

José María Cara Carmona

Adrián López Ibáñez

MODELO ECONOMÉTRICO

Explicación del desempleo

Desarrollaremos un modelo econométrico para intentar predecir el desempleo. Trataremos los diversos problemas que se nos presente para que nuestros estimadores sean eficaces

1. Presentación de las variables
2. Presentación de los datos
3. Representación del modelo
4. Problemas
5. Modelo final
6. Predicciones
7. Conclusiones
8. Bibliografía

Presentación de las variables

Para la elaboración de nuestro modelo, el cual tiene la siguiente ecuación (Desempleo = $\beta_1 + \beta_2 \text{ PIB} + \beta_3 \text{ IPC} + \beta_4 \text{ SMI} + \beta_5 \text{ FBC} + u$) hemos utilizado las variables:

- Desempleo: que es la variable explicada o dependiente. Se define como la situación de quienes pudiendo y queriendo trabajar, pierden su empleo o ven reducida temporalmente su jornada ordinaria de trabajo y salario.

Las demás variables son las explicativas o independientes:

- PIB: Producto Interior Bruto. Es una medida macroeconómica que expresa el valor monetario de la producción de bienes y servicios de demanda final de un país (o una región) durante un período determinado de tiempo (normalmente un año).
- IPC: Evolución de los precios de una cesta modelo que consume la población residente en España.
- SMI: fija la **cuantía retributiva mínima** que percibirá el trabajador referida a la jornada legal de trabajo, sin distinción de sexo u edad de los trabajadores, sean fijos, eventuales o temporeros.
- FBC: inversión interna bruta.

De forma más intuitiva:

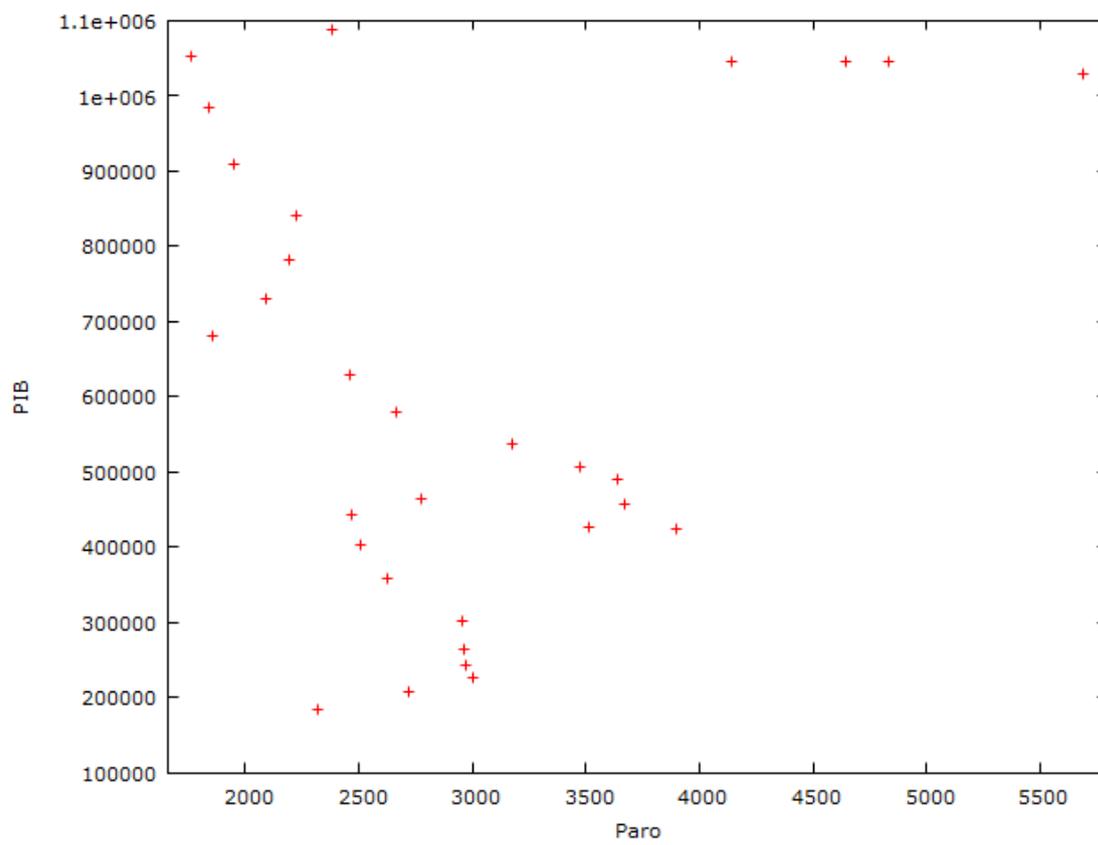
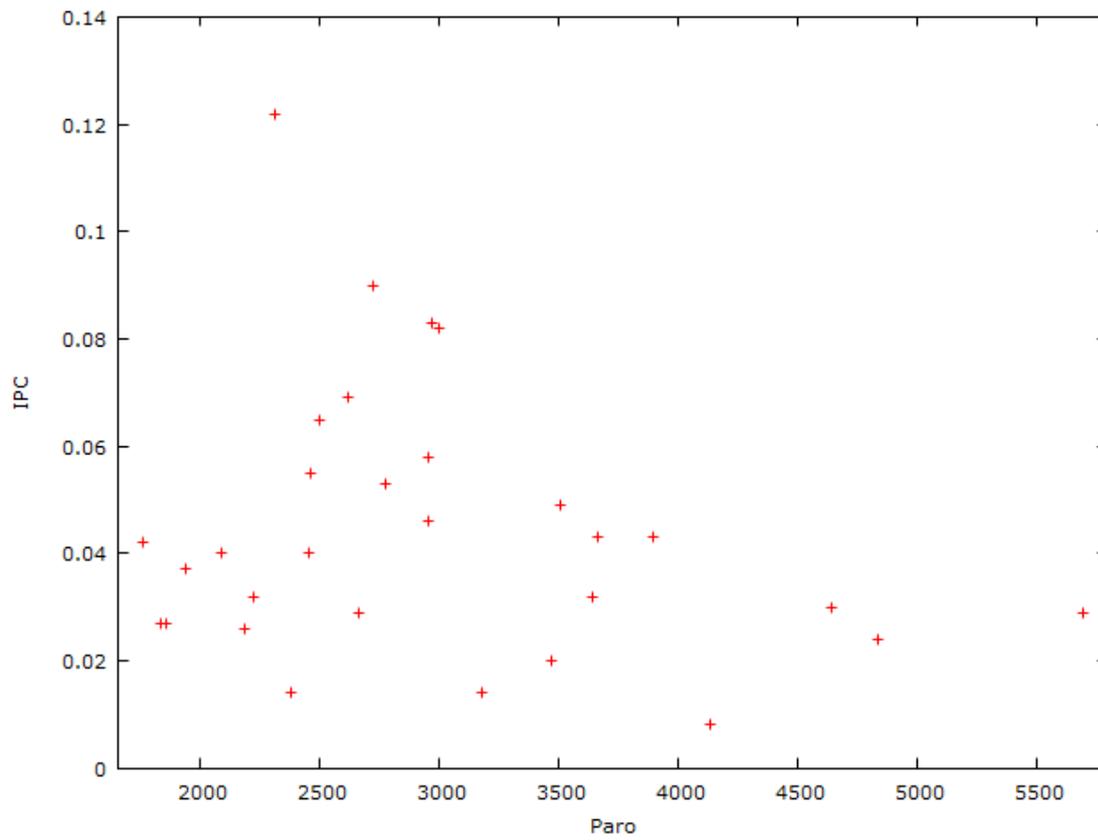
Variable	Tipo	Unidades	Signo esperado
Desempleo	Explicada Cuantitativa	Miles de personas	
PIB	Explicativa Cuantitativa	Millones de €	Negativo
IPC	Explicativa Cuantitativa	Porcentaje	Positivo
SMI	Explicativa Cuantitativa	€	Negativo
FBC	Explicativa Cuantitativa	Millones de €	Negativo

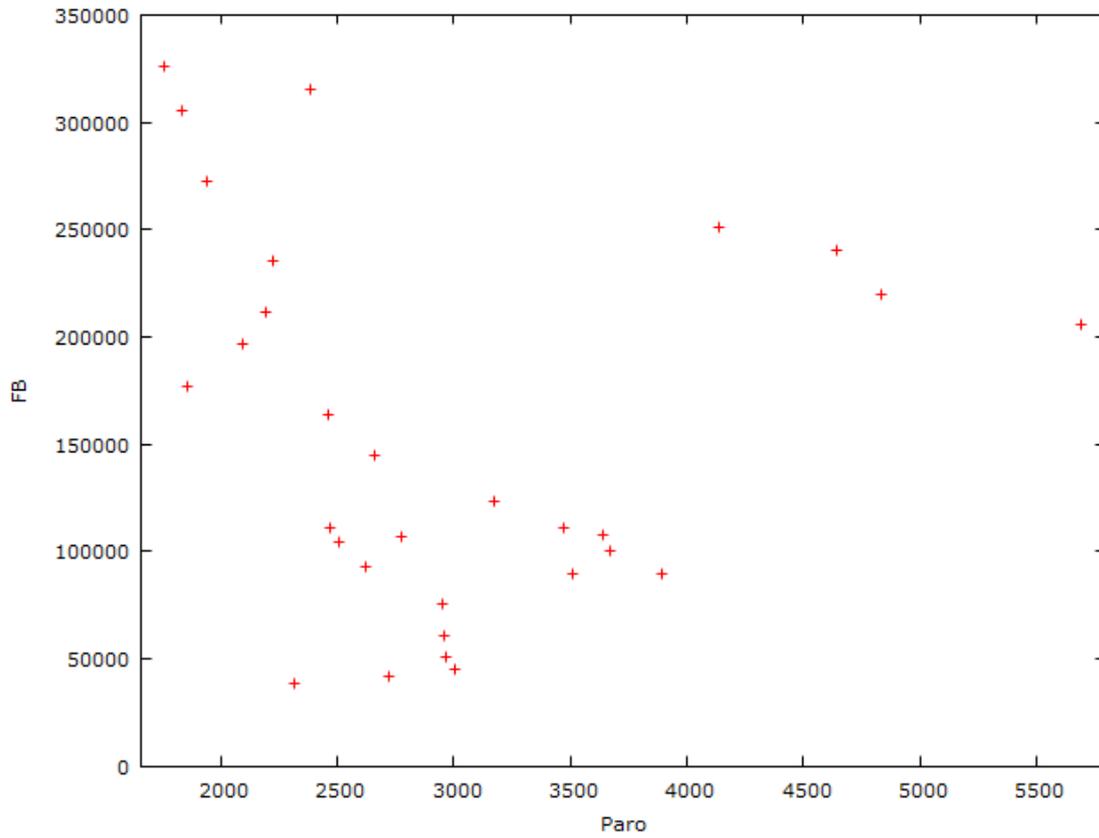
Una vez definidas las variables comenzamos con los primeros análisis.

Variable	Media	Mediana	Mínimo	Máximo
PIB	612682.	521178.	182781.	1,09E+11
Desempleo	2979.53	2749.50	1760.00	5693.00
IPC	0.0443000	0.0400000	0.00800000	0.122000
SMI	411.324	404.690	193.290	641.400
FB	153876.	117344.	38384.0	326480.
Variable	Desv. Típica.	C.V.	Asimetría	Exc. de curtosis
PIB	300840.	0.491021	0.290107	-129.038
Desempleo	950.941	0.319158	105.100	0.722682
IPC	0.0253597	0.572453	117.748	132.248
SMI	137.449	0.334162	0.230513	-0.991800
FB	86902.9	0.564759	0.497849	-0.929661
Variable	Porc. 5%	Porc. 95%	Rango IQ	Observaciones ausentes
PIB	196528.	1,07E+11	537625.	0
Desempleo	1802.35	5220.55	1247.75	0
IPC	0.0113000	0.104400	0.0287500	0
SMI	201.815	641.400	224.410	0
FB	40128.2	320418.	134285.	0

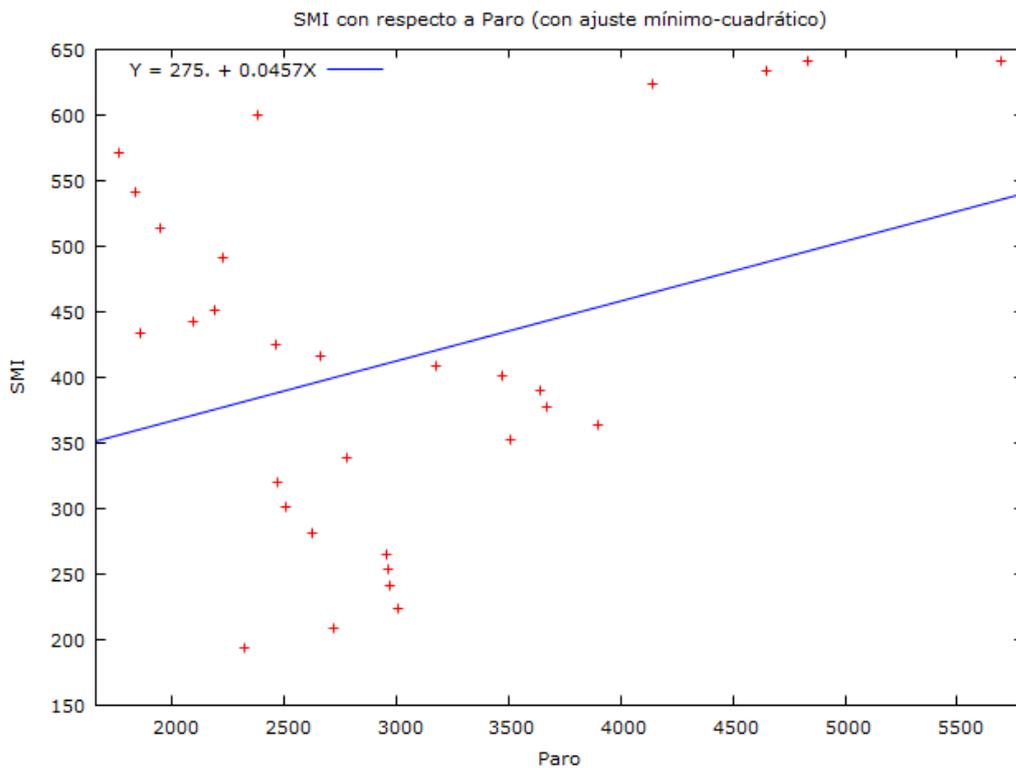
Correlación de las variables

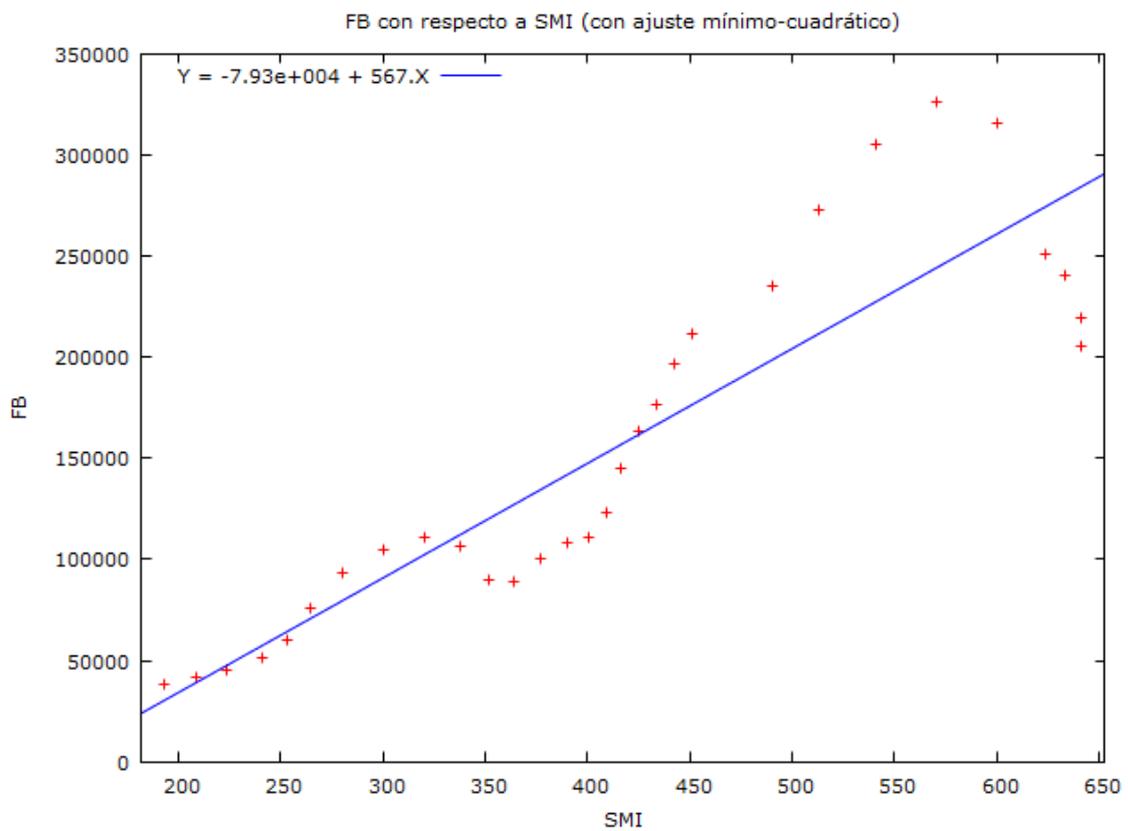
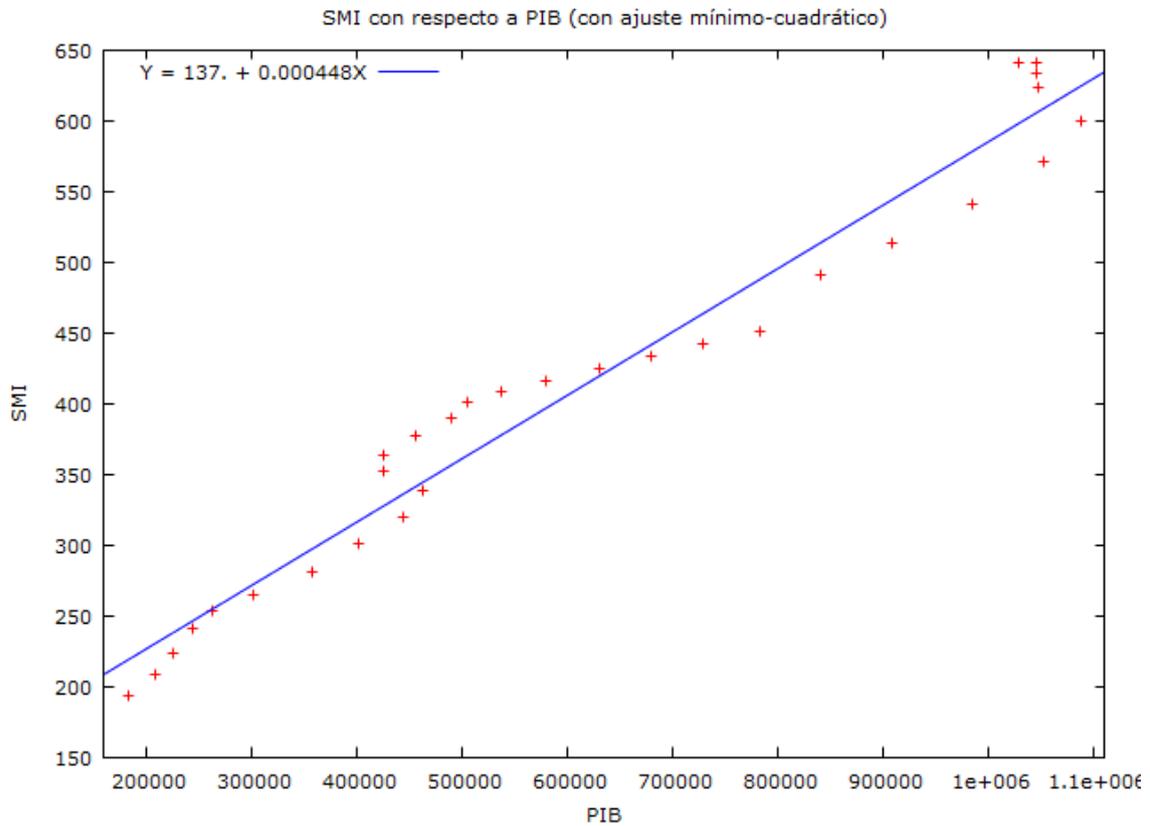
PIB	desempleo	IPC	SMI	FB	
1	0.1565	-0.7310	0.9812	0.9592	PIB
	1	-0.2249	0.3159	-0.1084	desempleo
		1	-0.7851	-0.6676	IPC
			1	0.8965	SMI
				1	FB





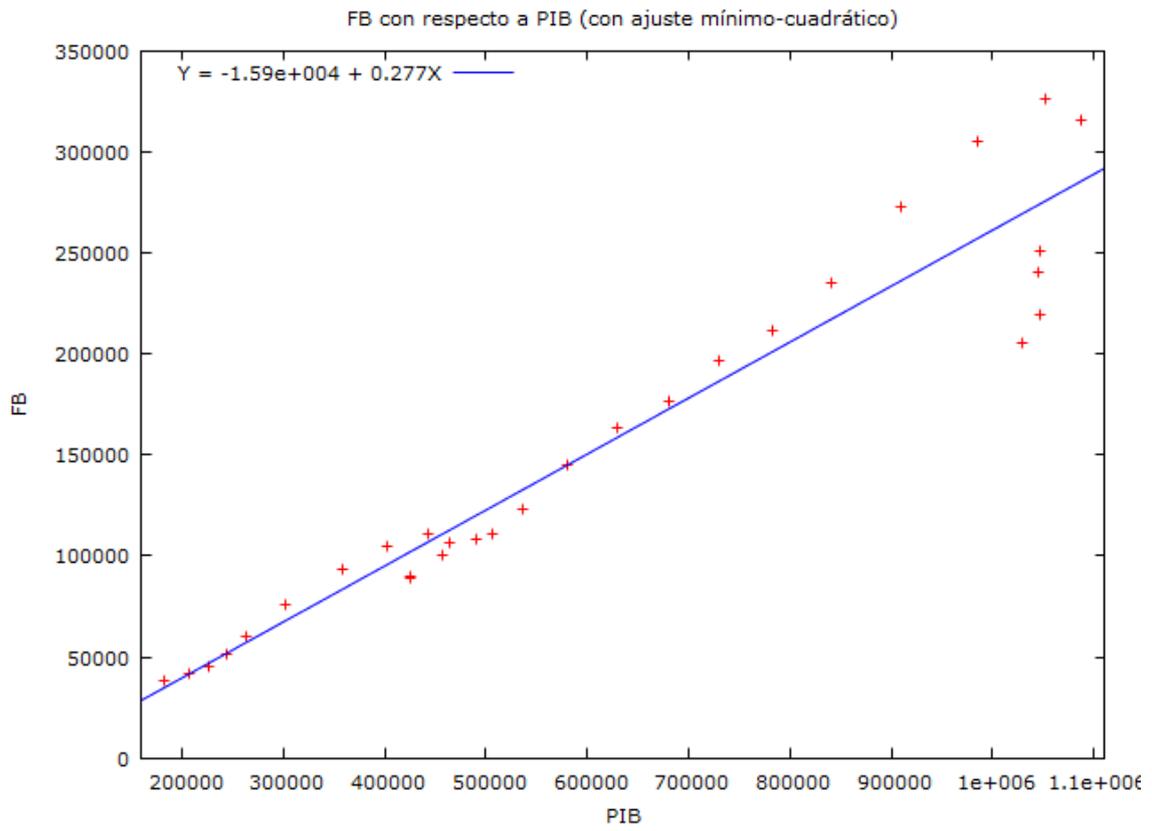
Estas gráficas presentan la ausencia de correlación entre las variables, para llegar a esta conclusión, nos podemos fijar en la falta de una línea de ajuste y también en los valores numéricos, ya que tiene valores inferiores en términos absolutos al 0,36.

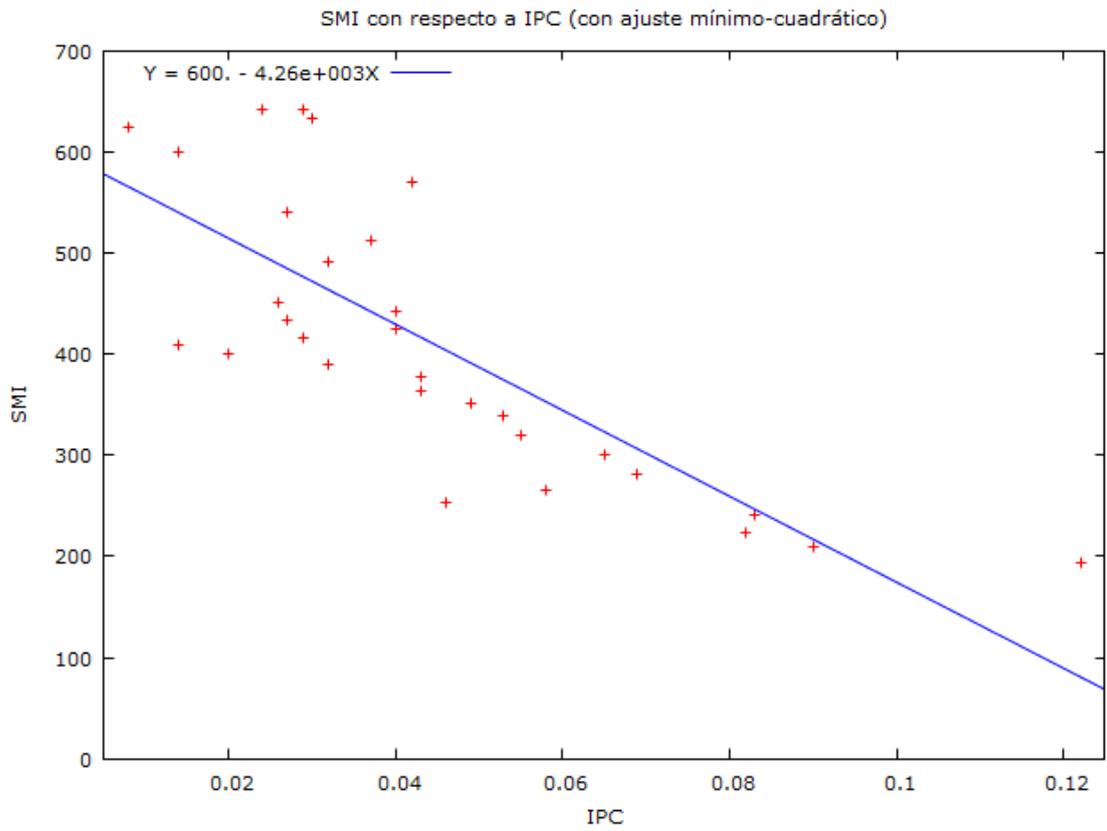
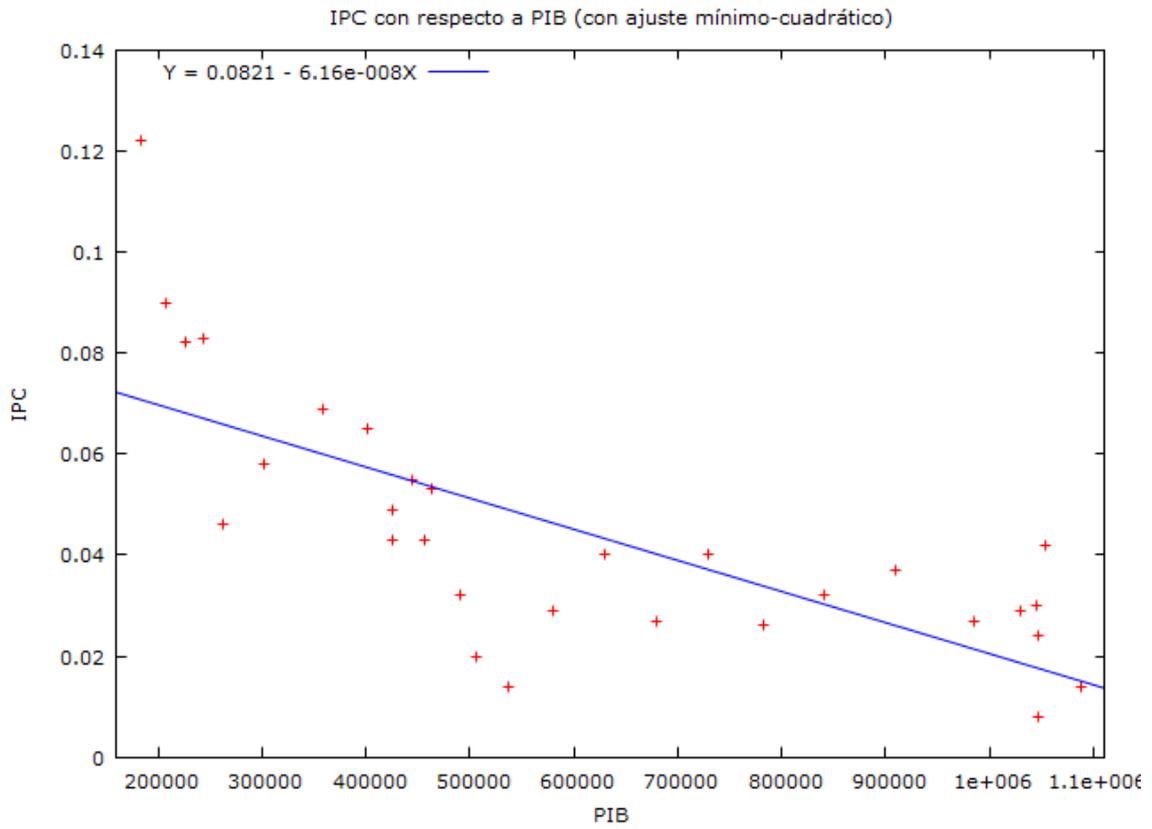


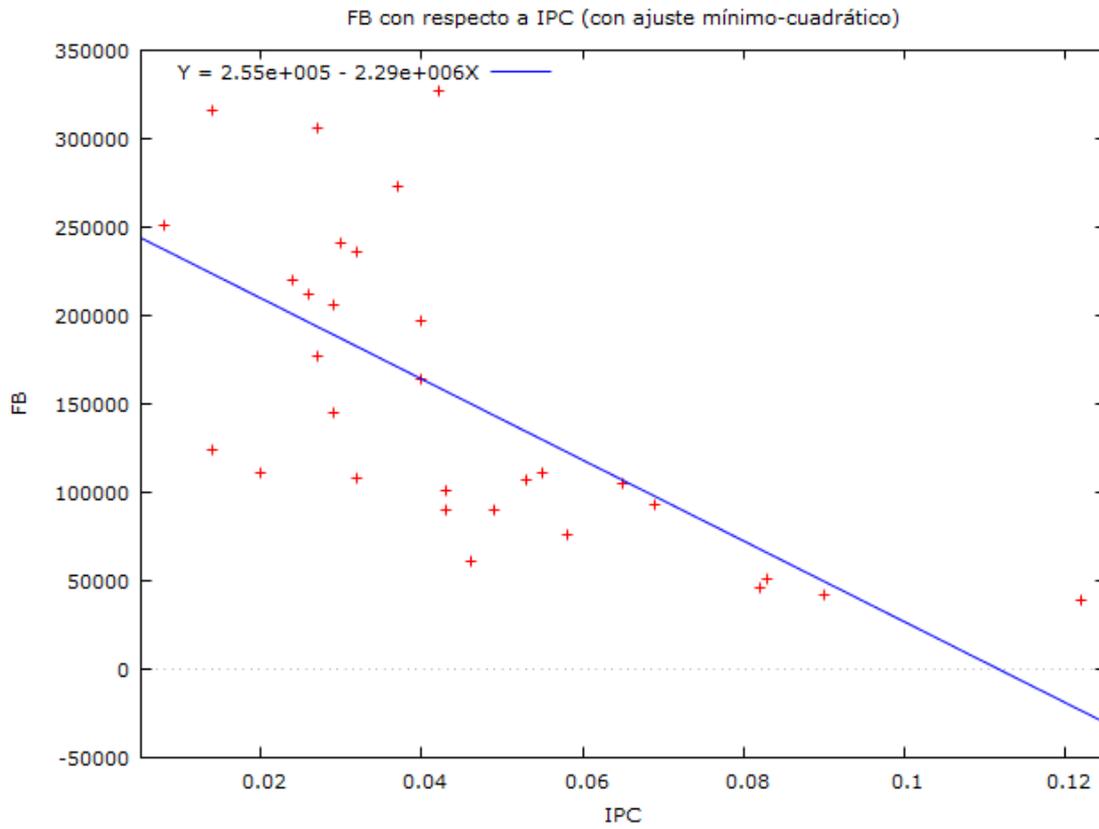


Estas gráficas muestran una correlación positiva, por la pendiente de la línea de ajuste. También podemos llegar a la misma conclusión, porque los valores son mayores y con signo

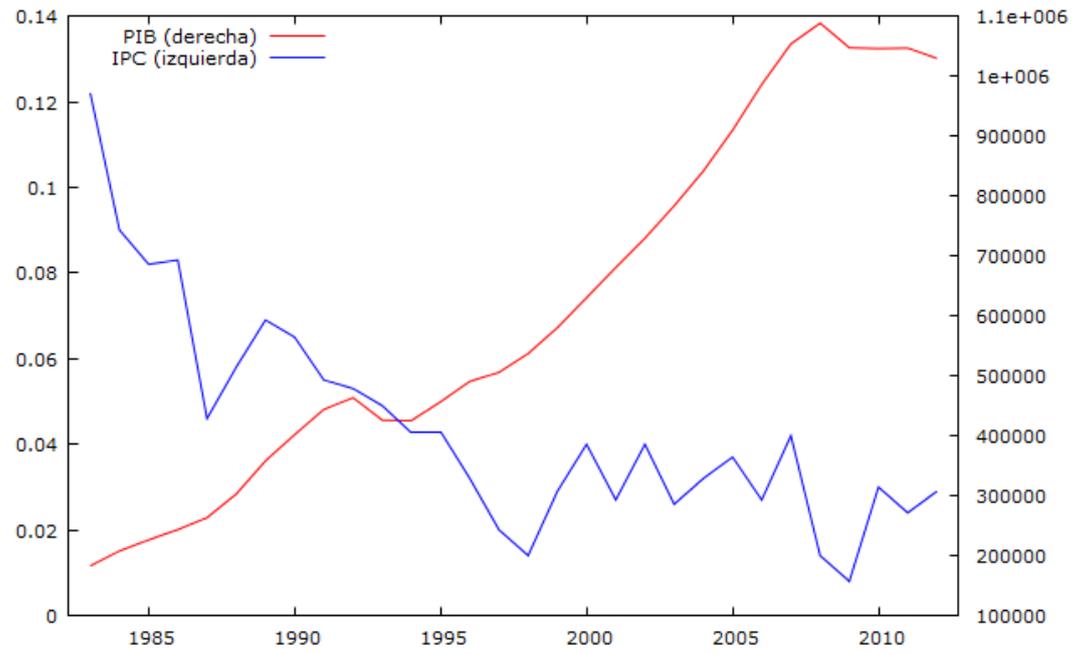
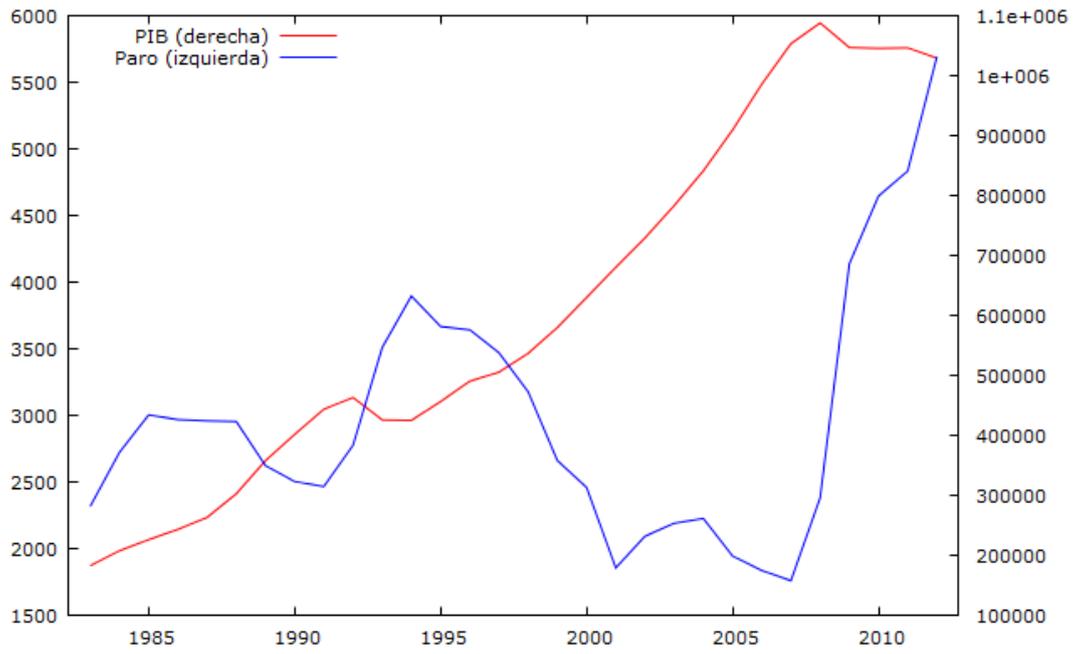
positivo.

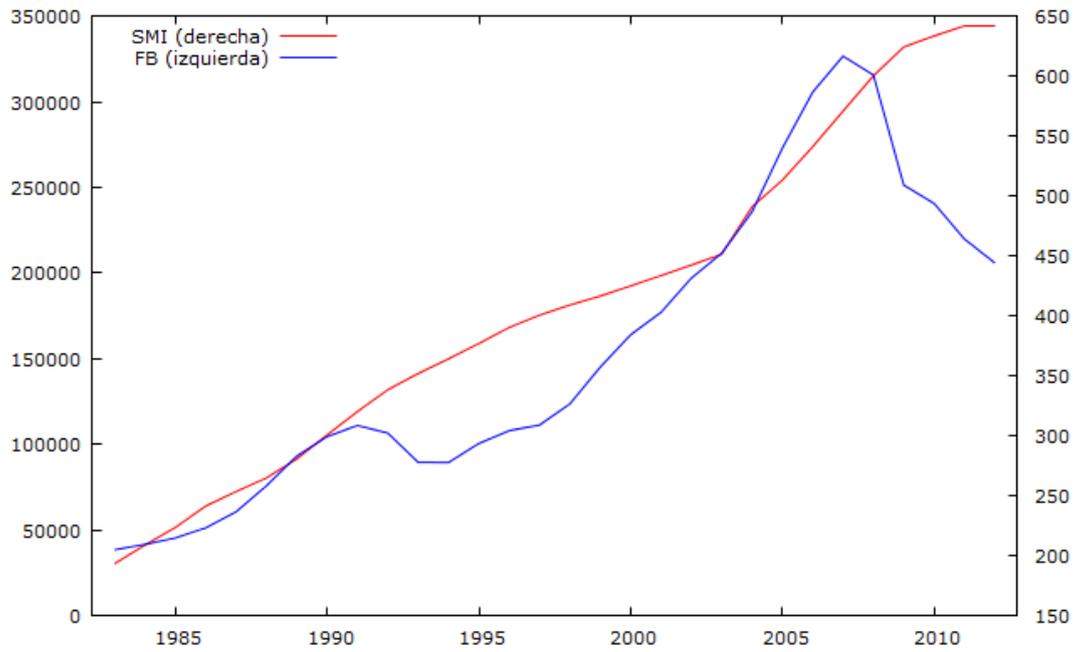
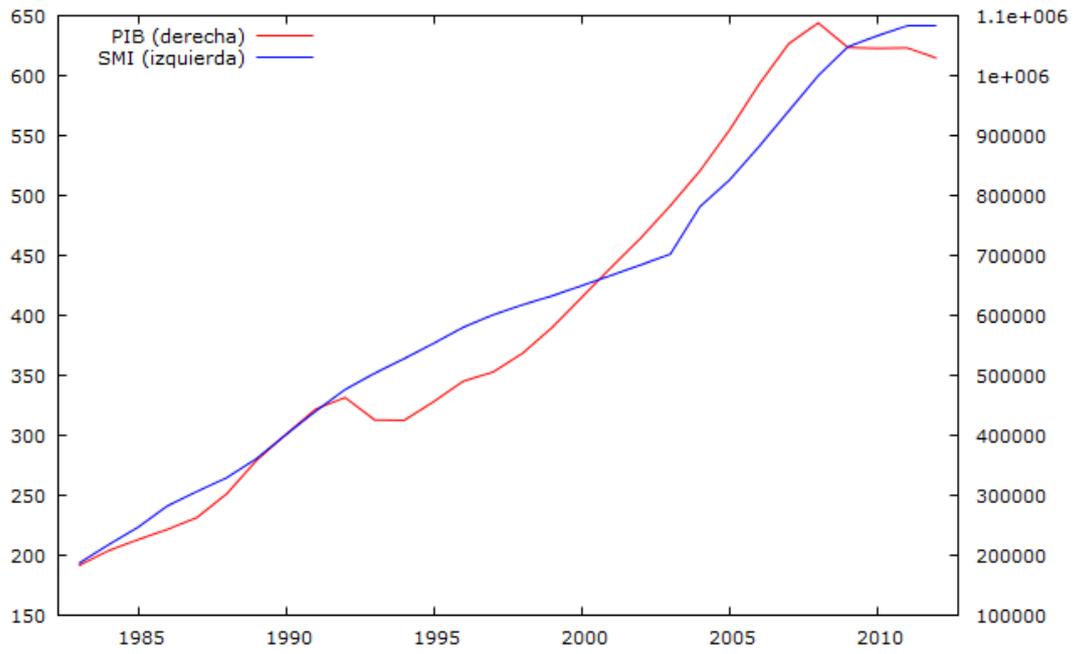


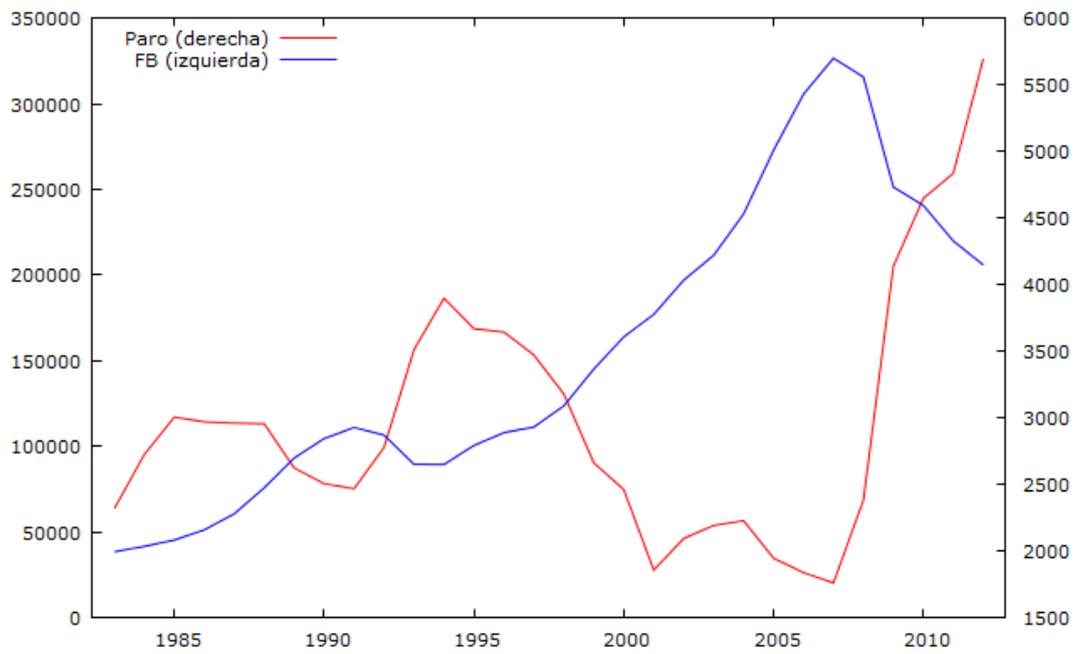
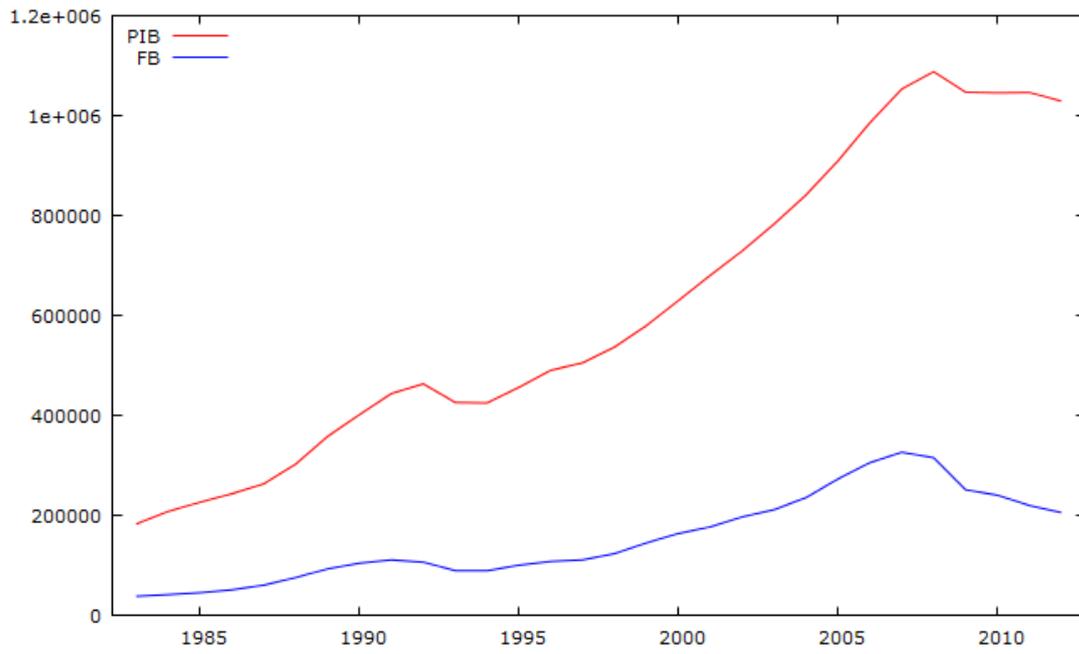


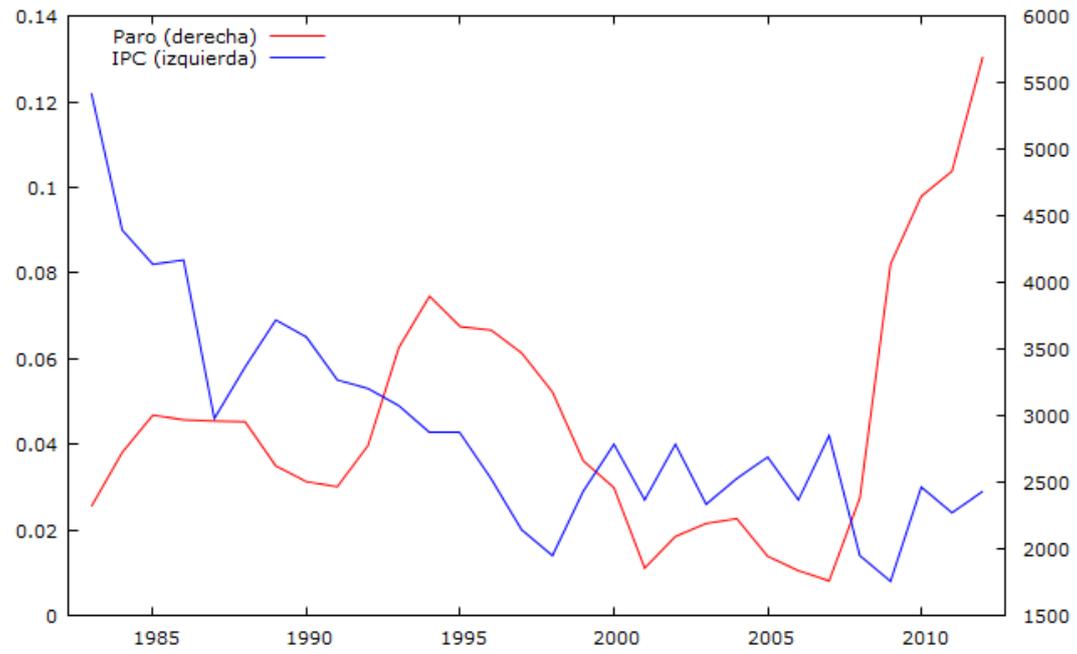
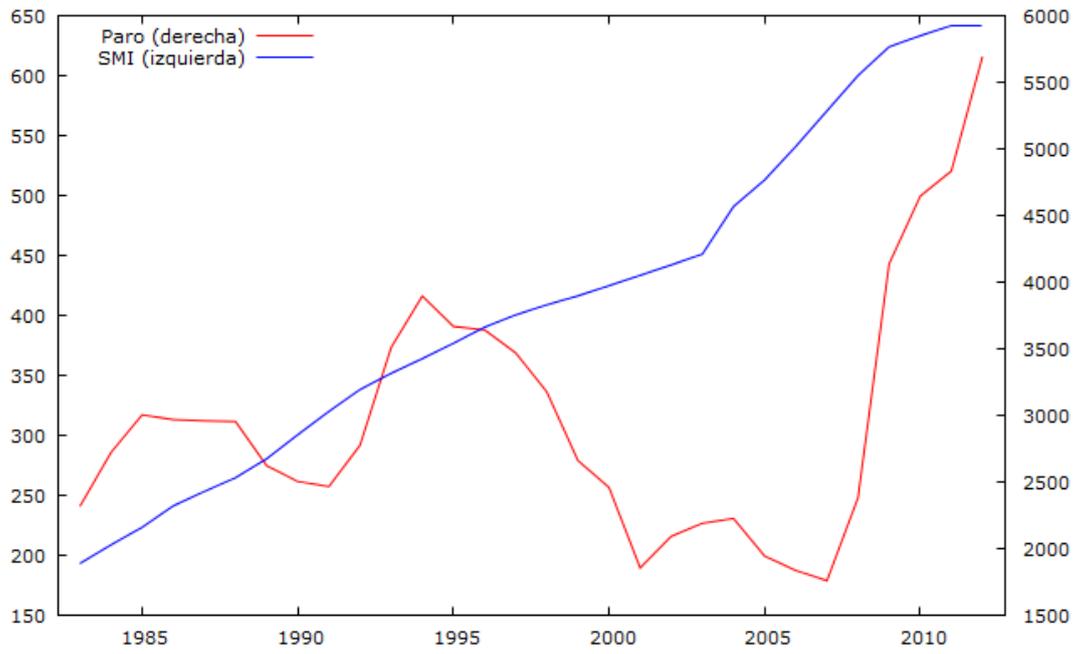


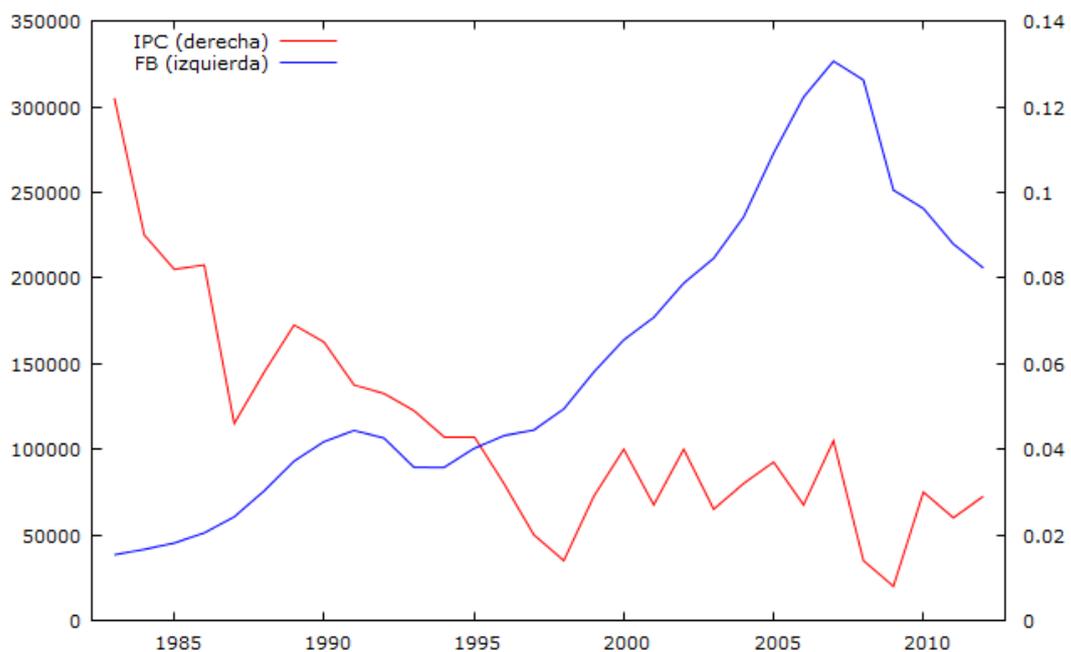
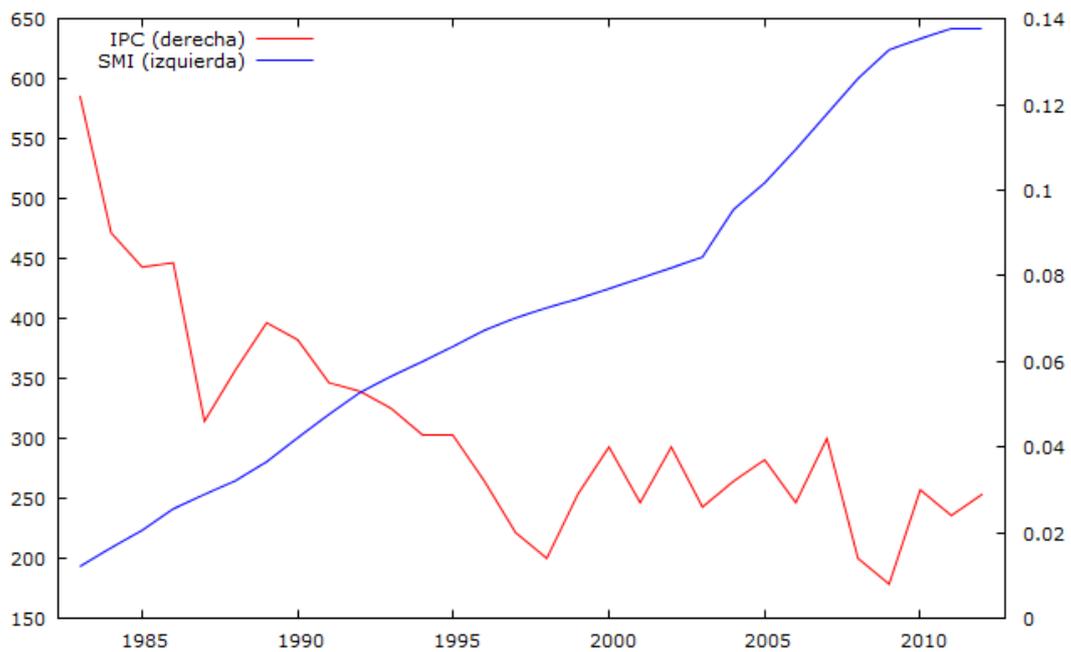
Estas gráficas muestran la correlación negativa a través de la pendiente y fijándonos en los datos que tienen más valor a la hora de llegar a una conclusión, nos muestran que son mayores que el 0,36 pero con signo negativo.











Estas gráficas concluyen con las mismas ideas que en con las gráficas anteriores. Podemos observar una correlación directa en los casos en los que las fluctuaciones son del mismo tipo, es decir, cuando una aumenta la otra variable también. También se da el caso de correlaciones inversas en los casos en los que una variable aumenta, la otra disminuye y viceversa. Por último, observamos que en algunos casos no se observa una correlación entre las variables, esto se percibe porque cada variable fluctúa a su libre albedrío sin relación con la otra a la que se le enfrenta.

Destacar una curiosidad en la comparación de estas gráficas con las anteriores, que es en el caso de SMI VS PARO. Este es un caso curioso, porque en los valores no se observa relación, pero al presentar estas dos variables enfrentadas aparece la línea de ajuste y ahora, nuevamente, nos dice que no hay relación el gráfico de series temporales, en el cual podemos

destacar que se aprecia una cierta relación en los últimos años. Esto se puede deber al nivel de confianza con la que cada una trabaja.

Presentación del modelo.

Una vez definidas las variables comenzamos con los primeros análisis.

	Coeficiente	Desv. Típica	Estadístico t	Valor p	
Const	-891.997	932.887	-0.9562	0.34814	
PIB	-	0.000880225	-0.2641	0.79385	
IPC	10178.4	4555.26	22.344	0.03463	**
SMI	175.401	509.351	34.436	0.00203	***
FB	-0.0211516	0.00478958	-44.162	0.00017	***

Media de la vble. dep.	2.979.533
Suma de cuad. Residuos	2471115
R-cuadrado	0.905770
F(4, 25)	6.007.723
Log-verosimilitud	-2.123.529
Criterio de Schwarz	4.417.118
Rho	0.572238

D.T. de la vble. dep.	9.509.406
D.T. de la regresión	3.143.956
R-cuadrado corregido	0.890693
Valor p (de F)	1.85e-12
Criterio de Akaike	4.347.058
Crit. de Hannan-Quinn	4.369.471
Durbin-Watson	0.779034

Estimando el modelo por mínimos cuadrados ordinarios, obtenemos que 3 de los 5 parámetros son significativos, es decir solo influye en el modelo las variables IPC, SMI, FB, mientras que PIB y la constante no influyen en el modelo.

Obtenemos una bondad de ajuste muy alta, entorno al 90%, que sumado a las correlaciones obtenidas anteriormente, puede presentar graves problemas de multicolinealidad, aunque eso ya lo trataremos en los siguientes apartados.

Prestando atención a los coeficientes;

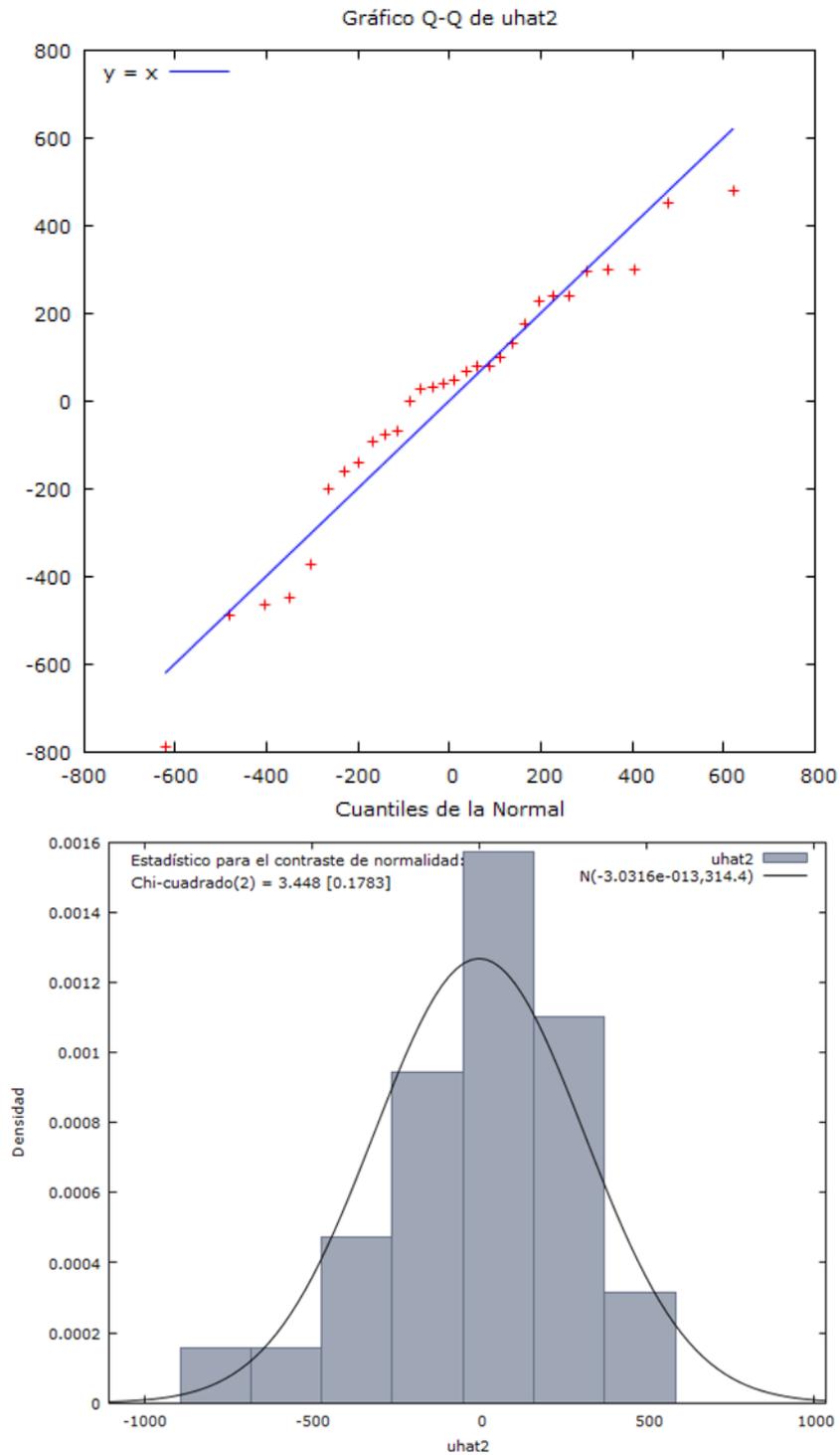
- Por cada unidad que suba el IPC (1%), el desempleo aumenta en 10178 personas.
- Por cada euro que crezca el SMI, el desempleo aumenta en 175 personas.
- Por cada millón de euros que crezca la FB, el desempleo decrece 0,02 personas.

A la vista de los resultados, podemos decir que nuestro modelo sigue una línea lógica, ya que no hay ningún signo raro.

PROBLEMAS

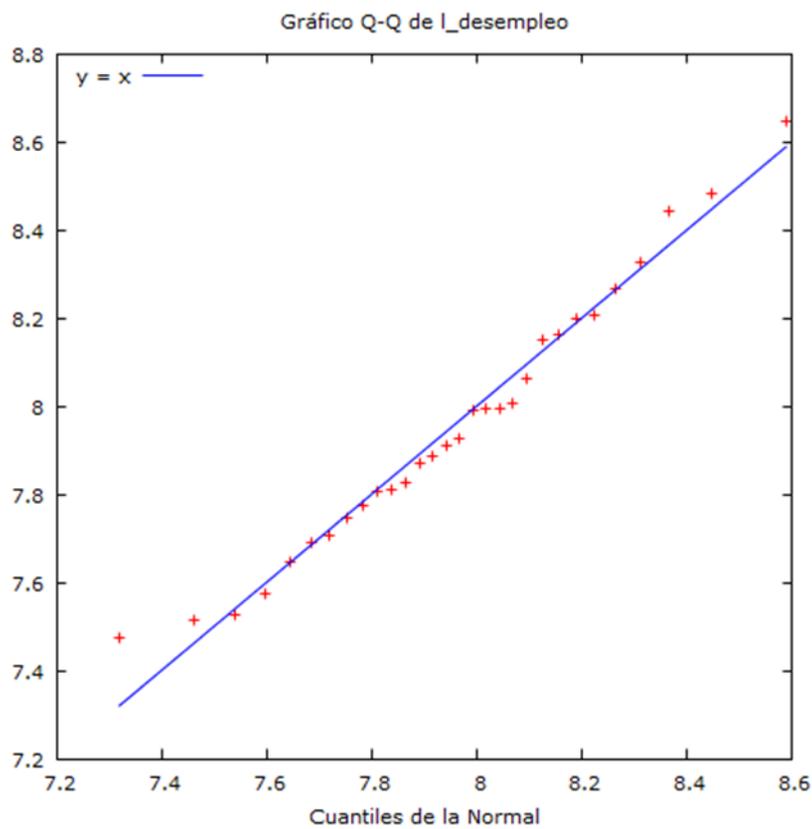
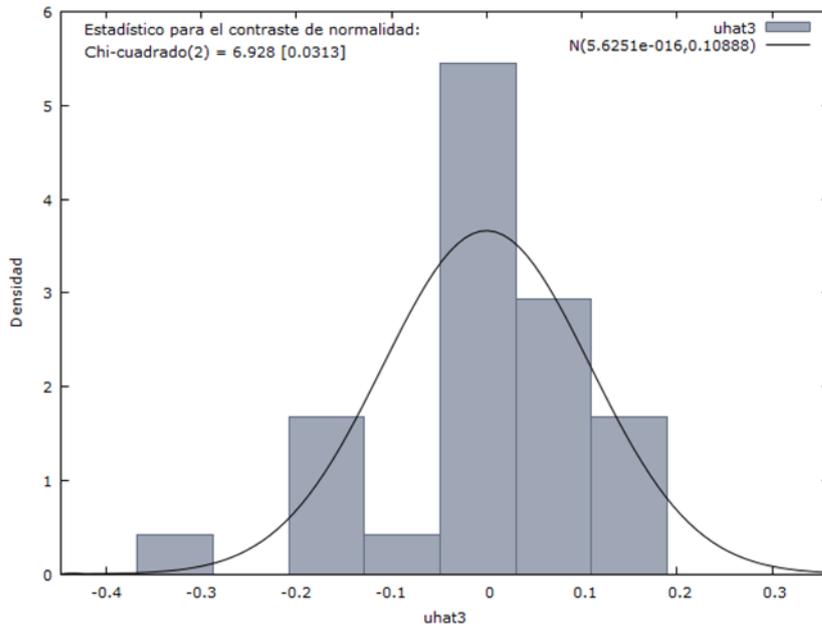
Normalidad

Para que el estimador sea eficiente, los residuos deben seguir una distribución normal.



Contraste de Jarque-Bera = 6.17588, con valor p **0.0455958**

Además de los gráficos, el contraste nos indica que no sigue una normal, por lo que tomamos logaritmos en la variable explicada, y después comprobamos si se ha solucionado el problema.



Contraste de Jarque-Bera = 1.07236, con valor p **0.584978**

Ahora sí, aceptamos la hipótesis nula, es decir, el modelo sigue una distribución normal.

Multicolinealidad.

Al realizar el contraste de multicolinealidad, nos dieron como resultados los siguientes:

Mínimo valor posible= 1	
Valores mayores que 10 pueden indicar un problema de colinealidad	
PIB	294.91
IPC	3.915
SMI	143.801
FB	50.829

Como habíamos dicho anteriormente, el modelo presenta graves problemas de multicolinealidad. Teníamos una bondad de ajuste de más del 90%, además de que 2 de 5 parámetros no son significativos, sumándole que nuestros datos son de tipo temporal, presentábamos todas las características para que apareciera este problema.

Tras probar diversas formas de solucionar la multicolinealidad, como la regresión alomada de Hoerl Kennar, optamos por eliminar el PIB. Al eliminar la variable que tiene un mayor VIF además de que no es significativa en el modelo, (PIB), los resultados obtenidos son mejores, eliminando posibles problemas de multicolinealidad.

Mínimo valor posible = 1.0	
Valores mayores que 10.0 pueden indicar un problema de colinealidad	
IPC	2.653
SMI	7.495
FB	5.187

Así nuestro nuevo modelo sería el siguiente:

$$\text{Desempleo} = \beta_1 + \beta_2 \text{IPC} + \beta_3 \text{SMI} + \beta_4 \text{FBC} + u$$

	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>Valor p</i>
const	-677.828	452.941	-14.965	0.14656
IPC	9495.28	3682.21	25.787	0.01593 **
SMI	162.303	114.183	142.143	<0.00001 ***
FB	-0.0223503	0.0015024	-148.764	<0.00001 ***
Media de la vble. dep.	2.979.533		D.T. de la vble. dep.	9.509.406
Suma de cuad. residuos	2478010		D.T. de la regresión	3.087.201
R-cuadrado	0.905507		R-cuadrado corregido	0.894604
F(3, 26)	8.305.116		Valor p (de F)	1.91e-13
Log-verosimilitud	-2.123.947		Criterio de Akaike	4.327.894
Criterio de Schwarz	4.383.942		Crit. de Hannan-Quinn	4.345.824
rho	0.582741		Durbin-Watson	0.761158

Con este modelo, todos los parámetros menos la constante, son significativos. La interpretación de los coeficientes sería:

Por cada subida porcentual del IPC, el desempleo aumenta 9495,28 personas

Por cada euro de subida en el SMI, el desempleo aumenta 162,303 personas

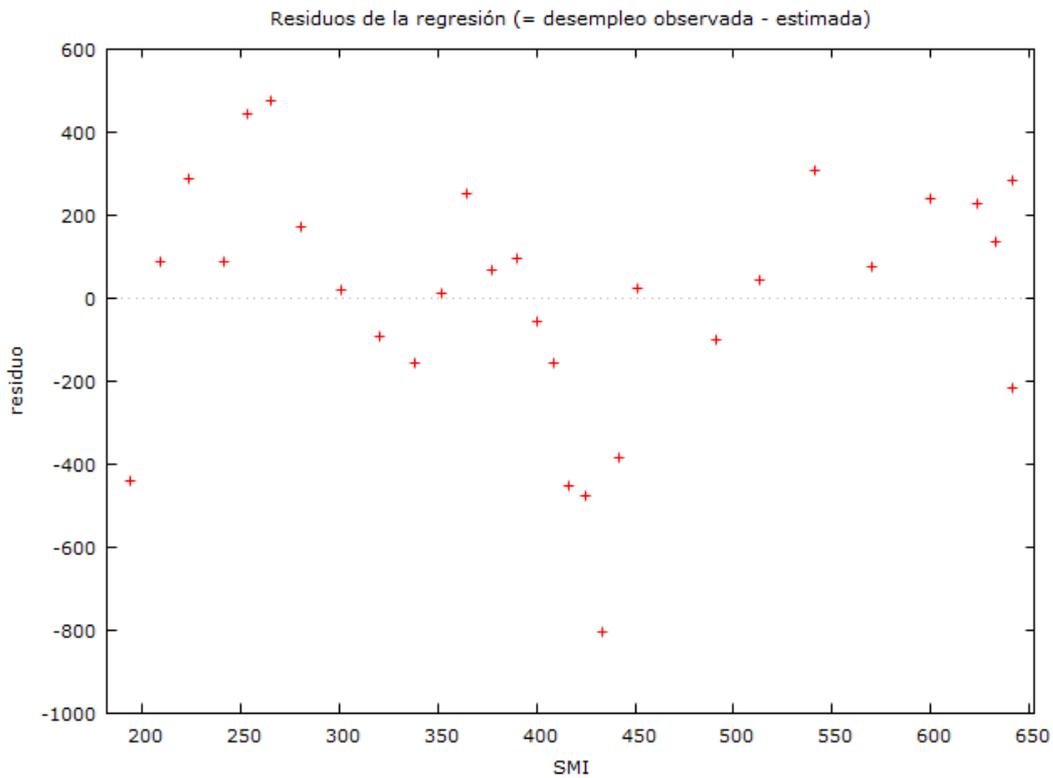
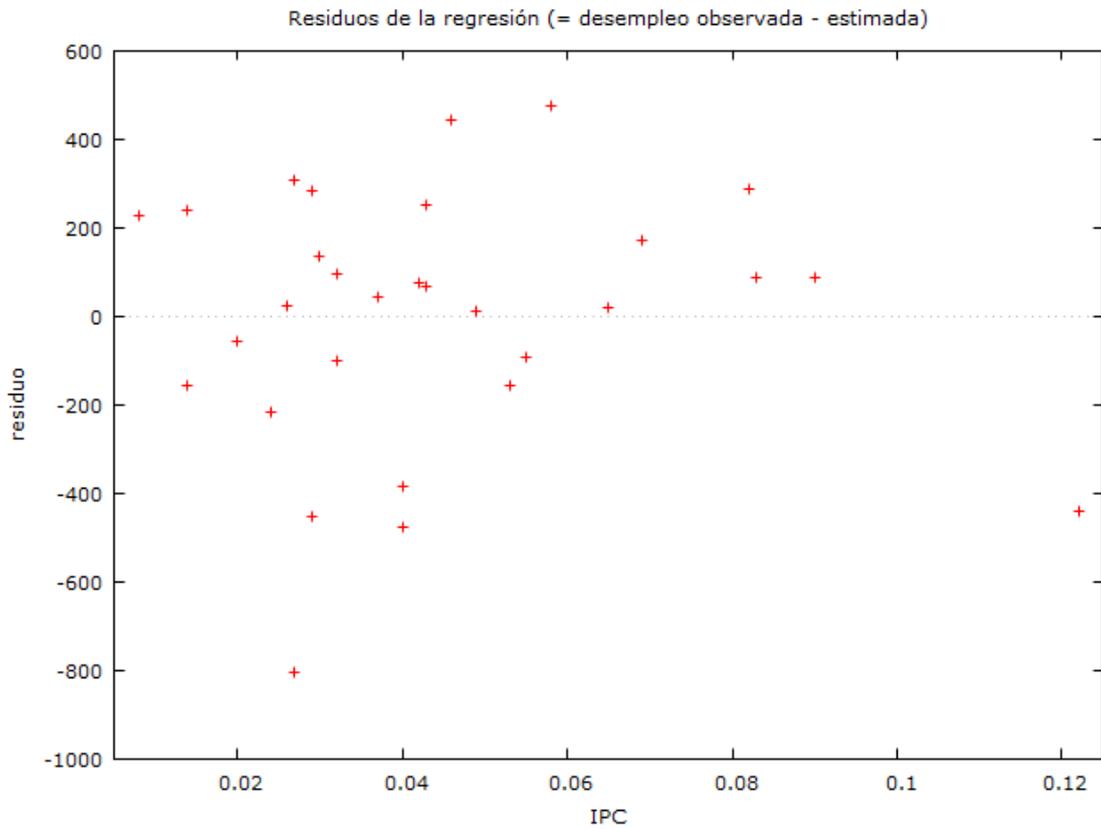
Por cada millón de euros de subida en el FB, el desempleo aumenta -0.0223503 personas

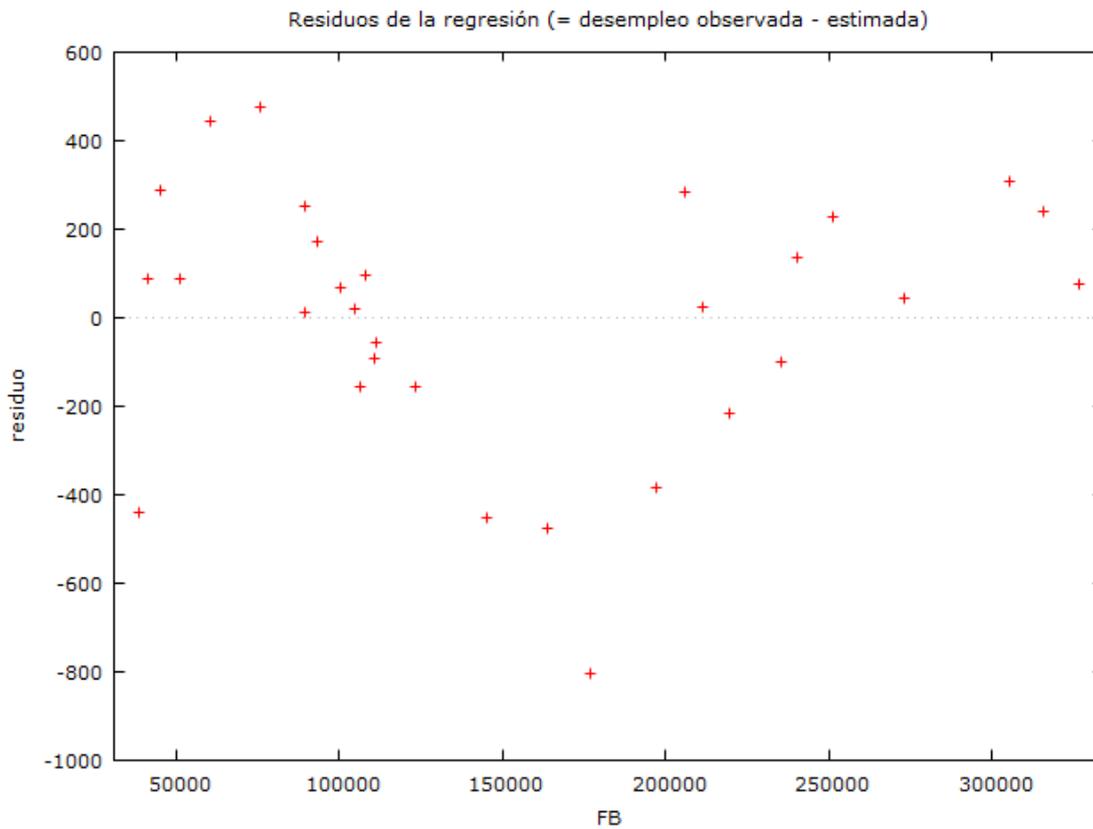
Comparando el R cuadrado corregido entre los modelos, no se puede sacar conclusiones, ya que los dos son prácticamente idénticos siendo 0,004 el nuevo.

Sin embargo, comprándolo con el criterio de Schwarz, es mejor modelo el anterior, aunque el nuevo modelo no presenta problemas de multicolinealidad.

Heterocedasticidad

A priori no deberíamos encontrarnos con este problema ya que trabajamos con datos de corte temporal





A simple vista, no encontramos signos de heterocedasticidad, vamos a comprobarlo empíricamente.

Contraste de heterocedasticidad de Breusch-Pagan -
Hipótesis nula: No hay heterocedasticidad
Estadístico de contraste: LM = 0.994851
con valor p = $P(\text{Chi-cuadrado}(3) > 0.994851) = 0.802498$

Bajo las hipótesis:

H_0 : Homocedasticidad

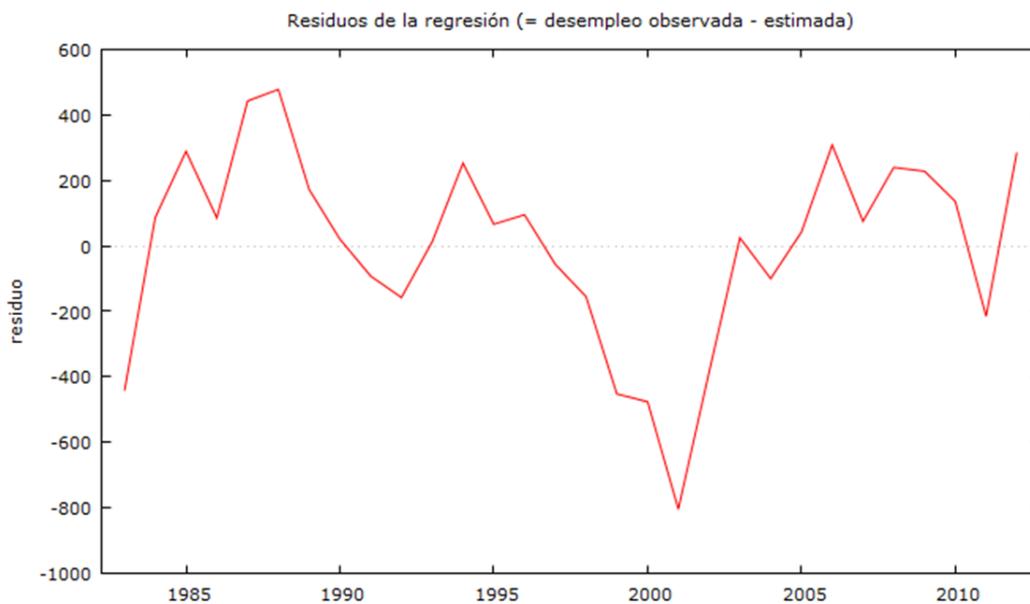
H_1 : Heterocedasticidad

Observando el p valor, aceptamos la hipótesis nula, por lo que las perturbaciones son homocedásticas.

Contraste de heterocedasticidad de White -
Hipótesis nula: No hay heterocedasticidad
Estadístico de contraste: LM = 6.68057
con valor p = $P(\text{Chi-cuadrado}(9) > 6.68057) = 0.670337$

Lo mismo ocurre al hacer el test de White. Visto que no hay problemas de heterocedasticidad, pasamos a detectar posibles problemas de autocorrelación.

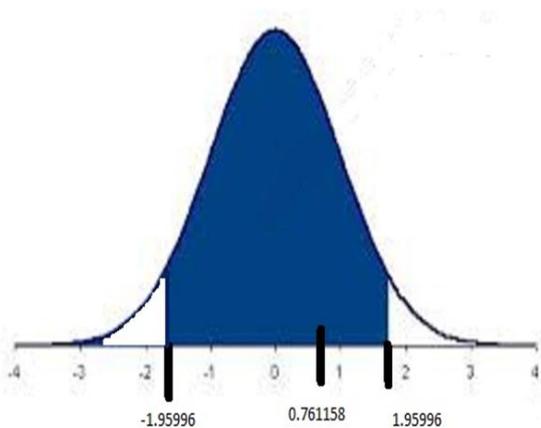
Autocorrelación.



A simple vista, no detectamos patrón que indique signos de autocorrelación.

Para contrastarlo, utilizamos la h de Durbin y la comparamos con la distribución normal y optemos el siguiente resultado:

Durbin-Watson	0.761158
----------------------	-----------------



Como la h de durbin(0,761158) cae dentro de la región de aceptación siendo H_0 no hay autocorrelación y H_1 : existe correlación, podemos aceptar H_0 y concluimos que no existe autocorrelación.

Modelo final

	<i>Coficiente</i>	<i>Desv. Típica</i>	<i>Estadístico t</i>	<i>Valor p</i>
const	-677.828	452.941	-14.965	0.14656
IPC	9495.28	3682.21	25.787	0.01593 **
SMI	162.303	114.183	142.143	<0.00001 ***
FB	-0.0223503	0.0015024	-148.764	<0.00001 ***
Media de la vble. dep.	2.979.533		D.T. de la vble. dep.	9.509.406
Suma de cuad. residuos	2478010		D.T. de la regresión	3.087.201
R-cuadrado	0.905507		R-cuadrado corregido	0.894604
F(3, 26)	8.305.116		Valor p (de F)	1.91e-13
Log-verosimilitud	-2.123.947		Criterio de Akaike	4.327.894
Criterio de Schwarz	4.383.942		Crit. de Hannan-Quinn	4.345.824
rho	0.582741		Durbin-Watson	0.761158

Después de analizar los posibles problemas, este sería nuestro modelo final.

Predicciones

Para el año 2013 con los valores de nuestras variables

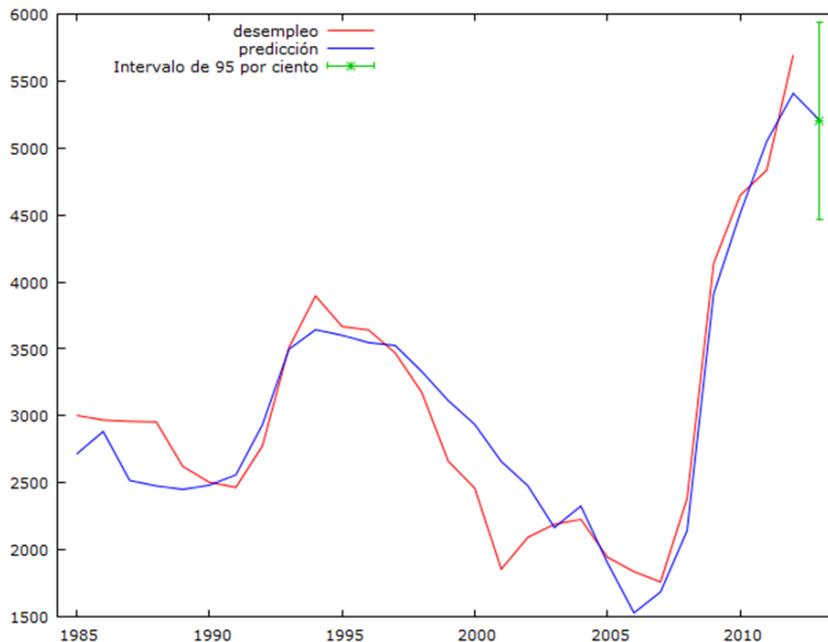
IPC	0.35
SMI	645
FB	220000

Observaciones	desempleo	predicción	Desv. Típica	Intervalo de 95%
2013	indefinido	5205.96	358.424	(4469.21, 5942.71)

El intervalo de confianza que obtenemos al 95% de confianza no incluye al dato real de desempleo que se produjo en el 2013. Esto se debe a que en el desempleo influyen muchas más variables que no hemos cuantificado en nuestro modelo. A esto, tenemos que sumarle, que los valores que hemos dado a las variables no son los del año 2013 reales.

Dicho dato es 6.200.000

Obtenemos el gráfico de predicción:



Conclusiones

Hemos conseguido un modelo con unos estimadores eficientes

Aunque hemos tenido varios problemas como el de multicolinealidad, lo hemos resuelto de forma eficaz.

Al modelo le faltan varias variables, para ser más completo.

Bibliografía

- <http://www.notasfiscales.com.mx/tasainflacion.html>
- <http://www.invertirenbolsa.info/historico-ipc-espana/datos-historicos-del-ipc-en-espana.htm>
- <http://www.invertia.com/noticias/articulo-final.asp?idNoticia=2678248>
- <http://datos.bancomundial.org/indicador/NE.GDI.TOTL.ZS?page=2>
- http://es.wikipedia.org/wiki/Anexo:Salario_m%C3%ADnimo_en_Espa%C3%B1a
- http://cvb.ehu.es/open_course_ware/castellano/social_juri/gretl/contenidos/tema-6.pdf
- <http://www.ugr.es/~jchica/Pagina2/Modelo/Modelo.htm>

- http://www.seg-social.es/Internet_1/Trabajadores/Trabajadoresdelmar/EmpleoyDesempleo/Desempleo/index.htm