



Econometría I

Ejemplo de regresión espúrea y multicolinealidad

Miguel Jerez y Sonia Sotoca

Universidad Complutense de Madrid

Diciembre, 2004

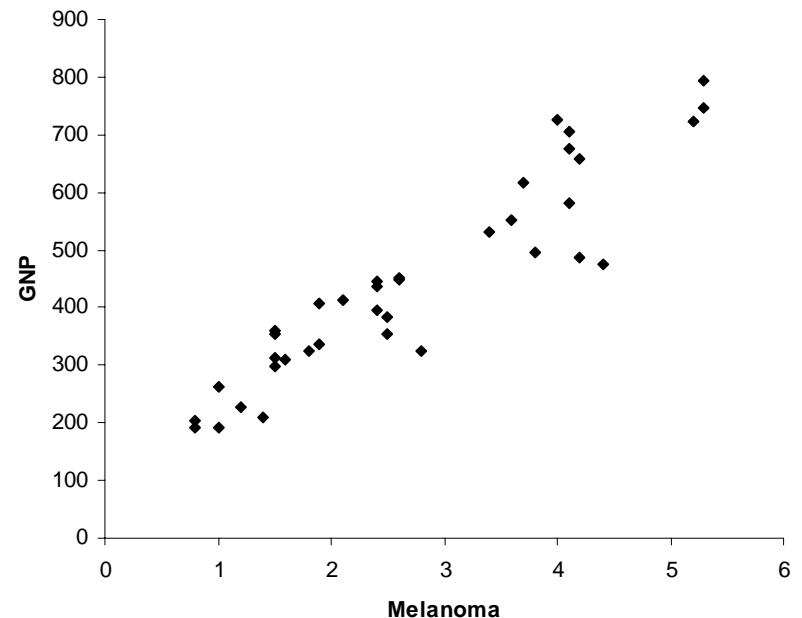
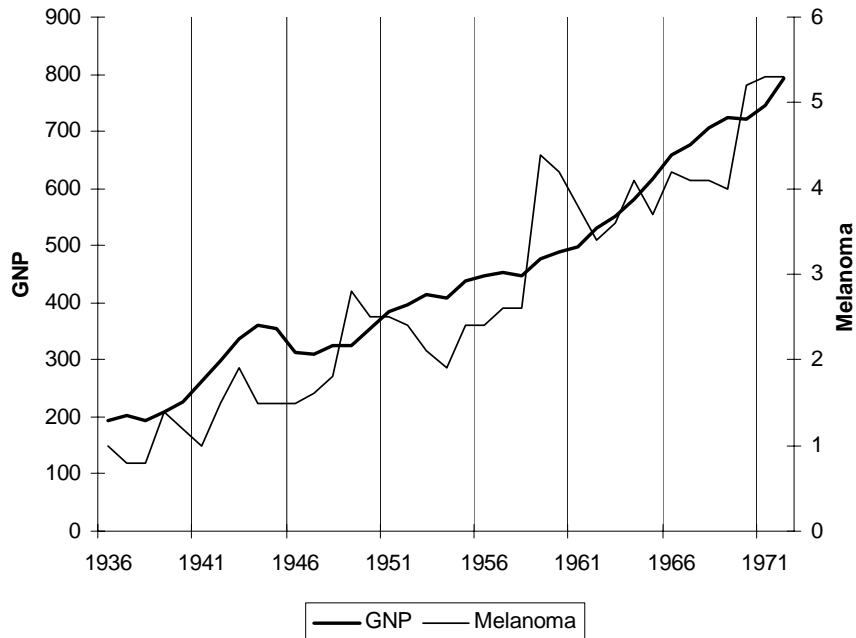
Los datos

Los gráficos muestran las series anuales (1936-1972) de:

- PNB nominal en Estados Unidos (datos en miles de millones de dólares), y
- Incidencia del melanoma en la población masculina (datos ajustados de edad) en el estado de Connecticut.

Fuente: <http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/>

Aparentemente, ambas series mantienen una relación lineal. Conceptualmente resulta absurdo relacionarlas.



Relación en niveles (I)

El cuadro muestra los resultados de una regresión en donde el PNB actúa como variable endógena y la incidencia de melanoma como variable explicativa.

Dependent Variable: GNP Method: Least Squares Date: 11/29/04 Time: 11:19 Sample: 1936 1972 Included observations: 37				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	118.5659	23.72897	4.996675	0.0000
MELANOMA	118.9808	7.814147	15.22633	0.0000
R-squared	0.868836	Mean dependent var		443.6730
Adjusted R-squared	0.865088	S.D. dependent var		171.4417
S.E. of regression	62.97110	Akaike info criterion		11.17577
Sum squared resid	138787.6	Schwarz criterion		11.26284
Log likelihood	-204.7517	F-statistic		231.8413
Durbin-Watson stat	0.879122	Prob(F-statistic)		0.000000

Resulta inmediato ver que:

- Todos los coeficientes son estadísticamente significativos, y
- El R^2 (86.7%) es muy elevado
- El coeficiente estimado implica que, si aumentara la incidencia de melanoma en un caso, cabría esperar un aumento del PNB de 119.000 millones de dólares.

Relación en niveles (II)

Resulta fácil mejorar el ajuste del modelo anterior. Por ejemplo, puede introducirse como variable explicativa adicional la incidencia del melanoma en el año anterior.

Dependent Variable: GNP Method: Least Squares Date: 11/29/04 Time: 11:21 Sample(adjusted): 1937 1972 Included observations: 36 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	114.1360	22.23928	5.132181	0.0000
MELANOMA	58.14551	20.19445	2.879281	0.0069
MELANOMA(-1)	65.69559	20.82531	3.154604	0.0034
R-squared	0.894331	Mean dependent var		450.6361
Adjusted R-squared	0.887927	S.D. dependent var		168.4839
S.E. of regression	56.40398	Akaike info criterion		10.98261
Sum squared resid	104986.5	Schwarz criterion		11.11457
Log likelihood	-194.6870	F-statistic		139.6477
Durbin-Watson stat	0.553467	Prob(F-statistic)		0.000000

Como puede apreciarse:

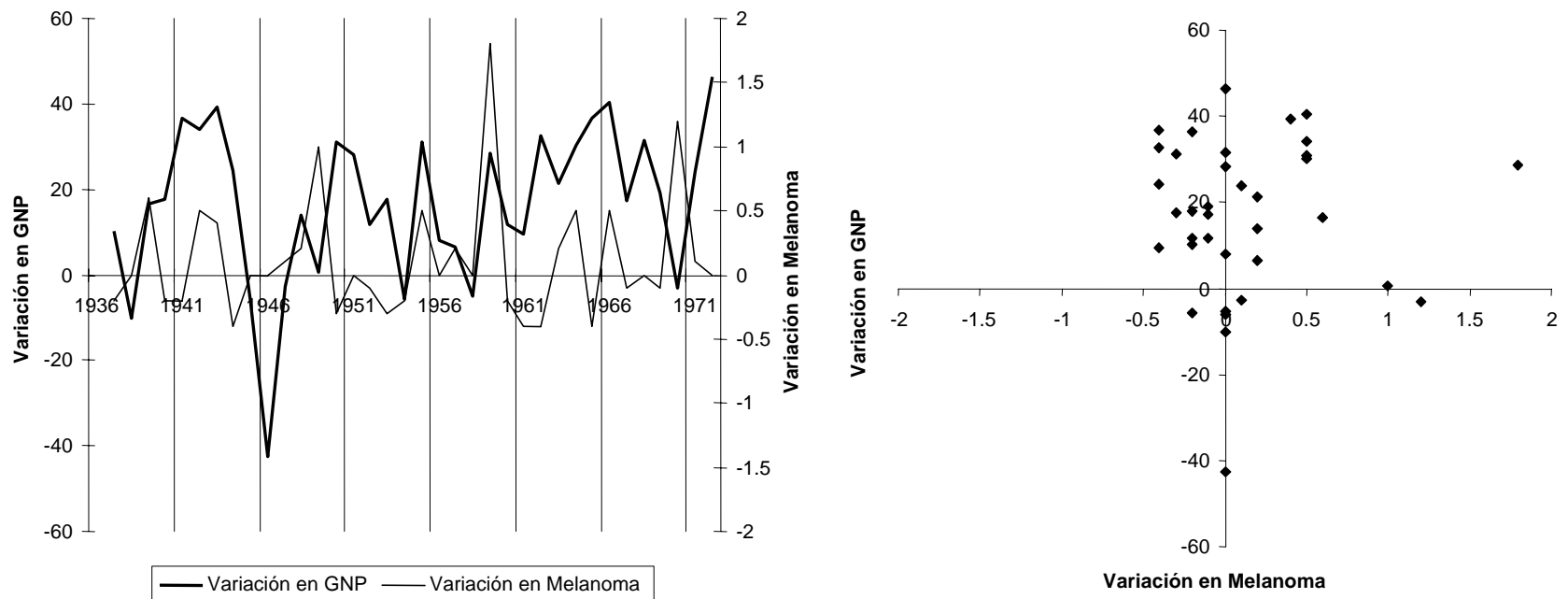
- Los coeficientes siguen siendo estadísticamente significativos y el R^2 mejora en dos puntos porcentuales
- Posiblemente la nueva variable crea un problema de colinealidad de grado, ya que la correlación entre las variables explicativas es de 0.93 y el cociente entre los autovalores mínimo y máximo de $|X^T X|$ es 5.7465e-003

Relación en primeras diferencias (I)

Si relacionamos las variables en primeras diferencias, esto es:

$$\nabla GNP_t = GNP_t - GNP_{t-1} ; \nabla Melanoma_t = Melanoma_t - Melanoma_{t-1}$$

la tendencia suele desaparecer y, con ella, la relación espúrea.



Relación en primeras diferencias (II)

Consecuentemente, la relación de regresión entre las variables transformadas no resulta significativa:

Dependent Variable: D(GNP,1) Method: Least Squares Date: 11/29/04 Time: 11:43 Sample(adjusted): 1937 1972 Included observations: 36 after adjusting endpoints				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	16.56841	3.179332	5.211287	0.0000
D(MELANOMA,1)	0.706295	6.585765	0.107246	0.9152
R-squared	0.000338	Mean dependent var		16.65278
Adjusted R-squared	-0.029064	S.D. dependent var		18.22001
S.E. of regression	18.48289	Akaike info criterion		8.725520
Sum squared resid	11614.98	Schwarz criterion		8.813493
Log likelihood	-155.0594	F-statistic		0.011502
Durbin-Watson stat	1.262415	Prob(F-statistic)		0.915224