Gujarati & Porter – Regresión por Mínimos Cuadrados – Ejemplos @

1. El Modelo de Asignación de Precios a los Activos de Capital (CAPM)

El objetivo de este ejercicio es ilustrar una aplicación del método de regresión por mínimos cuadrados en la esfera financiera. A tal efecto consideremos el modelo de asignación de precios a los activos de capital (CAPM, del inglés *capital asset pricing model*) de la teoría moderna de portafolios. Por *rendimiento excedente* se entiende la diferencia de rendimiento al que ofrece un activo sin riesgo, típicamente un bono del tesoro norteamericano (véase el modelo CAPM más abajo). La tabla siguiente presenta datos mensuales sobre los rendimientos excedentes Y_t (%) de un índice de 104 acciones del sector de bienes de *consumo cíclico* y los rendimientos excedentes X_t (%) del índice de *todo el mercado de valores* en el Reino Unido, correspondientes al periodo 1980-1999, para un total de 240 observaciones.[1]

OBS	Y	X	OBS	Y	X
1980:01	6.08022852	7.263448404	1984:12	3.52786616	3.191554763
1980:02	−0.924185461	6.339895504	1985:01	4.554587707	3.907838688
1980:03	−3.286174252	−9.285216834	1985:02	5.365478677	−1.708567484
1980:04	5.211976571	0.793290771	1985:03	4.525231564	0.435218492
1980:05	−16.16421111	−2.902420985	1985:04	2.944654344	0.958067845
1980:06	−1.054703649	8.613150875	1985:05	−0.268599528	1.095477375
1980:07	11.17237699	3.982062848	1985:06	−3.661040481	−6.816108909
1980:08	−11.06327551	−1.150170907	1985:07	−4.540505062	2.785054354
1980:09	−16.77699609	3.486125868	1985:08	9.195292816	3.900209023
1980:10	−7.021834032	4.329850278	1985:09	−1.894817019	−4.203004414
1980:11	−9.71684668	0.936875279	1985:10	12.00661274	5.60179802
1980:12	5.215705717	−5.202455846	1985:11	1.233987382	1.570093976
1981:01	−6.612000956	−2.082757509	1985:12	−1.446329607	−1.084427121
1981:02	4.264498443	2.728522893	1986:01	6.023618851	0.778669473
1981:03	4.916710821	0.653397106	1986:02	10.51235756	6.470651262
1981:04	22.20495946	6.436071962	1986:03	13.40071024	8.953781192
1981:05	−11.29868524	−4.259197932	1986:04	−7.796262998	−2.387761685
1981:06	−5.770507783	0.543909707	1986:05	0.211540446	−2.873838588
1981:07	−5.217764717	−0.486845933	1986:06	6.471111064	3.440269098
1981:08	16.19620175	2.843999508	1986:07	−9.037475168	−5.891053375
1981:09	−17.16995395	−16.4572142	1986:08	−5.47838091	6.375582004
1981:10	1.105334728	4.468938171	1986:09	−6.756881852	−5.734839396
1981:11	11.6853367	5.885519658	1986:10	−2.564960223	3.63088408
1981:12	−2.301451728	−0.390698164	1986:11	2.456599468	−1.31606687
1982:01	8.643728679	2.499567896	1986:12	1.476421303	3.521601216
1982:02	−11.12907503	−4.033607075	1987:01	17.0694004	8.673412896
1982:03	1.724627956	3.042525777	1987:02	7.565726727	6.914361923
1982:04	0.157879967	0.734564665	1987:03	−3.239325817	−0.460660854
1982:05	−1.875202616	2.779732288	1987:04	3.662578335	4.295976077
1982:06	−10.62481767	−5.900116576	1987:05	7.157455113	7.719692529
1982:07	−5.761135416	3.005344385	1987:06	4.774901623	3.039887622
1982:08	5.481432596	3.954990619	1987:07	4.23770166	2.510223804
1982:09	−17.02207459	2.547127067	1987:08	−0.881352219	−3.039443563
1982:10	7.625420708	4.329008106	1987:09	11.49688416	3.787092018
1982:11	−6.575721646	0.191940594	1987:10	−35.56617624	−27.86969311
1982:12	−2.372829861	−0.92167555	1987:11	−14.59137369	−9.956367094
1983:01	17.52374936	3.394682577	1987:12	14.87271664	7.975865948
1983:02	1.354655809	0.758714353	1988:01	1.748599294	3.936938398
1983:03	16.26861049	1.862073664	1988:02	−0.606016446	−0.32797064
1983:04	−6.074547158	6.797751341	1988:03	−6.078095523	−2.161544202
1983:05	−0.826650702	−1.699253628	1988:04	3.976153828	2.721787842
1983:06	3.807881996	4.092592402	1988:05	−1.050910058	−0.514825422
1983:07	0.57570091	−2.926299262	1988:06	3.317856956	3.128796482
1983:08	3.755563441	1.773424306	1988:07	0.407100105	0.181502075
1983:09	−5.365927271	−2.800815667	1988:08	−11.87932524	−7.892363786
1983:10	−3.750302815	−1.505394995	1988:09	−8.801026046	3.347081899
1983:11	4.898751703	4.18696284	1988:10	6.784211277	3.158592144
1983:12	4.379256151	1.201416981	1988:11	−10.20578119	−4.816470363
1984:01	16.56016188	6.769320788	1988:12	−6.73805381	−0.008549997
1984:02	1.523127464	−1.686027417	1989:01	12.83903643	13.46098219
1984:03	1.0206078	5.245806105	1989:02	3.302860922	−0.764474692
1984:04	−3.899307684	1.728710264	1989:03	−0.155918301	2.298491097
1984:05	−14.32501615	−7.279075595	1989:04	3.623090767	0.762074588
1984:06	3.056627177	−0.77947067	1989:05	−1.167680873	−0.495796117
1984:07	−0.02153592	−2.439634487	1989:06	−1.221603303	1.206636013
1984:08	3.355102212	8.445977813	1989:07	5.262902744	4.637026116
1984:09	0.100006778	1.221080129	1989:08	4.845013219	2.680874116
1984:10	1.691250318	2.733386772	1989:09	−5.069564838	−5.303858035
1984:11	8.20075301	5.12753329	1989:10	−13.57963526	−7.210655599

(continúa)

@ Extraído de Gujarati & Porter, Econometría, 5ª edición, 2010, punto 6, Extensiones del modelo de regresión lineal con dos variables.

[1] Estos datos, provenientes originalmente del banco de datos DataStream, se reproducen de Christiaan Heij et al., *Econometric Methods with Applications in Business and Economics*, Oxford University Press, Oxford, Reino Unido, 2004.

(continuación)

OBS	Y	X	OBS	Y	X
1989:11	1.100607603	5.350185944	1994:12	−4.225370964	0.264280259
1989:12	4.925083189	4.106245855	1995:01	−6.302392617	−2.420388431
1990:01	−2.532068851	−3.629547374	1995:02	1.27867637	0.138795213
1990:02	−6.601872876	−5.205804299	1995:03	10.90890516	3.231656585
1990:03	−1.023768943	−2.183244863	1995:04	2.497849434	2.215804682
1990:04	−7.097917266	−5.408563794	1995:05	2.891526594	3.856813589
1990:05	6.376626925	10.57599169	1995:06	−3.773000069	−0.952204306
1990:06	1.861974711	−0.338612099	1995:07	8.776288715	4.020036363
1990:07	−5.591527585	−2.21316202	1995:08	2.88256097	1.423600345
1990:08	−15.31758975	−8.476177427	1995:09	2.14691333	−0.037912571
1990:09	−10.17227358	−7.45941471	1995:10	−4.590104662	−1.17655329
1990:10	−2.217396045	−0.085887763	1995:11	−1.293255187	3.760277356
1990:11	5.974205798	5.034770534	1995:12	−4.244101531	0.434626357
1990:12	−0.857289036	−1.767714908	1996:01	6.647088904	1.906345103
1991:01	−3.780184589	0.189108456	1996:02	1.635900742	0.301898961
1991:02	20.64721437	10.38741504	1996:03	7.8581899	−0.314132324
1991:03	10.94068018	2.921913827	1996:04	0.789544896	3.034331741
1991:04	−3.145639589	0.971720188	1996:05	−0.907725397	−1.497346299
1991:05	−3.142887645	−0.4317819	1996:06	−0.392246948	−0.894676854
1991:06	−1.960866141	−3.342924986	1996:07	−1.035896351	−0.532816274
1991:07	7.330964031	5.242811509	1996:08	2.556816005	3.863737088
1991:08	7.854387926	2.880654691	1996:09	3.131830038	2.118254897
1991:09	2.539177843	−1.121472224	1996:10	−0.020947358	−0.853553262
1991:10	−1.233244642	−3.969577956	1996:11	−5.312287782	1.770340939
1991:11	−11.7460404	−5.707995062	1996:12	−5.196176326	1.702551635
1991:12	1.078226286	1.502567049	1997:01	−0.753247124	3.465753348
1992:01	5.937904622	2.599565094	1997:02	−2.474343938	1.115253221
1992:02	4.113184542	0.135881087	1997:03	2.47647802	−2.057818461
1992:03	−0.655199392	−6.146138064	1997:04	−1.119104196	3.57089955
1992:04	15.28430278	10.45736831	1997:05	3.352076269	1.953480438
1992:05	3.994517585	1.415987046	1997:06	−1.910172239	2.458700404
1992:06	−11.94450998	−8.261109424	1997:07	0.142814607	2.992341297
1992:07	−2.530701327	−3.778812167	1997:08	10.50199263	−0.457968038
1992:08	−9.842366221	−5.386818488	1997:09	12.98501943	8.111278967
1992:09	18.11573724	11.19436372	1997:10	−4.134761655	−6.967124504
1992:10	0.200950206	3.999870038	1997:11	−4.148579856	−0.155924791
1992:11	1.125853097	3.620674752	1997:12	−1.752478236	3.853283433
1992:12	7.639180786	2.887222251	1998:01	−3.349121498	7.379466014
1993:01	2.919569408	1.336746091	1998:02	14.07471304	4.299097886
1993:02	−1.062404105	1.240273846	1998:03	7.791650968	3.410780517
1993:03	1.292641409	0.407144312	1998:04	5.154679109	−0.081494993
1993:04	0.420241384	−1.734930047	1998:05	3.293686179	−1.613131159
1993:05	−2.514080553	1.111533687	1998:06	−13.25461802	−0.397288954
1993:06	0.419362276	1.354127742	1998:07	−7.714205916	−2.237365283
1993:07	4.374024535	1.943061568	1998:08	−15.26340483	−12.4631993
1993:08	1.733528075	4.961979827	1998:09	−15.22865141	−5.170734985
1993:09	−3.659808969	−1.618729936	1998:10	15.96218038	11.70544788
1993:10	5.85690764	4.215408608	1998:11	−8.684089113	−0.380200223
1993:11	−1.365550294	1.880360165	1998:12	17.13842369	4.986705187
1993:12	−1.346979017	5.826352413	1999:01	−1.468448611	2.493727994
1994:01	12.89578758	2.973540693	1999:02	8.5036	0.937105259
1994:02	−5.346700561	−5.479858563	1999:03	10.8943073	4.280082506
1994:03	−7.614726564	−5.784547088	1999:04	13.03497394	3.960824402
1994:04	10.22042923	1.157083438	1999:05	−5.654671597	−4.499198079
1994:05	−6.928422261	−6.356199493	1999:06	8.321969316	3.656745699
1994:06	−5.065919037	−0.843583888	1999:07	0.507652273	−2.503971473
1994:07	7.483498556	5.779953224	1999:08	−5.022980561	−0.121901923
1994:08	1.828762662	3.298130184	1999:09	−2.305448839	−5.388032432
1994:09	−5.69293279	−7.110010085	1999:10	−1.876879466	4.010989716
1994:10	−2.426962489	2.968005597	1999:11	1.348824769	6.265312975
1994:11	2.125100668	−1.531245158	1999:12	−2.64164938	4.045658427

La teoría moderna de composición del portafolios, en su versión de prima por riesgo, se expresa como:

$$(1) \qquad (ER_i - r_f) = \beta_i\,(ER_m - r_f)$$

donde ER_i = tasa esperada de rendimiento del título i.

ER_m = tasa esperada de rendimiento del portafolios del mercado como la representa, por ejemplo, el índice compuesto de acciones S&P 500.

r_f = tasa de rendimiento libre de riesgo, por ejemplo, el rendimiento de los bonos del Tesoro estadounidense a 90 días.

β_i = el coeficiente Beta, una medida de riesgo sistemático, es decir, el riesgo que no se ha eliminado mediante la diversificación de activos. Asimismo, es una medida del grado en el cual la i-ésima tasa de rendimiento del título se mueve con el mercado. Un $\beta_i > 1$ implica un título volátil o riesgoso, mientras que $\beta_i < 1$ es un título seguro. (Nota: No confundir esta β_i con el coeficiente de la pendiente de la regresión con dos variables.)

Si los mercados de capitales funcionan de manera eficiente, el CAPM postula que la prima esperada por el riesgo del título (= $ER_i - r_f$) es igual a ese coeficiente β del título multiplicado por la prima esperada del riesgo del mercado (= $ER_m - r_f$). Si el CAPM se mantiene se da la situación de la figura 1. La línea que aparece en la figura se conoce como línea del mercado de valores (*LMV*).

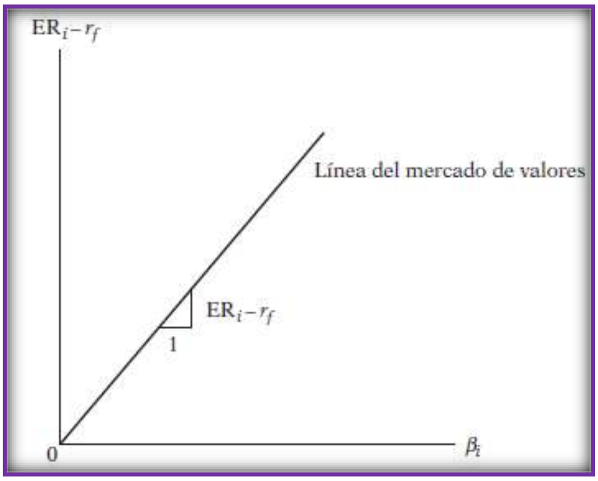

Figura 1 – Línea del mercado de valores. Riesgo sistemático

Para fines empíricos, (1) suele expresarse así:

(2) $R_i - r_f = \beta_i (R_m - r_f) + u_i$, o bien,

(3) $R_i - r_f = \alpha_i + \beta_i (R_m - r_f) + u_i$

Este último modelo se conoce como el <u>*Modelo del Mercado*</u>. Si el CAPM es válido, se espera que α_i sea cero. (Véase la figura 2.)

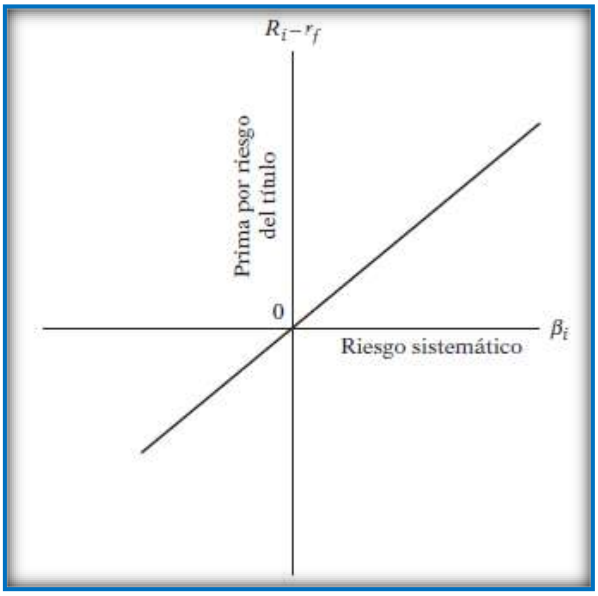

Figura 2 - El Modelo del Mercado de la teoría del portafolios (con $\alpha_i = 0$)

En primer lugar ajustamos el modelo (3) a los datos de pág.1-2. Con *EViews*$_6$ obtuvimos los siguientes resultados de regresión, que se presentan en el formato estándar de *EViews*.

Variable dependiente: Y
Método: mínimos cuadrados
Muestra: 1980 M01 1999 M12
Observaciones incluidas: 240

	Coeficiente	Error estándar	Estadístico t	Probabilidad
X	*1.155512*	*0.074396*	*15.53200*	*0.0000*

R cuadrada 0.500309	Media de la variable dependiente 0.499826
R cuadrada ajustada 0.500309	Desvío estándar de la variable dependiente 7.849594
Error estándar de regresión 5.548786	Estadístico de Durbin-Watson 1.972853
Suma de cuadrados de residuos 7 358.578	

Como muestran estos resultados, el coeficiente de la pendiente (el coeficiente Beta) es muy significativo, pues su p-valor es muy pequeño. La interpretación en este caso es que si la tasa excedente del mercado aumenta un punto porcentual, el rendimiento excedente del índice del sector de bienes de consumo aumentará alrededor de 1.15 puntos porcentuales. El coeficiente de la pendiente no es sólo estadísticamente significativo, sino que es significativamente mayor que 1 (¿pueden verificar esto?). Si un coeficiente β es mayor que 1, se dice que ese título (en este caso, un portafolios de 104 acciones, que podría ser un fondo común) es volátil; se mueve más que proporcionalmente con el índice general del mercado de valores. Sin embargo, este resultado no debe sorprender, porque en este ejemplo se consideran acciones del sector de bienes de consumo cíclico, como los bienes duraderos de uso doméstico, automóviles, textiles y equipo deportivo. Todos estos bienes suelen tener elevadas elasticidades-ingreso.

Observen que en (2) la variable dependiente, Y, es $(R_i - r_f)$, y la variable explicativa, X, es β_i, el coeficiente de volatilidad, y no $(R_m - r_f)$. Por consiguiente, para realizar la regresión (2), se debe estimar primero β_i, el cual se obtiene por lo general de la línea característica, según lo haremos a continuación.

A tal efecto, es conveniente realizar un análisis en dos etapas. *La primera etapa es una regresión de series de tiempo*. Para cada uno de los N títulos incluidos en la muestra efectuamos la siguiente regresión a través del tiempo:

$$(4) \qquad R_{it} = \alpha_i + \beta_i R_{mt} + e_{it},$$

donde R_{it} y R_{mt} son las tasas de rendimiento del i-ésimo título y del portafolios del mercado (por ejemplo, el S&P 500) en el año t; β_i, como ya vimos, es el coeficiente beta o coeficiente de volatilidad del mercado del i-ésimo título y e_{it} son los residuos. En total hay N regresiones, una para cada título, y se obtienen, por consiguiente, N valores estimados de β_i. La *línea característica* de la teoría moderna de la inversión (4), es simplemente la *regresión de la tasa de rendimiento de la acción i sobre la tasa de rendimiento del merca-*

do. El coeficiente de la pendiente, conocido como beta, es una medida de la volatilidad del rendimiento de la acción.

Etapa II *(Regresión transversal)*. En esta etapa calculamos la siguiente regresión para los N títulos:

$$R^{p}{}_{i} = gama_1 + gama_2\, \beta_i + u_i \qquad (5)$$

donde $R^{p}{}_{i}$ es el promedio o tasa media de rendimiento para el título i, calculado sobre el periodo muestral cubierto por la etapa I, β_i es el coeficiente beta estimado de la regresión de la primera etapa y u_i es el término residual.

Al comparar la regresión (5) de la segunda etapa con el CAPM, ecuación (1), escrita como

$$ER_i = r_f + \beta_i\,(ER_m - r_f), \qquad (6)$$

se aprecia que $gama_1$ es una estimación de r_f y que $gama_2$ es una estimación de $(ER_m - r_f)$, la prima del riesgo del mercado.

2. *Mortalidad Infantil y Crecimiento Económico*

Los modelos que ahora vamos a abordar se conocen como modelos *recíprocos*. Tómese por ejemplo la especificación siguiente:

$$Y_i = \beta_1 + \beta_2\left(\frac{1}{X_i}\right) + u_i \qquad (7)$$

A pesar de que este modelo es no lineal en la variable X porque entra inversamente o en forma recíproca, el modelo es lineal en β_1 y β_2, y, por consiguiente, es un modelo de regresión lineal.

Este modelo tiene las siguientes características: a medida que X aumenta indefinidamente, el término β2 (1/X) se acerca a cero (nota: β2 es una constante) e Y se aproxima al valor límite o asintótico β1. Por consiguiente, modelos como (7) contienen un valor asintótico o límite que tomará la variable dependiente cuando el valor de la variable X aumente indefinidamente.

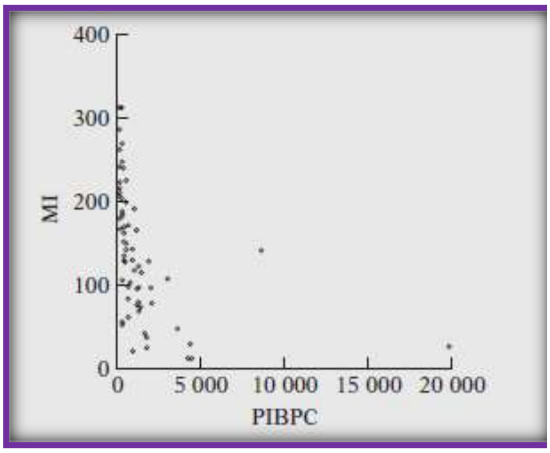

Figura 3 – Relación entre mortalidad infantil (MI) y PIB per cápita (PIBPC), en 66 países

Como ejemplo de un modelo de esta clase, consideren los datos de la tabla siguiente. Son datos transversales de 64 países sobre mortalidad infantil y otras variables. Por el momento, hay que concentrarse en las variables de mortalidad infantil (MI) y PIB per cápita (PIBPC), que se grafican en la figura 3.

Tabla de 64 países – Fecundidad, Mortalidad Infantil, y otros datos

Observación	MI	TAF	PIBPC	TFT	Observación	MI	TAF	PIBPC	TFT
1	128	37	1 870	6.66	33	142	50	8 640	7.17
2	204	22	130	6.15	34	104	62	350	6.60
3	202	16	310	7.00	35	287	31	230	7.00
4	197	65	570	6.25	36	41	66	1 620	3.91
5	96	76	2 050	3.81	37	312	11	190	6.70
6	209	26	200	6.44	38	77	88	2 090	4.20
7	170	45	670	6.19	39	142	22	900	5.43
8	240	29	300	5.89	40	262	22	230	6.50
9	241	11	120	5.89	41	215	12	140	6.25
10	55	55	290	2.36	42	246	9	330	7.10
11	75	87	1 180	3.93	43	191	31	1 010	7.10
12	129	55	900	5.99	44	182	19	300	7.00
13	24	93	1 730	3.50	45	37	88	1 730	3.46
14	165	31	1 150	7.41	46	103	35	780	5.66
15	94	77	1 160	4.21	47	67	85	1 300	4.82
16	96	80	1 270	5.00	48	143	78	930	5.00
17	148	30	580	5.27	49	83	85	690	4.74
18	98	69	660	5.21	50	223	33	200	8.49
19	161	43	420	6.50	51	240	19	450	6.50
20	118	47	1 080	6.12	52	312	21	280	6.50
21	269	17	290	6.19	53	12	79	4 430	1.69
22	189	35	270	5.05	54	52	83	270	3.25
23	126	58	560	6.16	55	79	43	1 340	7.17
24	12	81	4 240	1.80	56	61	88	670	3.52
25	167	29	240	4.75	57	168	28	410	6.09
26	135	65	430	4.10	58	28	95	4 370	2.86
27	107	87	3 020	6.66	59	121	41	1 310	4.88
28	72	63	1 420	7.28	60	115	62	1 470	3.89
29	128	49	420	8.12	61	186	45	300	6.90
30	27	63	19 830	5.23	62	47	85	3 630	4.10
31	152	84	420	5.79	63	178	45	220	6.09
32	224	23	530	6.50	64	142	67	560	7.20

Notas: MI = mortalidad infantil, el número de defunciones de niños menores de 5 años en un año por cada 1 000 nacidos vivos.
TAF = tasa de alfabetismo femenina (porcentaje).
PIBPC = PIB *per cápita* en 1980.
TFT = tasa de fecundidad total, 1980-1985, cantidad promedio de hijos por mujer, con tasas de fecundidad para edades específicas en un año determinado.
Fuente: Chandan Mukherjee, Howard White y Marc Whyte, *Econometrics and Data Analysis for Developing Countries*, Routledge, Londres, 1998, p. 456.

Si tratamos de ajustar el modelo recíproco (7), obtenemos los siguientes resultados:

$$(8) \qquad MI_t = 81{,}79436 + 27237{,}17\,(1/PBIPC)$$
$$\qquad\qquad (10{,}8321) \qquad (3759{,}999)$$
$$\qquad\qquad [7{,}5511] \qquad [7{,}2535]$$
$$R^2 = 0{,}459$$

Fueron indicados los errores estándar de los coeficientes debajo, entre paréntesis; y, entre corchetes, se ubicaron los coeficientes t-Student. Conforme el PIB per cápita se incrementa indefinidamente, ***la mortalidad infantil se acerca a su valor asintótico de casi 82 muertes por millar***. El valor positivo del coeficiente de (1/$PIBPC_t$) implica que la tasa de cambio de la MI respecto del PIBPC es negativa.

3. La Curva de Phillips

Otra aplicación importante de los modelos recíprocos es la conocida curva de Phillips en macroeconomía. Con base en los datos de tasa de variación porcentual de los salarios nominales (Y) y la tasa porcentual de desempleo (X) en el Reino Unido durante el periodo 1861 a 1957, Phillips obtuvo una curva cuya forma general se parece a la figura 4 siguiente.

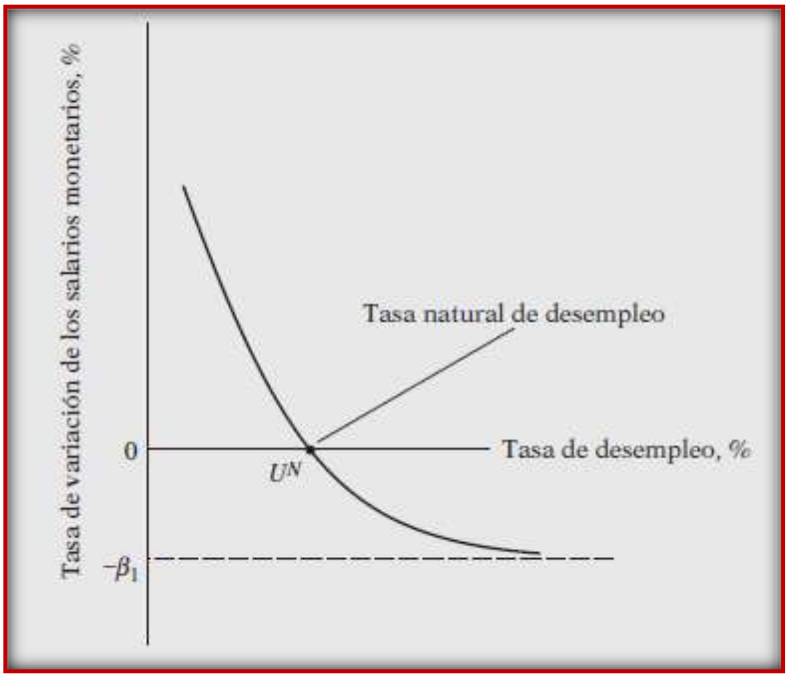

Figura 4 – Curva de Phillips

Como muestra la figura 4, existe asimetría en la respuesta de los cambios salariales al nivel de la tasa de desempleo: los salarios aumentan con mayor rapidez por cada unidad de cambio en el desempleo si la tasa de desempleo está por debajo de UN, denominada por los economistas *tasa natural de desempleo* (que es definida como la tasa de desempleo requerida para mantener constante la inflación [salarial]), y luego disminuyen suavemente por un cambio equivalente cuando la tasa de desempleo está por encima del nivel natural, UN, indicado por el piso asintótico, −β1, para el cambio salarial. Esta característica particular de la curva de Phillips puede deberse a factores institucionales, como poder de negociación de los sindicatos, salarios mínimos, compensaciones por desempleo, etcétera.

Desde la publicación del artículo de Phillips[2] se efectuó una muy extensa investigación sobre la curva de Phillips tanto a nivel teórico como empírico. La curva misma ha pasado por diversas representaciones. Una formulación comparativamente reciente la proporciona Olivier Blanchard.[3] Si π_t denota la tasa de inflación en el momento t, que se define como el cambio porcentual en el nivel de precios medido por un índice de precios representativo, como el índice de precios al consumidor (IPC), y si UN_t denota la tasa de desempleo en el tiempo t, entonces la versión moderna de la curva de Phillips se expresa según el siguiente formato:

---

[2] A. W. Phillips, *The Relationship between Unemployment and the Rate of Change of Money Wages in the United Kingdom, 1861-1957*, Economica, noviembre de 1958, vol. 25, pp. 283-299. Observen que la curva original no cruzaba el eje de la tasa de desempleo, pero la figura 4 representa una versión posterior de la curva.
[3] Olivier Blanchard, *Macroeconomics*, Prentice Hall, 1997, capítulo 17.

$$(9) \qquad \pi_t - \pi^e_t = \beta_2\,(UN_t - U^N) + u_t$$

donde π_t = tasa real de inflación en el momento t;

π^e_t = tasa de inflación esperada en el tiempo t, donde la expectativa se forma en el año (t – 1);

UN_t= tasa real de desempleo vigente en el tiempo t;

U^N = tasa natural de desempleo;

u_t= término de perturbación aleatoria.[4]

Como π^e_t no se puede observar de manera directa, en primer lugar se trabaja con el supuesto de que $\pi^e_t= \pi_{t-1}$; es decir, la inflación esperada este año es la tasa de inflación que prevaleció el año anterior; por supuesto, se pueden formular supuestos más complicados respecto de la formación de expectativas.

Al sustituir este supuesto en (9) y escribir el modelo de regresión en la forma estándar, obtenemos la siguiente ecuación de estimación:

$$(10) \qquad \pi_t - \pi_{t-1} = \beta_1 + \beta_2\, UN_t + u_t$$

donde β_1=-β_2 U^N. La ecuación (10) establece que el cambio en la tasa de inflación entre los dos periodos está linealmente relacionado con la tasa de desempleo real. A priori, se espera que β_2 sea negativa (¿por qué?) y β_1 positiva (porque β_2 es negativa y *UN* es positiva).

A manera de ilustración de la curva de Phillips modificada, en la tabla siguiente se presentan datos sobre la inflación medida por el porcentaje anual en el índice de precios al consumidor (IPC) y la tasa de desempleo de 1960 a 2006 en Estados Unidos. La tasa de desempleo representa la tasa de desempleo "civil". Con estos datos se obtuvo el cambio en la tasa de inflación (π_t – π_{t-1}) y se graficó respecto de la tasa de desempleo civil; se utiliza el IPC como medida de la inflación. La gráfica resultante aparece en la figura 5.

Como se esperaba, la relación entre el cambio en la tasa de inflación y la tasa de desempleo es negativa (un desempleo bajo provoca un incremento en la tasa de inflación y, por consiguiente, una aceleración del nivel de precios, de ahí el nombre de curva aceleradora de Phillips).

Al observar la figura 5, no resulta obvio si un modelo de regresión lineal (una línea recta) o un modelo recíproco sea el que se ajuste a los datos; tal vez haya una relación curvilínea entre las dos variables. Más adelante se presentan regresiones basadas en ambos modelos. Sin embargo, hay que tener presente que para el *modelo recíproco* (7) se espera que el término de la ordenada al origen sea negativo y la pendiente positiva.

[4] Los economistas creen que este término de error representa algún tipo de shock de oferta, como los embargos de petróleo de la OPEP en 1973 y 1979.

Tabla - Tasas de inflación y desempleo, Estados Unidos, 1960-2006
(Todos los consumidores urbanos; 1982-1984 = 100)

Año	TSINF	TSDES	Año	TSINF	TSDES
1960	1.718	5.5	1984	4.317	7.5
1961	1.014	6.7	1985	3.561	7.2
1962	1.003	5.5	1986	1.859	7.0
1963	1.325	5.7	1987	3.650	6.2
1964	1.307	5.2	1988	4.137	5.5
1965	1.613	4.5	1989	4.818	5.3
1966	2.857	3.8	1990	5.403	5.6
1967	3.086	3.8	1991	4.208	6.8
1968	4.192	3.6	1992	3.010	7.5
1969	5.460	3.5	1993	2.994	6.9
1970	5.722	4.9	1994	2.561	6.1
1971	4.381	5.9	1995	2.834	5.6
1972	3.210	5.6	1996	2.953	5.4
1973	6.220	4.9	1997	2.294	4.9
1974	11.036	5.6	1998	1.558	4.5
1975	9.128	8.5	1999	2.209	4.2
1976	5.762	7.7	2000	3.361	4.0
1977	6.503	7.1	2001	2.846	4.7
1978	7.591	6.1	2002	1.581	5.8
1979	11.350	5.8	2003	2.279	6.0
1980	13.499	7.1	2004	2.663	5.5
1981	10.316	7.6	2005	3.388	5.1
1982	6.161	9.7	2006	3.226	4.6
1983	3.212	9.6			

Nota: La tasa de inflación es el cambio porcentual anual en el IPC. La tasa de desempleo es la tasa de desempleo civil.

Modelo lineal: (11) $(\pi_t - \pi_{t-1}) = 3{,}7844 - 0{,}6385\, UN_t$

t (4,1912) (-4,2756) $R^2=0{,}2935$

Modelo recíproco: (12) $(\pi_t - \pi_{t-1}) = -3{,}0684 + 17{,}2077\,(1/UN_t)$

t (-3,1635) (3,2886) $R^2=0{,}1973$

Todos los coeficientes estimados en ambos modelos son significativos estadística e individualmente; además, todos los p-valores son inferiores a 0.005.

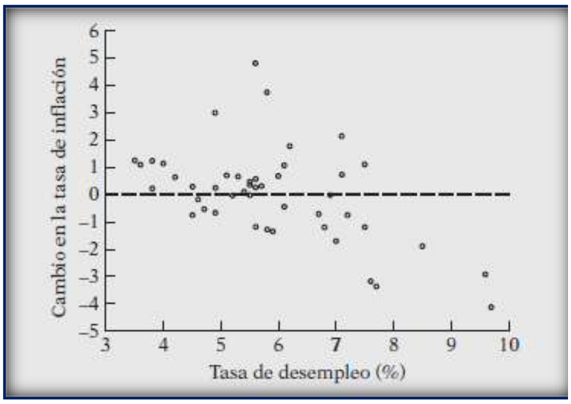

Figura 5 - Curva de Phillips modificada

El modelo (11) muestra que si la tasa de desempleo baja un punto porcentual, en promedio, el cambio en la tasa de inflación aumenta 0.64 puntos porcentuales, y viceversa. El modelo (12) revela que, aunque la tasa de desempleo se incrementara de manera indefinida, el máximo cambio en la tasa de inflación bajaría y sería de 3.07 puntos porcentuales. A propósito, de la ecuación (11) se puede calcular la *tasa de desempleo natural subyacente*, de la siguiente forma:

$$(13) \qquad U^N = \beta_1/{-\beta_2} = 3{,}7844/0{,}6385 = 5{,}9270.$$

Es decir, la tasa de desempleo natural es de casi 5.93%. Los economistas sitúan la tasa natural entre 5 y 6%, aunque en años recientes la tasa real de desempleo en Estados Unidos ha sido mucho más baja.

4. *Mortalidad infantil en relación con el PIB per cápita y la tasa de alfabetización de las mujeres*

Analizaremos ahora un modelo de regresión con más de una variable explicativa. Volvamos a los datos de la Sección 2 previa. Vimos que el PIBPC ejerce un impacto negativo sobre la MI, como era de esperarse. Ahora se presenta el alfabetismo femenino medido por la tasa de alfabetización de las mujeres (TAM). A priori, se espera que la TAM también ejerza un impacto negativo sobre la MI. Ahora, cuando se introducen ambas variables en el modelo, se requiere eliminar la influencia neta de cada regresora. Es decir, se requiere estimar los coeficientes de regresión (parcial) de cada regresora. Por tanto, el modelo es:

$$(14) \qquad MI_i = \beta_1 + \beta_2\, PIBPC_i + \beta_3\, TAM_i + u_i$$

Los datos necesarios se proporcionan en la tabla de la página 6. Tengan en cuenta que la MI es el número de fallecimientos de niños menores de 5 años por cada 1 000 nacidos vivos, el PIBPC es el PIB per cápita en 1980 y la TAM se mide en porcentaje. La muestra se realizó en 64 países.

Con el paquete estadístico *EViews$_6$*, para la parte sistemática de (14) se obtienen los siguientes resultados:

$$(15) \qquad \begin{aligned} MI_i &= \underset{(11{,}5932)}{263{,}6416} - \underset{(0{,}0019)}{0{,}0056}\, PIBPC_i - \underset{(0{,}2099)}{2{,}2316}\, TAM_i \\ R^2 &= 0{,}7077 \\ R^2_{aj} &= 0{,}6981 \end{aligned}$$

Las cifras entre paréntesis son los errores estándar estimados de los coeficientes de regresión. Ahora interpretemos estos coeficientes de regresión: −0.0056 es el coeficiente de regresión parcial del PIBPC e indica que, si se mantiene constante la influencia de la TAM, conforme el PIBPC se incrementa, por ejemplo en un dólar en promedio, la mortalidad infantil disminuye en 0.0056 unidades. Para interpretar esto desde el punto de vista económico, si el PIB per cápita se incrementara 1 000 dólares, en promedio, el número de muertes de niños menores de 5 años se reduciría en 5.6 por cada 1 000 nacidos vivos. El coeficiente −2.2316 señala que si la influencia del PIBPC se mantiene constante, el número de muertes de niños menores de 5 años disminuiría, en promedio, 2.23 por cada 1000 nacimientos vivos, si la tasa de alfabetización en las mujeres subiera un punto porcentual. El valor de la ordenada al origen de alrededor de 263, si se interpretara de una forma mecanicista, significaría que si los valores del PIBPC y de la TAM fuesen cero, la mortalidad infantil promedio sería de más o menos 263 muertes por cada 1 000 nacimidos vivos. Por supuesto, tal interpretación debe tomarse con mucho cuidado. *(Tal prudencia es aconsejable cuando se quiere predecir el valor que alcanzaría la variable endógena si las variables exógenas adoptaran valores muy alejados de los alcanzados en la muestra.)* Cualquiera puede inferir que si los dos regresoras tuviesen un valor cero, la mortalidad infantil sería muy alta, lo cual tiene sentido. El valor de R^2 de casi 0.71 significa que casi 71% de la variación en la mortalidad infantil se explica mediante el PIBPC y la TAM, lo cual es un gran porcentaje si se considera que el valor máximo que puede tener R^2 es 1. De todo lo dicho hasta aquí, los resultados de la regresión tienen sentido.

La R^2 ajustada En la ecuación (15) se ha incluido otro estadístico que suele ser usado en los ejercicios de regresión múltiple: la *R^2 ajustada o R^2_{aj}.* Se trata de un coeficiente de determinación alternativo. Tengan en cuenta que una propiedad importante de R^2 es que es una función no decreciente del número de variables explicativas o de regresoras presentes en el modelo; a medida que aumenta el número de regresoras, R^2 aumenta casi invariablemente y nunca disminuye. En vista de esto, al comparar dos modelos de regresión con la misma variable dependiente pero un número diferente de variables X, se debe tener mucho cuidado al escoger el modelo con la R^2 más alta. Para comparar dos términos R^2 se debe tener en cuenta el número de variables X presentes en el modelo. Esto se verifica con facilidad si consideramos un coeficiente de determinación alterno, que es el siguiente:

$$R^2_{aj} = 1 - [\sum u_i^2 / (n - k)/(\sum y_i^2(n - 1)] \qquad (16)$$

donde k = el número de parámetros en el modelo incluyendo el término de la ordenada al origen. (En la regresión con tres variables, k = 3. ¿Por qué?) R^2 definida así se conoce como R^2 ajustada, designada por R^2_{aj}. El término ajustado significa ajustado por los grados de libertad asociados a las sumas de cuadrados que se consideran en (16): $\sum u_i^2$ tiene $n - k$ grados de libertad en un modelo con k parámetros, incluyendo la ordenada al origen, y $\sum u_i^2$ tiene $n-1$ grados de libertad. (¿Por qué?) Para el caso de una ecuación de regresión de tres variables, sabemos que $\sum u_i^2$ tiene $n - 3$ grados de libertad. La relación entre R^2 y R^2_{aj} es simple de ver, ya que puede demostrarse que

$$R^2_{aj} = 1 - (1-R^2)\,[(n-1)/(n-k)] \qquad (17)$$

Por esta ecuación de inmediato se comprende que 1) para k > 1, $R^2_{aj} < R^2$, lo cual implica que, a medida que aumenta el número de variables X, R^2 ajustada aumenta menos que R^2 no ajustada; y que 2) R^2_{aj} puede ser negativa, aunque R^2 es necesariamente no negativa. En caso de que R^2_{aj} resulte ser negativa en una aplicación, su valor se toma como cero.

El econometrista holandés Henri Theil, en sus *Principios de Econometría,* sostuvo que *es una buena costumbre utilizar R^2_{aj} en lugar de R^2 porque R^2 tiende a dar una imagen demasiado optimista del ajuste de la regresión, en particular cuando el número de variables explicativas no es muy pequeño comparado con el de observaciones*.

Comparación de dos valores de R^2 Es de crucial importancia señalar que, al comparar dos modelos con base en el coeficiente de determinación, ajustado o no, el tamaño de la muestra n y la variable dependiente deben ser los mismos; las variables explicativas pueden adoptar cualquier forma. Así, para los modelos

$$ln\, Y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + u_i \qquad (18a)$$

$$Yi = \alpha_1 + \alpha_2 X_{2i} + \alpha_3 X_{3i} + u_i \qquad (18b)$$

no pueden compararse los términos R^2 calculados. La razón es la siguiente: por definición, R^2 mide la proporción de la variación en la variable dependiente explicada por la(s) variable(s) explicativa(s). Por consiguiente, en (18a), el R^2 mide la proporción de la variación en *ln Y* explicada por X_2 y X_3, mientras que en (18b), mide la proporción de la variación en *Y*, y las dos no son la misma variable: un cambio en *ln Y* da lugar a un cambio relativo o proporcional en Y, mientras que un cambio en Y da un cambio absoluto.[5] Por consiguiente, $var\ Y^{aj}_i / var\ Y_i$ no es igual a $var\ (ln\ Y^{aj}_i)/var\ (ln\ Y_i)$; es decir, los dos coeficientes de determinación no son lo mismo. Un ejemplo puede explicar mejor esta diferencia.

Consumo de Café en Estados Unidos

Consideren los datos de la tabla siguiente, los cuales se refieren al consumo de tazas de café por día (Y) y el precio al menudeo del café (X) en Estados Unidos de 1970 a 1980. Al aplicar mínimos cuadrados ordinarios a los datos se obtienen los resultados de la regresión de la ecuación (19), de la que se escribe la parte sistemática (entre paréntesis, los errores estándar de los coeficientes).

Tabla – Consumo de Café en Estados Unidos (Y) respecto al Precio promedio real al menudeo (X*) – 1970-1980

Año	Y, Tazas diarias por persona	X, $ por libra
1970	2.57	0.77
1971	2.50	0.74
1972	2.35	0.72
1973	2.30	0.73
1974	2.25	0.76
1975	2.20	0.75
1976	2.11	1.08
1977	1.94	1.81
1978	1.97	1.39
1979	2.06	1.20
1980	2.02	1.17

** Nota: El precio nominal fue dividido por el IPC de alimentos y bebidas, 1967=100.*
Fuente: Summary of National Coffee Drinking Study, Data Group, Elkins Park, Pensilvania, 1981, y los datos sobre X nominal (es decir, X a precios corrientes), de Nielsen Food Index, A. C. Nielsen, Nueva York, 1981.

$$Y_t = 2.6911 - 0.4795\, X_t \qquad (19)$$
$$(0.1216)\ (0.1140) \qquad SCR=0.1491 \quad R^2=0{,}663$$

Los resultados tienen sentido en el contexto económico: conforme se incrementa el precio del café, en promedio, su consumo disminuye casi media taza al día. El valor R^2 de 0.66 indica que el precio del café explica casi 66% de la variación en el consumo del café. El lector puede verificar con facilidad que el coeficiente del precio es estadísticamente significativo. Nota: SCR es la *suma de los cuadrados de los residuos de la regresión.*

[5] Observar el vínculo que tienen los logaritmos y los porcentajes: como $d(\ln X)/dX = 1/X$, o $d(\ln X)=dX/X$, para cambios muy pequeños, el cambio en ln X es igual al cambio relativo o proporcional en X. En la práctica, si el cambio en X es razonablemente pequeño, la relación anterior se escribe como el cambio en ln X ≈ al cambio relativo en X, donde ≈ significa aproximadamente igual.

A partir de los mismos datos, fue estimado el siguiente modelo doble logarítmico, o con elasticidad constante:

$$(20) \qquad \ln Y_t = 0.7774 - 0.2530 \ln X_t$$
$$(0.0152) \; (0.0494) \qquad SCR = 0.0226 \; R^2 = 0.745$$

Como es un modelo doble-logarítmico, el coeficiente de la pendiente proporciona un estimador directo del coeficiente de elasticidad-precio. En el ejemplo presente, indica que si el precio del café por libra se incrementa 1%, en promedio, su consumo diario disminuye casi 0.25%. Recuérdese que, en el modelo lineal (19), el coeficiente de la pendiente sólo señala la tasa de cambio del consumo del café respecto del precio. (¿Cómo estimaría la elasticidad-precio en el modelo lineal?) El valor R^2 de casi 0.745 significa que 74,5% de la variación en el logaritmo de la demanda de café se explica por la variación en el logaritmo del precio del café.

Como el valor R^2 (0.663) del modelo lineal es menor que el valor R^2 de 0.745 del modelo lineal logarítmico, se presenta la tentación de elegir este último modelo debido al alto valor de R^2. Sin embargo, por las razones expresadas, no es posible hacerlo así. No obstante, si se desea comparar ambos valores R^2, puede procederse de la siguiente manera:

1. Obtenga $\ln Y_t$ de (20) para cada observación; es decir, encuentre el valor estimado de cada observación a partir de este modelo. Tome el antilogaritmo de esos valores y después calcule R^2 entre dichos valores del antilogaritmo y la verdadera Y_t de la siguiente manera:

$$(21) \qquad R^2 = (\sum Y_t \, Y_t^{aj})^2 \,/\, (\sum Y_t)\,(\sum Y_t^{aj})$$

En esta expresión, Y_t = Y real; Y_t^{aj}= Y estimada (o ajustada). Este valor R^2 es comparable con el valor R^2 del modelo lineal (19).

2. Si los valores Y son positivos, otra forma consiste en calcular los logaritmos de los valores Y, $\ln Y$. Obtenga los valores estimados Y, Y_t^{aj} , del modelo lineal (19), calcule los logaritmos de dichos valores estimados Y (es decir, $\ln Y_t^{aj}$) y calcule la R^2 entre ($\ln Y_t$) y ($\ln Y_t^{aj}$) como en la ecuación (21). Este valor R^2 es comparable con el valor R^2 obtenido mediante (20).

Para el ejemplo del café, en la tabla siguiente se presentan los datos originales necesarios para calcular las R^2 comparables. A fin de comparar el valor R^2 del modelo lineal (19) con el de (20), primero obtenemos el logaritmo de (Y_t^{aj}) [dado en la columna (6) de la tabla], luego calculamos el logaritmo de los valores reales Y [dados en la columna (5) de la tabla] y por último calculamos R^2 entre estos dos conjuntos de valores mediante la ecuación (21). El resultado es un valor R^2 de 0.6779, el cual ahora se puede comparar con el valor R^2 de 0.7448 del modelo log-lineal. La diferencia entre ambos valores R^2 es aproximadamente 0.07.

Por otra parte, si deseamos comparar el valor R^2 del modelo log-lineal con el obtenido del modelo lineal, estimamos $\ln Y_t$ para cada observación de (20) [dadas en la columna (3) de la tabla], obtenemos sus valores antilog [dados en la columna (4) de la tabla] y por último calculamos R^2 entre estos valores antilog y los valores reales de Y observados mediante la fórmula (21). Esto da a R^2 un valor de 0.7187, el cual es un poco superior al valor obtenido del modelo lineal (19) de 0.663.

Ergo, con cualquier método, parece que el modelo log-lineal ofrece un ajuste ligeramente superior.

Tabla - Datos básicos para comparar dos valores de R^2

Año	Y_t (1)	$\hat{Y}_t$ (2)	$\widehat{\ln Y_t}$ (3)	Antilog de $\widehat{\ln Y_t}$ (4)	$\ln Y_t$ (5)	$\ln(\hat{Y}_t)$ (6)
1970	2.57	2.321887	0.843555	2.324616	0.943906	0.842380
1971	2.50	2.336272	0.853611	2.348111	0.916291	0.848557
1972	2.35	2.345863	0.860544	2.364447	0.854415	0.852653
1973	2.30	2.341068	0.857054	2.356209	0.832909	0.850607
1974	2.25	2.326682	0.846863	2.332318	0.810930	0.844443
1975	2.20	2.331477	0.850214	2.340149	0.788457	0.846502
1976	2.11	2.173233	0.757943	2.133882	0.746688	0.776216
1977	1.94	1.823176	0.627279	1.872508	0.662688	0.600580
1978	1.97	2.024579	0.694089	2.001884	0.678034	0.705362
1979	2.06	2.115689	0.731282	2.077742	0.722706	0.749381
1980	2.02	2.130075	0.737688	2.091096	0.703098	0.756157

Notas. Columna (1): Valores reales de Y de la tabla de página 12.
Columna (2): Valores estimados de Y del modelo lineal (19).
Columna (3): Valores estimados de log Y del modelo doble-log (20).
Columna (4): Antilog de valores de la columna (3).
Columna (5): Valores log de Y en la columna (1).
Columna (6): Valores log de Y estimado en la columna (2).