_random), 50),])
```

Cada una de las 20 columnas de `muestras` es una lista compuesta a su vez de 5 listas (una por cada variable) que contienen, individualmente, 50 elementos (las observaciones de cada variable). Posteriormente deberemos transformar estas listas (es decir, las 20 muestras) en data frames. El resultado de hacer esto último con una única columna de `muestras` es el siguiente:

```
head(as.data.frame(muestras[,1]))
```

```
##           Q_t           P_t    D_t    F_t A_t
## 1 104.51330  85.37660  96.4 110.5  17
## 2  97.65838 106.51812  96.7  99.1   3
## 3 100.31080 101.60174  99.8 110.8   5
## 4 107.26317  89.65828  88.9 100.6  10
## 5  97.91046  99.92180  87.4  98.0   1
## 6 100.59408  99.46091  99.8 110.8   5
```

Pasamos ahora a plantear el sistema de ecuaciones estructurales:

```
demanda_random <- Q_t ~ P_t + D_t
oferta_random <- Q_t ~ P_t + F_t + A_t
sistema_random <- list(demr = demanda_random, ofr = oferta_random)
instrum_random <- ~ D_t + F_t + A_t
```

Proseguimos con las 20 estimaciones por MCO del sistema estructural en base a cada una de las 20 muestras:


```
library(systemfit)

for (i in c(1:20)) {
  nombre <- paste("mco_", i, sep = "")
  assign(nombre, systemfit(sistema_random, method = "OLS",
                           data = as.data.frame(muestras[,i])))
}
```

Como para más adelante poder realizar los contrastes de Hausman se necesita que los datos que usamos para las estimaciones por MC2E tengan un nombre, no encontramos otra alternativa más que hacer cada una de las estimaciones de forma individual, habiendo primero asignado las 20 muestras de `muestras` a una variable independiente.

```
for (i in 1:20) {
  nombre <- paste("dataf_", i, sep = "")
  assign(nombre, as.data.frame(muestras[,i]))
}

mc2e_1 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_1)
mc2e_2 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_2)
mc2e_3 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_3)
mc2e_4 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_4)
mc2e_5 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_5)
mc2e_6 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_6)
mc2e_7 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_7)
mc2e_8 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_8)
mc2e_9 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_9)
mc2e_10 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_10)
mc2e_11 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_11)
mc2e_12 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_12)
mc2e_13 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_13)
mc2e_14 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_14)
```

```
mc2e_15 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_15)
mc2e_16 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_16)
mc2e_17 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_17)
mc2e_18 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_18)
mc2e_19 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_19)
mc2e_20 <- systemfit(sistema_random, method = "2SLS", inst = instrum_random,
                    data = dataf_20)
```

Por ejemplo, comparemos los resultados de la estimación por MCO y de la estimación por MC2E en base a la muestra número 12:

```
mco_12
```

```
##
## systemfit results
## method: OLS
##
## Coefficients:
## demr_(Intercept)      demr_P_t      demr_D_t  ofr_(Intercept)
##      105.9362739      -0.3876143      0.3449367      71.9702778
##      ofr_P_t      ofr_F_t      ofr_A_t
##      0.0415091      0.2130815      0.3755924
```

```
mc2e_12
```

```
##
## systemfit results
## method: 2SLS
##
## Coefficients:
## demr_(Intercept)      demr_P_t      demr_D_t  ofr_(Intercept)
##      101.9487369      -0.3248030      0.3214128      65.7231104
##      ofr_P_t      ofr_F_t      ofr_A_t
##      0.0986445      0.2183537      0.3783359
```

Ahora sí, ya estamos en condiciones de hacer los 20 tests de Hausman. Para ello es necesario en primera instancia crear dos listas que contengan cada una las 20 estimaciones correspondientes a los dos métodos empleados. De esta forma podremos hacer los tests de Hausman mediante un ciclo `for`:

```
mc2e_lista <- list(mc2e_1,mc2e_2,mc2e_3,mc2e_4,mc2e_5,mc2e_6,mc2e_7,mc2e_8,
                  mc2e_9,mc2e_10,mc2e_11,mc2e_12,mc2e_13,mc2e_14,mc2e_15,
                  mc2e_16,mc2e_17,mc2e_18,mc2e_19,mc2e_20)

mco_lista <- list(mco_1,mco_2,mco_3,mco_4,mco_5,mco_6,mco_7,mco_8,mco_9,
                 mco_10,mco_11,mco_12,mco_13,mco_14,mco_15,mco_16,mco_17,
                 mco_18,mco_19,mco_20)

for (i in 1:20) {
  nombre_h <- paste("hausman_", i, sep = "")
  assign(nombre_h, hausman.systemfit(mc2e_lista[[i]], mco_lista[[i]]))
}
```

Considerando los estimadores MCO y MC2E, usaremos el contraste de Hausman para para evaluar la consistencia de las estimaciones por MCO. Las hipótesis son las que se muestran a continuación:

- H_0 : Ambos estimadores son consistentes.
- H_1 : Al menos uno de los estimadores no es consistente.

El test calcula con una chi cuadrado las diferencias en las estimaciones de ambos modelos, a través de la siguiente fórmula:

$$H = (\beta_{MC2E} - \beta_{MCO})'(V_{MC2E} - V_{MCO})^{-1}(\beta_{MC2E} - \beta_{MCO}), \quad H \sim X_n^2$$

Donde los β_i son vectores de estimaciones, las V_i son matrices de varianzas y covarianzas y n es el número de variables incluida la constante (en caso de que exista).

Si las diferencias, aunque sean altas, no son sistemáticas ... entonces ambos estimadores son consistentes Si las diferencias son sistemáticas entonces nuestra hipótesis no se cumple, ambos no son consistentes y ahora tenemos un dilema: pensar que el modelo está mal especificado en ambos casos o quedarnos con el estimador consistente [el de variables instrumentales, MC2E] Si el valor de la prueba es alto (p.e. p-valor menor de 0.05) la hipótesis de diferencias no sistemáticas se rechaza, por lo que: o se reelabora el modelo o se elige al que se considera consistente. (Montero, 2005, p. 1)

De esta forma, y para dar un ejemplo, veamos el resultado de aplicar el contraste de Hausman a las estimaciones `mco_12` y `mc2e_12`:

```
hausman.systemfit(mc2e_12,mco_12)

##
## Hausman specification test for consistency of the 3SLS estimation
##
## data:  dataf_12
## Hausman = 36.516, df = 7, p-value = 5.793e-06
```

En este caso, como el p-value es menor a 0.05, rechazamos la Hipótesis Nula de que ambos estimadores son consistentes, y nos quedamos con el estimador consistente (el que surge de aplicar MC2E). Vale la pena mencionar que `hausman.systemfit` es una función planeada para probar la consistencia del estimador MC3E, pero se puede forzar para verificar la consistencia del estimador MCO.

Finalmente, tabulamos los resultados:

```
nombres_h <- c()
for (i in 1:20) {
  new <- paste('MC2E_',i, " vs MCO_", i, sep = "")
  nombres_h <- c(nombres_h,new)
}

pvalues_h <- c(hausman_1$p.value,hausman_2$p.value,hausman_3$p.value
,hausman_4$p.value,hausman_5$p.value,hausman_6$p.value
,hausman_7$p.value,hausman_8$p.value,hausman_9$p.value
,hausman_10$p.value,hausman_11$p.value,hausman_12$p.value
,hausman_13$p.value,hausman_14$p.value,hausman_15$p.value
,hausman_16$p.value,hausman_17$p.value,hausman_18$p.value
,hausman_19$p.value,hausman_20$p.value)
df_hausman <- data.frame("Test de Hausman"=nombres_h,
                        "p-value"=pvalues_h)

library(knitr)

kable(df_hausman)
```

Test.de.Hausman	p.value
MC2E_1 vs MCO_1	0.0001554
MC2E_2 vs MCO_2	0.0014295
MC2E_3 vs MCO_3	0.0000102
MC2E_4 vs MCO_4	0.0000551
MC2E_5 vs MCO_5	0.0000220
MC2E_6 vs MCO_6	0.0000976
MC2E_7 vs MCO_7	0.0003138

Test.de.Hausman	p.value
MC2E_8 vs MCO_8	0.0000195
MC2E_9 vs MCO_9	0.0002612
MC2E_10 vs MCO_10	0.0001787
MC2E_11 vs MCO_11	0.0002484
MC2E_12 vs MCO_12	0.0000058
MC2E_13 vs MCO_13	0.0064136
MC2E_14 vs MCO_14	0.0005203
MC2E_15 vs MCO_15	0.0604416
MC2E_16 vs MCO_16	0.0010750
MC2E_17 vs MCO_17	0.0000399
MC2E_18 vs MCO_18	0.0000331
MC2E_19 vs MCO_19	0.0000173
MC2E_20 vs MCO_20	0.0010668

Dado que el p-value en todos los casos es menor a 0.05 (a excepción de un único caso, MC2E_15 vs MCO_15), se concluye que el estimador MC2E es claramente preferido al de MCO, ya que el mismo es sistemáticamente categorizado como inconsistente por el contraste de Hausman. Este resultado era de esperar, debido a que por la teoría sabemos que a la hora de estimar ecuaciones simultáneas el estimador MCO es, efectivamente, sesgado e inconsistente, mientras que el estimador MC2E es consistente aunque asintóticamente ineficiente (ver Figura 1).

Ejercicio IV

Modelo Probit. En base al ejercicio 7.2. en Pérez López (2006), se le pide:

- a) Estimar el modelo planteado utilizando R. Identifique y explique detalladamente el algoritmo de cálculo numérico utilizado para la estimación.
- b) Comparar los resultados con los obtenidos en el ejercicio 7.1 del mismo libro, que fueron obtenidos a partir del empleo del modelo lineal de probabilidad.
- c) Indicar brevemente en qué consisten las diferencias entre los algoritmos Newton - Raphson, BHHH y Quadratic Hill Climbing y qué ventajas se pueden derivar de su empleo.
- d) Calcular la tabla de predicciones para cada una de las observaciones y analizarla cuidadosamente.
- e) Explicar detalladamente en qué consiste el método Heckit.

Respuesta:

Inciso a)

Pérez López (2006) en el ejercicio 7.1 de su libro, presenta una base de datos de 753 mujeres casadas para el año 1975. De estas mujeres, 428 mujeres participaron en el mercado laboral en algún momento de dicho año, mientras que el resto (325) no lo hizo.

Las variables que se consideraron como influyentes en la probabilidad de que la mujer trabaje (`inlf=1`) fueron las siguientes:

- `nwifeinc`: otras fuentes de ingresos.
- `educ`: años de educación.
- `exper` y `expersq`: años de experiencia laboral y su cuadrado.
- `age`: edad.
- `kidslt6`: número de hijos menores a 6 años.
- `kidsge6`: número de hijos mayores a 6 años.

Ahora bien, en base a lo presentado anteriormente, pasamos a estimar los modelos solicitados:

```
library(readxl)

mujeres <- read_xlsx(mujeres_path,
                     col_names = TRUE)

head(mujeres) # Podemos ver las primeras 6 observaciones de la base de datos

## # A tibble: 6 x 23
##   age  city  educ  exper  expersq  faminc  fatheduc  hours  husage  huseduc  hushrs
##   <dbl> <dbl> <dbl> <dbl>   <dbl>   <dbl>   <dbl> <dbl>  <dbl>   <dbl>   <dbl>
## 1   32    0    12    14     196   16310     7   1610    34     12   2708
## 2   30    1    12     5     25   21800     7   1656    30      9   2310
## 3   35    0    12    15    225   21040     7   1980    40     12   3072
## 4   34    0    12     6     36    7300     7    456    53     10   1920
## 5   31    1    14     7     49   27300    14   1568    32     12   2000
## 6   54    1    12    33   1089   19495     7   2032    57     11   1040
## # ... with 12 more variables: huswage <dbl>, inlf <dbl>, inlff <dbl>,
## #   kidsge6 <dbl>, kidslt6 <dbl>, lwage <dbl>, motheduc <dbl>, mtr <dbl>,
## #   nwifeinc <dbl>, repwage <dbl>, unem <dbl>, wage <dbl>

tail(mujeres) # Podemos ver las últimas 6 observaciones de la base de datos

## # A tibble: 6 x 23
##   age  city  educ  exper  expersq  faminc  fatheduc  hours  husage  huseduc  hushrs
##   <dbl> <dbl> <dbl> <dbl>   <dbl>   <dbl>   <dbl> <dbl>  <dbl>   <dbl>   <dbl>
## 1   36    0    12     4     16    5330    12     0    39     12   3120
## 2   40    1    13     5     25   28200    10     0    43     16   3020
## 3   31    0    12    14    196   10000    12     0    33     12   2056
## 4   43    0    12     4     16    9952     3     0    43     12   2383
## 5   60    1    12    15    225   24984    12     0    55      8   1705
## 6   39    1     9    12    144   28363     7     0    48     12   3120
## # ... with 12 more variables: huswage <dbl>, inlf <dbl>, inlff <dbl>,
## #   kidsge6 <dbl>, kidslt6 <dbl>, lwage <dbl>, motheduc <dbl>, mtr <dbl>,
## #   nwifeinc <dbl>, repwage <dbl>, unem <dbl>, wage <dbl>

# Estimamos el Modelo Lineal de Probabilidad que presenta Pérez López
# Porque luego utilizaremos sus resultados para realizar comparaciones

fit_mlp <- lm(inlf~1+nwifeinc+educ+exper+expersq+age+kidslt6+kidsge6,
             data = mujeres)

summary(fit_mlp) # Observamos los resultados

##
```

```
## Call:
## lm(formula = inlf ~ 1 + nwifeinc + educ + exper + expersq + age +
##      kidslt6 + kidsge6, data = mujeres)
##
## Residuals:
##      Min        1Q    Median        3Q        Max
## -0.93432 -0.37526  0.08833  0.34404  0.99417
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.5855192  0.1541780   3.798 0.000158 ***
## nwifeinc     -0.0034052  0.0014485  -2.351 0.018991 *
## educ         0.0379953  0.0073760   5.151 3.32e-07 ***
## exper        0.0394924  0.0056727   6.962 7.38e-12 ***
## expersq      -0.0005963  0.0001848  -3.227 0.001306 **
## age         -0.0160908  0.0024847  -6.476 1.71e-10 ***
## kidslt6     -0.2618105  0.0335058  -7.814 1.89e-14 ***
## kidsge6      0.0130122  0.0131960   0.986 0.324415
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4271 on 745 degrees of freedom
## Multiple R-squared:  0.2642, Adjusted R-squared:  0.2573
## F-statistic: 38.22 on 7 and 745 DF,  p-value: < 2.2e-16
```

Antes de realizar la estimación del modelo Probit (y el Logit también), se realiza una breve descripción teórica de estos modelos (con previa mención del MLP). Siguiendo a Moneta Pizarro (2021), para el MLP se comenzará suponiendo para cada individuo i una relación lineal entre la variable dependiente y un conjunto de variables explicativas, más un error aleatorio:

$$Y_i = X_i' \beta + \mu_i$$

Y suponiendo que $Y_i \sim \text{Bernoulli}(p_i)$, tomando valor esperado condicional a los valores de las X y bajo el supuesto de que $E(\mu_i) = 0$, entonces:

$$E(Y_i | X_i) = X_i' \beta$$

Y recordando que la esperanza de una variable que se distribuye Bernoulli es igual a p , entonces:

$$p_i = P(Y_i = 1 | X_i) = X_i' \beta$$

El cual es el modelo lineal de probabilidad, que tiene ciertas limitaciones, tales como:

1. Supone que existe una relación lineal entre la probabilidad de éxito y las variables explicativas.
2. Para ciertos valores de X , las probabilidades estimadas pueden ser negativas o mayores a 1, lo cual no cumple con la teoría estadística.

3. La varianza del error no es constante, estamos en presencia de heterocedasticidad.
4. Como se supuso que Y_i es Bernoulli, entonces no se cumple la hipótesis de normalidad necesaria para estimaciones por intervalos o pruebas de hipótesis, ni se puede utilizar el estimador por máxima verosimilitud bajo el supuesto de normalidad.

Por lo que aquí es donde aparecen los modelos Logit y Probit, los cuales no siguen funciones lineales, sino una relación funcional que evite los problemas del MLP. Establecen la siguiente relación:

$$p_i = P(Y_i = 1|X_i) = F(X_i'\beta)$$

En donde F es la función que toma valores estrictamente entre 0 y 1 para todos los reales, es decir:

- $F(-\infty) \rightarrow 0$
- $F(\infty) \rightarrow 1$

Por lo tanto, tendremos que:

- Logit utiliza la función de probabilidad acumulada logística:

$$p_i = F(X_i'\beta) = \frac{e^{X_i'\beta}}{1 + e^{X_i'\beta}}$$

- Probit utiliza la función de probabilidad acumulada normal estándar:

$$p_i = F(X_i'\beta) = \Phi(X_i'\beta) = \int_{-\infty}^{X_i'\beta} \phi(z) dz$$

En donde $\phi(z)$ es la función de densidad de la normal estándar y $\Phi(z)$ su correspondiente densidad acumulada.

Luego de la presentación del marco teórico, continuamos con la estimación de los modelos Probit y Logit:

```
# Estimamos el modelo Probit

fit_prob <- glm(inlf~1+nwifeinc+educ+exper+expersq+age+kidslt6+kidsge6,
               family = binomial(link = "probit"),
               data = mujeres)

summary(fit_prob) # Observamos los resultados

##
## Call:
## glm(formula = inlf ~ 1 + nwifeinc + educ + exper + expersq +
##      age + kidslt6 + kidsge6, family = binomial(link = "probit"),
```

```
##      data = mujeres)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q        Max
## -2.2156   -0.9151    0.4315    0.8653    2.4553
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.2700736  0.5080782   0.532  0.59503
## nwifeinc     -0.0120236  0.0049392  -2.434  0.01492 *
## educ          0.1309040  0.0253987   5.154 2.55e-07 ***
## exper         0.1233472  0.0187587   6.575 4.85e-11 ***
## expersq      -0.0018871  0.0005999  -3.145  0.00166 **
## age          -0.0528524  0.0084624  -6.246 4.22e-10 ***
## kidslt6      -0.8683247  0.1183773  -7.335 2.21e-13 ***
## kidsge6       0.0360056  0.0440303   0.818  0.41350
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1029.7  on 752  degrees of freedom
## Residual deviance:  802.6  on 745  degrees of freedom
## AIC: 818.6
##
## Number of Fisher Scoring iterations: 4
```

Los resultados nos muestran que casi todos los coeficientes son en general aceptables al 5%. Aquellos que no lo son, son la constante y la variable representativa de hijos entre 6 y 18 años.

Para analizar si el modelo posee un buen ajuste global, lo realizamos a partir de la siguiente prueba, comparando la “*Residual deviance*” con las siguientes hipótesis, utilizando un estadístico “*Chi Cuadrado*”:

- H_0 : el modelo presenta un buen ajuste global
- H_1 : el modelo no presenta un buen ajuste global

Por lo tanto:

```
# 745 son los grados de libertad
qchisq_fit_prob <- qchisq(0.95, 745); qchisq_fit_prob
```

```
## [1] 809.6087
```

```
rdfit_prob <- deviance(fit_prob); rdfit_prob
```

```
## [1] 802.6044
```

```
rdo_rd_prob <- if(rdfit_prob < qchisq_fit_prob){
  "No rechazo H_0"
}else{
  "Rechazo H_0"
}

print(rdo_rd_prob)
```

```
## [1] "No rechazo H_0"
```

Es un buen modelo ya que la “*Residual deviance*”= 802.6043863, un valor menor al valor crítico de la “*Chi Cuadrado*” (809.6087369).

Ahora bien, siguiendo a Pérez López (2006), se plantea el modelo Logit:

```
fit_log <- glm(inlf~1+nwifeinc+educ+exper+expersq+age+kidslt6+kidsge6,
  family = binomial(link = "logit"),
  data = mujeres)

summary(fit_log) # Observamos los resultados
```

```
##
## Call:
## glm(formula = inlf ~ 1 + nwifeinc + educ + exper + expersq +
##      age + kidslt6 + kidsge6, family = binomial(link = "logit"),
##      data = mujeres)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.1770  -0.9063   0.4473   0.8561   2.4032
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.425452   0.860365   0.495  0.62095
## nwifeinc    -0.021345   0.008421  -2.535  0.01126 *
## educ        0.221170   0.043439   5.091 3.55e-07 ***
## exper       0.205870   0.032057   6.422 1.34e-10 ***
## expersq     -0.003154   0.001016  -3.104  0.00191 **
## age        -0.088024   0.014573  -6.040 1.54e-09 ***
## kidslt6     -1.443354   0.203583  -7.090 1.34e-12 ***
## kidsge6      0.060112   0.074789   0.804  0.42154
```

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 1029.75  on 752  degrees of freedom
## Residual deviance:  803.53  on 745  degrees of freedom
## AIC: 819.53
##
## Number of Fisher Scoring iterations: 4
```

Observamos una vez más que los coeficientes de las mismas variables que no eran significativas en el modelo Probit, no lo son en el modelo Logit.

Realizamos ahora la prueba “*Chi Cuadrado*” para saber si es un buen modelo en el ajuste global:

```
qchisq_fit_log <- qchisq(0.95, 745); qchisq_fit_log
```

```
## [1] 809.6087
```

```
rdfit_log <- deviance(fit_log); rdfit_log
```

```
## [1] 803.5303
```

```
rdo_rd_log <- if(rdfit_log < qchisq_fit_log){
  "No rechazo H_0"
}else{
  "Rechazo H_0"
}

print(rdo_rd_log)
```

```
## [1] "No rechazo H_0"
```

Es un buen modelo ya que la “*Residual deviance*”= 803.5303022, es menor al valor crítico de la “*Chi Cuadrado*” 809.6087369.

Antes de pasar al siguiente inciso, se pueden comparar los modelos a través de un criterio de selección estadística, en este caso utilizaremos el criterio de Akaike, este será:

```
aic1 <- AIC(fit_prob)/nrow(mujeres); aic1
```

```
## [1] 1.087124
```

```
aic2 <- AIC(fit_log)/nrow(mujeres); aic2
```

```
## [1] 1.088354
```

```
prob_vs_log <- if(aic1 < aic2){
  "Probit"
}else{
  "Logit"
}
```

```
print(prob_vs_log)
```

```
## [1] "Probit"
```

Por lo tanto, es mejor el modelo Probit, ya que posee un menor valor del estadístico de Akaike.

Finalmente, se solicita explicar el algoritmo de cálculo numérico utilizado para la estimación. Como se mencionó antes, los modelos de tipo Probit y Logit son no lineales en los parámetros, por lo tanto no pueden ser estimados mediante MCO. Siguiendo a Moneta Pizarro (2021) se pueden estimar por MV, entonces, recurriendo a la función de verosimilitud de una Bernoulli:

$$L(\beta) = \prod_{i=1}^n p_i^{Y_i} (1 - p_i)^{1-Y_i}$$

Reemplazando a p_i por la función de distribución:

$$L(\beta) = \prod_{i=1}^n F\left[(X'_i\beta)^{Y_i} (1 - F(X'_i\beta))\right]^{1-Y_i}$$

Tomando logaritmo natural:

$$l(\beta) = \ln L(\beta) = \sum_{i=1}^n \left\{ Y_i \ln F(X'_i\beta) + (1 - Y_i) \ln [1 - F(X'_i\beta)] \right\}$$

Por lo que la condición de primer orden para maximizar esta función respecto al vector de parámetros β es:

$$\begin{aligned} \frac{\partial l}{\partial \beta} &= \sum_{i=1}^n \left\{ \frac{Y_i}{F(X'_i\beta)} f(X'_i\beta) + \frac{(1 - Y_i)}{1 - F(X'_i\beta)} [-f(X'_i\beta) X_i] \right\} \\ \frac{\partial l}{\partial \beta} &= \sum_{i=1}^n \left\{ \frac{Y_i - F(X'_i\beta)}{F(X'_i\beta) [1 - F(X'_i\beta)]} \right\} f(X'_i\beta) X_i \end{aligned}$$

En donde $F(z)$ es la función de probabilidad acumulada, la logística para el Logit y la normal estándar para el Probit, y $f(z)$ es la función de densidad correspondiente.

Por lo que, como se dijo, la no linealidad nos lleva a utilizar un algoritmo numérico iterativo, podría ser Newton-Raphson. En particular, R utiliza Fisher Scoring, el cual coincide con el recomendado por Pérez López en su libro. Es el siguiente:

$$\hat{\beta}^{k+1} = \hat{\beta}^k + [I(\hat{\beta}^k)]^{-1} S(\hat{\beta}^k)$$

En donde:

- $\hat{\beta}^{k+1}$ es el valor del vector β en la iteración $k + 1$ del algoritmo
- $\hat{\beta}^k$ es el valor del vector β en la iteración k del algoritmo
- $[I(\hat{\beta}^k)]^{-1}$ es la matriz inversa de varianzas y covarianzas asintótica, valuada en $\hat{\beta}^k$
- $S(\hat{\beta}^k)$ es el vector de primeras derivadas valuado en $\hat{\beta}^k$

Tiene un funcionamiento prácticamente igual que el algoritmo de Newton-Raphson (ya mencionado y explicado en detalle en el inciso **b**) del Ejercicio II), a saber:

1. Se escogen valores iniciales para β^k
2. A partir de dichos valores de β^k se estimarán $[I(\hat{\beta}^k)]^{-1}$ y $S(\hat{\beta}^k)$
3. Con los valores obtenidos anteriormente se calculará β^{k+1} a partir de la fórmula de Fisher Scoring
4. Con el valor obtenido de β^{k+1} se volverán a repetir los pasos anteriores, es decir, se iterará.

¿Hasta cuándo se seguirá iterando? Hasta cumplir un criterio de convergencia, en donde $\beta^{k+1} = \beta^k$

Siguiendo a Jorgensen (1983), Fisher Scoring equivale a resolver de forma iterativa un problema de mínimos cuadrados ponderados. Además, si consultamos en RStudio el método que utiliza en la función `glm`, nos dice que es IWLS (Iteratively Weighted Least-Squares), lo que coincide con la teoría.

?glm # RStudio detalla a glm, allí explica el método utilizado

IWLS es básicamente la suma de cuadrados residuales pero ponderados de la siguiente manera: cuanto más lejos el punto de datos del área de dispersión media, menor será el peso, y cuanto más cerca esté el punto de datos del área de dispersión media, mayor peso tendrá. El algoritmo de Fisher-Scoring da un ajuste considerando el peso relativo de los datos con respecto a la dispersión media.

La diferencia entre Newton-Raphson y Fisher Scoring es que reemplaza la matriz hessiana de Newton-Raphson por su valor esperado, la cual es conocida como la matriz de información de Fisher.

Inciso b)

En este inciso se compararán los resultados del Modelo Lineal de Probabilidad estimado por el autor con los modelos Probit y Logit que nosotros estimamos.

Comenzamos realizando un análisis de los coeficientes estimados en los modelos. Para ello utilizamos la función `coef` y la función `cbind` de la siguiente manera:

```
coefMLP <- coef(fit_mlp)
coefProbit <- coef(fit_prob)
coefLogit <- coef(fit_log)

cbind(coefMLP,coefProbit,coefLogit)
```

##	coefMLP	coefProbit	coefLogit
## (Intercept)	0.5855192245	0.270073572	0.425452376
## nwifeinc	-0.0034051689	-0.012023637	-0.021345175
## educ	0.0379953030	0.130903969	0.221170370
## exper	0.0394923895	0.123347168	0.205869531
## expersq	-0.0005963119	-0.001887067	-0.003154104
## age	-0.0160908061	-0.052852442	-0.088024375
## kidslt6	-0.2618104666	-0.868324680	-1.443354143
## kidsge6	0.0130122347	0.036005611	0.060112222

Lo primero que podemos observar es que en los 3 modelos estimados, los signos de los coeficientes de las variables coinciden siempre. Aquellos coeficientes menores a cero, son así en los 3 modelos, lo mismo ocurre con los coeficientes de signo positivo.

Por otro lado, según Pérez López (2006), existe la siguiente relación entre los coeficientes del MLP, Logit y Probit:

$$\hat{\beta}_{Logit} = 1,6\hat{\beta}_{Probit} = 4\hat{\beta}_{MLP}$$

En base a dicha relación, analizamos si se cumple en los coeficientes de las variables que obtuvimos al estimar los 3 modelos:

```
relacion_coef <- as.data.frame(cbind(4*coefMLP,1.6*coefProbit,coefLogit))
colnames(relacion_coef) <- c("MLP", "Probit", "Logit"); relacion_coef
```

##	MLP	Probit	Logit
## (Intercept)	2.342076898	0.432117715	0.425452376
## nwifeinc	-0.013620676	-0.019237819	-0.021345175
## educ	0.151981212	0.209446351	0.221170370
## exper	0.157969558	0.197355468	0.205869531
## expersq	-0.002385248	-0.003019308	-0.003154104
## age	-0.064363224	-0.084563907	-0.088024375
## kidslt6	-1.047241866	-1.389319488	-1.443354143
## kidsge6	0.052048939	0.057608977	0.060112222

En general, se cumple la relación entre los coeficientes, son bastante similares.

Además, podemos observar la relación de la siguiente manera:

```
coefLogit/coefProbit
```

```
## (Intercept)    nwifeinc      educ      exper      expersq      age
##    1.575320    1.775268    1.689562    1.669025    1.671432    1.665474
##      kidslt6      kidsge6
##    1.662229    1.669524
```

```
coefLogit/coefMLP
```

```
## (Intercept)    nwifeinc      educ      exper      expersq      age
##    0.7266241    6.2684628    5.8209924    5.2128912    5.2893528    5.4704764
##      kidslt6      kidsge6
##    5.5129734    4.6196694
```

```
meanLogProb <- mean(coefLogit/coefProbit); meanLogProb
```

```
## [1] 1.672229
```

```
meanLogMLP <- mean(coefLogit/coefMLP);meanLogMLP
```

```
## [1] 4.86518
```

Vemos que la relación entre el Logit y el Probit se cumple en mejor manera, y al tomar la media del cálculo de la relación de los coeficientes, esta es igual a 1.67.

En el caso de lo que acontece con los coeficientes del MLP y del Logit, la relación está un poco sobrestimada, además puede notarse que la media está más cercana a 5 que a 4.

Ahora bien, en el inciso anterior se utilizó el criterio de selección estadística de *Akaike*, el cual determinó que el modelo Probit era mejor, ya que presentaba un menor valor en este estadístico. Por lo tanto, ya que el modelo Probit fue mejor, comparamos este modelo con el MLP:

```
aic0 <- AIC(fit_mlp)/nrow(mujeres); aic0
```

```
## [1] 1.14978
```

```
prob_vs_MLP <- if(aic1 < aic0){
  "Probit"
}else{
  "MLP"
}
```

```
print(prob_vs_MLP)
```


[1] "Probit"

Lo cual nos señala que es mejor el modelo Probit, ya que posee un menor valor en el estadístico de *Akaike*, por lo que, al considerar los 3 modelos, el mejor de todos es el modelo Probit.

Lo último a realizar es el análisis que nos permita ver cuáles son los efectos parciales de las variaciones de cada variable sobre la probabilidad que las mujeres participen en el mercado laboral. Esto lo analizaremos en los modelos Logit y Probit.

Para el MLP, el análisis es diferente. Al ser un modelo lineal, cada coeficiente nos dice cuál es la variación de la probabilidad de que las mujeres participen en el mercado laboral ante una variación unitaria en alguna variable (considerando las demás constantes). Por ende, la interpretación de los modelos Logit y Probit son diferentes a las interpretaciones de las estimaciones del MLP.

Comenzaremos analizando los efectos parciales de las variables en los modelos Probit y Logit, pero antes de ello, se presenta una breve explicación teórica que nos permita comprender lo que se va a realizar.

En base a Pérez López (2006) el efecto de una variable continua X_j sobre la probabilidad de respuesta $P(Y = 1|X)$ será:

$$\frac{\partial P(Y = 1|X)}{\partial X_j} = g(X\beta) \beta_j$$

En donde $g(X\beta)$ no es más que la función de densidad logística del modelo Logit o la función de densidad normal estándar del modelo Probit, según sea el caso, por lo que lo anterior podemos expresarlo como:

$$\frac{\partial P(Y = 1|X)}{\partial X_j} = f(X'_i\beta) \beta_j$$

¿Cuál es el punto importante? Que la función depende de los valores de la variable X_j para el individuo i , tal es así que "... el efecto parcial de X_j sobre la variable de respuesta no es constante, sino que depende de las características del individuo i para todas las variables explicativas del modelo" (Moneta Pizarro, 2021, p. 159).

Esto lleva a que queden dos alternativas disponibles, una es calcular el efecto parcial para cada individuo y, en base a eso, calcular la media de los resultados obtenidos, mientras que la otra opción es la de simular un "individuo promedio" a partir de calcular los efectos parciales para la media o mediana de los valores de X_j .

Una vez dada esta introducción teórica, nosotros hemos decidido utilizar la última opción, trabajar con el individuo promedio. Lo primero a realizar, será calcular la media de cada una de las variables explicativas:

```
# Extraemos las columnas de las variables que necesitamos
# Es decir, mediante "mujeres[,c(20,3,4,5,1,16,15)]"
# Calcularemos las medias de "nwifeinc", "educ", y las variables
# del modelo
```

```
means <- sapply(mujeres[,c(20,3,4,5,1,16,15)],mean); means
```

```
##      nwifeinc      educ      exper      expersq      age      kidslt6
## 20.1289637 12.2868526 10.6308101 178.0385126 42.5378486 0.2377158
##      kidsge6
##      1.3532537
```

Ahora, lo que se hace es multiplicar el vector de coeficientes estimados por el vector de medias, con esto obtenemos el índice que será el “predictor lineal”:

```
# En ambos casos, trasponemos el vector de coeficientes
```

```
indexP <- t(coefProbit) %*% c(1, means); indexP
```

```
##           [,1]
## [1,] 0.2058387
```

```
indexL <- t(coefLogit) %*% c(1, means); indexL
```

```
##           [,1]
## [1,] 0.3341632
```

Ya que obtuvimos el índice como “predictor lineal”, ahora lo utilizamos como entrada en la función de densidad correspondiente (logística para el Logit y de densidad normal estándar para el Probit):

```
factorP <- dnorm(indexP); factorP
```

```
##           [,1]
## [1,] 0.3905797
```

```
factorL <- dlogis(indexP); factorL
```

```
##           [,1]
## [1,] 0.2473705
```

Ahora bien, lo último es obtener los buscados efectos parciales:

```
effectsP <- factorP %*% coefProbit[2:8]; effectsP
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,] -0.004696188 0.05112843 0.0481769 -0.0007370502 -0.02064309 -0.33915
##           [,7]
## [1,] 0.01406306
```

```
effectsL <- factorL %*% coefLogit[2:8]; effectsL
```

```
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,] -0.005280166 0.05471102 0.05092605 -0.0007802323 -0.02177463 -0.3570432
##           [,7]
## [1,] 0.01486999
```

Con los resultados listos, los presentamos en una salida conjunta para poder analizarlos:

```
effectsPL <- as.data.frame(cbind(c("nwifeinc", "educ", "exper", "expersq",
    "age", "kidslt6", "kidsge6"),
    t(effectsP),
    t(effectsL)))

colnames(effectsPL) <- c("Variables", "Probit", "Logit"); effectsPL
```

```
## Variables Probit Logit
## 1 nwifeinc -0.00469618816056175 -0.00528016630671594
## 2 educ 0.0511284287722782 0.0547110228679166
## 3 exper 0.0481768956700099 0.0509260468461189
## 4 expersq -0.000737050172022229 -0.000780232256680989
## 5 age -0.0206430890609143 -0.0217746326980931
## 6 kidslt6 -0.33914996464587 -0.357043221952433
## 7 kidsge6 0.0140630594115518 0.0148699897820292
```

A continuación, para analizar los resultados extraeremos algunos casos particulares

Primero verificamos el tipo de dato, son de tipo "character"

```
class(effectsPL[2,2])
```

```
## [1] "character"
```

```
class(effectsPL[2,3])
```

```
## [1] "character"
```

Por lo que creamos una variable para cada uno de ellos, donde sean "numeric"

```
educP <- as.numeric(effectsPL[2,2]); educP
```

```
## [1] 0.05112843
```

```
educL <- as.numeric(effectsPL[2,3]);educL
```

```
## [1] 0.05471102
```

```
edadP <- as.numeric(effectsPL[5,2]);edadP
```

```
## [1] -0.02064309
```

```
edadL <- as.numeric(effectsPL[5,3]);edadL
```

```
## [1] -0.02177463
```

```
# Verificamos si son de tipo "numeric"
```

```
class(educP)
```

```
## [1] "numeric"
```

```
class(educL)
```

```
## [1] "numeric"
```

```
class(edadP)
```

```
## [1] "numeric"
```

```
class(edadL)
```

```
## [1] "numeric"
```

Obtuvimos los efectos parciales de cada una de las variables en ambos modelos, Probit y Logit. Analicemos algunas variables:

- Por cada año de educación adicional, según el modelo Probit, la probabilidad de que la mujer participe del mercado laboral aumenta en un 0.0511, mientras que según el modelo Logit, la probabilidad aumenta en un 0.0547.
- Por cada año adicional en la edad de la mujer, la probabilidad de participar en el mercado laboral cae en un valor dado por -0.0206 y -0.0218, según el modelo Probit y Logit respectivamente.

Por su parte, se mencionó que en el MLP, al ser un modelo lineal, se analiza cada coeficiente de forma directa, por lo tanto:

coefMLP

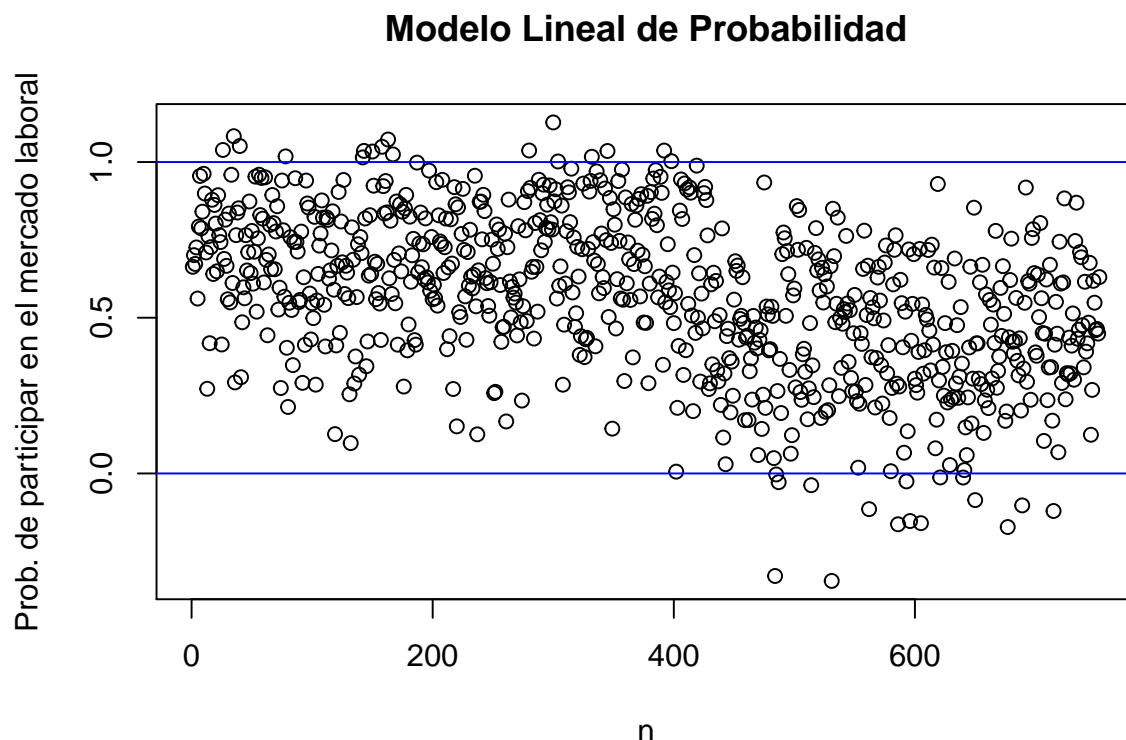
```
## (Intercept)      nwifeinc      educ      exper      expersq
## 0.5855192245 -0.0034051689 0.0379953030 0.0394923895 -0.0005963119
##          age      kidslt6      kidsge6
## -0.0160908061 -0.2618104666 0.0130122347
```

Nos muestra que, ante un aumento de un año en la educación de la mujer, la probabilidad de trabajar crece en un 0,038, mientras que, por cada año que pasa, la probabilidad de que la mujer participe en el mercado laboral tiene una variación del -0,016, siempre manteniendo las demás variables constantes.

Finalmente, podemos mencionar uno de los problemas del MLP, el cual se basa en que algunas de las combinaciones de las variables explicativas no se encuentran acotadas entre 0 y 1 (podemos obtener valores con probabilidad negativa y valores con probabilidad mayor a 1, lo cual sabemos que no existe en la realidad).

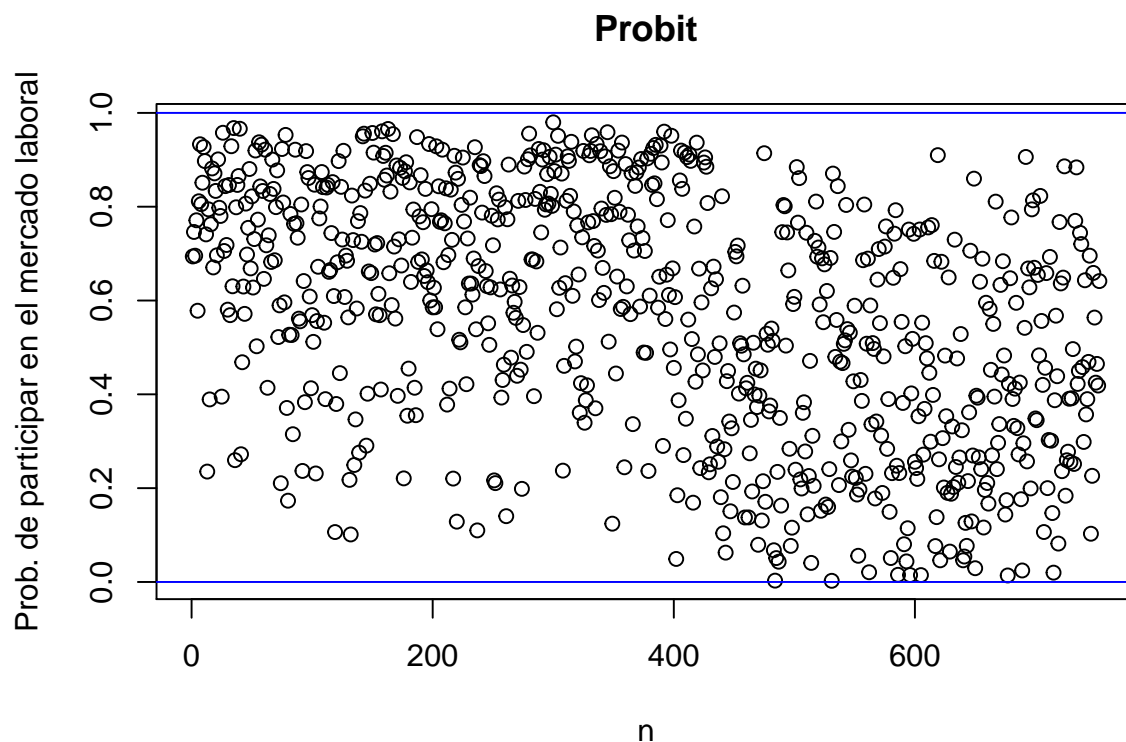
Replicaremos la gráfica 7-8 de la página 315 del libro de Pérez López (2006). Allí vemos los valores menores a cero y los valores superiores a la unidad:

```
plot(predict(fit_mlp),
      ylab="Prob. de participar en el mercado laboral",
      xlab="n",
      main="Modelo Lineal de Probabilidad")
abline(a=0,b=0,col="blue")
abline(a=1,b=0,col="blue")
```

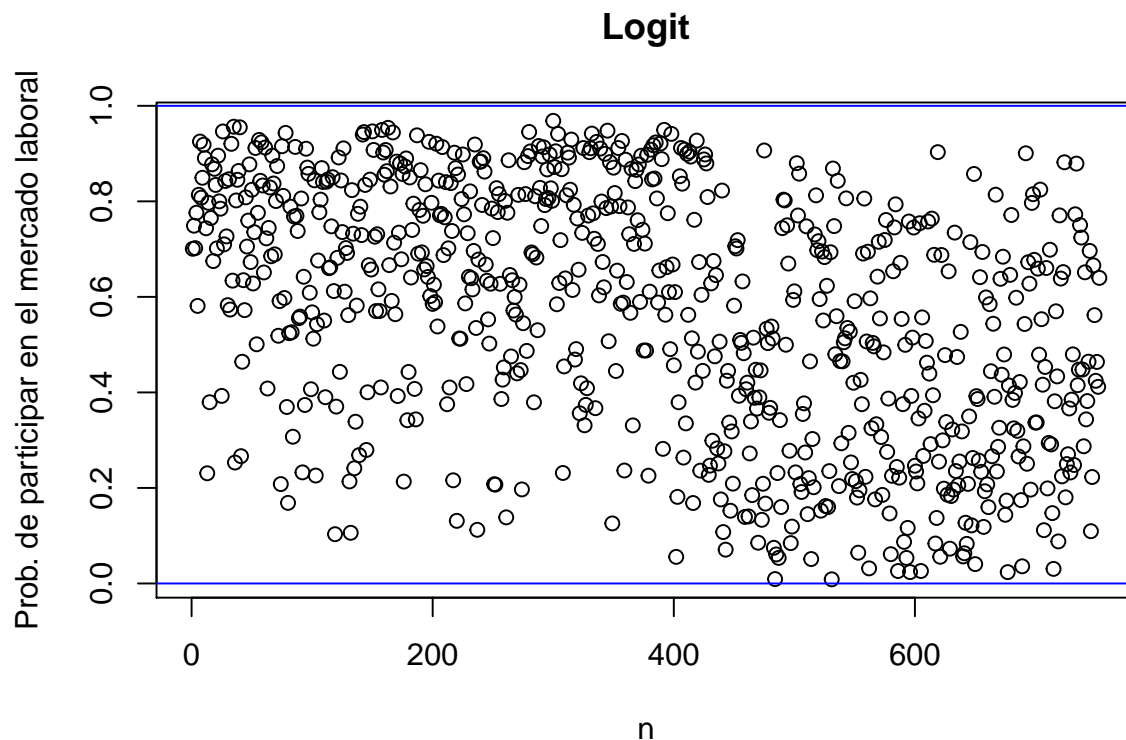


Mientras que para el modelo Probit y el Logit vemos que la probabilidad se encuentra acotada entre 0 y 1, lo cual condice con la teoría:

```
plot(pnorm(predict(fit_prob)),
     ylab="Prob. de participar en el mercado laboral",
     xlab="n",
     main="Probit")
abline(a=0,b=0,col="blue")
abline(a=1,b=0,col="blue")
```



```
plot(plogis(predict(fit_log)),
     ylab="Prob. de participar en el mercado laboral",
     xlab="n",
     main="Logit")
abline(a=0,b=0,col="blue")
abline(a=1,b=0,col="blue")
```



Inciso c)

Se pide realizar una breve comparación entre los algoritmos Newton-Raphson, BHHH, y Quadratic Hill Climbing.

Comenzamos con el marco teórico general de ellos, en los 3 casos estamos en presencia de algoritmos de optimización numérica, los cuales se utilizan para maximizar o minimizar una función objetivo, que llamaremos $F(\theta)$. La particularidad de la función objetivo se basa en que sus parámetros son no lineales, por lo tanto, para optimizarla se debe recurrir a este tipo de algoritmos numéricos.

En el caso estadístico y/o econométrico, se utilizan este tipo de algoritmos para poder maximizar funciones de log verosimilitud o minimizar funciones representativas de la suma de cuadrados residuales, en aquellos modelos que sean de tipo no lineal, ya sea en parámetros o en variables y parámetros.

Estos algoritmos tienen diferentes fórmulas de cálculo, pero en sí, comparten ciertas similitudes, tales como:

- Comenzar con valores iniciales para los parámetros a aproximar.
- Repetir el paso anterior, pero probando diferentes valores iniciales.
- Establecer cuándo se encontró el óptimo.

Por su parte, recordemos lo mencionado en el ejercicio II, es importante tener en cuenta 2 características de los algoritmos iterativos:

- **Velocidad:** cantidad de iteraciones que necesita el algoritmo para encontrar un extremo.
- **Confiabilidad:** la probabilidad del algoritmo de converger al óptimo global.

A su vez, los algoritmos de optimización numérica se pueden clasificar en:

- **Métodos de derivada primera:** encuentran candidatos a óptimos basados únicamente en la primer derivada de $F(\theta)$. Podemos mencionar aquí al algoritmo BHHH.
- **Métodos de derivada segunda:** encuentran candidatos a óptimos basados únicamente en la segunda derivada de $F(\theta)$, tales como Newton-Raphson y Gauss-Newton.

A su vez, podemos tomar la notación elegida por Judge et al. (1988). Ellos mencionan que en general los algoritmos comparten la misma estructura, la cual es:

$$\beta_{n+1} = \beta_n - t_n P_n \gamma_n$$

Donde:

- $\gamma_n = \frac{\partial S}{\partial \beta}|_{\beta_n}$ es el vector gradiente
- P_n es una matriz definida positiva conocida como la matriz de dirección
- t_n es el “step length”, el tamaño del paso, el cual es un número positivo

Lo que diferenciará a los algoritmos es la definición de la matriz P_n , ya que cada algoritmo tiene su propio procedimiento para determinar el tamaño del paso en cada iteración del proceso. Ahora bien, comenzaremos con cada uno de ellos:

Método de Newton-Raphson

Comenzando por el método de Newton-Raphson, también conocido como Newton-Fourier o simplemente método de Newton. Es un algoritmo que sirve para aproximar las raíces de una función. Rememorando lo expresado en el ejercicio II:

Caso de una sola variable:

$$x_{i+1} = x_i - \frac{f(X_i)}{f'(X_i)}$$

En donde:

- x_i es la estimación de la raíz en el paso i

- $f(X_i)$ es la función objetivo valuada en $X = X_i$
- $f'(X_i)$ es la primera derivada de la función objetivo valuada en $X = X_i$
- x_{i+1} es la estimación de la raíz que propone el algoritmo.

Caso de más variables, la fórmula es:

$$\vec{x}_{i+1} = \vec{x}_i - \nabla^{-1} f(\vec{x}_i) f'(\vec{x}_i)$$

En donde:

- \vec{x}_i es el vector solución propuesto en el paso i
- $\nabla^{-1} f(\vec{x}_i)$ es la inversa del vector gradiente (sus elementos son las derivadas primeras de la función respecto de cada una de las variables independientes contenidas en el vector \vec{x}) valuada en $\vec{x} = \vec{x}_i$
- \vec{x}_{i+1} es el vector solución propuesto para el próximo paso $i + 1$.

A su vez, el método de Newton-Raphson también puede utilizarse para encontrar los máximos o mínimos de una función a partir de la aproximación de los ceros o raíces de la derivada primera de la función objetivo, por lo tanto, la fórmula a utilizar será:

$$x_{i+1} = x_i - \frac{f'(X_i)}{f''(X_i)}$$

En donde:

- x_i es la estimación de la raíz en el paso i
- $f'(X_i)$ es la primera derivada de la función objetivo valuada en $X = X_i$
- $f''(X_i)$ es la derivada segunda de la función objetivo valuada en $X = X_i$
- x_{i+1} es la estimación de la raíz que propone el algoritmo.

Y para el caso de más variables:

$$\vec{x}_{i+1} = \vec{x}_i - [\nabla^2 f(\vec{x}_i)]^{-1} \nabla f(\vec{x}_i)$$

En donde:

- \vec{x}_i es el vector solución propuesto en el paso i
- $\nabla f(\vec{x}_i)$ es el vector gradiente (sus elementos son las derivadas primeras de la función respecto de cada una de las variables independientes contenidas en el vector \vec{x}) valuada en $\vec{x} = \vec{x}_i$

- $[\nabla^2 f(\vec{x}_i)]$ es la matriz hessiana, la matriz de derivadas segundas de la función respecto de cada una de las variables independientes contenidas en el vector \vec{x} valuada en $\vec{x} = \vec{x}_i$
- \vec{x}_{i+1} es el vector solución propuesto para el próximo paso $i + 1$.

Lo que se realiza en el método de Newton-Raphson se basa en los siguientes pasos:

1. Elegir un valor inicial para x_i si es una sola variable o los valores iniciales de \vec{x}_{i+1} en el caso de más de una variable.
2. A partir de dichos valores iniciales, se reemplazará en las funciones, se calcularán sus derivadas o se obtendrán la matriz hessiana en el caso que sea necesario.
3. Con los resultados anteriores se aplicará la fórmula de cálculo.
4. A partir del resultado obtenido en el *paso 3*, con dicho valor se retornará al *paso 2* y así sucesivamente se iterará hasta cumplir un criterio de convergencia.
5. Se repetirá probando distintos valores iniciales en el *paso 1*

Ahora bien, para no repetir de forma infinita el algoritmo o para no sobrepasarse, se pueden imponer criterios tales como terminar la iteración cuando la diferencia entre las aproximaciones de dos pasos consecutivos sea menor a un número ε , o elegir una cantidad predeterminada de iteraciones. Sino, hacer uso de la variable t_n para no sobrepasarse en la cantidad. Recordemos que esta variable funciona tal que $S(\beta_{n+1}) < S(\beta_n)$.

Antes de pasar al siguiente algoritmo numérico, es preciso mencionar que Newton-Raphson presenta el problema de que en cada iteración requiere que se calcule una nueva matriz hessiana, y si se considera la posibilidad de que puede darse que la matriz hessiana sea singular, es decir, no invertible, no podrá aplicarse el algoritmo. Además, si los valores iniciales planteados están lejos del verdadero óptimo, el algoritmo necesitará una gran cantidad de iteraciones para poder aproximarse al verdadero valor, por lo cual pierde eficiencia su aplicación, y relacionado a ello, elegir valores iniciales alejados del verdadero valor buscado puede llevar a que las iteraciones converjan hacia la dirección opuesta (ejemplo: al buscar un mínimo, el algoritmo termina convergiendo a un máximo).

Quadratic Hill-Climbing (Goldfeld-Quandt)

Siguiendo a Leyden (1991), el algoritmo de Quadratic Hill-Climbing es una combinación del método “steepest-ascent method” y del algoritmo de Newton-Raphson. El primero se basa en una rutina iterativa en donde los valores de los parámetros se modifican en base a las derivadas primeras en busca de mejorar el valor de la función optimizada para moverse en la dirección con mayor pendiente, mientras que el método de Newton-Raphson, como vimos anteriormente, se utiliza para aproximar las raíces de una función, por lo que sirve también para encontrar mínimos o máximos a partir de los ceros o raíces de la derivada primera de la función objetivo.

Entonces, cuando la función objetivo puede aproximarse por una función cuadrática, este último método tiene una mayor ponderación en la fórmula del algoritmo, pero si la aproximación no es buena, el steepest-ascent method adquiere una mayor ponderación. Por lo tanto, al combinar ambos métodos se obtiene la fórmula del Quadratic Hill-Climbing:

$$P_{t+1} = P_t - (S_t - \alpha I)^{-1} F_t$$

Donde:

- P_{t+1} cumple la misma función que β_{n+1}
- F_t representa las derivadas primeras de la función objetivo
- α es una función negativa del grado en el cual la función F_t puede aproximarse mediante una función cuadrática
- S_t es la matriz de derivadas segundas

El valor de α será quien determine el peso relativo de cada método dentro del análisis.

Berndt, Hall, Hall, and Hausman (BHHH)

El algoritmo BHHH es un algoritmo basado en la derivada primera de la función objetivo. Es similar al algoritmo de Newton-Raphson pero se diferencia en que este último utiliza la derivada segunda para encontrar el máximo o mínimo de la función. A su vez, en BHHH se reemplaza la matriz hessiana negativa observada por el producto exterior del gradiente.

Siguiendo a los autores, a diferencia de Newton-Raphson, este algoritmo (BHHH) necesita definir la función de verosimilitud para una observación y define a P_n de la siguiente manera:

$$\hat{P}_n = \left[\sum_{t=1}^T \left(\frac{\partial L_t}{\partial \beta} \right) \left(\frac{\partial L_t}{\partial \beta'} \right) \right]^{-1}_{|\beta_n, \sigma_n^2}$$

Como se dijo, es un método que requiere solamente el cálculo de la primera derivada de la función objetivo, pero su desventaja se basa en que puede no ser eficiente, es decir, si la aproximación no se encuentra cerca del verdadero óptimo, requerirá un gran número de iteraciones.

Inciso d)

En este inciso se solicita calcular la tabla de predicciones. Se calculará para el mejor modelo, el modelo Probit, y también para el otro modelo. Ahora bien, la tabla de predicciones señala cuántas unidades el modelo clasificó de forma correcta, esto es, a cuántos valores iguales a 0, les dio el valor de 0, y a cuántos valores iguales a 1, les dio dicho valor.

En otras palabras, con la tabla de predicciones, se analiza a cuántas mujeres que participaron del mercado laboral en 1975 el modelo Probit predijo de forma correcta que

participarían en el mercado laboral, y a cuántas que no lo hicieron, el modelo predijo de forma precisa.

```
table(true = mujeres$inlf,
      pred = round(fitted.values(fit_prob)))
```

```
##      pred
## true   0   1
##      0 205 120
##      1  80 348
```

```
prob0 <- (205/(205+120))*100
prob1 <- (348/(80+348))*100
probG <- ((205+348)/(205+120+80+348))*100
```

La tabla de predicciones nos muestra que el modelo Probit predijo un 63.08 % de las mujeres que no participan del mercado laboral, mientras que un 81.31 % de las que alguna vez participaron del mercado laboral. Finalmente, un 73.44 % para el modelo en general.

Por su parte, para el modelo Logit, se obtiene la siguiente tabla de predicciones:

```
table(true = mujeres$inlf,
      pred = round(fitted.values(fit_log)))
```

```
##      pred
## true   0   1
##      0 207 118
##      1  81 347
```

```
log0 <- (207/(207+118))*100
log1 <- (347/(81+347))*100
logG <- ((207+347)/(207+118+81+347))*100
```

Los resultados de esta tabla indican que este modelo predijo correctamente un 63.69 % de las mujeres que no trabajaron en 1975, mientras que un 81.07 % de las mujeres que participaron alguna vez del mercado laboral. Por último, si consideramos el modelo en general, predijo un 73.57 % de forma correcta.

Agregamos también la tabla de predicciones del MLP:

```
table(true = mujeres$inlf,
      pred = round(fitted.values(fit_mlp)))
```

```
##      pred
## true   0   1
##      0 203 122
##      1  78 350
```

```
mlp0 <- (203/(203+122))*100  
mlp1 <- (350/(78+350))*100  
mlpG <- ((203+350)/(203+122+78+350))*100
```

Considerando los resultados obtenidos de la tabla de predicciones del MLP, obtuvimos que predijo en forma correcta el 81.78 % de mujeres que fueron parte del mercado laboral en 1975, mientras que nos brindó un 62.46 % para mujeres que no lo hicieron. y por último, un 73.44 % en general fue predicho de forma correcta.

Entonces, con los resultados obtenidos:

- En el caso de las mujeres que no participaron del mercado laboral, el modelo con mayor porcentaje de predicción correcta fue el Logit, con un 63.08 %
- Por su parte, el modelo con mayor porcentaje de predicción correcta fue el MLP, con un 81.78 % para el caso de mujeres que participaron al menos una vez en el mercado laboral.
- Por último, en el caso general una vez más el Logit tuvo mejor desempeño con un 73.57 % de predicciones correctas.

Inciso e)

El método Heckit, también conocido como el método de Heckman, en referencia a su autor, es un método de corrección para poder realizar estimaciones en modelos en los que haya sesgo de selección, también llamado sesgo de selectividad. Es decir, el método de Heckit es un modelo estimable bajo sesgo de selectividad.

Por ende, antes de presentar el método Heckit propiamente dicho, se hace una referencia al sesgo de selección muestral. Este sesgo es el resultado de seleccionar muestras no aleatorias para estimar relaciones de comportamiento como un sesgo de especificación ordinario que surge debido a datos faltantes, pero se diferencia del problema de “variables omitidas” y errores de especificación, ya que es posible estimar las variables que, cuando se omiten en un análisis de regresión, dan lugar al error de especificación.

Siguiendo a Heckman (1979), el sesgo de selectividad puede surgir por dos razones:

1. Autoselección por parte de los individuos o las unidades de datos que se investigan.
2. Decisiones de selección por parte del analista.

El primer caso, el de la autoselección, según Wooldridge (2015) “... el término proviene del hecho de que las personas se autoseleccionan para participar en ciertas conductas o en ciertos programas: esta participación no es determinada de forma aleatoria” (p. 253). En este caso, Heckman expresa ejemplos relacionados con el mercado laboral y afirma que, en general, las funciones salariales estimadas en muestras seleccionadas no estiman las funciones salariales de la población (ejemplo: el salario de los inmigrantes, no representa

una buena estimación, en general, de lo que habrían ganado los no emigrantes si hubiesen emigrado).

Mientras que, la segunda razón, basada en las decisiones de selección de quien lleve a cabo la investigación, Heckman lo ejemplifica con aquellos estudios de datos de panel, en donde se suelen utilizar observaciones “intactas”, en donde un caso común suele ser el de considerar la estabilidad de la unidad familiar como requisito para poder ingresar las observaciones a la muestra. En ambos casos, se tiene el mismo efecto sobre las estimaciones estructurales: las funciones de regresión ajustadas confunden a los parámetros de comportamiento de interés con los parámetros de la función que determina la probabilidad de la muestra.

Procedemos a explicar el método de Heckman, siguiendo su propio desarrollo y ejemplificando cuando sea conveniente. Considerando una muestra de I observaciones y dos ecuaciones, las cuales son para cada individuo i :

$$(1a) \quad Y_{1i} = X_{1i}\beta_1 + U_{1i}$$

$$(1b) \quad Y_{2i}^* = X_{2i}\beta_2 + U_{2i}$$

En donde:

- X_{ji} es el vector de las variables exógenas, y es de dimensión $1 \times K_j$
- β_j es el vector de parámetros, y es de dimensión $K_j \times 1$.

A la primer ecuación, la (1a), se le conoce como función de regresión, mientras que a (1b) es llamada ecuación de selectividad. Además, podemos suponer una tercer variable, la cual funciona como un indicador binario, es decir:

$$y_{2i} = 1 [Y_{2i}^* > 0]$$

Un ejemplo para clarificar el análisis, es el siguiente, podemos considerar como:

- Y_{1i} = salarios
- Función de regresión: salarios en base a la productividad.
- Y_{2i}^* = utilidad neta del trabajo
- Ecuación de selectividad: decisión de trabajar.
- X_{1i} : determinantes de la productividad.
- X_{2i} : determinantes de la decisión de trabajar.

Cabe destacar que X_{1i} y X_{2i} pueden compartir elementos pero no necesariamente deben ser exactamente iguales, sino que pueden ser diferentes.

Los supuestos en los que se basa el modelo, son los siguientes:

1. (y_{2i}, x_{2i}) se observa para todos (siguiendo nuestro ejemplo, la utilidad positiva de trabajar y sus determinantes).
2. (y_{1i}, x_{1i}) se observa solamente si $y_{2i} = 1$, esto implica que la muestra que tenemos está sujeta a un mecanismo de selectividad (solamente observaremos los salarios y los determinantes de la productividad cuando la utilidad neta de trabajar sea positiva)
3. (U_{1i}, U_{2i}) es independiente de x_{2i} con esperanza nula.
4. $U_{2i} \sim N(0, \sigma_2^2)$
5. $E(U_{1i}|U_{2i}) = \gamma u_{2i}$: los no observables pueden estar relacionados. Este supuesto implica que como la esperanza es potencialmente diferente de cero, eso permite que haya una relación entre la función de regresión y la ecuación de selectividad.

Ahora bien, ya vistos los supuestos, se continua a partir de la estimación de la función de regresión, pero sabemos por supuesto del modelo que existen datos faltantes. Por lo tanto, se procede a presentar a la función de regresión de la población como:

$$E(Y_{1i}|X_{1i}) = X_{1i}\beta_1$$

La función de regresión de la submuestra con los datos disponibles es:

$$E(Y_{1i}|X_{1i}, y_{2i} = 1) = X_{1i}\beta_1 + E(U_{1i}|X_{1i}, y_{2i} = 1); i = 1, 2, \dots, I$$

Por convención, se considera que las primeras I_1 observaciones están disponibles.

Cabe destacar que, en el caso de que la esperanza condicional de U_{1i} sea igual a 0, la función de regresión de la muestra seleccionada es exactamente igual a la función de regresión de la población, por lo que podemos estimar a β_1 por MCO, siendo el único costo de tener una muestra incompleta la pérdida de eficiencia. Pero estamos en este caso trabajando con el caso contrario, por lo cual, continuamos con la ecuación recién presentada y trabajamos matemáticamente sobre ella:

La ecuación se basa en la esperanza condicional de Y_{1i} a sus determinantes y también condicional en el problema de selectividad:

$$E(Y_{1i}|X_{1i}, y_{2i} = 1) = X_{1i}\beta_1 + E(U_{1i}|X_{1i}, y_{2i} = 1)$$

En el segundo paso aplicamos la “ley de esperanzas iteradas”, la cual se basa en que se puede aplicar la esperanza de U_{1i} en 1 paso o sino en 2 pasos. Si es en 2 pasos, puedo condicionarla en cualquier variable, en este caso a U_{2i} , y luego se calcula la esperanza de la esperanza.

$$E(Y_{1i}|X_{1i}, y_{2i} = 1) = X_{1i}\beta_1 + E[E(U_{1i}|U_{2i})|X_{1i}, y_{2i} = 1]$$

Lo que sigue, se basa en tener en cuenta el supuesto de $E(U_{1i}|U_{2i}) = \gamma u_{2i}$, por lo que:

$$E(Y_{1i}|X_{1i}, y_{2i} = 1) = X_{1i}\beta_1 + E(\gamma u_{2i}|X_{1i}, y_{2i} = 1)$$

El siguiente paso se deriva de la relación que hay entre la variable y_{2i} y la indicadora binaria (en términos del ejemplo, es lo mismo que una persona trabaje a que esta persona tenga utilidad marginal neta de trabajar positiva), y por propiedad de esperanza, la constante sale hacia afuera:

$$E(Y_{1i}|X_{1i}, y_{2i} = 1) = X_{1i}\beta_1 + \gamma E(u_{2i}|X_{1i}, Y_{2i}^* > 0)$$

Luego, se reemplaza a Y_{2i}^* y se trabaja sobre la desigualdad:

$$E(Y_{1i}|X_{1i}, y_{2i} = 1) = X_{1i}\beta_1 + \gamma E(u_{2i}|X_{1i}, X_{2i}\beta_2 + U_{2i} > 0)$$

$$E(Y_{1i}|X_{1i}, y_{2i} = 1) = X_{1i}\beta_1 + \gamma E(u_{2i}|X_{1i}, U_{2i} > -X_{2i}\beta_2)$$

$$E(Y_{1i}|X_{1i}, y_{2i} = 1) = X_{1i}\beta_1 + \gamma E(u_{2i}|X_{1i}, U_{2i} < X_{2i}\beta_2)$$

Por lo que la función de regresión de la muestra seleccionada depende de X_{1i} y X_{2i} , lo cual es diferente a la ecuación de regresión inicial que solamente dependía de X_{1i} ignorando el término $\gamma E(u_{2i}|X_{1i}, U_{2i} < X_{2i}\beta_2)$ como regresor, siendo el sesgo de selectividad.

Por su parte, antes de continuar con el análisis matemático, Heckit realiza unas aclaraciones a tener en cuenta:

1. Si la única variable en el vector regresor X_{2i} que determina la selección de la muestra es 1, de modo que la probabilidad de inclusión de la muestra sea la misma para todas las observaciones, la media condicional de U_{1i} es una constante y el único sesgo en β_1 que resulta del uso de muestras seleccionadas para estimar la ecuación estructural de la población surge en la estimación del intercepto.
2. El estimador de mínimos cuadrados de la varianza poblacional σ_{11} se encuentra sesgado hacia abajo.
3. Variables que pertenecen a X_{2i} y no a X_{1i} pueden parecer estadísticamente significativas para Y_{1i} cuando se trabajan con muestras seleccionadas.

Ahora bien, continuamos con el desarrollo matemático, asumiendo que $h(U_{1i}, U_{2i})$ es una densidad normal bivariada, y sabiendo que $E(u_{2i}|X_{1i}, U_{2i} < X_{2i}\beta_2)$ es la esperanza de una variable aleatoria normal (supuesto de normalidad para u_{2i}) pero truncada a la derecha ($U_{2i} < X_{2i}\beta_2$).

Por lo tanto, se aplica la fórmula de una esperanza de una variable aleatoria normal truncada a la derecha, la cual es igual a la esperanza de la variable original (en este caso es igual a 0) menos el desvío estándar multiplicado por la inversa de la razón de Mills⁴ (λ), el cociente entre la función de densidad y la función de distribución de una normal estándar, por lo tanto:

$$E(Y_{1i}|X_{1i}, y_{2i} = 1) = X_{1i}\beta_1 - \gamma\sigma_2\lambda\left(X_{2i}\frac{\beta_2}{\sigma_2}\right)$$

El siguiente paso es determinar a $z_i \equiv \lambda\left(X_{2i}\frac{\beta_2}{\sigma_2}\right)$, por lo que la esperanza de Y_{1i} dado X_{1i} , y dado que opera el mecanismo de selectividad, es:

$$E(Y_{1i}|X_{1i}, y_{2i} = 1) = X_{1i}\beta_1 - \gamma\sigma_2 z_i$$

⁴Es una función de probabilidad monótona decreciente de que una observación sea elegida en la muestra

Como el último término no es 0, la esperanza analizada no es $X_{1i}\beta_1$, por lo que utilizar MCO con la muestra seleccionada conduce a estimaciones inconsistentes. Siguiendo a Sosa Escudero (2020), la inconsistencia se da según Heckman por “omitir” el término z_i al realizar una especificación incorrecta, y la fuente de la inconsistencia es que $\gamma \neq 0$. Hay correlación entre U_{1i} y U_{2i} , la correlación que hay entre el mecanismo de selectividad y el mecanismo de regresión a través de los no observables.

Entonces, se obtiene lo siguiente:

$$u_1^* \equiv Y_{1i} - X_{1i}\beta_1 - \gamma^* z_i, \gamma^* \equiv -\gamma\sigma_2$$

Despejando y descomponiendo a U_{1i} en $\gamma^* z_i + u_1^*$:

$$Y_{1i} = X_{1i}\beta_1 + \gamma^* z_i + u_1^*$$

Por construcción:

$$E(U_1^* | X_{1i}, Y_{2i} = 1) = 0$$

Por lo que Heckman concluye que:

- Si X_{1i} y z_i fuesen observables cuando $Y_{2i} = 1$, por lo tanto, las estimaciones por MCO de Y_{1i} en X_{1i} y z_i utilizando la muestra seleccionada son consistentes en β_1 y γ^* , pero, ¿por qué se utiliza la muestra seleccionada? Porque en ese caso el término de error del modelo recién definido tiene esperanza nula.
- Pero el problema es que $z_i \equiv \lambda \left(X_{2i} \frac{\beta_2}{\sigma_2} \right)$ es no observable, depende de β_2 y de σ_2 , pero a pesar de ello, podemos estimar a λ de la siguiente manera:

Recordando que $U_{2i} \sim N(0, \sigma_2^2)$, y siguiendo nuestro ejemplo, la probabilidad de que una persona trabaje es igual a la probabilidad de que la utilidad neta del trabajo sea positiva, entonces:

$$P(Y_{2i} = 1) = P(Y_{2i}^* > 0) = P\left(\frac{U_{2i}}{\sigma_2} < X_{2i} \frac{\beta_2}{\sigma_2}\right) = \Phi(X_{2i}\delta)$$

Cabe destacar que U_{2i} tiene distribución normal, por lo tanto $\frac{U_{2i}}{\sigma_2}$ tiene distribución normal estándar, por lo que esa desigualdad está dada por la función de distribución acumulada para una normal estándar hasta el punto $\Phi(X_{2i}\delta)$, ¿qué conclusiones se obtienen?

- $P(Y_{2i} = 1)$ es un modelo Probit con coeficiente desconocido δ
- Siguiendo el ejemplo, al observar la decisión de trabajar y sus determinantes, es decir X_{2i} y Y_{2i} , entonces se puede estimar por máxima verosimilitud vía Probit a δ
- No se pueden estimar β_2 y σ_2 al no estar identificados, pero si se puede estimar el cociente entre ellos, es decir, δ

Por lo tanto, ahora resta seguir el método bietápico de Heckman:

- **1° etapa:** mediante un modelo Probit $P(Y_{2i} = 1) = \Phi(X_{2i}\delta)$ se estima $\hat{\delta}$ y se utiliza toda la muestra, esto porque se necesitan las variables binarias en el modelo Probit. En nuestro ejemplo, esto implica que deben utilizarse tanto las personas que trabajan como aquellas que no trabajan. Se regresa el indicador binario en base a sus determinantes, se obtiene $\hat{\delta}$ y con ello se obtiene la estimación $\hat{z}_i = \lambda(X_{2i}\hat{\delta})$
- **2° etapa:** se realiza la regresión de Y_{1i} en X_{1i} y \hat{z}_i usando la muestra seleccionada porque en este caso la esperanza de U_{1i}^* es nula, esto estima de forma consistente a β_1 y a γ^* .

Vale destacar que $\hat{z}_i = \lambda(X_{2i}\hat{\delta})$ y que λ es una función monótona, por lo que al realizar la regresión de Y_{1i} en base a X_{1i} y a \hat{z}_i , por lo monotonicidad de λ , \hat{z}_i es casi una función monótona de X_{2i} . Por lo tanto, en aquellos casos en que X_{1i} y X_{2i} son muy similares, se está casi que regresando Y_{1i} en función de X_{1i} y de \hat{z}_i , que es prácticamente una función de X_{1i} , por lo que la segunda etapa del método de Heckman está sujeta a un problema de multicolinealidad.

Ejercicio V

Resumir, en no más de 1000 palabras el trabajo de Redrado, M.; Carrera, J.; Bastourre, D. e Ibarlucia, J. (2006), “La política económica de la acumulación de reservas: nueva evidencia internacional”, *Investigaciones Económicas*, Banco Central de la República Argentina, Serie Estudios BCRA, 2, en donde los autores estiman un modelo econométrico de los determinantes del nivel de reservas utilizando un modelo dinámico de datos de panel. Opcional: Relacionar este trabajo identificando coincidencias y diferencias con la aplicación desarrollada en R por Croissant y Millo (2008).

Respuesta:

Este trabajo busca identificar las razones por las cuales el último decenio ha sido testigo de un crecimiento explosivo en el nivel global de reservas internacionales, motorizado principalmente por las economías emergentes.

Desde Bretton Woods se distinguen **3 períodos**. La tasa de acumulación de reservas en cada etapa se ha acelerado.

- *Patrón dólar fijo* (1948-1973): todas las monedas se pegaban al dólar. La oferta de liquidez mundial pasó a estar determinada por la posición del balance de pagos norteamericano.
- *Patrón dólar flotante* (1974-1998): flotación de las monedas más importantes y erradicación progresiva de controles sobre las cuentas corriente y capital. Las economías aceleraron fuertemente su acumulación de reservas a pesar de la mayor flexibilidad cambiaria.
- *Desequilibrios en la economía mundial* (1998-2004): déficit de cuenta corriente en Estados Unidos y surgimiento de una periferia en Asia que sigue una estrategia de crecimiento guiado por exportaciones y viene acumulando muchas reservas.

Teorías de la demanda de reservas:

- *Visión tradicional*: concentrada en las transacciones comerciales y la cuenta corriente. Se demandan porque pueden evitar ajustes macroeconómicos innecesarios ante escenarios adversos de balance de pagos, aunque también hay un motivo transacción.
- *Visión moderna*: focalizada en la demanda precautoria frente a las crisis financieras (potencial inconveniente de la desregulación y liberalización de la cuenta capital).

- *Demanda derivada del régimen cambiario y monetario*: endogeneidad de las reservas respecto a otras variables relevantes, como el tipo de cambio.

Evidencia internacional:

Existe un proceso de convergencia: los emergentes van acercándose a las tenencias de reservas de los desarrollados a medida que crece su nivel de vida. Un amplio stock de reservas tiene más valor para los primeros ante shocks externos negativos, ya que suplen mecanismos de autorregulación que no poseen.

Agrupando los países por ubicación geográfica, se observa cierta heterogeneidad entre regiones, lo cual podría ser un indicio de imitación regional.

La estrategia de acumulación sostenida de reservas es posible bajo cualquier régimen de política cambiaria y monetaria. Estos regímenes suelen ser estables a lo largo de los procesos de acumulación. En estos últimos, existe una preferencia por esterilizar parcialmente el incremento de activos externos. Se percibe una relación negativa entre el grado de esterilización y la flexibilidad cambiaria, mientras que la esterilización es máxima bajo un régimen monetario de metas de inflación plenas.

Reservas en Argentina:

El período 1948-1972 se caracteriza por una caída entre puntas de las reservas con alta volatilidad. Desde 1973 la trayectoria de las mismas es muy inestable. Durante “La Tablita”, e impulsadas por los flujos de capital, las reservas llegan a un máximo en 1979, particularmente alto en términos históricos si se las normaliza por importaciones o PBI. Luego, déficit de cuenta corriente y fuga de capitales mediante, se abandona el programa. Prosiguió una etapa de ajuste desordenado y drenaje de reservas.

Durante la hiperinflación las reservas alcanzaron un nuevo mínimo. En 1991 se instauró el Plan de Convertibilidad. Producto de la llegada de capitales en grandes magnitudes, las reservas iniciaron un ascenso vertiginoso, hasta alcanzar un máximo en 1999, con la recesión ya iniciada. El drenaje durante el 2001 fue incesante. Desde el 2002 los activos externos del BCRA retomaron un sendero creciente acompañando primero la recuperación y luego el crecimiento económico.

Indicadores de adecuación de reservas:

Intentan responder qué cantidad de reservas debería poseer un Banco Central en función de las características de su economía.

- *Criterio comercial*: si permiten financiar entre tres y seis meses de importaciones.
- *Criterio financiero*: si permiten cubrir la deuda externa que amortiza en un plazo menor o igual a un año.

Armando un indicador basado en un *criterio aditivo*, que consiste en la suma lineal de los otros dos, se observa que las economías latinoamericanas analizadas se encuentran por debajo de los niveles apropiados, mientras que los asiáticos mayormente por encima.

Análisis econométrico de los determinantes del nivel de reservas:

Se estima una demanda de reservas para un panel no balanceado de 139 países (industrializados y en desarrollo) para el período 1973-2003 con datos anuales. La variable dependiente es el logaritmo del ratio de stock de reservas a PBI.

Metodología econométrica:

Como se incluye la variable dependiente rezagada un período para incorporar el grado de inercia de la demanda de reservas, una especificación adecuada sería:

$$y_{i,t} = \gamma y_{i,t-1} + x'_{i,t} \beta + \eta_i + \varepsilon_{i,t}$$

Donde η_i capta la heterogeneidad no observable entre países. Estimar un modelo de efectos fijos dinámicos por OLS (DFE) subestima el efecto de $y_{i,t-1}$ cuando $\gamma > 0$, mientras que estimar por OLS sesgaría γ hacia arriba debido a la correlación entre la variable rezagada y η_i . Así, una estimación consistente de γ debería ubicarse entre el DFE y el OLS.

Se puede obtener un estimador que solucione estas dificultades removiendo el efecto fijo (diferenciando la ecuación) y usando como instrumento $y_{i,t-2}$. Luego se propone el estimador *System GMM*, que consiste en adicionar a la ecuación en diferencias (instrumentada con niveles) una regresión en niveles (instrumentada con diferencia rezagadas). Esta metodología soluciona los problemas de especificación dinámica, de endogeneidad de los regresores y de instrumentación débil cuando hay alta persistencia.

Las variables explicativas significativas son:

- Coeficiente de inercia: se encuentra entre los coeficientes sesgados estimados mediante pooled OLS y efectos fijos, un indicio de consistencia.
- Imitación regional: cociente entre países de una región que incrementaron sus reservas el año anterior y el total de países de la zona geográfica.
- PIB por habitante en relación al promedio mundial de cada año, y esta misma variable al cuadrado: incorporadas para testear la relación cuadrática entre reservas y grado de desarrollo, por lo que los mayores ratios de reservas a producto los poseen países en estado intermedio de desarrollo.
- Los niveles de apertura comercial (importaciones/PIB) y financiera (influjo de capitales/PIB): ambos con signo positivo. Se interpreta que los comportamientos precautorios aumentan a mayor exposición externa.
- *Dummy* que toma valor a partir de 1990: para verificar un cambio estructural en la demanda de reservas en esos años, caracterizados por mayor apertura financiera. Es positivo, lo cual tiene sentido dado lo dicho previamente.

Coincidencias y diferencias con Croissant y Millo (2008)

El trabajo de estos dos autores se exploya sobre un paquete de R, llamado **plm**, que resulta particularmente útil para trabajar con datos de panel económicos y para lidiar con los inconvenientes que estos suelen exhibir, como la heterogeneidad no observable y el hecho de que los mismos no suelen estar balanceados. Estos problemas, entre otros, están

presentes en el modelo propuesto en Redrado et al. (2006) para estimar la demanda de reservas.

En este último paper académico los autores emplean diferentes métodos de estimación. Seguidamente presentan los resultados de los diferentes modelos obtenidos: estático con efectos fijos por OLS (transformación *within*), dinámico con efectos fijos por OLS, dinámico mediante *pooled* OLS, con la metodología Dif. GMM (estimaciones a dos pasos) y finalmente, con la metodología econométrica considerada más apropiada: System GMM (también a dos pasos). El paquete **plm** parece proveer las herramientas requeridas para llevar a cabo todas estas estimaciones (aunque no hace mención explícita del último método indicado): para la estimación de modelos de panel básicos, se emplea la función **plm**. En cambio, **pgmm** es usada para modelos GMM. Dentro de los argumentos de estas funciones se destacan **effect** (por si se quieren agregar efectos individuales, con "individual", por ejemplo), **model** (como "within" o "pooling" para modelos básicos, o "onestep" o "twosteps" para modelos GMM), e **instruments** (para definir las variables instrumentales). Además, se debe definir la función, es decir, la especificación del modelo. En el caso de que la misma incluya diferenciaciones (**diff**), variables con retardo (**lag**) y/o transformaciones logarítmicas (**log**), resulta conveniente emplear la función **dynformula**, que facilita esta tarea.

Ejercicio VI

Resumir en no más de 1000 palabras el siguiente trabajo: Lanteri, L. (2008), “Respuesta a precios del área sembrada de soja en la Argentina”, Investigaciones Económicas, Banco Central de la República Argentina, vol. 52, octubre - diciembre, en el que se comparan los resultados obtenidos a partir del enfoque econométrico basado en el modelo de corrección del error con la utilización de una metodología econométrica más antigua.

Respuesta:

En este trabajo se analiza la respuesta del área sembrada de soja (*AS*) en Argentina frente a los cambios en los incentivos de precios y otras variables relevantes en la producción agrícola a partir de Modelos de Corrección del Error (VEC) para algunas de las principales provincias productoras en el período 1974 - 2006.

El autor destaca el aumento en los precios de los commodities, lo cual benefició a Argentina. La soja se expandió, en conjunto con la industria procesadora. También tuvo un rol importante el aumento de la demanda por parte de China, India y otras economías en desarrollo. Influyó además la reducción de la superficie dedicada a otros cultivos, tierras ocupadas por la ganadería, la expansión de la frontera agropecuaria y deforestación de montes nativos. La biotecnología tuvo su participación, y en las últimas dos décadas, los “pool” de siembra.

En cuanto al modelo, para estimar la oferta agropecuaria agregada se eligió el área sembrada como su indicador, mientras que, como variables explicativas, los precios relativos respecto de productos sustitutos, insumos y otras variables.

Por su parte, para no llegar a estimaciones espurias, debido a no estacionariedad de alguna variable, se recurrió modelos VEC que suponen que las variables son integradas del mismo orden y se basan en un análisis previo de la estacionariedad de las series.

Considerando un VEC en base a trabajos de Johansen (1988) y Johansen y Juselius (1990) se intentó explicar la relación entre la participación del (*AS*), incentivos de precios y demás variables, buscando estimar ajustes dinámicos de corto plazo (*CP*) y relaciones de largo plazo (*LP*).

El modelo:

$$\Delta X_t = \Pi X_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta X_{t-i} + \varepsilon_t$$

Donde:

- X_t : vector de k variables endógenas no estacionarias $I(1)$.

- II: matriz de coeficientes de largo plazo que incluye los vectores de cointegración
- Γ_i : matriz de coeficientes de corto plazo
- ε_t : vector de errores aleatorios (ruido blanco)
- Para determinar el número de relaciones de cointegración r , la metodología de Johansen proporciona dos test: λ_{trace} y el de λ_{max} , el primero testea la hipótesis nula (H_0) de r relaciones, donde k indica el número de variables endógenas, para $r = 0, 1, \dots, k - 1$; el otro la (H_0) de r vectores de cointegración contra la alternativa de $r + 1$.

Las variables consideradas fueron:

- $\frac{\text{Área soja}}{\text{área soja} + \text{área sustitutos}}$
- $\frac{\text{Área soja}}{\text{Área granos}}$
- $\frac{\text{Precio relativo soja}}{\text{Precio relativo sustitutos}}$: precio de la soja respecto al promedio ponderado de los precios del maíz y del girasol

$$P_t^0 = P_{t-1}^0 + \lambda (P_{t-1} - P_{t-1}^0)$$

En donde λ es una fracción del error en la formación de expectativas del período previo y varía entre 0 y 1.

- Área sembrada de trigo
- Consumo de fertilizantes
- Riesgo: desvío estándar móvil de los retornos para los m períodos anteriores a la campaña t :

$$\sigma_t = \sqrt{\frac{1}{m-1} \sum_{j=1}^m (R^* - R_{t-j})^2}$$

El primer paso consiste en identificar el orden de integración de cada una de las series. Para ello utilizó los test de Phillips-Perron, Dickey-Fuller aumentada y DF Rolling. Los test llevaron a la misma conclusión, las series son $I(1)$, ya que ninguno pudo rechazar la hipótesis nula de existencia de raíz unitaria.

Después se realizaron los test de cointegración para determinar la existencia de relaciones de (LP) entre las variables y el número de relaciones de cointegración. Se utilizaron los estadísticos de λ_{trace} y λ_{max} , ambos indicaron la existencia de una relación de cointegración.

Mientras que, los tests LM de correlación serial y de heterocedasticidad (White) no permitieron rechazar las (H_0) de ausencia de correlación serial y ausencia de heterocedasticidad.

A partir de la estimación de los VEC se obtuvo la cointegración para cada uno de ellos, y dado que las variables estaban en logaritmos, los parámetros pueden interpretarse como elasticidades de cada variable respecto a la participación del (*AS*).

En todos los casos, la velocidad de ajuste de la variable dependiente resultó estadísticamente significativa señalando que ella no sería débilmente exógena y no habría un proceso explosivo.

Los coeficientes de la variable de precios relativos soja/sustitutos resultan en todos los modelos significativos y con el signo positivo esperado, muestran una importante respuesta de las decisiones de siembra de los productores a los incentivos de precios relativos.

Los resultados de las elasticidades área - precio de (*LP*) están en línea con estimaciones antes realizadas para la provincia de Santa Fe aunque con valores algo inferiores.

Se intentó incluir el stock de ganado bovino pero no se obtuvieron buenos resultados en las pruebas de cointegración. También se incluyó el precio real de los fertilizantes, fue significativo pero con signo inverso al esperado.

El coeficiente de la variable de la superficie sembrada de trigo fue significativo y con signo positivo.

Se observó que un aumento en el riesgo conlleva un incremento en la participación del (*AS*), por ende puede responder a que los productores no presentan aversión a los riesgos o que al ser la soja un cultivo de segunda siembra, los aumentos en los riesgos no alterarían los planes de producción.

Los test de ADF para determinar estacionariedad de los residuos permiten rechazar la (H_0) de existencia de raíz unitaria al 5 %, por ende los modelos estimados son razonablemente correctos.

En cuanto a las funciones de impulso-respuesta, indican el comportamiento de la participación del (*AS*) frente a un shock aleatorio en distintos momentos, de sí misma u otra variable. Se utilizó la transformación de Pesara y Shin (1988), para corregir la correlación de los errores. Se obtuvo que shocks positivos en los precios inducen una respuesta positiva y permanente en la participación del (*AS*), así como shocks positivos en el área sembrada de trigo. Pero, shocks de riesgos generan una respuesta positiva y permanente en la participación del área de soja, salvo en el muy (*CP*).

Finalmente, el análisis de descomposición de la varianza, considera la importancia relativa de cada shock aleatorio en el VEC. Los resultados mostraron que shocks en precios relativos explican alrededor de un 46,6 % de la variabilidad de la participación del (*AS*) en Santa Fe y un 23,4 % para el país.

Para cerrar, el autor señala que podría inferirse que los productores rurales no serían insensibles a los estímulos económicos.

Ejercicio VII

En el archivo “GDP_Arg_Bra_International_Financial_Statistics.txt” se puede encontrar información sobre los índices de Producto Bruto Interno trimestrales de Argentina y Brasil para el período 1995:Q1-2016:Q2. Se le pide:

a) A partir del logaritmo natural de la serie correspondiente a Brasil, aplique la metodología de Box y Jenkins para seleccionar el mejor modelo ARIMA estacional. Para una primera aproximación al tema se puede consultar el siguiente link: <https://online.stat.psu.edu/stat510/lesson/4/4.1>

b) Explique claramente la prueba HEGY desarrollada por Hylleberg et al. (1990). Para este inciso y el que sigue, una referencia de fácil lectura es Alonso & Semaán (2010).

c) Aplique la prueba HEGY sobre la serie analizada en el inciso a) e interprete en forma detallada los resultados.

Respuesta:

Inciso a)

Antes de poder aplicar la metodología de Box-Jenkins debemos extraer la serie del GDP de Brasil del archivo .txt, tomar logaritmo natural y luego utilizar la función `ts()` para que R la identifique como una serie de tiempo:

```
GDP_Arg_Bra <- read.delim(path)

ln_GDP_Bra <- ts(as.data.frame(log(GDP_Arg_Bra[,3])),
                 start = c(1995,1), frequency = 4)
colnames(ln_GDP_Bra) <- c("ln(GDP_Brazil)")
rownames(ln_GDP_Bra) <- GDP_Arg_Bra[,1]

head(ln_GDP_Bra)
```

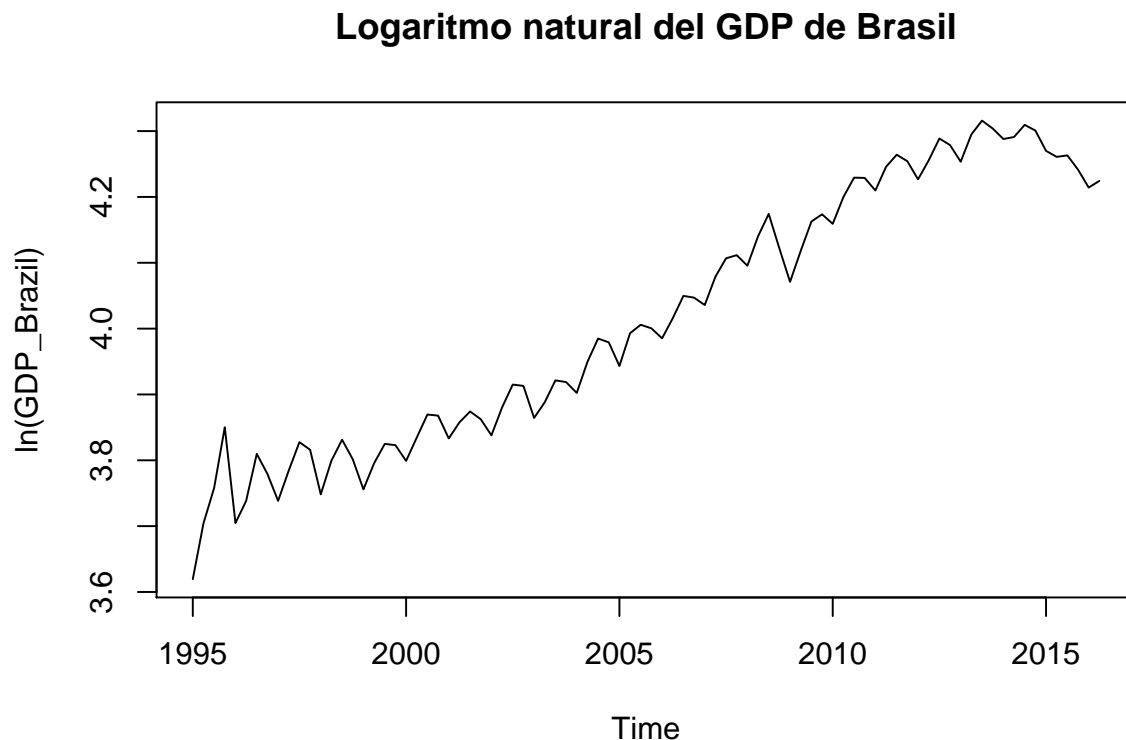
```
##          ln(GDP_Brazil)
## Q1 1995          3.619529
## Q2 1995          3.704014
## Q3 1995          3.757939
## Q4 1995          3.850148
## Q1 1996          3.704753
```

Q2 1996 3.738146

1) Identificación

a) Representación gráfica de la serie:

```
plot(ln_GDP_Bra, main = "Logaritmo natural del GDP de Brasil")
```



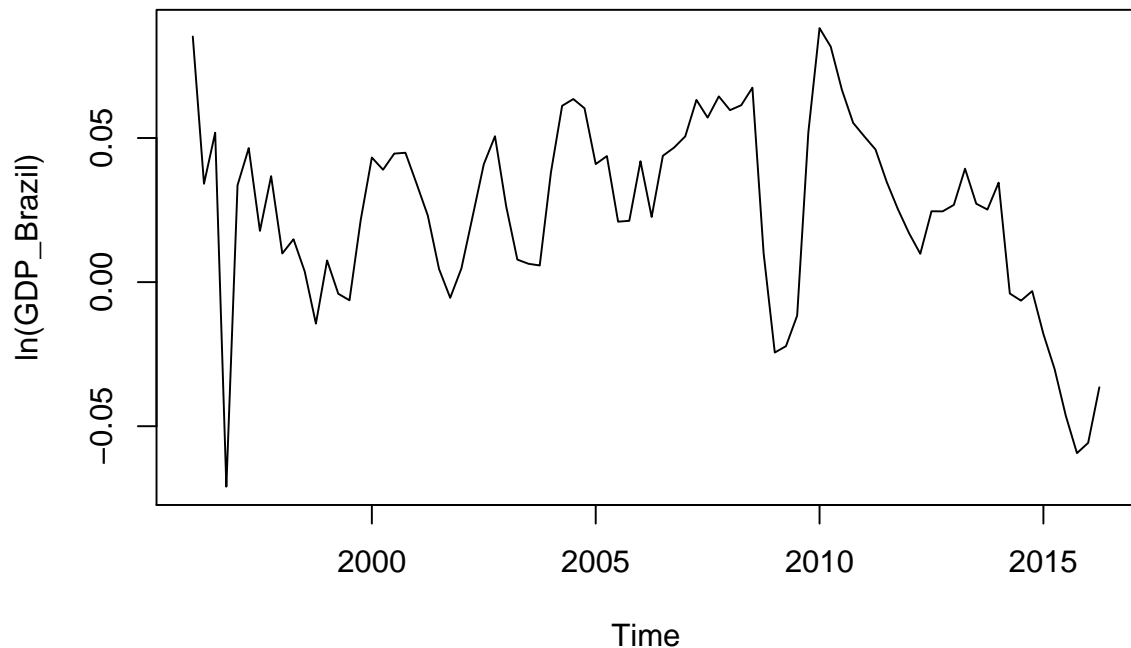
Se percibe con claridad que la serie no es estacionaria en media, aunque posiblemente sí en varianza. La tendencia que se observa es positiva hasta el año 2014, y a partir de allí se vuelve negativa. El comportamiento estacional periódico parece darse en lapsos de 1 año (4 trimestres), donde dentro de los mismos los valores más altos se dan en algunos trimestres y los más bajos en otros, sistemáticamente.

b) **Transformación y eliminación de tendencia:** dado que tenemos tanto estacionalidad como tendencia, prevemos que tendremos que tomar cuatro diferencias para eliminar la estacionalidad (diferenciación estacional) y luego primeras diferencias nuevamente para terminar de eliminar la tendencia (diferenciación no estacional). Si B es el operador de rezagos y x_t una observación, el resultado de aplicar ambas es:

$$(1 - B^4)(1 - B)x_t = (x_t - x_{t-1}) - (x_{t-4} - x_{t-5})$$

Veamos primero el resultado de aplicar 4 diferencias:

```
dif_4 <- diff(ln_GDP_Bra, 4)
plot(dif_4)
```



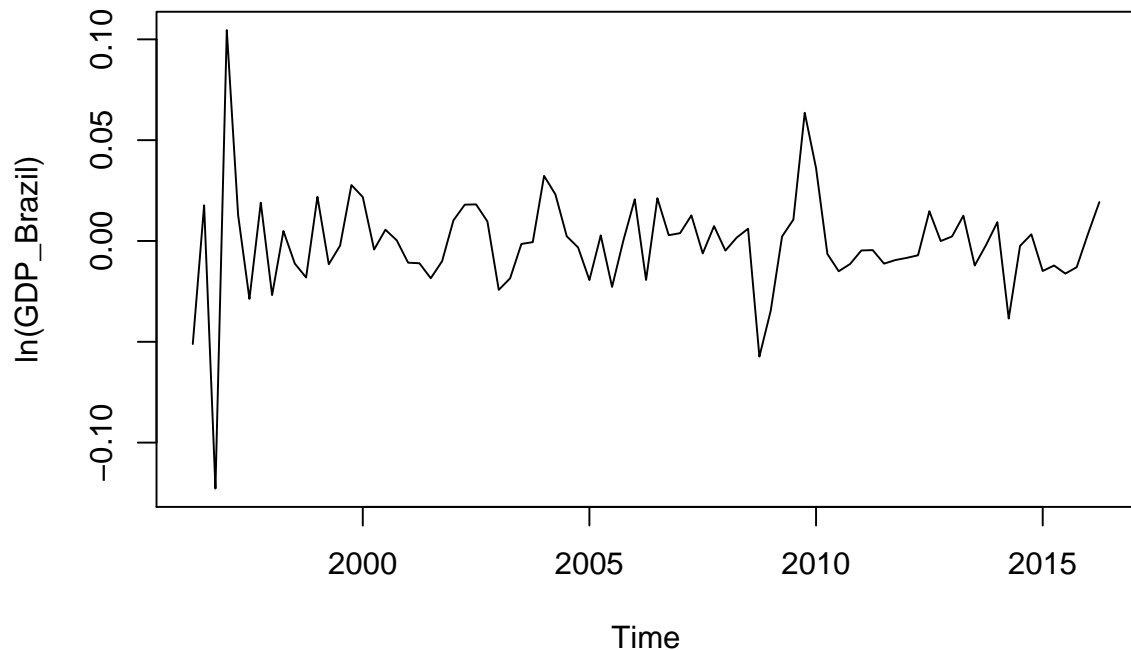
Ya no se percibe estacionalidad determinística. También parece haberse eliminado la tendencia, aunque habría que verificarlo con el test de Dickey-Fuller Aumentado, cuya hipótesis nula es que la serie es no estacionaria (tiene raíz unitaria). No rechazarla sugiere que deberíamos diferenciar una vez más:

```
library(tseries)
adf.test(dif_4)

##
## Augmented Dickey-Fuller Test
##
## data: dif_4
## Dickey-Fuller = -1.9287, Lag order = 4, p-value = 0.605
## alternative hypothesis: stationary
```

Como el p-value es mayor a 0.05 no rechazamos la hipótesis nula (concluimos que la serie sigue siendo no estacionaria), por lo que pasamos a diferenciar la serie (que ya fue diferenciada 4 veces) una vez más:

```
dif_1y4 <- diff(dif_4,1)
plot(dif_1y4)
```



Ahora sí parece obvio que se logró eliminar la tendencia. Una vez más, lo verificamos:

```
adf.test(dif_1y4)
```

```
##
## Augmented Dickey-Fuller Test
##
## data: dif_1y4
## Dickey-Fuller = -5.0843, Lag order = 4, p-value = 0.01
## alternative hypothesis: stationary
```

Efectivamente, la serie sí es estacionaria en esta instancia (p-value menor a 0.05).

c) **Inspección de correlogramas:** en este paso se analizan los diagramas que representan la función de autocorrelación (FAC o ACF en inglés) y la función de autocorrelación parcial (FAP o PACF en inglés) para estimar el orden de los procesos AR y MA. Antes de hacer eso corresponde explicar que se va a trabajar con un modelo ARIMA estacional, el cual incorpora factores tanto estacionales como no estacionales en un modelo multiplicativo cuya notación simplificada es la siguiente:

$$\text{ARIMA } (p, d, q) \times (P, D, Q)_S$$

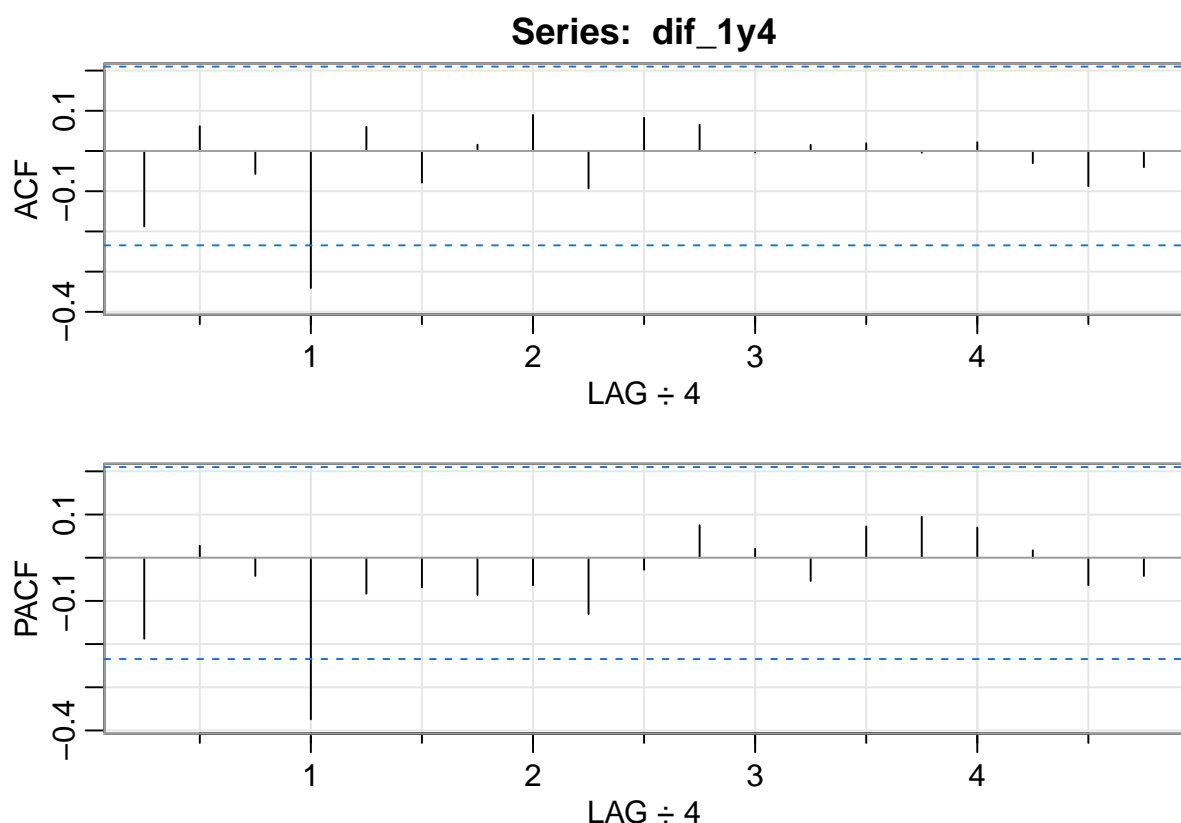
Donde p es el orden del proceso AR no estacional, d es el orden de integración no estacional (número de diferencias que debemos aplicarle al proceso para convertirlo en estacionario), q es el orden del proceso MA no estacional, P es el orden del proceso AR estacional, D es el orden de integración estacional, Q es el orden del proceso MA estacional y S es el número de períodos de tiempo hasta que el patrón estacional se repite (en nuestro caso 4 trimestres). Más formalmente, y sin considerar los órdenes de integración, el modelo responde a:

$$\Phi(B^S)\phi(B)(x_t - \mu) = \Theta(B^S)\theta(B)w_t$$

Donde $\Phi(B^S) = 1 - \Phi_1 B^S - \dots - \Phi_P B^{PS}$ es el proceso AR estacional, $\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$ el proceso AR no estacional, $\Theta(B^S) = 1 + \Theta_1 B^S + \dots + \Theta_Q B^{QS}$ el proceso MA estacional, $\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$ el proceso MA no estacional, w_t el término de error aleatorio (tal que $w_t \stackrel{iid}{\sim} N(0, \sigma_w^2)$), x_t las observaciones y $\mu = E(x_t) = E(x_{t-1}) = \dots = E(x_{t-p})$ bajo supuesto de estacionariedad.

Una vez hecha esta aclaración, pasamos a analizar los correlogramas:

```
library(astsa)
acf2(dif_1y4)
```



```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
## ACF  -0.19 0.06 -0.06 -0.34  0.06 -0.08  0.02  0.09 -0.09  0.08  0.07  0.00
## PACF  -0.19 0.03 -0.04 -0.37 -0.08 -0.07 -0.09 -0.06 -0.13 -0.03  0.08  0.02
##      [,13] [,14] [,15] [,16] [,17] [,18] [,19]
```

```
## ACF    0.02  0.02  0.00  0.02 -0.03 -0.09 -0.04
## PACF  -0.05  0.07  0.09  0.07  0.02 -0.06 -0.04
```

Tanto la FAC como la FACP son decrecientes y oscilatorias con un retardo múltiplo de S significativo (x_{t-4}) y ningún retardo significativo de entre los primeros 3. Dado que la significancia de los primeros retardos se suele usar para determinar el orden de los procesos no estacionales, y que la significancia de los retardos múltiplos de S se suele emplear para determinar el orden de los procesos estacionales, los modelos que proponemos para estimar son:

$$\text{ARIMA } (0, 1, 0) \times (1, 1, 0)_4$$

Es decir, un modelo con un proceso AR(1) estacional, un orden de integración estacional (dado que $S = 4$, implica 4 diferencias), un orden de integración no estacional (una diferenciación adicional), y ningún proceso AR o MA no estacional. Más formalmente, y sin los órdenes de integración, este modelo se expresa:

$$\begin{aligned}\Phi(B^4)z_t &= w_t \\ (1 - \Phi_1 B^4)z_t &= w_t \\ z_t - \Phi_1 z_{t-4} &= w_t \\ z_t &= \Phi_1 z_{t-4} + w_t\end{aligned}$$

Donde $z_t = (x_t - \mu)$. También proponemos:

$$\text{ARIMA } (0, 1, 0) \times (0, 1, 1)_4$$

Es decir, un modelo con un proceso MA(1) estacional, un orden de integración estacional (como $S = 4$, implica 4 diferencias), un orden de integración no estacional, y ningún proceso AR o MA no estacional. Formalmente:

$$\begin{aligned}z_t &= \Theta(B^4)w_t \\ z_t &= (1 + \Theta_1 B^4)w_t \\ z_t &= w_t + \Theta_1 w_{t-4}\end{aligned}$$

2) Estimación

a) Ajuste del modelo:

Ahora estimamos los 2 modelos propuestos con la función `sarima()` del paquete `astsa`. La misma usa como argumentos la serie de tiempo (los datos) y los órdenes del modelo ARIMA estacional de la siguiente forma: `sarima(data,p,d,q,P,D,Q,S)`. Entonces:

```
AR_estacional <- sarima(ln_GDP_Bra,0,1,0,1,1,0,4,details = F, Model = F)
MA_estacional <- sarima(ln_GDP_Bra,0,1,0,0,1,1,4, details = F, Model = F)
```

b) Diagnóstico del modelo:

Primero vemos si los coeficientes estimados son significativos analizando los contrastes t de significación individual. Comenzamos por el modelo $\text{ARIMA } (0, 1, 0) \times (1, 1, 0)_4$:

```
AR_estacional$ttable
```

```
##      Estimate      SE t.value p.value
## sar1  -0.6567 0.1235 -5.3164      0
```

Como el p-value es menor a 0.05, queda claro que el coeficiente AR(1) estacional (Φ_1) es significativamente distinto de 0. Ahora seguimos con el modelo ARIMA $(0, 1, 0) \times (0, 1, 1)_4$:

```
MA_estacional$ttable
```

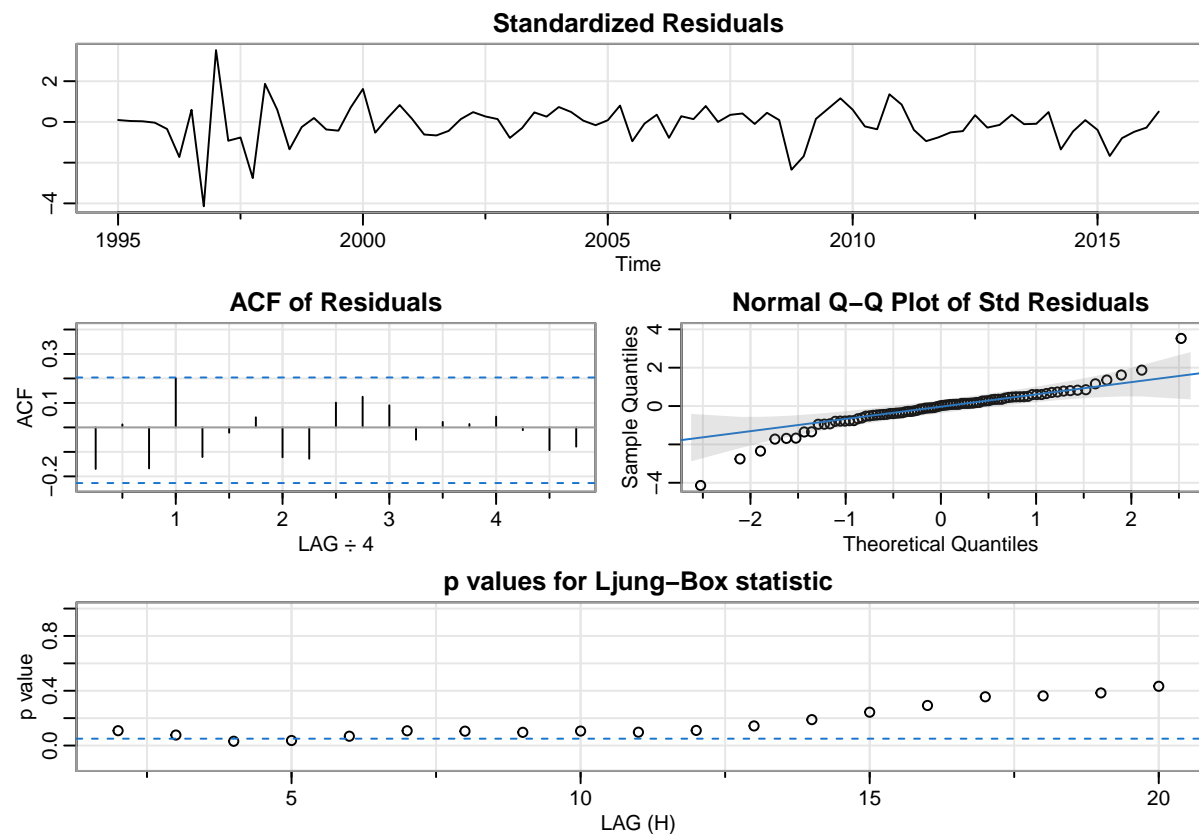
```
##      Estimate      SE t.value p.value
## sma1  -0.6283 0.1175 -5.3477      0
```

El coeficiente MA(1) estacional (Θ_1) también es significativo. Hasta ahora, ambos modelos estimados son válidos.

Continuamos con el análisis de los residuos. Para ello recurrimos al gráfico resumen provisto por la función `sarima()`. Comenzamos por el modelo ARIMA $(0, 1, 0) \times (1, 1, 0)_4$:

```
AR_estacional <- sarima(ln_GDP_Bra,0,1,0,1,1,0,4, Model = F)
```

```
## initial  value -4.015555
## iter    2 value -4.106786
## iter    3 value -4.136457
## iter    4 value -4.158900
## iter    5 value -4.158924
## iter    6 value -4.158930
## iter    6 value -4.158930
## iter    6 value -4.158930
## final   value -4.158930
## converged
## initial  value -3.752034
## iter    2 value -3.776259
## iter    3 value -3.785800
## iter    4 value -3.786434
## iter    5 value -3.786453
## iter    6 value -3.786476
## iter    6 value -3.786476
## iter    6 value -3.786476
## final   value -3.786476
## converged
```

Al parecer los residuos se distribuyen con media cero. En el correlograma de los residuos no se perciben señales de autocorrelación dado que ningún coeficiente de autocorrelación es significativo, aunque el cuarto se encuentra en el límite. El Q-Q plot indica que se puede suponer normalidad de los residuos, ya que sólo algunos pocos residuos estandarizados extremos se encuentran por fuera de los intervalos de confianza.

En cuanto al test de Ljung-Box, aquí sí se perciben problemas. Viendo el gráfico, pareciera haber, con un 95 % de confianza, autocorrelación de orden 4 y de orden 5. Esto se debe a que para los coeficientes de autocorrelación correspondientes, el p-value es menor a 0,05, por lo que se rechaza la hipótesis nula de ausencia de autocorrelación. Si no fuera por esto, se concluiría que los residuos son ruido blanco gaussiano. Ante estas conclusiones, correspondería rechazar el modelo. Aun así, el gráfico no resulta lo suficientemente claro, por lo que sería conveniente corroborar la significancia de los residuos 4 y 5 individualmente. Podemos valernos de la función `Box.test` para esto:

```
LJ_4 <- Box.test(AR_estacional$fit$residuals,4,type = c('Ljung-Box'))
LJ_5 <- Box.test(AR_estacional$fit$residuals,5,type = c('Ljung-Box'))
LJ_4$p.value
```

```
## [1] 0.06481629
```

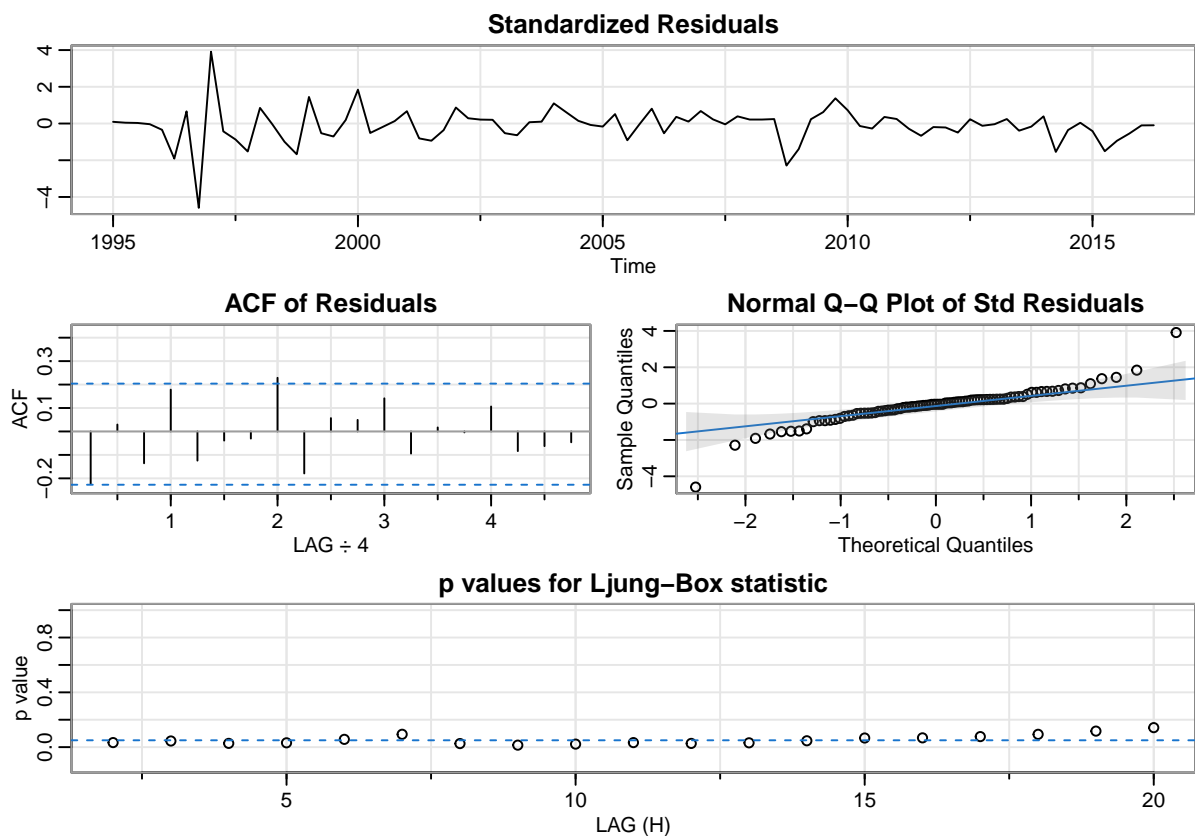
```
LJ_5$p.value
```

```
## [1] 0.06919594
```

Resulta que en realidad no se rechaza la hipótesis nula de autocorrelación de orden 4 ni de orden 5. En conclusión, los residuos son ruido blanco gaussiano y se acepta el modelo. Veamos ahora el modelo $\text{ARIMA}(0, 1, 0) \times (0, 1, 1)_4$:

```
MA_estacional <- sarima(ln_GDP_Bra,0,1,0,0,1,1,4, Model = F)
```

```
## initial value -3.656755
## iter 2 value -3.713113
## iter 3 value -3.713397
## iter 4 value -3.713418
## iter 4 value -3.713418
## final value -3.713418
## converged
## initial value -3.741632
## iter 2 value -3.770382
## iter 3 value -3.774779
## iter 4 value -3.775621
## iter 5 value -3.775645
## iter 6 value -3.775646
## iter 6 value -3.775646
## iter 6 value -3.775646
## final value -3.775646
## converged
```



En este caso, nuevamente, los residuos se distribuyen con media cero y a partir del Q-Q plot se puede suponer normalidad de los residuos. Aun así, y a diferencia del modelo anterior, el correlograma ya muestra señales de autocorrelación si se observa el retardo 8. Ljung-Box no hace más que confirmarlo, esta vez con mayor énfasis, ya que se pueden ver numerosos coeficientes de autocorrelación a los que les corresponden p-values menores a 0.05. Claramente se debería rechazar este modelo.

c) Selección del modelo:

Con base en el paso anterior, nos quedaríamos únicamente con un modelo (AR_estacional), ya que el otro bajo consideración fue rechazado. De cualquier forma, recurriremos a la comparación de los estadísticos de Akaike (AIC) de cada uno de los dos modelos estimados para seleccionar el mejor de ellos, obviando los problemas de autocorrelación de MA_estacional con fines demostrativos:

```
df_aic <- data.frame("Modelo SARIMA"=c("(0,1,0)x(1,1,0)4",
                                         "(0,1,0)x(0,1,1)4"),
                    "AIC"=c(AR_estacional$AIC,MA_estacional$AIC))
library(knitr)
kable(df_aic)
```

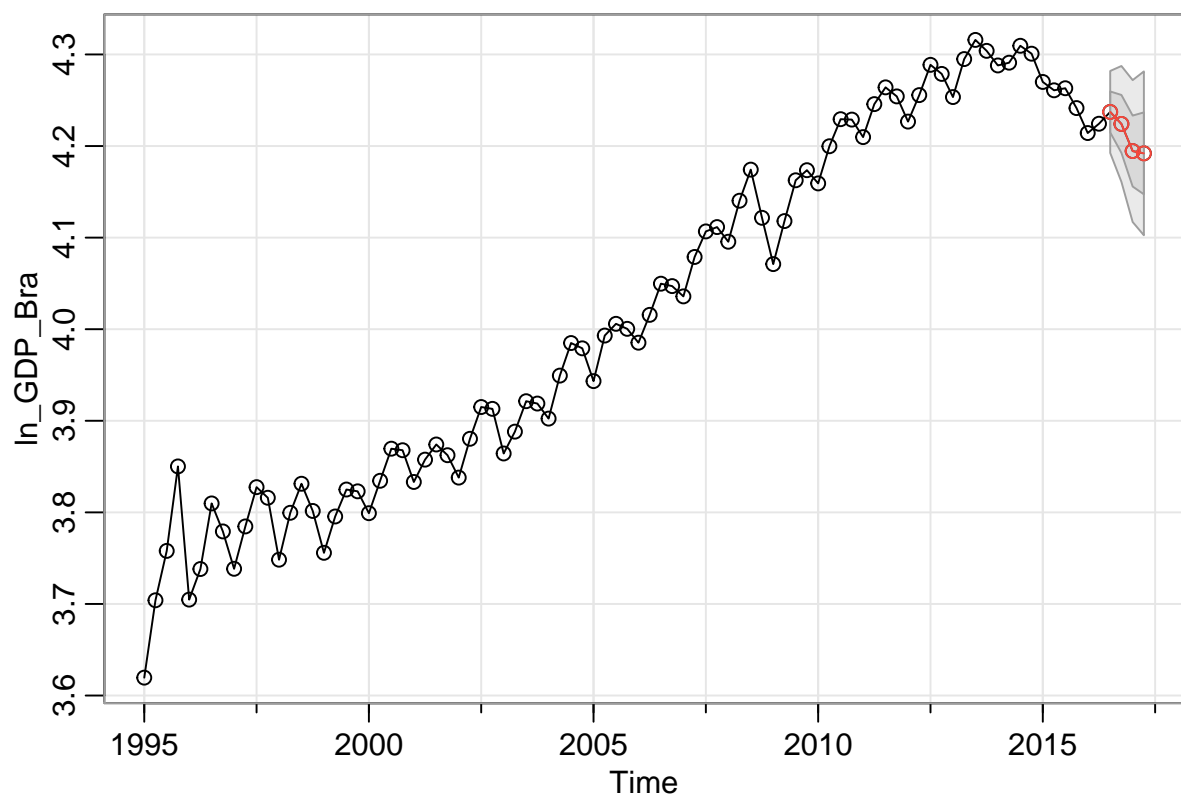
Modelo.SARIMA	AIC
(0,1,0)x(1,1,0)4	-4.685692
(0,1,0)x(0,1,1)4	-4.664032

Dado que el valor del estadístico para el modelo ARIMA $(0,1,0) \times (1,1,0)_4$ (AR_estacional) es menor, se lo elige por sobre el otro, al ser de mayor calidad. Es decir, incluso obviando los problemas de ARIMA $(0,1,0) \times (0,1,1)_4$ (MA_estacional) lo hubiéramos rechazado en favor de AR_estacional.

3) Predicción

Continuando con la suposición de que con un nivel de confianza del 95 % el modelo ARIMA $(0,1,0) \times (1,1,0)_4$ no exhibe autocorrelación y que por ende es válido, sumado a que el Criterio de Información de Akaike confirma que se trata de un modelo superior al otro propuesto, proseguimos con el último paso de la metodología de Box-Jenkins, que consiste en utilizar el modelo seleccionado para hacer proyecciones a futuro. Con la función `sarima.for()` esto se puede llevar a cabo. Lo hacemos para los cuatro trimestres siguientes al último observado, con el argumento `n.ahead`. El gráfico generado muestra que continuaría la tendencia negativa que comenzó a sufrir el GDP de Brasil en el año 2014, por un año más por lo menos.

```
sarima.for(ln_GDP_Bra, n.ahead = 4, 0,1,0,1,1,0,4)
```



```
## $pred
##           Qtr1      Qtr2      Qtr3      Qtr4
## 2016                4.237239 4.224139
## 2017 4.194525 4.192013
##
## $se
##           Qtr1      Qtr2      Qtr3      Qtr4
## 2016                0.02236160 0.03162407
## 2017 0.03873142 0.04472319
```

Inciso b)

La prueba desarrollada por Hylleberg, Engle, Granger y Yoo (HEGY) se emplea para encontrar una serie estacionaria cuando hay presencia de una o varias raíces unitarias en series de tiempo con frecuencia mayor al año (particularmente mensuales o trimestrales). Siguiendo a Alonso y Semaán (2010), vamos a centrarnos en el caso de muestras trimestrales ($s = 4$) para explicar la prueba. Comenzamos recordando que la estacionalidad suele ser tratada empleando variables dummy, modelos ARMA estacionales (SARMA) o integración estacional. Es correcto el uso de la primera cuando el proceso generador de datos (DGP) es *estacional determinístico*, de la siguiente manera:

$$y_t = \mu + \sum_{i=1}^3 \alpha_i D_{i,t} + \varepsilon_t$$

Donde $D_{i,t}$ es una variable dummy que asume el valor 1 cuando la observación corresponde al trimestre i . Si el DGP incluye una tendencia determinística:

$$y_t = \mu + \sum_{i=1}^3 \alpha_i D_{i,t} + \sum_{i=1}^3 \beta_i D_{i,t} t + \varepsilon_t$$

Por el contrario, si el DGP corresponde a un proceso *estacionario estacional*, lo óptimo será estimar un modelo SARMA. Si además tenemos tendencia, nos quedaría:

$$y_t = \mu + \theta y_{t-4} + \beta t + \varepsilon_t$$

Por último tenemos la integración estacional, que es una aproximación correcta cuando el DGP implica un proceso *no estacional estacionario*. Una serie y_t con frecuencia de s períodos en un año será integrada de orden d_0 y d_s si $y_t = \Delta^{d_0} [S(L)]^{d_s} y_t$ es estacionario, donde $S(L) = 1 + L + L^2 + \dots + L^{s-1}$, L es el polinomio de rezagos y $\Delta_s = 1 - L^s$.

Teniendo en cuenta los casos vistos hasta aquí, pasamos a desarrollar la prueba HEGY propiamente dicha. Para datos trimestrales, el operador de diferencia trimestral se puede escribir:

$$\Delta_4 y_t = y_t - y_{t-4} = (1 - L^4) y_t$$

Factorizando:

$$(1 - L^4) y_t = (1^2 - (L^2)^2) y_t = (1 - L^2)(1 + L^2) y_t = (1 - L)(1 + L)(1 + L^2) y_t$$

Si $\Delta_4 y_t$ es estacionario, al menos una de las siguientes condiciones aplica:

1. $(1 - L)y_t = y_t - y_{t-1} = 0$, por lo que $y_t - y_{t-1}$ será estacionario. Esto implica $y_t \approx y_{t-1}$. Es decir, existe una *raíz unitaria no estacional*.
2. $(1 + L)y_t = y_t + y_{t-1} = 0$, por lo que $y_t + y_{t-1}$ será estacionario. Esto implica $y_t \approx -y_{t-1} \approx y_{t-2}$. Es decir, existe una *raíz unitaria semestral*.
3. $(1 + L^2)y_t = y_t + y_{t-2} = 0$, por lo que $y_t + y_{t-2}$ será estacionario. Esto implica $y_t \approx -y_{t-2} \approx y_{t-4}$. Es decir, existe una *raíz unitaria anual*.

Si se quiere probar si la Condición 1 se cumple, se debe determinar si la pendiente es o no igual a cero al estimar y_t en función de y_{t-1} . Uno se vería tentado a estimar entonces:

$$(1 - L)y_t = \mu + \pi y_{t-1} + \varepsilon_t$$

Pero al hacer esto se corre el riesgo de estar empleando series no estacionarias y de incurrir en el problema de la regresión espuria. Para evitar esto se puede multiplicar a ambos lados por $(1 + L)(1 + L^2)$, obteniendo la regresión de $\Delta_4 y_t$ con respecto a $(1 + L)(1 + L^2)y_t$ (y ahí se analiza la pendiente). Entonces, para probar si la Condición 2 se cumple, se corre la regresión de $\Delta_4 y_t$ con respecto a $(1 - L)(1 + L^2)y_t$, y para probar la Condición 3 se corre $\Delta_4 y_t$ con respecto a $(1 - L)(1 + L)y_t$ (para probar los ciclos con pico en los trimestres 1 y 3) o con respecto a $(1 - L)(1 + L)y_{t-2}$ (para picos en los trimestres 2 y 4).

La prueba HEGY emplea estadísticos que permiten comprobar las 3 condiciones conjuntamente y de forma individual, definiendo las siguientes variables:

$$\begin{aligned} y_{1t} &\equiv (1 + L)(1 + L^2)y_t = y_t + y_{t-1} + y_{t-2} + y_{t-3} \\ y_{2t} &\equiv -(1 - L)(1 + L^2)y_t = -(y_t - y_{t-1} + y_{t-2} - y_{t-3}) \\ y_{3t} &\equiv (1 - L)(1 + L)y_t = y_t - y_{t-2} \\ y_{4t} &\equiv (1 - L^4)y_t = y_t - y_{t-4} \end{aligned}$$

Con estas variable (el signo negativo en y_{2t} es para asegurar la dirección usual de la prueba t de hipótesis), y teniendo en cuenta la posibilidad de que exista una tendencia determinística y una estacionalidad determinística (correspondería emplear en ese caso variables dummy, como ya vimos), la prueba HEGY implica correr la siguiente regresión:

$$y_{4t} = \mu_t + \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{4,t-1} + \sum_{i=1}^p y_{4,t-i} + \varepsilon_t \quad (11)$$

Donde $\mu_t = \delta + \beta t + \sum_{i=1}^3 \alpha D_{it}$. Los rezagos de y_{4t} se incluyen para eliminar posible autocorrelación en el error. A partir de (11) se pueden comprobar las siguientes hipótesis nulas:

- $H_A : \pi_1 = 0$ (presencia de una *raíz unitaria no estacional*)
- $H_B : \pi_2 = 0$ (presencia de una *raíz unitaria semestral*)
- $H_C : \pi_3 = \pi_4 = 0$ (presencia de una *raíz unitaria anual*)

H_A y H_B son probadas por medio de pruebas tipo t de una sola cola, mientras que para H_C se deberá emplear una prueba tipo F . Los resultados de esta prueba nos permitirán determinar qué diferenciaciones realizar (en caso de que sea necesario) para que la serie sea estacionaria. Se puede recurrir a la tabla 2 como guía.

En términos prácticos el primer paso de la prueba es determinar qué versión del modelo (11) se debe usar. Particularmente, si se va a emplear un intercepto (δ), una tendencia (βt) y/o variables dummy ($\sum_{i=1}^3 \alpha D_{it}$). Como el DGP es desconocido, se comienza observando el gráfico de la serie para estimar si se requiere el uso de una o varias de las 3 opciones recién consideradas. Por ejemplo, si la serie crece a lo largo del tiempo, se incluye la tendencia. Si se aprecia estacionalidad determinística, se incluyen variables dummy. Si esto no es seguro en un principio, se las incluye y luego se determina su significancia. Si no son significativas, se las elimina.

Una vez establecido el modelo a usar, se prosigue con la determinación del número de rezagos óptimos p . Para ello se emplean criterios de información (Akaike, Schwarz). El número de rezagos que minimiza el criterio de información utilizado será p . Habiendo hecho esto se estima el modelo y se analizan los p-values arrojados por la prueba, que nos dicen qué hipótesis se rechazan y cuál/cuáles no. Finalmente recurrimos a la tabla 2. Luego de hacer la diferenciación propuesta allí, se verifica si la serie es ahora estacionaria, es decir, si las 3 hipótesis se rechazan (no hay raíces unitarias).

Hipótesis que <i>NO</i> se rechazan	Hipótesis que se rechazan	Tipo de raíz unitaria presente	Transformación para obtener variable estacionaria
H_A, H_B, H_C	-	no estacional semestral anual	$\Delta_4 y_t = y_t - y_{t-4}$
H_A, H_B	H_C	no estacional semestral	$\Delta_2 y_t = y_t - y_{t-2}$
H_A, H_C	H_B	no estacional anual	$(1 - L)(1 + L^2) y_t$ $y_t - y_{t-1} + y_{t-2} - y_{t-3}$
H_B, H_C	H_A	semestral anual	$(1 + L)(1 + L^2) y_t$ $y_t + y_{t-1} + y_{t-2} + y_{t-3}$
H_A	H_B, H_C	no estacional	$\Delta_1 y_t = y_t - y_{t-1}$
H_B	H_A, H_C	semestral	$(1 + L) y_t = y_t + y_{t-1}$
H_C	H_A, H_B	anual	$(1 + L^2) y_t = y_t + y_{t-2}$
-	H_A, H_B, H_C	no hay	ninguna

Figura 2: Interpretación de los resultados de la prueba HEGY trimestral

Inciso c)

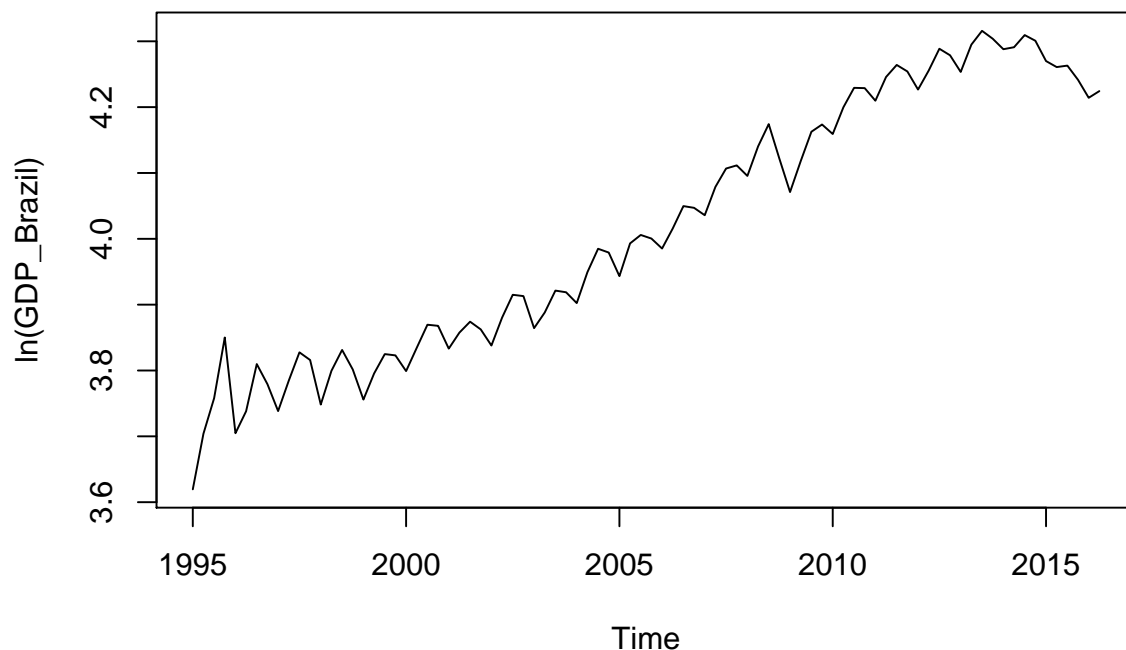
Comenzamos descargando la versión del paquete `uroot` usada por Alonso y Semaán (2010). Para ello primero descargamos el paquete `devtools` y luego hacemos: `install_version("uroot", version = "1.4-1")`. Ahora sí pasamos a aplicar el método HEGY a la serie analizada en el inciso **a)** del logaritmo natural del GDP de Brasil. Dado que se trata de datos trimestrales, el primero paso es determinar qué versión del siguiente modelo utilizar:

$$y_{4t} = \delta + \beta t + \sum_{i=1}^3 \alpha D_{it} + \pi_1 y_{1,t-1} + \pi_2 y_{2,t-1} + \pi_3 y_{3,t-2} + \pi_4 y_{4,t-1} + \sum_{i=1}^p y_{4,t-1} + \varepsilon_t$$

Para ello lo que hacemos es observar el gráfico de la serie para determinar si corresponde incluir tendencia (βt) y/o variables dummy ($\sum_{i=1}^3 \alpha D_{it}$). Veamos nuevamente el plot de `ln_GDP_Bra`:

```
plot(ln_GDP_Bra, main = "Logaritmo natural del GDP de Brasil")
```

Logaritmo natural del GDP de Brasil



Ya que la serie crece a lo largo del tiempo, corresponde incluir la tendencia en el modelo. Por otro lado, para que sea correcto incluir las variables dummy deberíamos poder apreciar en el gráfico estacionalidad no estocástica. Este último pareciera ser el caso, aunque no es del todo seguro. Así, decidimos comenzar con el modelo más general que sí incluye las variables dummy trimestrales para luego determinar su significancia. En consecuencia, usamos la función `HEGY.test()` para estimar el modelo inicial. El problema es que si usamos nuestra serie `ln_GDP_Bra` en `wts` el código devuelve error. Esto se debe a que esta versión de `uroot` requiere que los datos de la serie tengan pocos decimales. Transformamos la serie para que tenga sólo dos decimales:

```
lnGDP <- log(GDP_Arg_Bra[,3])
lnGDP2 <- format(round(lnGDP,2),nsmall=2)
lnGDP3 <- as.numeric(lnGDP2)
ln_GDP_Bra2 <- ts(lnGDP3,
                  start = c(1995,1), frequency = 4)
head(ln_GDP_Bra2)
```

```
## [1] 3.62 3.70 3.76 3.85 3.70 3.74
```

Ahora sí, estimamos el primer modelo HEGY. El argumento `itsd` es una lista de 3 elementos. El primero corresponde al intercepto, el segundo a la tendencia y el tercero a las variables dummy. Un 1 implica que se incluye ese elemento en el modelo, y 0 lo contrario. Donde se lee `mode` se especifica el criterio de información con el que se quiere

que se determine el número óptimo de rezagos p . En este caso usamos el criterio de Akaike (aic):

```
library(uroot)
hegy1 <- HEGY.test(wts = ln_GDP_Bra2, itsd = c(1,1,c(1:3)),
                  regvar = 0, selectlags = list(mode="aic", Pmax=NULL))
summary(hegy1)
```

```
##      ----
##      HEGY test
##      ----
##
## ----
##      Deterministic regressors estimates:
##
##              Estimate Std. Error t value Pr(>|t|)
## Intercept          0.326      0.234   1.392   0.169
## Trend              0.001      0.001   1.238   0.220
## SeasDummy.1       -0.021      0.010  -2.041   0.045
## SeasDummy.2        0.027      0.013   2.062   0.043
## SeasDummy.3        0.019      0.010   1.859   0.068
##
## ----
##      HEGY regressors estimates:
##
##              Estimate Std. Error t value Pr(>|t|)
## Ypi1       -0.023      0.016  -1.401   0.10
## Ypi2       -0.584      0.136  -4.292   0.01
## Ypi3       -0.386      0.112  -3.458   0.06
## Ypi4       -0.347      0.115  -3.016   0.01
##
## ----
##      Selected lags estimates:
##
##              Estimate Std. Error t value Pr(>|t|)
## Lag.1        0.255      0.127   2.018   0.048
## Lag.3       -0.134      0.079  -1.705   0.093
## Lag.5        0.182      0.067   2.714   0.008
##
##      Number of available observations: 77
```

Con un nivel de significancia del 1 %, las variables dummy trimestrales (SeasDummy) no son significativas, por lo que reestimamos el modelo sin incluirlas:

```
hegy2 <- HEGY.test(wts = ln_GDP_Bra2, itsd = c(1,1,c(0)),
                  regvar = 0, selectlags = list(mode="aic", Pmax=NULL))
```

Debemos analizar los p-values de las pruebas de significación para determinar si rechazamos o no las 3 hipótesis planteadas en el inciso **b)**:

```
hegy2

##      ---- ----
##      HEGY test
##      ---- ----
##
##      Null hypothesis: Unit root.
##      Alternative hypothesis: Stationarity.
##
## ----
##      HEGY statistics:
##
##              Stat. p-value
## tpi_1      -1.373    0.100
## tpi_2      -2.472    0.015
## Fpi_3:4     1.201    0.100
## Fpi_2:4     2.865      NA
## Fpi_1:4     2.714      NA
##
##      Lag orders: 1 4 5 11
##      Number of available observations: 71
```

tpi_1 se refiere a H_A ($\pi_1 = 0$), **tpi_2** se refiere a H_B ($\pi_2 = 0$) y **Fpi_3:4** se refiere a H_C ($\pi_3 = \pi_4 = 0$). Las primeras dos son pruebas t , mientras que la tercera es una prueba F , de ahí los nombres. Al 1 % de significancia no se rechaza ninguna de las 3 hipótesis, y al 5 % sólo se rechaza H_B . Se llega a las mismas conclusiones usando el criterio de Akaike ("aic"), el criterio de Schwarz ("bic") y el criterio de significancia ("signf") para determinar el número de rezagos óptimos (el que minimiza el criterio de información correspondiente). Vamos a estimar los modelos que se derivan de cada caso, siguiendo las transformaciones propuestas en Alonso y Semaán (2010) para obtener una variable estacionaria (tabla 2).

Comenzamos por el caso en que no rechazamos ni H_A ni H_C y sí rechazamos H_B , es decir, $\pi_1 = \pi_3 = \pi_4 = 0$ y $\pi_2 \neq 0$. Se interpreta que asumimos raíz unitaria no estacional y anual. La transformación que se aplica a la serie es:

$$(1 - L)(1 + L^2)y_t = y_t - y_{t-1} + y_{t-2} - y_{t-3}$$

Una vez hecho esto, estimamos el modelo con la nueva serie:

```
dif_B <- (ln_GDP_Bra2 - lag(ln_GDP_Bra2,-1) + lag(ln_GDP_Bra2,-2) -
          lag(ln_GDP_Bra2,-3))
hegy3 <- HEGY.test(wts = dif_B, itsd = c(1,1,0),
                  regvar = 0, selectlags = list(mode="aic", Pmax=NULL))
hegy3
```

```
## ---- ----
## HEGY test
## ---- ----
##
## Null hypothesis: Unit root.
## Alternative hypothesis: Stationarity.
##
## ----
## HEGY statistics:
##
##          Stat. p-value
## tpi_1    -1.261    0.100
## tpi_2    -1.646    0.092
## Fpi_3:4  49.222    0.010
## Fpi_2:4  36.418      NA
## Fpi_1:4  31.015      NA
##
## Lag orders: 2 7 8
## Number of available observations: 71
```

Con la transformación efectuada no se obtuvo una serie estacionaria. Para que eso ocurra se deberían rechazar las 3 hipótesis (no existen raíces unitarias), pero aquí se da que al 1% de significancia no se rechaza ninguna y al 5% se logra rechazar sólo H_C ($\pi_3 = \pi_4 = 0$). Veamos que sucede en el otro caso propuesto, que es aquel en el que no rechazamos ninguna de las 3 hipótesis (hay presencia de raíz unitaria no estacional, semestral y anual). La transformación pasa a ser:

$$(1 - L^4) = \Delta_4 y_t = y_t - y_{t-4}$$

```
dif_ABC <- ln_GDP_Bra2 - lag(ln_GDP_Bra2,-4)
hegy4 <- HEGY.test(wts = dif_ABC, itsd = c(1,1,0),
                  regvar = 0, selectlags = list(mode="aic", Pmax=NULL))
hegy4
```

```
## ---- ----
## HEGY test
## ---- ----
##
## Null hypothesis: Unit root.
```

```
## Alternative hypothesis: Stationarity.
##
## ----
## HEGY statistics:
##
##          Stat. p-value
## tpi_1    1.189    0.10
## tpi_2   -7.118    0.01
## Fpi_3:4 47.219    0.01
## Fpi_2:4 47.068     NA
## Fpi_1:4 51.505     NA
##
## Lag orders: 1 5 7 8 10
## Number of available observations: 68
```

Lo que sucede es que tanto al 1% como al 5% de significancia se rechazan H_B y H_C pero no se rechaza H_A . Por ende, con esta transformación tampoco se obtuvo una variable estacionaria. Este último resultado nos da pie para aplicar el test de HEGY a la serie obtenida en el inciso a), que surgía de primero tomar 4 diferencias para eliminar estacionalidad y luego tomar primeras diferencias nuevamente para terminar de eliminar la tendencia. Hacer esto equivale a seguir el procedimiento del test de HEGY nuevamente. Dado que H_A es la única que no se rechaza, corresponde hacer $\Delta_1 y_t = y_t - y_{t-1}$ sobre la serie transformada previamente (dif_ABC , que surgió de aplicar $y_t - y_{t-4}$). Recordemos que esto implica:

$$(1 - L^4)(1 - L)y_t = (y_t - y_{t-1}) - (y_{t-4} - y_{t-5})$$

Entonces:

```
dif_ABCy1 <- dif_ABC - lag(dif_ABC,-1)
hegy5 <- HEGY.test(wts = dif_ABCy1, itsd = c(1,1,0),
                  regvar = 0, selectlags = list(mode="aic", Pmax=NULL))
hegy5
```

```
## ---- ----
## HEGY test
## ---- ----
##
## Null hypothesis: Unit root.
## Alternative hypothesis: Stationarity.
##
## ----
## HEGY statistics:
##
##          Stat. p-value
## tpi_1   -5.549    0.01
```

```
## tpi_2    -3.002    0.01
## Fpi_3:4  35.509    0.01
## Fpi_2:4  24.705     NA
## Fpi_1:4  27.345     NA
##
##   Lag orders: 1 3 4 5 6
##   Number of available observations: 71
```

Ahora sí se rechazaron todas las hipótesis con un nivel de confianza del 95%. Por ende, con esta última transformación la serie no presenta raíces unitarias estacionales ni no estacionales (es estacionaria). Por lo tanto, se llegó a la misma conclusión que con Dickey-Fuller en el inciso [a\)](#).

Ejercicio VIII

En Econometría I, al estudiar el problema de la heterocedasticidad, se encontró que una de sus posibles soluciones consistía en obtener por Mínimos Cuadrados Ordinarios (MCO) las estimaciones puntuales de los coeficientes y luego, para construir los intervalos de confianza o realizar pruebas de hipótesis, emplear estimadores robustos de los errores estándar, entre otros los planteados por White (1980). Se le pide:

a) Describa brevemente en qué consisten los estimadores de errores estándar robustos (HC). Encontrará información útil en el capítulo 4 de Kleiber & Zeileis (2008).

b) A partir de la base de datos “PublicSchools” de la biblioteca AER:

- I. Estime por MCO una regresión usando la variable ingreso (“Expenditure”) como variable dependiente y a ingreso (“Income”) y su cuadrado como variables independientes.
- II. Docime la significatividad del coeficiente asociado al cuadrado del ingreso sin corregir su error estándar.
- III. Docime la significatividad del coeficiente asociado al cuadrado del ingreso usando el estimador “HC4”. Compare con lo obtenido en b) y concluya.

Inciso a)

Siguiendo a Kleiber y Zeileis (2008), los estimadores de errores estándar robustos son estimadores que permiten realizar inferencias válidas en modelos en los que se está en presencia de heterocedasticidad y/o autocorrelación. Además, según White (1980), cuando los errores son heterocedásticos, las estimaciones de los parámetros del modelo son consistentes pero ineficientes y la estimación de la matriz de covarianzas es inconsistente, lo cual reafirma lo dicho por Kleiber y Zeileis en relación a que se extraerán inferencias erróneas al realizar pruebas de hipótesis.

Por lo dicho anteriormente, los estimadores de errores estándar robustos se constituyen como importantes ya que, en general, los datos económicos siguen procesos no homocedásticos y/o autocorrelacionados, y se necesitan realizar inferencias válidas en los test de hipótesis.

Existen 2 tipos de estimadores estándar de errores robustos, un tipo es para poder lograr estimar matrices de varianzas y covarianzas consistentes con heterocedasticidad (HC), mientras que, por otro lado están los HAC, para estimar de manera más general

a la matriz de varianzas y covarianzas cuando los datos presentan heterocedasticidad y autocorrelación.

Procedemos ahora a explicar brevemente ambos tipos de estimadores, comenzando por el primer tipo, Heterokedasticity-Consistent Standard Errors (HC). En general, cuando se trabajan con datos de corte transversal, se toma que la matriz Ω , la cual es igual a la varianza del vector de coeficientes, esto es, $Var(\hat{\beta}|X) = \Omega$, es una matriz diagonal con elementos que no son constantes, por lo tanto, en general se utiliza una matriz diagonal para estimar a Ω , es decir:

$$\hat{\Omega} = (\omega_1, \dots, \omega_n)$$

Ahora bien, en la literatura se han sugerido diversas opciones para los elementos w_i , los que se encuentran en la diagonal principal de la matriz que se desea estimar, estos son:

- Estimador estándar cuando los errores se comportan de forma homocedástica:

$$w_i = \hat{\sigma}^2$$

- HC_0 desarrollado por Eicker (1963) y popularizado por White (1980):

$$w_i = \hat{\varepsilon}_i^2$$

- HC_1 por MacKinnon y White (1985) para muestras pequeñas:

$$w_i = \frac{n}{n-k} \hat{\varepsilon}_i^2$$

- HC_2 también de MacKinnon y White (1985) para muestras pequeñas:

$$w_i = \frac{\hat{\varepsilon}_i^2}{1 - h_{ii}}$$

- HC_3 , una vez más MacKinnon y White (1985) para muestras pequeñas, Long y Ervin (2000) concluyeron que HC_3 proporciona el mejor rendimiento para muestras pequeñas:

$$w_i = \frac{\hat{\varepsilon}_i^2}{(1 - h_{ii})^2}$$

- HC_4 sugerido por Cribari-Neto (2004):

$$w_i = \frac{\hat{\varepsilon}_i^2}{(1 - h_{ii})^{\delta_i}}$$

En donde h_{ii} son los elementos diagonales de la matriz Ω , \bar{h} es la media y $\delta_i = \min\left(4, \frac{h_{ii}}{\bar{h}}\right)$

Algo también a destacar, siguiendo a Cribari-Neto (2004), es que en modelos en donde no hay heterocedasticidad, pero con residuos que presentan valores atípicos u observaciones influyentes, los estimadores HC también arrojan buenos resultados.

Ahora bien, pasando al otro tipo de estimadores, los HAC (Heterokedasticity and Autocorrelation Consistent) resuelven el problema surgido en las series de tiempo cuando los términos de error ε_i se encuentran correlacionados, por lo que la matriz Ω no es diagonal y solamente puede estimarse añadiendo suposiciones adicionales sobre su estructura.

¿Qué se hace en estos casos? Se calculan sumas ponderadas de las autocorrelaciones empíricas de $\hat{\varepsilon}_i x_i$.

Debe tenerse en cuenta que los estimadores sugeridos, difieren con respecto a las ponderaciones utilizadas.

Inciso b)

Ítem I

En este inciso, se estima una regresión lineal, tomando como variable dependiente al ingreso (**Expenditure**) y como variables independientes (explicativas) al ingreso (**Income**) y a su cuadrado. Se estimará el modelo por MCO de la siguiente manera:

```
library(AER) # Se carga el paquete "AER" (Applied Econometrics with R)

data("PublicSchools") # Tomamos la base de datos "PublicSchools"

ps <- PublicSchools # Creamos el Data Frame

head(ps) # Aquí podemos observar las primeras 6 observaciones
```

##	Expenditure	Income
## Alabama	275	6247
## Alaska	821	10851
## Arizona	339	7374
## Arkansas	275	6183
## California	387	8850
## Colorado	452	8001

```
tail(ps) # Aquí podemos observar las últimas 6 observaciones
```

##	Expenditure	Income
## Virginia	356	7624
## Washington	415	8450
## Washington DC	428	10022
## West Virginia	320	6456
## Wisconsin	NA	7597
## Wyoming	500	9096

Ahora procedemos a estimar la regresión:


```
fit_ps <- lm(Expenditure~Income+I(Income^2),
            data = ps) # Estimamos el modelo antes mencionado
```

Ya estimamos el modelo, antes de pasar al siguiente inciso, analicemos los resultados obtenidos:

```
summary(fit_ps) # Observamos las salidas del modelo
```

```
##
## Call:
## lm(formula = Expenditure ~ Income + I(Income^2), data = ps)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -160.709  -36.896   -4.551   37.290  109.729
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.329e+02  3.273e+02   2.545  0.01428 *
## Income      -1.834e-01  8.290e-02  -2.213  0.03182 *
## I(Income^2)  1.587e-05  5.191e-06   3.057  0.00368 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 56.68 on 47 degrees of freedom
## (1 observation deleted due to missingness)
## Multiple R-squared:  0.6553, Adjusted R-squared:  0.6407
## F-statistic: 44.68 on 2 and 47 DF,  p-value: 1.345e-11
```

Podemos notar que todas las variables explicativas (constante, income y su cuadrado) son estadísticamente significativas al 5% por lo menos. Por otro lado, si tomamos el p-valor de la prueba F de validez global del modelo, al ser menor que cualquier nivel de significación convencional que se elija, nos dice que el modelo presenta un buen ajuste global.

Ítem II

En este inciso, se prueba formalmente en R si la variable explicativa `income2` es significativamente distinta de cero. Cabe destacar que hemos analizado en el inciso anterior la prueba t de significación individual de cada coeficiente, y esta variable se encontró significativamente distinta de 0. A pesar de ello, probamos de las siguientes maneras:

Primero, utilizamos la función `linearHypothesis` y expresamos la restricción que queremos probar, por lo tanto, nuestras hipótesis serán:

- $H_0: \beta_2 = 0$

$$\blacksquare H_1: \beta_2 \neq 0$$

A partir de ella, obtendremos dos modelos, uno restringido y uno no restringido y con sus sumas de cuadrados residuales, obtendremos un estadístico F . Este se comparará con un valor crítico $F_{(1,47,0.95)}^*$, en donde 1 representa la restricción, 47 son los grados de libertad del modelo no restringido, y 0,95 el nivel de confianza elegido. Observemos la salida de R:

```
linearHypothesis(fit_ps, "I(Income^2)=0") # Planteamos la hipótesis
```

```
## Linear hypothesis test
##
## Hypothesis:
## I(Income^2) = 0
##
## Model 1: restricted model
## Model 2: Expenditure ~ Income + I(Income^2)
##
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1      48 181015
## 2      47 150986   1      30030 9.3479 0.003677 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

La salida nos dice que el p-valor es menor a cualquier nivel de significación convencional (1 %, 5 %, 10 %), por lo cual se rechaza la hipótesis nula de que el coeficiente asociado a la variable Income^2 no es significativamente diferente de cero.

Al mismo resultado se arriba si se toma el valor F obtenido, igual a 9,34 y se lo compara con el valor crítico $F^* = 4.0470999$, dado que el valor del estadístico es mayor al valor crítico, se rechaza la hipótesis nula y se concluye con un nivel de significación del 5 % que el coeficiente asociado a la variable Income^2 es significativamente distinto de cero.

Ítem III

Ahora bien, a pesar de haber obtenido que el coeficiente asociado a la variable a analizar es significativamente distinto de cero, hemos estado trabajando con un error estándar inconsistente, lo cual lleva a realizar inferencias erróneas. Por esa razón, procedemos a utilizar un estimador de los errores estándar robusto, en este caso, utilizamos el estimador HC_4 sugerido por Cribar-Neto (2004):

$$w_i = \frac{\hat{\varepsilon}_i^2}{(1 - h_{ii})^{\delta_i}}$$

En base a este estimador, docimamos la significatividad del cuadrado del ingreso, y en este caso los resultados obtenidos son diferentes. Veamos:

```
# Realizamos la prueba t para los coeficientes
coeftest(fit_ps, vcov = vcovHC(fit_ps, type = "HC4"))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  8.3291e+02 3.0080e+03  0.2769   0.7831
## Income      -1.8342e-01 8.1832e-01 -0.2241   0.8236
## I(Income^2)  1.5870e-05 5.4889e-05  0.2891   0.7737
```

Los resultados del test nos muestran que en este caso, ninguna variable explicativa es estadísticamente significativa, es decir, los resultados son diferentes a los que se obtenían al estimar el modelo por MCO ignorando el problema de heterocedasticidad o de valores atípicos (el estado de Alabama, según Kleiber y Zeilis).

Por lo tanto, la conclusión es que, es importante tener presente el problema de heterocedasticidad y autocorrelación porque, como mencionamos antes al explicar teóricamente la situación, al ser los estimadores de los errores estándar inconsistentes, las pruebas de hipótesis nos brindan resultados erróneos (lo cual ocurrió en el inciso 2), y al utilizar un estimador robusto de los errores estándar (en este caso fue seleccionado el HC_4), el resultado fue diferente.

Anexo

Demostración Slutsky

Partiendo de:

$$\frac{\partial q_i}{\partial p_j} + q_j \frac{\partial q_i}{\partial y} = \frac{\partial q_j}{\partial p_i} + q_i \frac{\partial q_j}{\partial y}$$

Multiplico ambos miembros por $y \frac{p_j}{q_i} \frac{p_i}{q_j}$

$$y \frac{p_j}{q_i} \frac{p_i}{q_j} \frac{\partial q_i}{\partial p_j} + y \frac{p_j}{q_i} \frac{p_i}{q_j} q_j \frac{\partial q_i}{\partial y} = y \frac{p_j}{q_i} \frac{p_i}{q_j} \frac{\partial q_j}{\partial p_i} + y \frac{p_j}{q_i} \frac{p_i}{q_j} q_i \frac{\partial q_j}{\partial y}$$

Donde:

$$\frac{p_j}{q_i} \frac{\partial q_i}{\partial p_j} = \beta_{ij} \quad \frac{p_i}{q_j} \frac{\partial q_j}{\partial p_i} = \beta_{ji} \quad \frac{y}{q_i} \frac{\partial q_i}{\partial y} = \eta_i = \beta_{i4}$$

Entonces:

$$\begin{aligned} y \frac{p_i}{q_j} \beta_{ij} + p_j p_i \beta_{i4} &= y \frac{p_j}{q_i} \beta_{ji} + p_j p_i \beta_{i4} \\ p_i \left(y \frac{\beta_{ij}}{q_j} \right) + p_j \beta_{i4} &= p_j \left(y \frac{\beta_{ji}}{q_i} \right) + p_i \beta_{i4} \\ \frac{y}{p_j q_j} \beta_{ij} + \beta_{i4} &= \frac{y}{p_i q_i} \beta_{ji} + \beta_{i4} \\ \frac{\beta_{ij}}{w_j} + \beta_{i4} &= \frac{\beta_{ji}}{w_i} + \beta_{i4} \end{aligned}$$

Estimaciones Ejercicio 1 inciso b)

```
summls
```

```
##
## systemfit results
## method: OLS
##
##           N DF      SSR  detRCov   OLS-R2 McElroy-R2
## system  90  75  9.75021  0.000131  0.714334   0.982294
```

```
##
##      N DF      SSR      MSE      RMSE      R2      Adj R2
## d1 30 25 3.82956 0.153182 0.391385 0.655657 0.600562
## d2 30 25 5.23639 0.209456 0.457663 0.561813 0.491703
## d3 30 25 0.68426 0.027370 0.165440 0.938132 0.928233
##
##
## Coefficients:
##              Estimate Std. Error   t value   Pr(>|t|)
## d1_(Intercept) -4.8850224  1.5024859  -3.25129 0.00327610 **
## d1_log(p1)      -0.5676041  0.2194958  -2.58595 0.01592729 *
## d1_log(p2)       0.0729461  0.1679463   0.43434 0.66776594
## d1_log(p3)      -0.2031487  0.1379349  -1.47279 0.15328796
## d1_log(y)        1.4290952  0.2271721   6.29081 1.3925e-06 ***
## d2_(Intercept) -4.3107064  1.7569198  -2.45356 0.02146019 *
## d2_log(p1)       0.1907108  0.2566656   0.74303 0.46438780
## d2_log(p2)      -0.6323996  0.1963866  -3.22018 0.00353654 **
## d2_log(p3)       0.1351623  0.1612930   0.83799 0.40997562
## d2_log(y)        1.1406341  0.2656418   4.29388 0.00023224 ***
## d3_(Intercept)  1.0165828  0.6351063   1.60065 0.12201693
## d3_log(p1)      -0.2503605  0.0927817  -2.69838 0.01230418 *
## d3_log(p2)      -0.1442449  0.0709915  -2.03186 0.05292355 .
## d3_log(p3)      -0.9652175  0.0583056 -16.55446 5.5511e-15 ***
## d3_log(y)        0.8796938  0.0960265   9.16095 1.8232e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
summsur
```

```
##
## systemfit results
## method: SUR
##
##      N DF      SSR detRCov  OLS-R2 McElroy-R2
## system 90 75 9.75021 0.000131 0.714334  0.982294
##
##      N DF      SSR      MSE      RMSE      R2      Adj R2
## d1 30 25 3.82956 0.153182 0.391385 0.655657 0.600562
## d2 30 25 5.23639 0.209456 0.457663 0.561813 0.491703
## d3 30 25 0.68426 0.027370 0.165440 0.938132 0.928233
##
##
## Coefficients:
##              Estimate Std. Error   t value   Pr(>|t|)
## d1_(Intercept) -4.8850224  1.5024859  -3.25129 0.00327610 **
## d1_log(p1)      -0.5676041  0.2194958  -2.58595 0.01592729 *
```

```
## d1_log(p2)      0.0729461  0.1679463   0.43434  0.66776594
## d1_log(p3)     -0.2031487  0.1379349  -1.47279  0.15328796
## d1_log(y)       1.4290952  0.2271721   6.29081  1.3925e-06 ***
## d2_(Intercept) -4.3107064  1.7569198  -2.45356  0.02146019 *
## d2_log(p1)      0.1907108  0.2566656   0.74303  0.46438780
## d2_log(p2)     -0.6323996  0.1963866  -3.22018  0.00353654 **
## d2_log(p3)      0.1351623  0.1612930   0.83799  0.40997562
## d2_log(y)       1.1406341  0.2656418   4.29388  0.00023224 ***
## d3_(Intercept)  1.0165828  0.6351063   1.60065  0.12201693
## d3_log(p1)     -0.2503605  0.0927817  -2.69838  0.01230418 *
## d3_log(p2)     -0.1442449  0.0709915  -2.03186  0.05292355 .
## d3_log(p3)     -0.9652175  0.0583056 -16.55446  5.5511e-15 ***
## d3_log(y)       0.8796938  0.0960265   9.16095  1.8232e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

summolSR

```
##
## systemfit results
## method: OLS
##
##          N DF      SSR  detRCov   OLS-R2 McElroy-R2
## system  90 82 11.1669 0.000189 0.672827   0.977971
##
##      N DF      SSR      MSE      RMSE      R2  Adj R2
## d1 30 25 4.341540 0.166982 0.408635 0.609621 0.564578
## d2 30 25 5.902647 0.227025 0.476471 0.506060 0.449067
## d3 30 25 0.922714 0.035489 0.188385 0.916572 0.906946
##
##
## Coefficients:
##              Estimate Std. Error  t value  Pr(>|t|)
## d1_(Intercept) -2.46778032  0.66599825  -3.70539 0.00038198 ***
## d1_log(p1)     -0.80408607  0.14980484  -5.36756 7.2625e-07 ***
## d1_log(p2)     -0.09436549  0.07970368  -1.18395 0.23985201
## d1_log(p3)     -0.28876206  0.10958267  -2.63511 0.01005484 *
## d1_log(y)       1.18721362  0.16753024   7.08656 4.3563e-10 ***
## d2_(Intercept) -1.52757958  0.67255289  -2.27132 0.02574841 *
## d2_log(p1)     -0.07573713  0.12814922  -0.59101 0.55614110
## d2_log(p2)     -0.80947435  0.13239060  -6.11429 3.1379e-08 ***
## d2_log(p3)      0.02690637  0.11678592   0.23039 0.81836215
## d2_log(y)       0.85830512  0.16801924   5.10837 2.0778e-06 ***
## d3_(Intercept) -0.30660090  0.25251944  -1.21417 0.22816901
## d3_log(p1)     -0.04454892  0.04752063  -0.93746 0.35127248
## d3_log(p2)     -0.00976168  0.03124881  -0.31239 0.75554035
```

```
## d3_log(p3)      -0.91726723  0.04123670 -22.24396 < 2.22e-16 ***
## d3_log(y)       0.97157784  0.06136446  15.83291 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

summsurR

```
##
## systemfit results
## method: SUR
##
##          N DF      SSR detRCov   OLS-R2 McElroy-R2
## system 90 82 11.2246 0.000187 0.671137   0.978536
##
##      N DF      SSR      MSE      RMSE      R2   Adj R2
## d1 30 25 4.365748 0.167913 0.409772 0.607445 0.562150
## d2 30 25 5.906975 0.227191 0.476646 0.505698 0.448663
## d3 30 25 0.951875 0.036611 0.191339 0.913935 0.904005
##
##
## Coefficients:
##              Estimate Std. Error  t value  Pr(>|t|)
## d1_(Intercept) -2.78276801  0.76361941  -3.64418 0.00046916 ***
## d1_log(p1)      -0.86450681  0.16945525  -5.10168 2.1343e-06 ***
## d1_log(p2)      -0.12518943  0.09780931  -1.27993 0.20417774
## d1_log(p3)      -0.27739752  0.12649243  -2.19300 0.03114023 *
## d1_log(y)        1.26709375  0.19205608   6.59752 3.8243e-09 ***
## d2_(Intercept) -1.75771299  0.87873197  -2.00028 0.04877942 *
## d2_log(p1)      -0.11786415  0.15813393  -0.74534 0.45819504
## d2_log(p2)      -0.83557472  0.17104824  -4.88502 5.0434e-06 ***
## d2_log(p3)       0.03687038  0.15313923   0.24076 0.81033916
## d2_log(y)        0.91656849  0.21938986   4.17781 7.3078e-05 ***
## d3_(Intercept) -0.16373845  0.30398448  -0.53864 0.59159400
## d3_log(p1)      -0.01755511  0.05613171  -0.31275 0.75526568
## d3_log(p2)       0.00494286  0.04101331   0.12052 0.90436728
## d3_log(p3)      -0.92275588  0.05156872 -17.89371 < 2.22e-16 ***
## d3_log(y)        0.93536814  0.07607651  12.29510 < 2.22e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Referencias

- ALONSO, J. C. & SEMAÁN, P. (2010). *Prueba de HEGY en R: una guía*. Apuntes de Economía, 23.
- BERNDT, E. R., HALL, B. H., HALL, R. E., & HAUSMAN, J. A. (1974). *Estimation and inference in nonlinear structural models*. In Annals of Economic and Social Measurement, Volume 3, number 4 (pp. 653-665). NBER.
- CROISSANT, Y. & MILLO, G. (2008). *Panel data econometrics in R: The plm package*. Journal of statistical software, 27(2).
- DHRYMES, P. J. (1973). *Small sample and asymptotic relations between maximum likelihood and three stage least squares estimators*. Econometrica: Journal of the Econometric Society, 357-364.
- FIGUEROA, C., VÁZQUEZ, P. C., MARTÍN, I. T., & DEL REY, S. (2012). *James Heckman, el sesgo de selección muestral*.
- FOX, J. (2006). *Teacher's corner: structural equation modeling with the sem package in R*. Structural equation modeling, 13(3), 465-486.
- HECKMAN, J. J. (1979). *Sample selection bias as a specification error*. Econometrica: Journal of the econometric society, 153-161.
- HENNINGSEN, A., & HAMANN, J. D. (2008). *systemfit: A package for estimating systems of simultaneous equations in R*. Journal of statistical software, 23, 1-40.
- HYLLEBERG, S., ENGLE, R. F., GRANGER, C. W., & YOO, B. S. (1990). *Seasonal integration and cointegration*. Journal of Econometrics, 44(1-2), 215-238.
- JUDGE, G. G., W. E. GRIFFITHS, R. C. HILL, H. LUETKEPOHL & T. C. LEE (1985). *Theory and Practice of Econometrics*, Second Edition. Wiley. New York.
- JUDGE, GEORGE G., HILL, R. CARTER, GRIFFITHS, WILLIAM E., LUETKEPOHL, HELMUT AND LEE, TSOUNG CHAO (1988). *Introduction to the Theory and Practice of Econometrics*, Second Edition. Wiley. New York.
- KLEIBER, C., & ZEILEIS, A. (2008). *Applied econometrics with R*. Springer Science & Business Media.
- KMENTA, JAN (1986). *Elements of Econometrics*, Second Edition. Macmillan. New York.

- LANTERI, L. (2009). *Respuesta a precios del área sembrada de soja en la Argentina (No. 2009/44)*. Working Paper.
- LAYARD, P.R.G. & WALTERS, A.A. (1978). *Microeconomic Theory*. McGrawHill. New York. Cap. 5.
- LEYDEN, D. P. (1991). *Modified quadratic hill-climbing with SAS/IML*. Computer Science in Economics and Management, 4(1), 15-31.
- MONETA PIZARRO, A.M. (2021). *Econometría*. Editorial Económicas.
- MONTERO, R. (2005). *Test de Hausman*. Documentos de Trabajo en Economía Aplicada. Universidad de Granada. España.
- OSBORNE, M. R. (1992). *Fisher's method of scoring*. International Statistical Review/Revue Internationale de Statistique, 99-117.
- PÉREZ LÓPEZ, C. (2006). *Problemas resueltos de econometría*. Editorial Paraninfo.
- PHILLIPS, P.C.B. & WICKENS, M.R. (1978). *Exercises in Econometrics – Volume two*. Philip Allan. Oxford.
- REDRADO, M., CARRERA, J., BASTOURRE, D. E IBARLUCIA, J. (2006). “*La política económica de la acumulación de reservas: nueva evidencia internacional*”. Investigaciones Económicas. Banco Central de la República Argentina. Serie Estudios BCRA, 2.
- SCHMIDT, P. (1976). *Econometrics*. Statistics: Textbooks and Monographs, Volume 18. CRC Press.
- SCHWORER, A., & HOVEY, P. (2004). *Newton-raphson versus fisher scoring algorithms in calculating maximum likelihood estimates*.
- WHITE, H. (1980). *A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity*. Econometrica: journal of the Econometric Society, 817-838.
- WOOLDRIDGE, J. M. (2015). *Introducción a la econometría: Un enfoque moderno*.
- ZEILEIS, A. (2004). *Econometric computing with HC and HAC covariance matrix estimators*.