Los problemas econométricos más comunes del MCRL son:

- ✓ Multicolinealidad.
- ✓ Heterocedasticidad de residuos (datos corte transversal).
- ✓ Autocorrelación de residuos (datos series de tiempo).
- ✓ Error de especificación.
- ✓ No estacionaridad de las series de tiempo (Este tema se abordará en la unidad V)

LOS PROBLEMAS ECONOMÉTRICOS SE RELACIONAN CON LA VIOLACIÓN DE LOS SUPUESTOS BÁSICOS DEL MCRL

1. MULTICOLINEALIDAD:

SUPUESTO DEL MCRL:
NO DEBE EXISTIR MULTICOLINEALIDAD PERFECTA ENTRE VARIABLES EXPLICATIVAS.

El término multicolinealidad se atribuye a Ragnar Frisch (1934), hace referencia a la existencia de una relación lineal entre algunas o todas las variables explicativas[1].

Multicolinealidad perfecta:

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k = 0$$

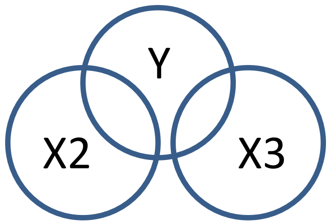
Multicolinealidad no perfecta o menos que perfecta:

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k + v_i = 0$$

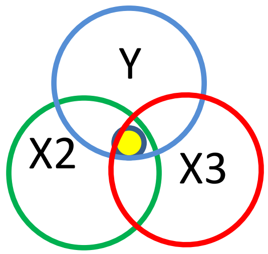
¿CUÁNDO NO EXISTE MULTICOLINEALIDAD DE NINGÚN TIPO?

[1] "Estrictamente hablando, la multicolinealidad se refiere a la existencia de más de una relación lineal exacta y colinealidad se refiere a la existencia de una sola relación lineal. Pero esta distinción raramente se mantiene la práctica haciéndose entonces referencia a multicolinealidad en ambos casos". D. Gujarati y D. Porter (2010)

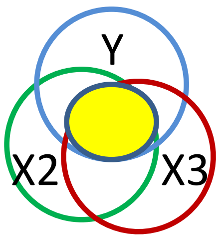
¿CÓMO ES LA MULTICOLINEALIDAD?



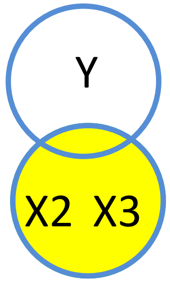
No multicolinealidad



Baja multicolinealidad



Alta colinealidad



Multicolinealidad perfecta

CAUSAS MÁS COMUNES DE MULTICOLINEALIDAD:

- Método de recolección de información empleado (manipulación).
- Restricciones en el modelo o en la población objeto de muestreo.
- Incorrecta especificación del modelo.
- Sobredeterminación *($k>n$) o sobreidentificación.*
- Regresoras comparten una tendencia común (datos series de tiempo).

ESTIMACIÓN EN PRESENCIA DE MULTICOLINEALIDAD PERFECTA O EXACTA:

$$\lambda_1 X_1 + \lambda_2 X_2 + \cdots + \lambda_k X_k = 0$$

Si la multicolinalidad es perfecta los coeficientes de regresión de las variables explicativas son indeterminados y sus errores estándar son infinitos.

Suponga: $$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{\varepsilon}_i$$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\sum {x_{3i}}^2) - (\sum y_i x_{3i})(x_{2i} x_{3i})}{(\sum {x_{2i}}^2)(\sum {x_{3i}}^2) - (\sum x_{2i}\, x_{3i})^2}$$

$$\hat{\beta}_3 = \frac{(\sum y_i x_{3i})(\sum {x_{2i}}^2) - (\sum y_i x_{2i})(x_{2i} x_{3i})}{(\sum {x_{2i}}^2)(\sum {x_{3i}}^2) - (\sum x_{2i}\, x_{3i})^2}$$

Suponga que $x_{3i} = \lambda x_{2i}$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\lambda^2 \sum {x_{2i}}^2) - (\lambda \sum y_i x_{2i})(\lambda {x_{2i}}^2)}{(\sum {x_{2i}}^2)(\lambda^2 \sum {x_{2i}}^2) - \lambda^2 (\sum {x_{2i}}^2)^2} = \frac{0}{0}$$

Lo mismo ocurre para $\hat{\beta}_3$.

$$y_i = \hat{\beta}_2 x_{2i} + \hat{\beta}_3 x_{3i} + \hat{\varepsilon}_i$$

$$= \hat{\beta}_2 x_{2i} + \hat{\beta}_3 (\lambda x_{2i}) + \hat{\varepsilon}_i$$

$$= (\hat{\beta}_2 + \lambda \hat{\beta}_3) x_{2i} + \hat{\varepsilon}_i$$

$= \hat{\alpha} x_{2i} + \hat{\varepsilon}_i \longrightarrow$ No se puede separar el efecto de las V.E no se pueden estimar los betas

Haciendo uso del enfoque matricial no existe la inversa $(X'X)^{-1}$ por tanto no puede estimarse $\hat{\beta} = (X'X)^{-1}X'Y$. Los $\hat{\beta}$ resultan indeterminados y de varianza infinita. Su matriz de covarianzas que viene dada por $Var(\hat{\beta}) = \sigma_u^2 (X'X)^{-1}$ no puede estimarse.

ESTIMACIÓN EN PRESENCIA DE MULTICOLINEALIDAD "ALTA" PERO NO PERFECTA:

Si: $x_{3i} = \lambda x_{2i} + v_i$

$\lambda_i \neq 0$

$$\hat{\beta}_2 = \frac{(\sum y_i x_{2i})(\lambda^2 \sum {x_{2i}}^2 + \sum {v_i}^2) - (\lambda \sum y_i x_{2i} + \sum y_i v_i)(\lambda {x_{2i}}^2)}{(\sum {x_{2i}}^2)(\lambda^2 \sum {x_{2i}}^2 + \sum {v_i}^2) - (\lambda \sum {x_{2i}}^2)^2}$$

De igual forma se puede derivar $\hat{\beta}_3$.

En resumen:

- En multicolinealidad perfecta $|X'X| = 0$

- En multicolinealidad menos que perfecta $|X'X| \sim 0$

CONSECUENCIAS PRÁCTICAS DE LA MULTICOLINEALIDAD:

- La multicolinealidad genera pérdida de precisión. Un ejemplo[2]

Considere el siguiente modelo:

$$\hat{y}_t = 8 + 5x_{2t} - 3x_{3t}$$

Donde $E(\varepsilon_t) = 0;\quad E(\varepsilon_t, \varepsilon_s) = 0 \;\; para\ s \neq t\ ;\ E(\varepsilon_t^2) = 25.$

[2] Ejemplo tomado de Novales, A. Páginas 346-349.

	Correlación =0.0	Correlación = 0.90	Correlación =0.99
$\rho(x_{2t}, x_{3t})$	0.009	0.898	0.99
$\lvert X'X \rvert$	0.992	0.194	0.021
$\hat{\beta}_1$	8.03	7.93	7.96
$\hat{\beta}_2$	5.1	5.13	5.25
$\hat{\beta}_3$	-3	-3.18	-3.28
$\hat{\sigma}_u^2$	25.3	26.2	24.13
$Var(\hat{\beta}_2)$	0.233	1.29	11.22
$Cov(\hat{\beta}_2, \hat{\beta}_3)$	-0.001	-1.05	-11.01
$Var(\hat{\beta}_3)$	0.232	1.06	11.03

- Aun cuando los estimadores MCO son MELI, presenta varianzas y covarianzas grandes que dificultan una estimación precisa.

$$Var(\hat{\beta}_j) = \frac{\sigma^2}{\sum x_j^2 (1 - R_j^2)}$$

- Los intervalos de confianza tienden a ser más amplios y los estadísticos t a ser más pequeños, por ello aumenta la posibilidad de no rechazar Ho.

- Aun cuando los t sean estadísticamente no significativos el R^2 puede ser muy alto.

- Los estimadores MCO y sus errores estándar son sensibles a pequeños cambios en la información.

$$\overset{X'X}{\begin{pmatrix} 1 & 0.90 \\ 0.90 & 1 \end{pmatrix}} \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \overset{X'y}{\begin{pmatrix} 10 \\ 12 \end{pmatrix}} \qquad \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} -4.21 \\ 15.79 \end{pmatrix}$$

$$\overset{X'X}{\begin{pmatrix} 1 & 0.99 \\ 0.99 & 1 \end{pmatrix}} \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \overset{X'y}{\begin{pmatrix} 10 \\ 12 \end{pmatrix}} \qquad \begin{pmatrix} \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix} = \begin{pmatrix} -9.40 \\ 105 \end{pmatrix}$$

DETECCIÓN DE LA MULTICOLINEALIDAD:

Antes de determinar cómo se descubre la presencia multicolinealidad se debe tener en cuenta lo que argumenta Kmenta:

1) La multicolinealidad es un problema de grado y no de clase.

2) Como la multicolinealidad se refiere a la condición de las variables explicativas que no son estocásticas, es una característica de la muestra y no de la población.

No se tiene un método específico para detectar la multicolinealidad, lo que existe en realidad son ciertas reglas prácticas formales e informales como:

- Un R^2 elevado pero pocas razones t significativas.

$$\hat{y}_t = 10.81 - 2.92x_{2t} - 0.54x_{3t}$$
$$\begin{array}{llll} t & (1.6) & (0.42) & (0.21) \end{array}$$
$$R^2 = 0.92 \qquad \hat{\sigma}_u^2 = 2.09$$

- Alta correlación de orden cero entre parejas de regresoras[3]. Un coeficiente de correlación entre las variables explicativas superior a 0.8 da indicios de la presencia de multicolinealidad pero no es una regla clara.

- Regresiones auxiliares de las variables explicativas: si el R^2 de la regresión auxiliar es superior al R^2 de la regresión original se sospecha de multicolinealidad. ***Regla Práctica de Klein.***

- Valores propios[4] e índice de condición:

$$IC = \sqrt{\frac{Máximo\ valor\ propio}{Mínimo\ valor\ propio}} = \sqrt{k}$$

[3] Las correlaciones de orden cero elevadas son una condición suficiente pero no necesaria para la existencia de multicolinealidad debido a que ésta puede existir a pesar de que dichas correlaciones sean bajas.

[4] El vector propio o eigenvector $\vec{v}$ de una matriz A_{nxn} es una matriz de orden $n\ x1$ tal que $A\vec{v} = \lambda\vec{v}$ donde λ es un valor escalar real que recibe el nombre de valor propio o eigenvalor.

Si k está entre 100 y 1000 existe una colinealidad de moderada a fuerte. Si excede a 1000 es una multicolinealidad severa.

- Factor inflador de varianza (FIV) y tolerancia (TOL):

$$FIV = \frac{1}{(1 - R_{23}{}^{2})}$$

Si FIV > 10 se dice que la variable es altamente colineal.

$$TOL = \left(1 - R_{23}{}^{2}\right) = \frac{1}{FIV}$$

Mientras más cerca este la tolerancia de cero, mayor será el grado de colinealidad de esas regresoras.

Ejemplo: Importaciones de Estados Unidos (IMPORTS), PIB (GDP) e IPC (CPI), 1975-2005.

```
## Call:
## lm(formula = Imports ~ CPI + GDP, x = T)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -67470 -27980  -6985  20883 101913
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 160054.69   48279.22   3.315  0.00254 **
## CPI          -7468.18     948.94  -7.870 1.43e-08 ***
## GDP            230.73      12.62  18.285  < 2e-16 ***
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 43130 on 28 degrees of freedom
## Multiple R-squared:  0.9906, Adjusted R-squared:   0.99
## F-statistic:  1479 on 2 and 28 DF,  p-value: < 2.2e-16
```

Test de normalidad:

```
##  Shapiro-Wilk normality test
## data:  res
## W = 0.93721, p-value = 0.06909
```

Matriz de correlaciones:

```
##           Imports       CPI       GDP
## Imports 1.0000000 0.9373526 0.9848229
## CPI     0.9373526 1.0000000 0.9805229
## GDP     0.9848229 0.9805229 1.0000000
```

Regresión auxiliar:

```
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 47.599657   3.335938   14.27 1.21e-14 ***
## GDP          0.013038   0.000485   26.89  < 2e-16 ***
## Residual standard error: 8.44 on 29 degrees of freedom
## Multiple R-squared:  0.9614, Adjusted R-squared:  0.9601
## F-statistic: 722.8 on 1 and 29 DF,  p-value: < 2.2e-16
```

FIV:

```
##      CPI      GDP
## 25.92361 25.92361
```

IC:

```
42884.69
```

PRINCIPALES MEDIDAS CORRECTIVAS:

- Datos nuevos o adicionales.
- Información a priori.
- Combinación de información de corte transversal y de series de tiempo. (DATOS PANEL)
- Eliminación de variables. ***ESTO PUEDE GENERAR SESGO DE ESPECIFICACIÓN.***
- Transformación de variables.
- Regresiones polinomiales.
- Regresión en cresta*[5]. $\hat{\beta}_c = (X'X + cI_k)^{-1}X'Y$. Sin embargo, se puede generar un estimador sesgado. Ahora si se elige adecuadamente el valor de c, su matriz de varianzas $\sigma_u^2(X'X + cI_k)^{-1}X'X(X'X + cI_k)^{-1}$ puede ser menor que la del estimador MCO.
- Utilizar componentes principales*.

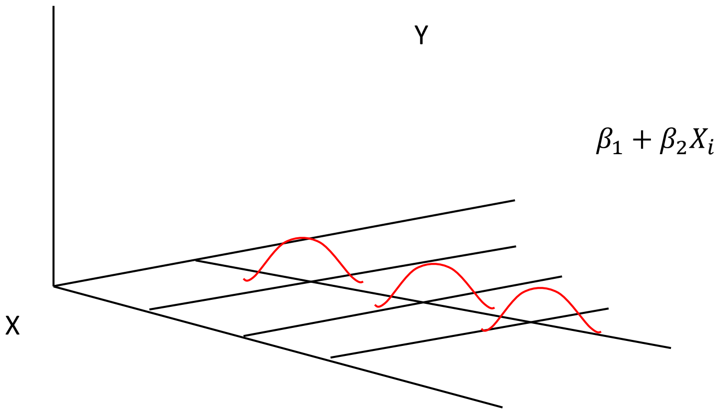
* Son técnicas no aplicables a nivel del curso de econometría de la maestría.

2. HETEROSCEDASTICIDAD: ¿QUÉ PASA CUANDO LA VARIANZA DEL ERROR NO ES CONSTANTE?

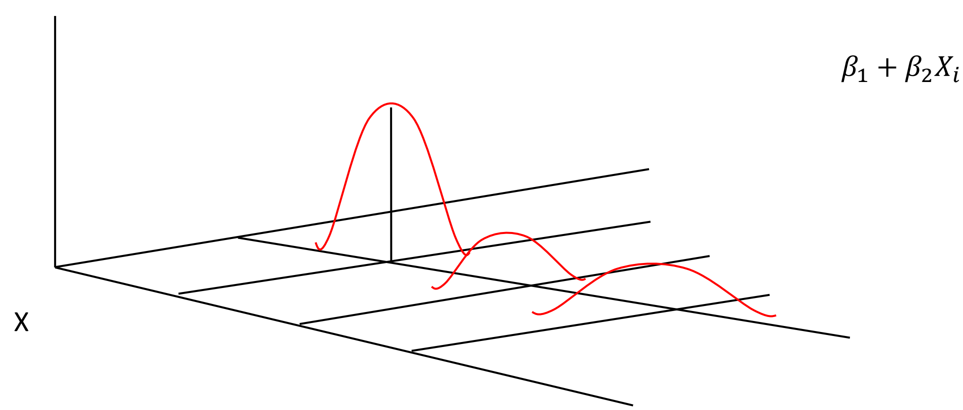
SUPUESTO: la varianza de cada término de error u_i, condicional a los valores seleccionados de las variables explicativas, es algún número constante igual a σ^2.

$$E({\varepsilon_i}^2) = \sigma^2 \quad \text{i= 1,2,3,4...n}$$

HOMOSCEDASTICIDAD:



HETEROSCEDASTICIDAD: $E({\varepsilon_i}^2) = {\sigma^2}_i$ **i= 1,2,3,4…n**



PRINCIPALES CAUSAS DE LA HETEROSCEDASTICIDAD:

- Modelos de aprendizaje sobre errores, a medida que la gente aprende, sus errores de comportamiento se hacen menores, con el tiempo. Se espera que la varianza se reduzca.

- Ingreso discrecional. A medida que aumentan los ingresos la gente tiene mayores posibilidades de selección con la forma de disponer de su ingreso. Las personas de menores ingresos tienden a gastar todo su ingreso mientras que quienes poseen mayores ingresos tienen más alternativas para decidir cómo gastar su ingreso. La varianza crece.

- A medida que mejoran las técnicas de recolección de información es probable que la varianza $(\boldsymbol{\sigma^2}_{\boldsymbol{i}})$ se reduzca.

- Presencia de factores atípicos u outliers.

- Mala especificación del modelo de regresión: omisión de variables relevantes o forma funcional incorrecta

- Asimetría en la distribución de una o varias regresoras.

- Incorrecta transformación de los datos. (D. Hendry)

EL PROBLEMA DE LA HETEROSCEDASTICIDAD ES PROBABLEMENTE MÁS COMÚN EN DATOS DE CORTE TRANSVERSAL (INFORMACIÓN DE INDIVIDUOS, FAMILIAS, EMPRESAS, ETC) QUE EN LAS SERIES DE TIEMPO (LAS VARIABLES TIENDEN A SER DE ÓRDENES DE MAGNITUD SIMILARES).

ESTIMACIÓN EN PRESENCIA DE HETEROSCEDASTICIDAD:

$$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$$

$\hat{\beta}_2 = \frac{\sum x_i y_i}{\sum x_i^2}$ pero su varianza viene dada por:

$$Var(\hat{\beta}_2) = \frac{\sum x_i^2}{(\sum x_i^2)^2}\ \sigma_i^2$$

Resulta que los estimadores MCO son lineales, insesgados y consistentes pero no eficientes, su varianza no es mínima, es decir, dejan de ser MELI, por tanto las pruebas de hipótesis no serán confiables.

SI LOS ESTIMADORES MCO NO SIRVEN PARA ESTIMAR EL MODELO PORQUE NO SON MELI, ¿QUÉ MÉTODO DEBEMOS UTILIZAR?
R: EL MÉTODO DE MÍNIMOS CUADRADOS GENERALIZADOS (MCG)

MÉTODO DE MÍNIMOS CUADRADOS GENERALIZADOS (MCG)

A diferencia del MCO este método toma en cuenta la información contenida en la variabilidad desigual de la variable dependiente.

$Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$ para facilitar los cálculos supóngase:

$Y_i = \beta_1 X_{0i} + \beta_2 X_i + \varepsilon_i$ donde $X_{0i} = 1$

Suponga también que las varianzas heteroscedásticas ${\sigma_i}^2$ son conocidas:

$$\frac{Y_i}{{\sigma_i}^2} = \beta_1 \left(\frac{X_{0i}}{{\sigma_i}^2}\right) + \beta_2 \left(\frac{X_i}{{\sigma_i}^2}\right) + \left(\frac{\varepsilon_i}{{\sigma_i}^2}\right)$$

Así:

$$Y_i^* = \beta_1^* X_{0i}^* + \beta_2^* X_i^* + \varepsilon_i^* \qquad \text{modelo transformado}$$

$$Var(\varepsilon_i^*) = E(\varepsilon_i^*)^2 = E\left(\frac{\varepsilon_i}{\sigma_i}\right)^2$$

$$= \frac{1}{\sigma_i^2} E(\varepsilon_i^2) \qquad \text{como } \sigma_i^2 \text{ es conocida}$$

$$= \frac{1}{\sigma_i^2} (\sigma_i^2) = 1$$

Esta transformación hace la varianza constante, HOMOSCEDÁSTICA y así los estimadores MCG son MELI.

MCG es MCO sobre las variables transformadas que satisfacen los supuestos estándar de mínimos cuadrados. Así:

Como se puede observar MCG pondera las observaciones, esto se conoce como mínimos cuadrados ponderados (MCP).

¿QUÉ DIFERENCIA EXISTE ENTRE LOS MCO Y LOS MCG?

MCO ⇒ $\sum \hat{\varepsilon}_i^{\,2} = \sum (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2$

MCG ⇒ $\sum w_i \hat{\varepsilon}_i^{\,2} = \sum w_i (Y_i - \hat{\beta}_1^{\,*} - \hat{\beta}_2^{\,*} X_i)^2$

Donde $w_i = 1/\sigma_i^{\,2}$

En MCG el peso asignado a las observaciones es inversamente proporcional a σ_i , esto significa que darán menos peso a las observaciones que provienen de una población con mayor variabilidad y más peso a aquellas que provienen de observaciones con menos variabilidad.

Ejemplo: Datos sobre 81 automóviles respecto de su MPG (millas promedio por galón), CF (caballos de fuerza de su motor), VOL (pies cúbicos de su cabina), VM (velocidad máxima en millas por hora) y su PS (peso del vehículo en cientos de lb).

```
## Call:
## lm(formula = MPG ~ VM + CF + PS + VOL)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.1178 -2.8494  0.0746  1.6949 12.0265
##
## Coefficients:
##              Estimate Std. Error t value   Pr(>|t|)
## (Intercept) 188.70352   22.74704   8.296   2.99e-12 ***
## VM           -1.25282    0.23669  -5.293   1.13e-06 ***
## CF            0.38058    0.07867   4.838   6.72e-06 ***
## PS           -1.85528    0.20584  -9.013   1.26e-13 ***
## VOL          -0.01211    0.02206  -0.549   0.585
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 3.524 on 76 degrees of freedom
## Multiple R-squared:  0.8833, Adjusted R-squared:  0.8772
## F-statistic: 143.8 on 4 and 76 DF,  p-value: < 2.2e-16
```

MÉTODOS PARA DETECTAR LA HETEROSCEDASTICIDAD

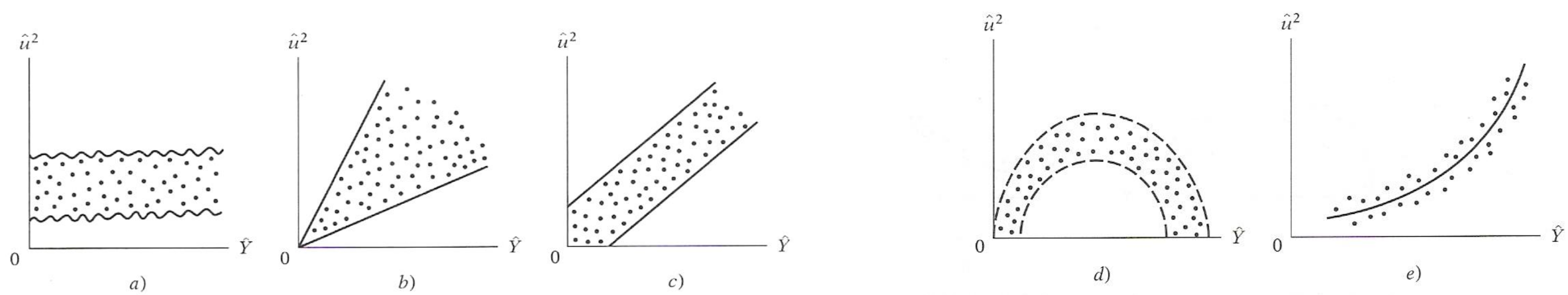
a) Métodos informales:

- ***Naturaleza del problema:*** los datos de corte transversal son altamente propensos a presentar heteroscedasticidad, mientras más heterogénea sea la muestra.

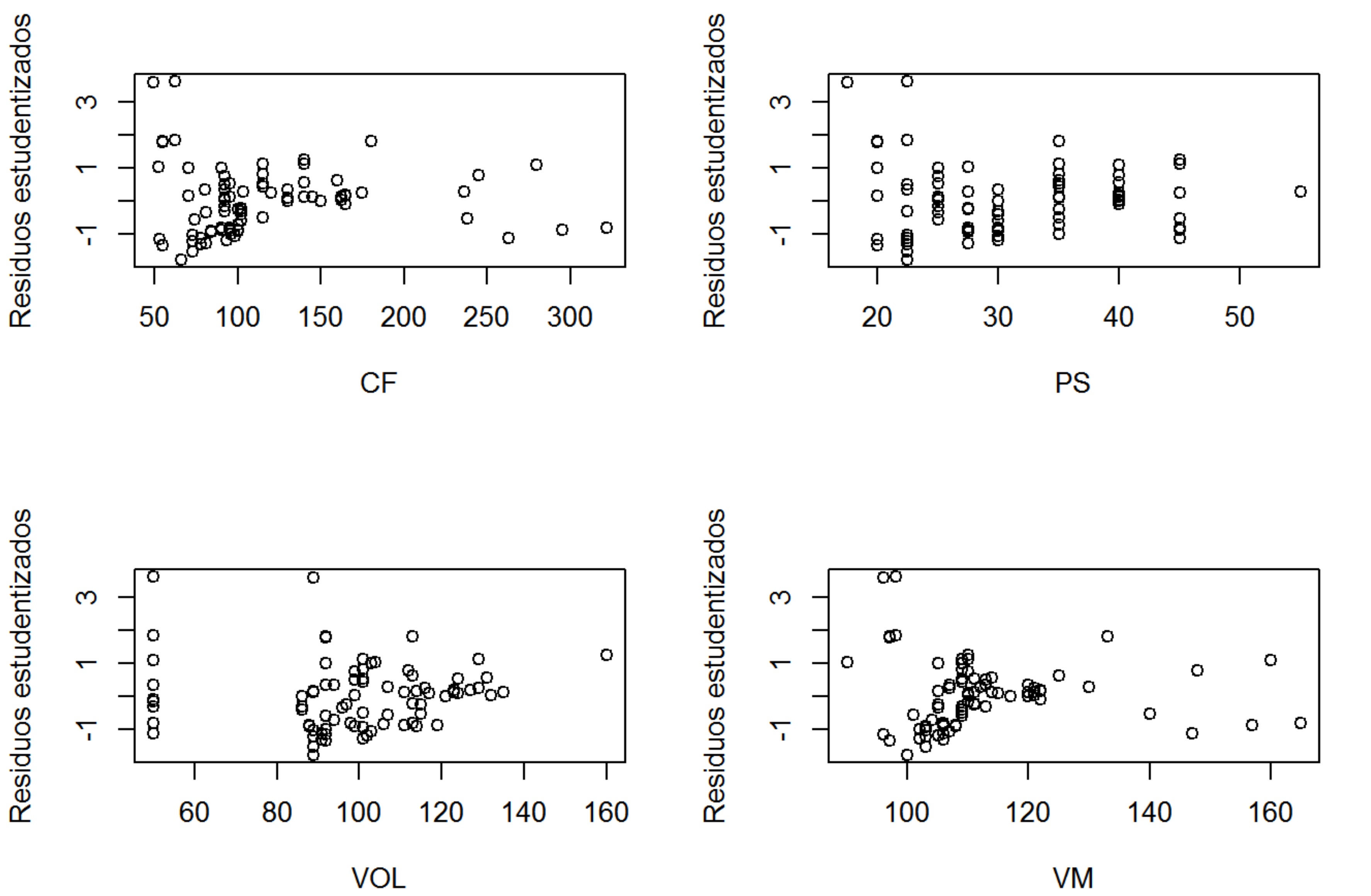
- ***Método gráfico:*** se lleva a cabo un análisis de regresión bajo el supuesto de no heteroscedasticidad (homoscedasticidad) en residuos y luego se hace un análisis post mortem de éstos elevados al cuadrado.

$$\hat{u}_i{}^2 \quad Vs. \quad \hat{Y}_i$$

$$\hat{u}_i{}^2 \quad Vs. \quad X_i$$



Del modelo estimado:



b) Métodos formales:

Todas las pruebas se van a realizar para realizar el siguiente contraste de hipótesis:

H_0 : Los residuos son homoscedásticos.

H_1: Los residuos son heteroscedásticos.

Prueba de Park
Prueba de Glejser
Prueba de Goldfeld y Quand
Prueba de Breusch-Pagan y Godfrey (BPG)[6]**,**

Prueba BPG:

```
##  studentized Breusch-Pagan test
## data:  mh
## BP = 26.098, df = 4, p-value = 3.023e-05
```

[6] Esta prueba es sensible al supuesto de normalidad.

Prueba general de White: es uno de los test más utilizados actualmente

Procedimiento:

Paso 1: estimar la regresión

$$Y_i = \beta_1 + \beta_2 X_{i2} + \beta_3 X_{i3} + \varepsilon_i$$

Paso 2: estimar una regresión auxiliar con los residuos como variable dependiente para obtener el R^2 , quien se denominará R^2 auxiliar

$$\hat{\varepsilon}_i{}^2 = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + \alpha_4 X_{2i}{}^2 + \alpha_5 X_{3i}{}^2 + \alpha_3 X_{2i} X_{3i} + v_i$$

Paso 3: calcular el estadístico de prueba, con el R^2 auxiliar

$$n * R^2 \sim X^2{}_{k-1}$$

Si $X^2{}_c > X^2{}_t$ se rechaza la hipótesis nula de homoscedasticidad

En un modelo con muchas regresoras se pierden muchos grados de libertad rápidamente, por ello se debe tener cuidado al usar la prueba.

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 188.70352   22.74704   8.296 2.99e-12 ***
## VM           -1.25282    0.23669  -5.293 1.13e-06 ***
## CF            0.38058    0.07867   4.838 6.72e-06 ***
## PS           -1.85528    0.20584  -9.013 1.26e-13 ***
## VOL          -0.01211    0.02206  -0.549    0.585
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.524 on 76 degrees of freedom
## Multiple R-squared:  0.8833, Adjusted R-squared:  0.8772
## F-statistic: 143.8 on 4 and 76 DF,  p-value: < 2.2e-16
```

El test de White:

```
### Prueba de White
reswhite=residuals(mh)
resw2=(reswhite^2)
W<- summary(lm(resw2~VM+CF+PS+VOL+I(VM^2)+I(CF^2)+I(PS^2)+ I(VOL^2)+I(VM*CF)+I(VM*PS)
+I(VM*VOL)+I(CF*PS)+I(CF*VOL)+I(PS*VOL)))$r.squared
###Con el comando anterior se extrae el R2 auxiliar
W
## [1] 0.6142561
Chi.White<- nrow(datos2)*W
p.value <- 1-pchisq(Chi.White,14)
p.value
## [1] 6.71153e-06
```

Prueba de Koenker-Basset (KB): al igual que las pruebas anteriores se basa en los residuos estimados al cuadrado pero a diferencia utiliza la variable estimada al cuadrado como función de los residuos

$$Y_i = \beta_1 + \beta_2 X_{2i} + ... + \beta_k X_{ki} + u_i$$

$$\hat{u}_i{}^2 = \alpha_1 + \alpha_3 \widehat{Y}_i + \alpha_3 \left(\widehat{Y}_i\right)^2 + v_i$$

Para determinar si existe heteroscedasticidad se realiza una prueba de significancia individual para α_2, mediante el procedimiento tradicional. Si α_2 es estadísticamente significativo se concluye que el modelo presenta heteroscedasticidad. Esta prueba no requiere del supuesto de normalidad de residuos.

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 90.68232   23.50921   3.857 0.000235 ***
## y           -6.37889    1.46137  -4.365 3.86e-05 ***
## I(y^2)       0.11095    0.02162   5.132 2.05e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 18.35 on 78 degrees of freedom
## Multiple R-squared:  0.3866, Adjusted R-squared:  0.3709
## F-statistic: 24.58 on 2 and 78 DF,  p-value: 5.275e-09
```

MEDIDAS PARA CORREGIR EL PROBLEMA DE HETEROSCEDASTICIDAD:

1) Aplicar MCG.

a) Cuando $\sigma_i{}^2$ es conocida, emplear MCP. Esto garantiza estimadores MELI.

b) Cuando $\sigma_i{}^2$ no es conocida, emplear Mínimos Cuadrados Factibles (MCF o MCGE).

```
## Generalized least squares fit by REML
##   Model: MPG ~ VM + CF + PS + VOL
##   Data: NULL
##        AIC      BIC    logLik
##   458.9731 472.9575 -223.4866
## Coefficients:
##                 Value Std.Error   t-value p-value
## (Intercept) 188.70352 22.747037  8.295741  0.0000
## VM           -1.25282  0.236692 -5.293036  0.0000
## CF            0.38058  0.078668  4.837804  0.0000
## PS           -1.85528  0.205843 -9.013077  0.0000
## VOL          -0.01211  0.022062 -0.549029  0.5846
##  Correlation:
##     (Intr) VM     CF     PS
## VM  -0.996
## CF   0.981 -0.984
## PS  -0.865  0.854 -0.922
## VOL  0.101 -0.145  0.228 -0.425
## Standardized residuals:
##         Min          Q1         Med          Q3         Max
## -1.73609522 -0.80859420  0.02117774  0.48096919  3.41283425
## Residual standard error: 3.523895
## Degrees of freedom: 81 total; 76 residual
```

2) Aplicar transformaciones logarítmicas de las variables para reducir la diferencia entre sus valores.

```
## lm(formula = lMPG ~ lVM + lCF + lPS + lVOL)
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.59088    2.72900   2.049   0.0439 *
## lVM          0.45510    0.71195   0.639   0.5246
## lCF         -0.46623    0.30191  -1.544   0.1267
## lPS         -0.58142    0.22268  -2.611   0.0109 *
## lVOL        -0.02297    0.04220  -0.544   0.5879
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08343 on 76 degrees of freedom
## Multiple R-squared:  0.9294, Adjusted R-squared:  0.9257
## F-statistic: 250.2 on 4 and 76 DF,  p-value: < 2.2e-16
```

```r
### Prueba de White


reswhite1=residuals(mh1)

resw1=(reswhite1^2)

W1<- summary(lm(resw2~lVM+lCF+lPS+lVOL+I(lVM^2)+I(lCF^2)+I(lPS^2)+
I(lVOL^2)+I(lVM*lCF)+I(lVM*lPS)+I(lVM*lVOL)+I(lCF*lPS)+I(lCF*lVOL)
+I(lPS*lVOL)))$r.squared

###Con el comando anterior se extrae el R2 auxiliar

W1
```

```
## [1] 0.5957224
```

```r
Chi.White1<- nrow(datos2)*W1

p.value1 <- 1-pchisq(Chi.White1,14)

p.value1
```

```
## [1] 1.19341e-05
```

```r
### Según el test de White el modelo tiene problemas de heterocedasticidad
```

```r
### Test de Koenker-Basset
u22 <- mh1$residuals^2
y2 <- fitted(mh1)
y22 <- y2^2
KB2 <- (lm(u2 ~ y2 + I(y2^2)))
summary(KB2)
```

```
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1117.61     270.45   4.132 8.96e-05 ***
## y2           -677.91     157.37  -4.308 4.76e-05 ***
## I(y2^2)       102.76      22.78   4.511 2.24e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.13 on 78 degrees of freedom
## Multiple R-squared:  0.3331, Adjusted R-squared:  0.316
## F-statistic: 19.48 on 2 and 78 DF,  p-value: 1.373e-07
```

3) Errores estándar robustos de White (regresión robusta): técnicamente se conocen como *estimadores de matriz de covarianza consistentes con heteroscedasticidad (EMCCH).* White sugiere la transformación de las variables para corregir el problema de la heteroscedasticidad mediante el siguiente procedimiento:

Considere la regresión: $Y_i = \beta_1 + \beta_2 X_i + \varepsilon_i$; $Var(\varepsilon_i) = \sigma_i^2$

Recuerde que $Var(\hat{\beta}_2) = \frac{\sum x_i^2 \sigma_i^2}{(\sum x_i^2)^2}$ (1)

Como σ_i^2 no es directamente observable, White sugirió emplear $\hat{u}_i^2$ en su lugar, por ello:

$Var(\hat{\beta}_2) = \frac{\sum x_i^2}{(\sum x_i^2)^2} \hat{\varepsilon}_i^2$ (2)

Tenga presente que (2) es un estimador consistente de (1). De allí se deriva que para un modelo de k variables la varianza homoscedástica de cualquier coeficiente de regresión parcial es:

$$Var(\hat{\beta}_i) = \frac{\sum \widehat{w}_{ji}^2 \ \hat{\varepsilon}_i^2}{\left(\sum \widehat{w}_{ji}^2\right)^2}$$

$\hat{\varepsilon}_i^2$: son los residuos estimados al cuadrado de la regresión original.

$\widehat{w}_{ji}^2$: son los residuos de la regresión auxiliar entre las regresoras.

Ejemplo: REGRESIÓN ROBUSTA

#	Estimate	Std. Error	t value	Pr(>|t|)	
## (Intercept)	188.70352	22.74704	8.296	2.99e-12	***
## VM	-1.25282	0.23669	-5.293	1.13e-06	***
## CF	0.38058	0.07867	4.838	6.72e-06	***
## PS	-1.85528	0.20584	-9.013	1.26e-13	***
## VOL	-0.01211	0.02206	-0.549	0.585	

Matriz de covarianza consistentes con heteroscedasticidad (EMCCH)

##	(Intercept)	VM	CF	PS	VOL
## (Intercept)	**1344.8290753**	-1.326431e+01	4.271413e+00	-10.772601439	1.143375e-01
## VM	-13.2643068	**1.312357e-01**	-4.231336e-02	0.106752130	7.487357e-04
## CF	4.2714126	-4.231336e-02	**1.389208e-02**	-0.036046474	2.561414e-05
## PS	-10.7726014	1.067521e-01	-3.604647e-02	**0.099868683**	-1.721337e-03
## VOL	-0.1143375	7.487357e-04	2.561414e-05	-0.001721337	**7.949531e-04**

		MCO (no corregidos)		Estimadores corregidos (Robustos)	
	Parámetro estimado	**Error Estándar**	**Estadístico t**	**Error Estándar**	**Estadístico t**
Intercepto	188.70352	22.74704	8.296	36.6719	5.145
VM	-1.25282	0.23669	-5.293	0.36226	-3.458
CF	0.38058	0.07867	4.838	0.11786	3.204
PS	-1.85528	0.20584	-9.013	0.31602	-5.870
VOL	-0.01211	0.02206	-0.549	0.02819	-0.429

3) AUTOCORRELACIÓN ¿QUÉ PASA CUANDO LOS ERRORES ESTÁN AUTOCORRELACIONADOS?

Espacial (datos cross section)

Autocorrelación

Serial (datos series de tiempo)

La autocorrelación se define como la correlación entre miembros de series de observaciones ordenadas en el tiempo (ST) o en el espacio (CS); en el modelo de regresión lineal se dice que hay autocorrelación cuando, el término de error de un modelo econométrico está correlacionado consigo mismo a través del tiempo.

De los supuestos básicos del modelo clásico de regresión lineal se deriva que:

$$
\begin{aligned}
Cov\left(\varepsilon_i, \varepsilon_j \middle| X_i, X_j\right) &= E\{[\varepsilon_i - E(\varepsilon_i)] | X_i\}\left\{\left[\varepsilon_j - E\left(\varepsilon_j\right)\right] \middle| X_j\right\} \\
&= E(\varepsilon_i | X_i)\left(\varepsilon_j \middle| X_j\right)
\end{aligned}
$$

$$E(\varepsilon_i, \varepsilon_j) = 0 \qquad\qquad i \neq j$$

No correlación serial de residuos: $Cov(\varepsilon_i, \varepsilon_j) = 0$

$$E(\varepsilon\varepsilon') = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix}$$

Cuando existe autocorrelación:

$$E(\varepsilon_i, \varepsilon_j) \neq 0 \qquad\qquad i \neq j$$

$$E(\varepsilon\varepsilon') = \begin{bmatrix} E({\varepsilon_1}^2) & E(\varepsilon_1\varepsilon_2) & \dots & E(\varepsilon_1\varepsilon_n) \\ E(\varepsilon_2\varepsilon_1) & E({\varepsilon_2}^2) & \dots & E(\varepsilon_2\varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ E(\varepsilon_n\varepsilon_1) & E(\varepsilon_n\varepsilon_2) & \dots & E({\varepsilon_n}^2) \end{bmatrix}$$

Tenga presente que autocorrelación y correlación serial, si bien es cierto se emplean como sinónimos[7], según algunos autores como Tintner no lo son.

[7] En efecto, durante este curso se trabajará con ellas como sinónimas.

¿POR QUÉ OCURRE LA AUTOCORRELACIÓN SERIAL?

- **Inercia o lentitud de la serie**. Muchas de las variables presentan ciclos y tendencias.

- **Sesgo o error de especificación por variables relevantes omitidas**

- **Sesgo o error de especificación por forma funcional incorrecta**

- **Rezagos (relaciones dinámicas o autoregresion):** la variable explicada depende de los valores que ella misma ha tomado en periodos anteriores:

$$\pi_t = \beta_1 + \beta_2 m_t + \beta_3 \pi_{t-1} + \varepsilon_t$$

- **Manipulación de los datos:** interpolación o extrapolación.

- **Transformación de datos:**

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \qquad (1) \qquad\qquad Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \varepsilon_{t-1} \qquad (2)$$

Restando (2) de (1) resulta: $\Delta Y_t = \beta_2 \Delta X_t + \Delta \varepsilon_t \longrightarrow$ Modelos dinámicos de regresión

- **No estacionariedad[8]:** los primeros momentos estadísticos de la serie (media, varianza y covarianza) no son constantes en el tiempo.

---

[8] No confunda estacionariedad con estacionalidad, son dos cosas totalmente distintas.

ESTIMACIÓN DE MCO EN PRESENCIA DE AUTOCORRELACIÓN

Dado un modelo de regresión simple:

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$$

Cuando existe autocorrelación $E(\varepsilon_t, \varepsilon_{t-s}) \neq 0 \;\; ; \quad s \neq 0$

Supóngase que los términos de error se generan de la siguiente manera:

$$\varepsilon_t = \rho \varepsilon_{t-1} + \epsilon_t \qquad\qquad -1 < \rho < 1$$

Esquema autorregresivo de primer orden de Markov, AR(1)

ρ: Coeficiente de autocovarianza
ϵ_t: Perturbación estocástica (ruido blanco).

$$E(\epsilon_t) = 0 \qquad\qquad Var(\epsilon_t) = {\sigma_\varepsilon}^2$$

$$Cov(\epsilon_t, \epsilon_{t+s}) = 0$$

Del esquema AR(1) se debe tener presente que:

$$Var(\varepsilon_t) = E({\varepsilon_t}^2) = \frac{{\sigma_\epsilon}^2}{1-\rho^2}$$

$$Cov(\varepsilon_t, \varepsilon_{t+s}) = E\left(\varepsilon_{t,}\varepsilon_{t+s}\right) = \rho^s \frac{{\sigma_\epsilon}^2}{1-\rho^2} \qquad Correlación(u_t, u_{t+s}) = \rho^s$$

Recuerde que en un modelo de regresión simple $\hat{\beta}_2 = \frac{\sum x_t y_t}{\sum {x_t}^2}$ y que su varianza es $Var\left(\hat{\beta}_2\right) = \frac{\sigma^2}{\sum {x_t}^2}$.

Cuando se emplea un esquema AR(1) la varianza del parámetro de estimación $\hat{\beta}_2$ viene dada por:

$$Var\left(\hat{\beta}_{AR1}\right) = \frac{\sigma^2}{\sum {x_t}^2}\left[1 + 2\rho\frac{\sum x_t x_{t-1}}{\sum {x_t}^2} + 2\rho^2\frac{\sum x_t x_{t-2}}{\sum {x_t}^2} + \cdots + 2\rho^{n-1}\frac{\sum x_t x_n}{\sum {x_t}^2}\right]$$

Esta varianza muestra que los estimadores MCO son lineales e insesgados pero no tienen varianza mínima (no eficientes) por tanto, dejan de ser MELI y como consecuencia las pruebas de hipótesis no serán confiables.

En un modelo de regresión simple un estimador MELI de β_2 viene dado por:

$$\hat{\beta}_2{}^{MCG} = \frac{\sum_{t=2}^{n}(x_t-\rho x_{t-1})(y_t-\rho y_{t-1})}{\sum_{t=2}^{n}(x_t-\rho x_{t-1})^2} + C \qquad Var\left(\hat{\beta}_2{}^{MCG}\right) = \frac{\sigma^2}{\sum_{t=2}^{n}(x_t-\rho x_{t-1})^2} + D$$

C y D: factores de corrección

PARA TENER PRESENTE:

- Es muy probable que la varianza de los residuos estimada $\hat{\sigma}^2 = \frac{\sum \hat{\varepsilon}_t{}^2}{(n-k)}$ subestime la verdadera varianza σ^2.

- R^2 tiende a estar sobreestimado.

- Aun cuando σ^2 no este subestimada, la $Var(\hat{\beta}_2)$ podría subestimar $Var(\hat{\beta}_{2\ AR1})$.

$$Var(\hat{\beta}_2) < Var(\hat{\beta}_{2\ AR1})$$

- Cuando se usa MCO en presencia de autocorrelación, los intervalos de confianza tienden a ser más amplios y los estadísticos t y F a ser más pequeños, aumentando así la probabilidad de no rechazar Ho siendo falsa (cometer error tipo II).

MÉTODOS PARA DETECTAR AUTOCORRELACIÓN

- ***Método gráfico:*** consiste en realizar un gráfico de los residuos generados en la estimación del modelo respecto ($\hat{\varepsilon}_t$) al tiempo, conocido como **gráfica de secuencia de tiempo,** también se pueden graficar los residuos estandarizados respecto al tiempo mediante un gráfico de dispersión. Una vez realizada la gráfica se busca patrones sistemáticos para determinar si existe autocorrelación.

H_o: No autocorrelación de residuos.
H_1: Autocorrelación de residuos

Prueba de Durbin – Watson (DW): supone que el termino de error sigue un esquema autorregresivo de primer orden

$$\varepsilon_t = \rho\varepsilon_{t-1} + \epsilon_t$$

$$\epsilon_t \sim NID(0,1) \;\; Rudio\ blanco\ (white\ noise)$$

El estadístico *d* de Durbin Watson viene dado por:

$$d = \frac{\sum_{t=2}^{t=n}(\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=2}^{t=n}(\hat{\varepsilon}_t)^2} \approx 2\left(1 - \frac{\sum \hat{\varepsilon}_t\hat{\varepsilon}_{t-1}}{\sum \hat{\varepsilon}_t{}^2}\right) \approx 2(1-\hat{\rho})$$

$-1 < \hat{\rho} < 1 \Rightarrow \quad 0 < DW < 4 \begin{cases} \hat{\rho} = 1 \text{ Autocorrelación positiva} \\ \hat{\rho} = 0 \text{ No Autocorrelación} \\ \hat{\rho} = -1 \text{ Autocorrelación negativa} \end{cases}$

Supuesto en los que se basa el estadístico DW:

- ✓ El modelo de regresión incluye el término de intercepto.
- ✓ Las variables explicativas no son estocásticas, son fijas.
- ✓ Las perturbaciones se generan mediante un esquema autorregresivo de primer orden, no se usa DW para esquemas autorregresivos de orden superior.
- ✓ Se supone que el término de error está normalmente distribuido.
- ✓ El modelo no incluye valores rezagados de la variable dependiente como variable explicativa.

DW NO SE PUEDE APLICAR A MODELOS AUTORREGRESIVOS

$$Y_t = \beta_1 + \beta_2 X_{2t} + \cdots + \beta_k X_{kt} + \delta Y_{t-1} + \varepsilon_t$$

- ✓ No faltan observaciones en los datos.

Ejemplo:

```
## Coefficients:
##               Estimate  Std. Error  t value     Pr(>|t|)
## (Intercept) 160054.69   48279.22     3.315      0.00254 **
## CPI          -7468.18     948.94    -7.870      1.43e-08 ***
## GDP            230.73      12.62    18.285      < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43130 on 28 degrees of freedom
## Multiple R-squared:  0.9906, Adjusted R-squared:   0.99
## F-statistic:  1479 on 2 and 28 DF,  p-value: < 2.2e-16
```

```
##
##  Durbin-Watson test
##
## data:  modelo
## DW = 0.95062, p-value = 0.0001341
## alternative hypothesis: true autocorrelation is greater than 0
```

Prueba Breusch-Godfrey (BG): conocida como el test del Multiplicador de Lagrange, permite la inclusión de valores rezagados de la variable regresada como variable explicativa, el uso de esquemas autorregresivos de orden superior a 1, AR(2), AR(3), etc, y el uso de promedios móviles (MA) de términos de error ruido blanco de cualquier orden.

Suponga: $Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$

Donde: $\varepsilon_t = \rho_1 \varepsilon_{t-1} + \rho_2 \varepsilon_{t-2} + \cdots + \rho_p \varepsilon_{t-p} + \mu_t$

$H_0: \rho_1 = \rho_2 = \ldots = \rho_p = 0$ No autocorrelación de residuos

$H_1: al\ menos\ un\ \rho_i \neq 0$ Autocorrelación de residuos

Procedimiento:

Paso 1: estime la regresión original mediante MCO.

Pasó 2: estime la regresión de $\hat{u}_t$ sobre la o las variables regresoras incluyendo los valores rezagados de los residuos estimados en la regresión original y obténgase el R^2 de esa regresión auxiliar. El número de rezagos de los residuos dependerá del orden del esquema autorregresivo.

$$\hat{\varepsilon}_t = \alpha_1 + \alpha_2 X_t + \hat{\rho}_1 \hat{\varepsilon}_{t-1} + \hat{\rho}_2 \hat{\varepsilon}_{t-2} + \cdots + \hat{\rho}_p \hat{\varepsilon}_{t-p} + \hat{\mu}_t$$

Paso 3: calcule el estadístico de prueba:

$$(n-p) * R^2 \sim {\chi_p}^2$$

Si ${\chi_c}^2 > {\chi_p}^2$ rechazar Ho.

Si $Pv > \alpha$ No rechazar Ho.

```
##  Breusch-Godfrey test for serial correlation of order up to 2
##
## data:  modelo
## LM test = 9.9486, df = 2, p-value = 0.006913
```

MEDIDAS PARA CORREGIR EL PROBLEMA DE AUTOCORELACIÓN

- Averiguar si la autocorrelación es pura y no es resultado de un problema de sesgo de especificación.

- **Aplicar Mínimos Cuadrados Generalizados (MCG).**

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t$$

$$\varepsilon_t = \rho \varepsilon_{t-1} + \epsilon_t \qquad -1 < \rho < 1$$

- ✓ **Cuando se conoce $\boldsymbol{\rho}$:**

$$Y_t = \beta_1 + \beta_2 X_t + \varepsilon_t \qquad (1)$$

$$Y_{t-1} = \beta_1 + \beta_2 X_{t-1} + \varepsilon_{t-1} \qquad (2)$$

Multiplicando (2) por ρ:

$$\rho Y_{t-1} = \rho\beta_1 + \rho\beta_2 X_{t-1} + \rho\varepsilon_{t-1} \qquad (3)$$

Restando (3) de (1)

$$(Y_t - \rho Y_{t-1}) = \beta_1(1-\rho) + \beta_2(X_t - \rho X_{t-1}) + \mu_t \qquad (4)$$

Como μ_t satisface las suposiciones de MCO, se obtienen estimadores MELI. Son MCP, ponderados por *(1-p)*

La ecuación (4) se conoce como ***generalizada, cuasi generalizada o ecuación en cuasi diferencia***. Se debe tener cuidado al trabajar con este tipo de ecuaciones porque se pierde la primera observación debido a que no tiene antecedentes, para evitar esa pérdida se realiza una transformación ***Prais-Winsten***, donde:

$$Y_o = Y_1\sqrt{1-\rho^2}$$

$$X_o = X_1\sqrt{1-\rho^2}$$

ES OBLIGATORIO REALIZAR LA TRANSFORMACIÓN PRAIS - WINSTEN PORQUE SI NO LA VARIANZA DE LOS RESIDUOS NO SERÁ HOMOSCEDASTICA

- ✓ **Cuando no se conoce ρ:** es lo más común porque en la práctica ocasionalmente ρ se conoce.

Método de la primera diferencia: como generalmente $\boldsymbol{\rho}$ se encuentra entre 0 y ± 1, se parte de la posición extrema de que $\rho = 1$, sustituyendo ese valor en (4),

$$(Y_t - Y_{t-1}) = \beta_2(X_t - X_{t-1}) + (\varepsilon_t - \varepsilon_{t-1})$$

$$\Delta Y_t = \beta_2 \Delta X_t + \mu_t$$

Regla práctica de Maddala: "utilice la forma de primera diferencia siempre que $DW < R^2$"

***ρ* basada en el estadístico DW:**

$$d = 2(1 - \hat{\rho}) \quad \Longrightarrow \quad \hat{\rho} = 1 - \frac{d}{2}$$

Y se calcula la ecuación (4)

***ρ* estimada a partir de los residuos:**

$$\hat{\varepsilon}_t = \hat{\rho} * \hat{\varepsilon}_t + v_t$$

Métodos iterativos para estimar ρ:

- Procedimiento iterativo de Cochrane Orcutt.
- Procedimiento de dos pasos de Cochrane Orcutt.
- Procedimiento Durbin de dos pasos.
- Procedimiento de rastreo o búsqueda Hildreth-Lu.

Todos los métodos descritos para estimar ρ se conocen como MCG factibles (MCGF) o MCG estimados (MCGE). Tenga en cuenta que estos métodos tienen propiedades óptimas en forma asintótica.

- **Metodo de Newey-West:** se utiliza para obtener los errores estándar de los estimadores MCO corregido para autocorrelación. Estos errores estándar corregidos se conocen como ***CHA (consistentes con heteroscedasticidad y autocorrelación).*** Al igual que MCG, este método trabaja con muestras grandes.

Modelo en diferencia: Importaciones de Estados Unidos (IMPORTS), PIB (GDP) e IPC (CPI), 1975-2005.

```
## lm(formula = dImports ~ dCPI + dGDP)
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -84933.94   24280.44  -3.498  0.00164 **
## dCPI          1548.12    3677.45   0.421  0.67710
## dGDP           361.28      43.27   8.350 5.84e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## Residual standard error: 34130 on 27 degrees of freedom
## Multiple R-squared:  0.7212, Adjusted R-squared:  0.7005
## F-statistic: 34.91 on 2 and 27 DF,  p-value: 3.254e-08

##  Durbin-Watson test
## data:  modelo3
## DW = 2.1242, p-value = 0.5237
## alternative hypothesis: true autocorrelation is greater than 0

##  Breusch-Godfrey test for serial correlation of order up to 2
##
## data:  modelo3
## LM test = 1.3275, df = 2, p-value = 0.5149
```

Agregar una variable de tendencia:

```
## lm(formula = Imports ~ CPI + GDP + Year)
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.498e+08  2.988e+07   5.012 2.95e-05 ***
## CPI          2.843e+03  2.174e+03   1.308    0.202
## GDP          3.122e+02  1.871e+01  16.681 9.57e-16 ***
## Year        -7.609e+04  1.520e+04  -5.006 2.99e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Residual standard error: 31630 on 27 degrees of freedom
## Multiple R-squared:  0.9951, Adjusted R-squared:  0.9946
## F-statistic:  1841 on 3 and 27 DF,  p-value: < 2.2e-16
##  Durbin-Watson test
##
## data:  modelo2
## DW = 1.2343, p-value = 0.002335
## alternative hypothesis: true autocorrelation is greater than 0
##  Breusch-Godfrey test for serial correlation of order up to 2
##
## data:  modelo2
## LM test = 5.8818, df = 2, p-value = 0.05282
```

Método de Newey-West:

```
## Esta es la matriz de varianzas covarianzas corregidas según NW
##                (Intercept)         CPI          GDP
## (Intercept) 31064922005  -784549322 11031090.955
## CPI          -784549322   20281742  -286994.952
## GDP            11031091    -286995     4069.402
```

		MCO (no corregidos)		Estimadores corregidos (Robustos)	
	Parámetro estimado	Error Estándar	Estadístico t	Error Estándar	Estadístico t
Intercepto	160054.69	48279.22	3.315	176252.438	0.9080991
CPI	-7468.18	948.94	-7.870	4503.525	-1.6582956
GDP	230.73	12.62	18.285	63.79186	3.6169325

4) SESGO DE ESPECIFICACIÓN:

ERRORES DE ESPEFICACIÓN ≠ ERRORES DE MALA ESPECIFICACIÓN

¿CUÁLES SON LOS CRITERIOS PARA SELECCIONAR UN MODELO?

Según Hendry y Richard (1983) el análisis empírico para la selección de un modelo debe cumplir lo siguiente:

- Ser aceptable según los datos.
- Ser consistente con la teoría.
- Tener represoras débilmente exógenas.
- Mostrar constancia paramétrica.
- Exhibir coherencia en los datos.
- Ser inclusivo.

CLASIFICACIÓN DE LOS ERRORES DE ESPECIFICACIÓN

- **Omisión de una variable relevante:**

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$ modelo verdadero

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + v_i$ modelo ajustado

$v_i = \beta_4 X_{4i} + \varepsilon_i$

- **Inclusión de una variable irrelevante:**

$Y_i = \beta_1 + \beta_2 X_2 + \beta_3 X_{3i} + \varepsilon_i$ modelo verdadero $\beta_4 = 0$

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + v_i$ modelo ajustado

$v_i = \varepsilon_i - \beta_4 X_{4i}$

Adopción de una forma funcional incorrecta:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 {X_{2i}}^2 + \beta_4 {X_{2i}}^3 + \varepsilon_i$$ modelo verdadero

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 {X_{2i}}^2 + v_i$$ modelo ajustado

$$v_i = \varepsilon_i + \beta_4 {X_{2i}}^3$$

Sesgo por errores de medición:

$$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \varepsilon_i$$ modelo verdadero

$$Y_i^* = {\beta_1}^* + {\beta_2}^* {X_{2i}}^* + {\beta_3}^* {X_{3i}}^* + {u_i}^*$$ modelo ajustado

$$Y_i^* = Y_i + \tau_i$$

$$X_i^* = X_i + w_i$$

τ_t y w_t: errores de medición.

❖ **Especificación incorrecta del término de error estocástico:**

$$Y_t = \beta_1 X_1 u_t$$

$$Y_t = \alpha_1 X_1 + u_t$$

¿Será $E(\widehat{\alpha_1}) = \beta_1$?

CONSECUENCIAS DE COMETER ERRORES DE ESPECIFICACIÓN

1. Omisión de una variable relevante (especificación insuficiente):

$Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i$ modelo verdadero

$Y_i = \alpha_1 + \alpha_2 X_{2i} + v_i$ modelo ajustado

❖ Si la variable X_3 está correlacionada con la variable X_2 $(r_{23} \neq 0)$ los parámetros estimados del modelo ajustado $\hat{\alpha}_1$ y $\hat{\alpha}_2$ son ***SEGADOS E INCONSISTENTES***. El sesgo no desaparece independientemente de qué tan grande sea la muestra.

- Aunque X_2 y X_3 no estén correlacionadas $(r_{23} = 0)$, $\hat{\alpha}_1$ será sesgado no obstante $\hat{\alpha}_2$ será insesgado.

- La varianza de los residuos y la varianza de $\hat{\alpha}_2$ $\left(Var(\hat{\alpha}_2) = \sigma^2 / \sum {x_{2i}}^2\right)$ estarán incorrectamente estimada.

- En resultado, los intervalos de confianza y las pruebas de hipótesis conducirán a resultados errados sobre la significancia estadística de los parámetros, así los pronósticos basados en esos modelos tampoco son confiables.

$$E(\hat{\alpha}_2) = \beta_2 + \beta_3 b_{32}$$

Término del sesgo

b_{32}: *pendiente de la regresión de* X_3 *sobre* X_2.

$$Var(\hat{\alpha}_2) = \sigma^2 / \sum {x_{2i}}^2 \qquad Var(\hat{\beta}_2) = \frac{\sigma^2}{\sum {x_{2i}}^2} FIV$$

A MENOS QUE EXISTA UNA RAZÓN TEÓRICA MUY FUERTE QUE LO SUSTENTE, NUNCA ELIMINE UNA VARIABLE DEL MODELO

2. *Inclusión de una variable irrelevante (sobreespecificación):*

$Y_i = \beta_1 + \beta_2 X_{2i} + u_t$ modelo verdadero

$Y_i = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + v_t$ modelo ajustado

- Los estimadores MCO de los parámetros del modelo ajustado son ***INSESGADOS Y CONSISTENTES.***
- La varianza de los residuos está correctamente estimada.
- Las α estimadas serán ***INEFICIENTES*** en la mayoría de los casos, por tanto, sus varianzas serán más grandes que las de los parámetros del verdadero modelo. Así, los intervalos de confianza y las pruebas de hipótesis no serán del todo válidas.

$$Var(\hat{\beta}_2) = \sigma^2 / \sum {x_{2i}}^2 \qquad Var(\hat{\alpha}_2) = \frac{\sigma^2}{\sum {x_{2i}}^2} FIV$$

EL MEJOR ENFOQUE ES INCLUIR SOLAMENTE LAS VARIABLES EXPLICATIVAS QUE, TEÓRICAMENTE, INFLUYAN *DIRECTAMENTE* SOBRE LA VARIABLE EXPLICADA Y QUE NO SEAN TENIDAS EN CUENTA EN OTRAS VARIABLES INCLUIDAS. M. Intrilligator (1978).

3. Errores de medición en la variable explicada (Y):

$$Y_i^{\,*} = \alpha_1 + \beta X_2 + u_i$$

Como $Y_i^{\,*}$ no puede ser medido directamente, se utiliza $Y_i = Y_i^{\,*} + \varepsilon_i$.

$$Y_i = \alpha_1 + \beta X_2 + u_i + \varepsilon_i$$

$$E(u_i) = E(\varepsilon_i) = 0; \quad Cov(u_i, \varepsilon_i) = 0; \quad Var(\hat{\beta}) = \frac{\sigma_u^{\,2} + \sigma_\varepsilon^{\,2}}{\sum x_{2i}^{\,2}}$$

- Los parámetros estimados serán ***INSESGADOS Y CONSISTENTES*** pero las varianzas estimadas son ahora más grandes.

4. Errores de medición en la variable explicativa X:

$$Y_i = \alpha_1 + \beta X_i^{\,*} + u_i$$

Donde X_i^* no puede ser observado por ello se utiliza $X_i = X_i^* + w_i$

$$Y_i = \alpha_1 + \beta(X_i - w_i) + u_i$$

$$Y_i = \alpha_1 + \beta X_i + z_i$$

$Cov(z_i, X_i) = -\beta\sigma^2{}_w$ → Autocorrelación

- Los estimadores MCO son ***SESGADOS E INCONSISTENTES***.

PRUEBAS PARA DETECTAR ERRORES DE ESPECIFICACIÓN

- Detección de la presencia de variables innecesarias. Usar una prueba t , aplicar una ***búsqueda exhaustiva de datos***[9]**.**

- Revisar el R^2, el $\bar{R}^2$, las razones t estimadas, los signos de los coeficientes y el estadístico D-W, para detectar la omisión de variables relevantes o forma funcional incorrecta.

- Examen de los residuos, para buscar patrones distinguibles (especialmente datos cross section).

[9] Enfoque ascendente, regresión al tanteo, exhumación de datos, rastreo de datos y trituración de datos.

❖ **Prueba RESET de Ramsey**, es una prueba general de errores de especificación:

1. Regrese la regresión original y obtenga $R^2{}_{viejo}$ y $\hat{Y}_i$.

2. Vuelva a estimar la regresión original introduciendo la variable estimada $(\hat{Y}_i)$ como variable independiente y obtenga $R^2{}_{nuevo}$:

$$Y_i = \beta_1 + \beta_2 X_2 + \beta_3 \hat{Y}_i + \beta_4 \hat{Y}_i^{\,2} + \beta_5 \hat{Y}_i^{\,3} + u_i$$

3. Planteamiento de la hipótesis de prueba:

H_0: $El\ modelo\ está\ correctamente\ especificado$

H_1: $El\ modelo\ \boldsymbol{NO}\ está\ correctamente\ especificado$

4. Calcule el estadístico de prueba:

$$\boldsymbol{F_c = \frac{\left(R^2{}_{nueva} - R^2{}_{vieja}\right)/m}{(1 - R^2{}_{nueva})/[n - número\ de\ parámetros\ del\ modelo\ nuevo\ (k)]} \sim F_{m,n-k}}$$

5. Aplique la regla de decisión:

Fc > Ft RECHAZAR Ho.
Fc < Ft NO RECHAZAR Ho.
Si se tiene el P-value a mano:

P-v > α NO RECHAZAR Ho.
P-v < α RECHAZAR Ho.

Otras pruebas serán: Prueba del Multiplicador de Langrange (ML), Prueba J de Davison y MacKinnon[10].

[10] Existen otras pruebas para seleccionar modelos como la prueba de Cox, la prueba JA, la P y la prueba de incluision de Mizon – Richard.

CRITERIOS PARA LA SELECCIÓN DE MODELOS

Los criterios que se estudiaran sirven para elegir entre modelos rivales y/o contrastar modelos con fines de pronóstico, para esto último se debe distinguir entre ***pronóstico dentro de la muestra y pronóstico fuera de la muestra.*** El primero hace referencia a cómo elegir el modelo que mejor se ajusta a los datos de determinada muestra y el segundo permite determinar como un modelo ajustado pronostica valores futuros de la variable explicada, conocidos valores de las variables explicativas.

- **Criterio del $\mathbf{R^2}$:** mientras más cercano a 1 se presentará un mejor ajuste. Sin embargo, presenta algunos problemas: mide la bondad de ajuste dentro de la muestra por tanto, observaciones fuera de la muestra no tienen garantía de ser pronosticadas bien. No se olvide que al comparar los R^2 de dos o más modelos la variable explicada tiene que ser la misma y por último, recuérdese que el objetivo no es maximizar el R^2 añadiendo variables.

- **Criterio del $\mathbf{R^2}$ ajustado:** este coeficiente penaliza más la introducción de nuevas regresoras al modelo. Para pronósticos comparativos R^2 ajustado es una mejor medida de bondad de ajuste, tal como lo explica Theil.

- **Criterio de información de Akaike (CIA):** impone una mayor penalización a la incorporación de nuevas variables al modelo.

$$CIA = e^{2k/n}\frac{\sum \hat{u}_i^{\,2}}{n} = e^{2k/n}\frac{SCE}{n}$$

$$\underbrace{lnCIA = \left(\frac{2k}{n}\right)}_{\text{Factor de penalización}} + ln\left(\frac{SCE}{n}\right)$$

Cuando se comparan modelos, bajo el criterio CIA se preferirá aquel cuyo valor del CIA sea menor. Es ventajoso para predecir dentro o fuera de la muestra.

- **Criterio de información Schwarz (CIS):** su fundamento el similar al CIA, se define como:

$$CIS = n^{k/n} \frac{\sum \hat{u}^2}{n} = n^{k/n} \left(\frac{SCE}{n}\right)$$

$$ln\, CIS = \frac{k}{n} ln(n) + ln\left(\frac{SCE}{n}\right)$$

Factor de penalización

- **Criterio Cp de Mallow:** si en un modelo de ***k*** regresoras se han elegido solamente ***p*** de ellas (***p*** < ***k)*** se obtiene la SCE*p* y se calcula:

$$Cp = \frac{SCEp}{\hat{\sigma}^2} - (n - 2p)$$

Al elegir el modelo bajo el criterio Cp se debe elegir aquel con el valor Cp más bajo, que se aproxime a ***p***.

LOS CRITERIOS PARA LA SELECCIÓN DE MODELOS, DEBEN SER CONSIDERADOS COMO UN COMPLEMENTO DE LAS PRUEBAS DE ESPECIFICACIÓN DEL MODELO Y NO COMO PRUEBAS FOMALES PORQUE NO LO SON, SON CRITERIOS SIMPLEMENTE DESCRIPTIVOS.

```
## Coefficients:
##              Estimate  Std. Error  t value    Pr(>|t|)
## (Intercept) 188.70352   22.74704   8.296     2.99e-12 ***
## VM           -1.25282    0.23669  -5.293     1.13e-06 ***
## CF            0.38058    0.07867   4.838     6.72e-06 ***
## PS           -1.85528    0.20584  -9.013     1.26e-13 ***
## VOL          -0.01211    0.02206  -0.549     0.585
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.524 on 76 degrees of freedom
## Multiple R-squared:  0.8833, Adjusted R-squared:  0.8772
## F-statistic: 143.8 on 4 and 76 DF,  p-value: < 2.2e-16
```

```
##  RESET test
##
## data:  mh
## RESET = 18.884, df1 = 2, df2 = 74, p-value = 2.365e-07
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 160054.69   48279.22   3.315  0.00254 **
## CPI          -7468.18     948.94  -7.870 1.43e-08 ***
## GDP            230.73      12.62  18.285  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 43130 on 28 degrees of freedom
## Multiple R-squared:  0.9906, Adjusted R-squared:   0.99
## F-statistic:  1479 on 2 and 28 DF,  p-value: < 2.2e-16
```

```
##  RESET test
##
## data:  modelo
## RESET = 8.6341, df1 = 2, df2 = 26, p-value = 0.001332
```

```
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -84933.94   24280.44  -3.498  0.00164 **
## dCPI          1548.12    3677.45   0.421  0.67710
## dGDP           361.28      43.27   8.350 5.84e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 34130 on 27 degrees of freedom
## Multiple R-squared:  0.7212, Adjusted R-squared:  0.7005
## F-statistic: 34.91 on 2 and 27 DF,  p-value: 3.254e-08
```

```
##  RESET test
##
## data:  modelo3
## RESET = 2.5496, df1 = 2, df2 = 25, p-value = 0.09825
```

```
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.498e+08  2.988e+07   5.012 2.95e-05 ***
## CPI          2.843e+03  2.174e+03   1.308    0.202
## GDP          3.122e+02  1.871e+01  16.681 9.57e-16 ***
## Year        -7.609e+04  1.520e+04  -5.006 2.99e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 31630 on 27 degrees of freedom
## Multiple R-squared:  0.9951, Adjusted R-squared:  0.9946
## F-statistic:  1841 on 3 and 27 DF,  p-value: < 2.2e-16

##  RESET test
##
## data:  modelo2
## RESET = 0.42647, df1 = 2, df2 = 25, p-value = 0.6575
```

Lecturas recomendadas:

- Gujarati, D. y Porter, D. (2010). Econometría. 5ta. Edición. McGraw Hill Interamericana.

- Novales, A. (1993). Econometría. 2da. Edición. McGraw Hill Interamericana.

- Wooldridge, Jeffrey M. (2010). Introducción a la econometría. Un enfoque moderno. 4ta Edición. CENGAGE Learning.