ESTADÍSTICA ESPAÑOLA
núm. 108, 1985, págs. 5 a 17

Nuevo procedimiento para el ajuste de la curva logística: aplicación a la población española

por ERNESTO VERES FERRER
Consellería de Industria,
Comercio y Turismo.
VALENCIA

RESUMEN

En este artículo se presenta un procedimiento de ajuste de la curva logística, basado en el conocimiento previo del punto de inflexión. Dicho procedimiento proporciona como estimación las asíntotas superior e inferior de la curva. Se finaliza el trabajo con su aplicación a la población española.

Palabras clave: Ajuste; curva logística; punto de inflexión; tasa de crecimiento intercensal.

El presente trabajo expone un nuevo procedimiento de ajuste de la curva logística a un conjunto de datos que, empíricamente, se supone siguen tal tendencia. Así pues, a los métodos ya clásicos —el de Verhulst y el de Pearl y Reed— añadimos éste que se basa, primordialmente, en el conocimiento previo del punto de inflexión de la curva. De esta forma, en la tradicional expresión de la logística

$$y = y_o + \frac{K}{1 + C.\ e^{-ht}} \quad h > 0 \tag{1}$$

ya no será necesario estimar con anterioridad el parámetro y_o (ajuste de Verhulst) o los parámetros y_o y K (ajuste de Pearl y Reed), sino que éstos vendrán dados por el mismo proceso de ajuste. En efecto, el conocimiento del punto de inflexión lleva consecuentemente a que los parámetros desconocidos a estimar por el ajuste queden reducidos a dos (K y h):

$$y = y_I + \frac{K}{2} \cdot \frac{1 - e^{h(t_I - t)}}{1 + e^{h(t_I - t)}} \tag{2}$$

donde el par (t_I, y_I) representa las coordenadas del punto de inflexión conocido de antemano. Finalizamos el artículo aplicando el procedimiento para estimar la logística del actual ciclo de evolución de la población española.

1. AJUSTE DE LA LOGISTICA SIENDO CONOCIDO EL PUNTO DE INFLEXIÓN

Sea (t_i, y_i) i=1, 2, ..., n un conjunto de pares de valores crecientes en sus dos componentes y que siguen una tendencia aproximadamente logística que se pretende estimar. Sea (t_I, y_I) el punto de inflexión conocido para la misma, de coordenadas

$$t_I = \ln C / h \qquad y_I = y_o + \frac{K}{2} \tag{3}$$

Haciendo pasar la curva (1) por el punto de coordenadas (3) resulta finalmente la expresión (2). Esta expresión presenta como asíntotas

$$y_I - \frac{K}{2} \qquad y_I + \frac{K}{2},$$

es creciente y tiene como punto de inflexión —que actúa como centro de simetría— el ya conocido (t_I, y_I).

Para efectuar el ajuste deben considerarse dos puntos (t_I, y_I) y (t_2, y_2) cuyas abcisas t_I y t_2 estén —en valor absoluto— una doble distancia de la otra respecto de la abcisa t_I. En definitiva, deberán verificar la condición

$$| t_I - t_I | = 2 | t_2 - t_I |$$

donde supondremos, en lo que sigue, que t_2 es la abcisa más cercana a t_I. La decisión sobre si los dos puntos elegidos deben quedar ambos al mismo lado del punto de inflexión o dejar a éste en una posición central es irrelevante a los efectos teóricos del ajuste, lo que le proporciona gran versatilidad según el conjunto real de datos de los que se disponga. Sin embargo, en la medida de lo posible y para las aplicaciones, resultará siempre interesante que el punto de inflexión quede situado en el centro, a fin de recoger la información sobre el comportamiento de las dos ramas de la curva. En todo caso —insistimos— la elección deberá tomarse según la estructura de los datos disponibles.

Describamos el proceso de ajuste:

a) *Estimación del parámetro h.* Realicemos un cambio de origen —el nuevo tendrá como abcisa la del punto de inflexión, manteniendo la misma ordenada— y de medida, según

$$t' = (t_I - t) \,/\, |\, t_I - t_2 \,|$$

Con este cambio, la logística (2) admite como nueva expresión:

$$y - y_I = \frac{K}{2} \cdot \frac{1 - e^{h' t'}}{1 + e^{h' t'}} \qquad (4)$$

con $h' = |\, t_I - t_2 \,|.h$. Haciendo pasar (4) por los puntos (t_1, y_1) y (t_2, y_2) se obtienen las dos condiciones siguientes:

$$y_2 - y_I = \frac{K}{2} \cdot \frac{1 - e^{sig\,(t_I - t_2).h'}}{1 + e^{sig\,(t_I - t_2).h'}}$$

$$y_1 - y_I = \frac{K}{2} \cdot \frac{1 - e^{sig\,(t_I - t_1).2h'}}{1 + e^{sig\,(t_I - t_1).2h'}}$$

esto es,

$$y_2 - y_I = \frac{K}{2} \,.\, sig\,(t_I - t_2) \,.\, \frac{1 - e^{h'}}{1 + e^{h'}}$$

$$y_1 - y_I = \frac{K}{2} \,.\, sig\,(t_I - t_1) \,.\, \frac{(1 - e^{h'})\ (1 + e^{h'})}{1 + e^{2h'}}$$

donde *"sig"* es la función signo. Dividiendo miembro a miembro las dos condiciones anteriores se simplifica el parámetro K, quedando una única función en el parámetro *h*:

$$L = \frac{|\, y_2 - y_I \,|}{|\, y_1 - y_I \,|} = \frac{1 + e^{2h'}}{1 + 2e^{h'} + e^{2h'}}$$

que a su vez se transforma en una ecuación de segundo grado en $a = e^{h'}$:

$$(L - 1).a^2 + 2L.a + (L - 1) = 0 \tag{5}$$

cuyo discriminante es no negativo siempre que

$$2 \mid y_2 - y_1 \mid > \mid y_1 - y_1 \mid \tag{6}$$

Esta última condición —siempre que los datos a ajustar sigan una tendencia aproximada a la logística— será siempre satisfecha por los puntos (t_i, y_i) i=1, 2, por misma definición de (1). En todo caso, bastará elegir un valor para t_2 lo suficientemente alejado de t_1 para asegurar el cumplimiento de (6).

La ecuación (5) proporciona dos soluciones:

$$a_{1,2} = \frac{-L \pm \sqrt{2L - 1}}{L - 1}$$

El hecho de que los valores (t_i, y_i), i=1, 2, ..., n sean crecientes y estén sobre la logística (4) implica el que $L < 1$. Por tanto, $a_2 > 1$ y $a_1 < 1$, de forma que sólo debe considerarse como solución a_2, toda vez que $h' = \ln a$ debe ser esencialmente positivo. Finalmente, pues, se obtendrá una estimación para h según:

$$h = \ln a_2 / \mid t_1 - t_2 \mid \tag{7}$$

b) *Estimación del parámetro K.* A fin de involucrar a todos los puntos(t_i, y_i), i=1, 2, ..., n —y no sólo a los (t_1, y_1) y (t_2, y_2) considerados hasta ahora— se procede a la obtención del parámetro K mediante una estimación mínimo-cuadrática. Para ello —y según (2)— basta obtener K tras un ajuste por mínimos cuadrados de una recta por el origen de pendiente K, en la nube de puntos de coordenadas

$$\left(\frac{1 - e^{h(t_1 - t_i)}}{1 + e^{h(t_1 - t_i)}}, 2 (y_i - y_1) \right) \qquad i=1, 2, \ldots, n$$

y en donde h viene dada por la estimación (7). Se obtiene, pues, finalmente que

$$K = \sum_{i=1}^{n} 2 (y_i - y_1). \frac{1 - e^{h(t_1 - t_i)}}{1 + e^{h(t_1 - t_i)}} \Big/ \sum_{i=1}^{n} \left(\frac{1 - e^{h(t_1 - t_i)}}{1 + e^{h(t_1 - t_i)}} \right)^2$$

valor con el que queda ajustada la logarística a los datos empíricos considerados.

2. APLICACION A LA POBLACION ESPAÑOLA

La aplicación del ajuste descrito a la población española debe partir inicialmente de la estimación del punto de inflexión para el actual ciclo de evolución de sus efectivos humanos. Para realizarla, nos basaremos en un estudio previo de las tasas de crecimiento intercensales experimentadas por nuestra población a lo largo de la reciente historia de sus Censos de Población.

I. *Determinación del punto de inflexión.* Los datos origen para nuestro estudio quedan recogidos en la Tabla I. En ella aparecen los stocks de población existentes en cada fecha de referencia, que corresponden a las distintas inscripciones censales efectuadas desde 1587 hasta nuestros días, junto al Padrón de 1975 cuyas cifras también se consideraron como oficiales tras la inspección y supervisión efectuada por el I.N.E. Las poblaciones que contiene la Tabla están expresadas en miles y en ellas no están incluídas —a fin de homogeneizar todas las cifras— las poblaciones de las ciudades de Ceuta y de Melilla.

TABLA I

FECHA DE REFERENCIA	POBLACION DE HECHO (en miles)
21-05-1857	15.455
25-12-1860	15.645
31-12-1877	16.622
31-12-1887	17.550
31-12-1897	18.109
31-12-1900	18.594
31-12-1910	19.927
31-12-1920	21.303
31-12-1930	23.564
31-12-1940	25.878
31-12-1950	27.977
31-12-1960	10.431
31-12-1970	33.824
31-12-1975	35.899
28-02-1981	37.617

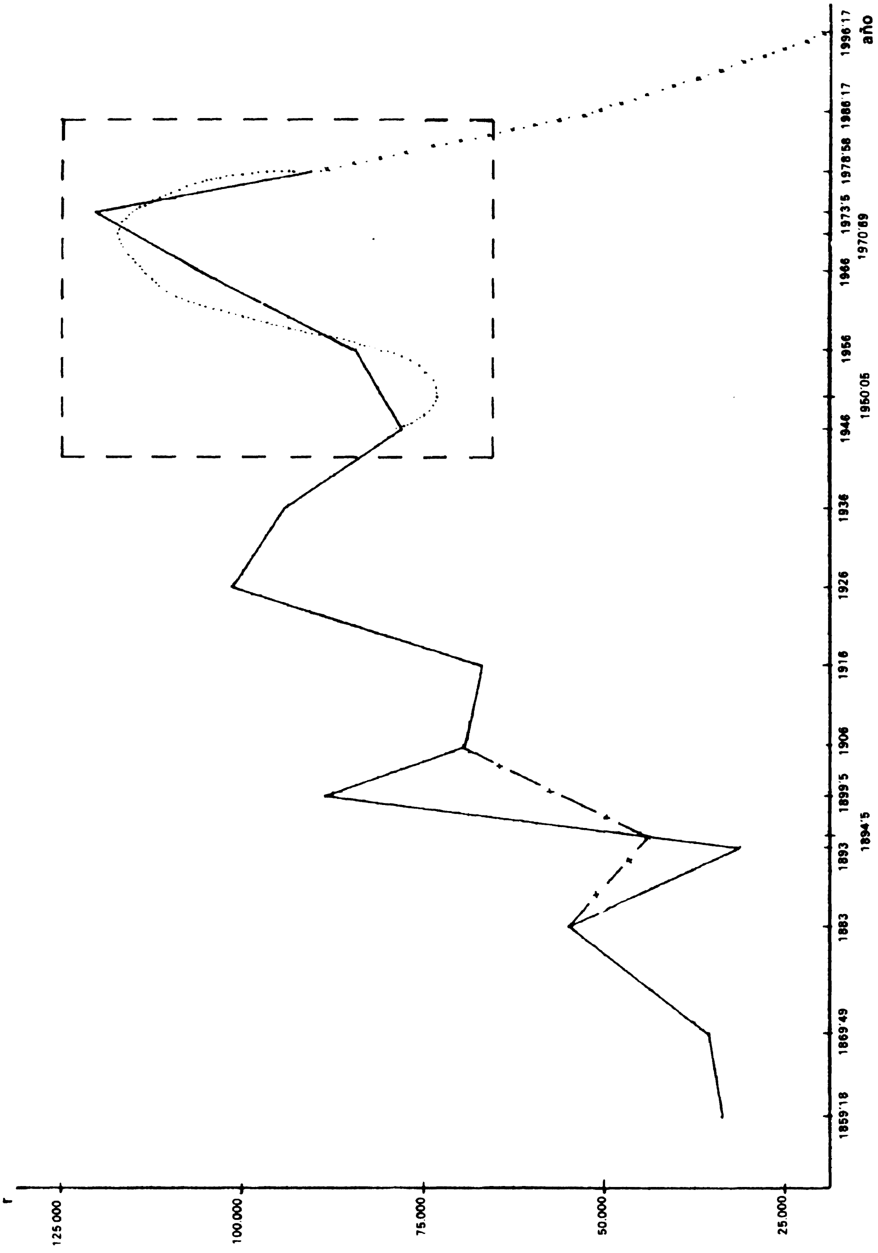
A partir de los datos anteriores pueden calcularse las tasas de crecimiento medio para los períodos intercensales. Por corresponder a una tasa media supondremos que coinciden precisamente con la tasa de crecimiento existente en el punto medio del período intercensal correspondiente. Así se obtiene la siguiente Tabla, donde las tasas aparecen multiplicadas por diez millones:

TABLA II

Punto de referencia	r = Tasa de crecimiento x 10^7
1859'18	34.051
1869'49	35.675
1883'0	54.475
1893'0	31.404
1899'5	88.489
1906'0	69.477
1916'0	66.996
1926'0	101.383
1936'0	94.113
1946'0	78.294
1956'0	84.433
1966'0	106.269
1973'5	119.789
1978'58	90.829

La figura adjunta recoge la representación gráfica de los datos anteriores y en la que los distintos valores de r están unidos mediante una línea contínua. Empezando por el origen, destaca el descenso habido en el período intercensal 1887–1897, y la fuerte recuperación de la tasa media en los tres años siguientes, 1897–1900. Dado el corto período de tiempo que representan esos tres años parece que ese súbito incremento no obedezca a consideraciones demográficas internas a la población existente en aquel momento, sino que haya que buscar explicaciones externas a la misma: tal vez, una inscripción censal (la del año 1897) no realizada con las suficientes garantías y con bajos valores finales; o, acaso, un fuerte incremento poblacional derivado del retorno de españoles a la Península provenientes de las últimas colonias. En todo caso, prescindiendo de la inscripción censal de 1897, la tasa de crecimiento medio para el período 1887–1900 —con año medio representativo 1894'5— sería de 0'0044549, que, gráficamente, ha quedado recogida en la figura mediante la línea de trazos discontinuos.

Para una población perfectamente logística, las tasas de crecimiento medio para iguales períodos —esencialmente positivas— presentarían una tendencia creciente cada


r
125.000
100.000
75.000
50.000
25.000
1859'18
1869'49
1883
1893
1894'5
1899'5
1906
1916
1926
1936
1946
1950'05
1956
1966
1970'69
1973'5
1978'58
1986'17
1996'17
año

vez más acusada hasta alcanzar un máximo, cuya abcisa correspondería a la del punto de inflexión de la curva. A partir de ese momento, simétricamente, los crecimientos medios irían disminuyendo, siguiendo un comportamiento asintótico hacia la parte positiva del eje de abcisas. Observando nuestra figura, en sus trazos continuos y discontinuos —prescindiendo, pues, del pico existente en 1899'5—, puede apreciarse para la población española que desde 1857 existe una tendencia al alza de las tasas de crecimiento medio intercensales, alcanzando un primer máximo en 1926. Los moderados decrecimientos existentes en 1894'5 y 1916 no pueden considerarse como representativos de un cambio en la tendencia general ante su mínima amplitud (10 años como máximo), obedeciendo a circunstancias coyunturales como la epidemia de gripe de 1918 para la tasa media centrada en 1916. Por otra parte, a partir del máximo de 1926, las tasas medias decrecen durante dos períodos intercensales consecutivos para iniciar más adelante un rápido y acelerado crecimiento hasta alcanzar un nuevo máximo centrado en 1973'5. Todo ello —y con la evidente limitación introducida por el escaso número de datos— permite considerar en el conjunto de años estudiados la existencia de un doble ciclo de evolución para nuestra población. Ambos ciclos —dentro de una tendencia general creciente— abarcan, respectivamente, desde 1860 hasta 1950 y desde esta fecha en adelante. Esta hipótesis ya aparece recogida por Alcaide (1974) cuando a su primera estimación logística (1955) —que explica con gran exactitud la evolución seguida por la población hasta 1950— superpone una segunda estimación, con enlace en este último año, justificada por "el cambio estructural de nuestra evolución demográfica, puesta de manifiesto por el espectacular crecimiento de la esperanza de vida al nacer y el brusco descenso de la mortalidad infantil, debido, todo ello, al indudable progreso de la medicina y, especialmente, a la comercialización de los antibióticos a partir de los años próximos a 1950".

Así pues, para proceder a una estimación logística de nuestra población que tenga valor predictivo deberá tenerse en cuenta, fundamentalmente, los datos del actual ciclo de su evolución, pues la consideración de los anteriores podría sesgar a la baja las estimaciones obtenidas. Por otra parte, a fin de utilizar el procedimiento presentado en la primera parte, es necesario estimar el punto de inflexión del ciclo. Nuevamente la observación de la figura ilustra el camino a seguir. Podemos suponer que la tasa media de crecimiento para el decenio 1941–1951 (centrada en 1946) representa la última observación conocida de la rama descendente correspondiente al primer ciclo de su evolución. A partir de ella, las tasas medias de los años 1956 y 1966 están ya situadas en la rama ascendente correspondiente al segundo de los ciclos. Esta tendencia creciente se trunca a partir de la tasa de 1973'5 —que es máxima— para descender bruscamente según los datos obtenidos en el quinquenio 1976–1981. De la observación de estos hechos se deduce que una buena representación para la tendencia de las tasas de crecimiento existentes en cada momento puede venir expresada a través de una parábola de tercer grado, con un mínimo cercano a 1946 y un máximo cercano a 1973'5 (ver recuadro a trazos en la figura). El mínimo de la curva ajustada marcará el inicio del ciclo actual; el máximo corresponderá al punto de inflexión a utilizar en el proceso de ajuste de la logística.

El ajuste por mínimos cuadrados del polinomio

$$r = a + b.t + c.t^2 + d.t^3$$

a los puntos (1946, 78294), (1956, 84433), (1966, 106269), (1973'5, 119789) y (1978'58, 90829) conduce a la parábola ajustada:

$$r = 111404'42 + 22496'738.t - 16960'23.t^2 - 10033'097.t^3$$

con el cambio de origen y escala $t = (t' - 1966)/10$. Esta curva presenta su mínimo en la abcisa 1950'0457 (siendo la ordenada 73086) y su máximo en la abcisa 1970'69 (de ordenada igual a 117190). El resultado del ajuste queda recogido, gráficamente, en la línea punteada del recuadro de la figura. Por tanto, y a fin de estimar la curva logística para la población española, partiremos como hipótesis fundamentales el que el actual ciclo de su evolución nace en 1950'05 y que presenta un punto de inflexión en sus tasas de crecimiento en el año 1970'69. La fuerte caída de las tasas de natalidad experimentada en los últimos años avala precisamente el hecho de una inflexión en el comportamiento demográfico de nuestra población.

II. *Ajuste de la logística a la población española.* Siendo 1970'69 la abcisa del punto de inflexión, su ordenada será:

$$y_I = 30431.(1'0106269)^{9'69} = 33713$$

de forma que

$$(t_I, y_I) = (1970'69, 33713)$$

Consideremos el año 1981'17, último del que se dispone de información censal. Sea t_2 = 1981'17, siendo por tanto

$$| t_2 - t_I | = 10'48$$

Sea $t_1 = t_I - 2 \mid t_I - t_2 \mid$, esto es, t_1 = 1949'73 (que prácticamente puede considerarse dentro del ciclo de evolución estudiado). Siendo y_i la población del año i, las correspondientes ordenadas para los puntos de abcisa t_1 y t_2 serán:

$y_1 = 25878.(1'0078294)^{8'73} = 27701$
$y_2 = 37617$ (inscripción censal del 01–03–1981)

Así pues, quedan definidos los puntos (t_1, y_1) y (t_2, y_2) a considerar en el ajuste que verifican: 1.° dejan al punto de inflexión entre ellos; 2.° ambos corresponden al ciclo de evolución estudiado (con la aclaración efectuada para t_1). Sean por tanto los tres puntos siguientes:

$$\begin{aligned} (t_1, y_1) &= (1949'73,\ 27701) \\ (t_2, y_2) &= (1981'17,\ 37617) \\ (t_I, y_I) &= (1970'69,\ 33713) \end{aligned} \tag{8}$$

Realizando el cambio

$t' = (1970'69 - t)/10'48$

y haciendo pasar (4) por los puntos (8) se obtienen las dos condiciones siguientes:

$$37617 - 33713 = \frac{K}{2} \cdot \frac{1 - e^{-h'}}{1 + e^{-h'}}$$

$$27701 - 33713 = \frac{K}{2} \cdot \frac{1 - e^{2h'}}{1 + e^{2h'}}$$

que dividiéndolas miembro a miembro conducen a la ecuación de segundo grado en $e^{h'}$ $(h' = h.10'48)$:

$$0'539959.e^{2h'} - 2.e^{h'} + 0'539959 = 0$$

cuya solución mayor que la unidad proporciona el siguiente valor para h:

$$h = 0'117075$$

Conocido este parámetro, K se determinará mediante el ajuste de una recta por el origen y por mínimos cuadrados a los datos:

$2\,(y_i - y_I)$	$(1 - e^{h(t_I - t_i)})/(1 + e^{h(t_I - t_i)})$
−11472	−0'8186147
−6564	−0'5133209
222	0'0181446
4372	0'3011957
7808	0'5465672

obteniendo la pendiente de dicha recta como estimación de K:

$$K = 13865$$

Finalmente, la curva logística resultante tendrá como ecuación

$$y = 33713 + 6932'5 \cdot \frac{1 - e^{0'117075.(1970'69 - t)}}{1 + e^{0'117075.(1970'69 - t)}}$$

o en su expresión clásica

$$y = 26780'5 + \frac{13865}{1 + e^{230'71853 - 0'117075.t}} \qquad (9)$$

y en la que t viene expresada en años reales. Por tanto, el tope mínimo (asíntota inferior) para el actual ciclo de nuestra población es de 26780500 personas, mientras que el tope máximo al que podría llegar nuestra población en sus actuales circunstancias sería (asíntota superior) de 40645500 habitantes.

La bondad del ajuste efectuado queda recogido en la Tabla III en la que se aprecia una diferencia máxima relativa entre las poblaciones censadas y ajustada del –0'9%, por lo que el ajuste puede considerarse como satisfactorio a efectos de descripción del pasado. En lo que respecta al futuro, las predicciones basadas en (9) también quedan recogidas en la Tabla, con un horizonte máximo de 25 años:

TABLA III

Año	Población censada	Población ajustada	Diferencia %
1951	27977	28037	0'2
1961	30431	30155	–0'9
1971	33824	33839	0
1976	35899	35801	–0'3
1981'17	37617	37502	–0'3
1986'17	–	38700	–
1991'17	–	39490	–
2001'17	–	40265	–
2011'17	–	40525	–

III. *Evaluación de resultados: comentarios.* Dejando a un lado las consideraciones de si la logística proporciona o no buenas previsiones sobre la cuantía futura de las poblaciones humanas, es indudable su importancia histórica en el campo de la demografía. La principal ventaja del procedimiento de ajuste que ahora hemos presentado —aparte de su evidente rapidez y operatividad, comparable a las de los procedimientos clásicos— reside en su adaptabilidad a las posibles modificaciones que se producen con la introducción de nuevos datos censales. En efecto, cada nuevo resultado censal puede confirmar o no el ajuste realizado para la obtención del punto de inflexión, dando lugar a una posible modificación de la logística para adaptarla mejor a la nueva situación. La ruptura de este proceso se producirá, evidentemente, cuando existan indicios razonables de una nueva evolución de la población bajo distintos ordenamientos demográficos. En todo caso, el análisis previo del ciclo en el que se encuentra la población es fundamental para proceder a ligeras modificaciones de la logística ya conocida o al ajuste de otra curva bajo hipótesis completamente distintas.

El estudio del ciclo aparece, pues, como fundamental. En nuestra aplicación concreta, el escaso número de datos disponibles obliga a centrar la perspectiva en períodos cortos de tiempo. No obstante, la disponibilidad de un número mayor de datos podría difuminar los dos ciclos que se han considerado, quedando enmarcados dentro de otro de amplitud más grande, lo que podría traer consigo la consideración de criterios diferentes para la obtención del punto de inflexión. En todo caso, el procedimiento de ajuste desarrollado evita la estimación a priori de las asíntotas, sustituyéndola por la del punto de inflexión, siendo la naturaleza de los datos disponibles la que determine —en último término— la conveniencia o no del empleo del ajuste que se ha considerado.

Volviendo a la estimación concreta que hemos realizado, podría pensarse en describir la evolución experimentada por la población española mediante dos curvas logísticas que abarcarían cada uno de los dos ciclos tratados en el análisis. La primera -por su magnífica descripción— puede ser la calculada por Alcaide (1955), de expresión

$$y = 14756 + \frac{21244}{1 + e^{1'316 - 0'040344.t}}$$

(con origen del tiempo en el año 1905) y validez hasta 1950. A partir de este año, con enlace en él, consideraríamos la curva ajustada en el presente trabajo (9). Por otra parte, comparando la logística (9) con la obtenida por Alcaide (1974) puede apreciarse la mayor moderación en sus previsiones de aquélla. La explicación hay que buscarla en la acelerada caída de la tasa de natalidad que, si bien fue vislumbrada y comentada por dicho autor, no la cuantificó en su total magnitud.

Precisamente, en los últimos años nuestra población ha asistido a un hecho demográfico que puede marcar una inflexión, con perspectiva a largo plazo, en la tendencia general de su evolución: la caída de la natalidad a unas tasas comparables a las de otras sociedades del máximo desarrollo industrial y económico. De persistir este hecho podríamos asistir a un auténtico vuelco en las hipótesis a utilizar en los próximos años e, incluso, podría pensarse en un futuro de situaciones inversas a las hasta ahora conocidas, con tasas de crecimiento medio negativas y disminución, por tanto, de la población en valores absolutos. Lo cierto es que la evolución que siga nuestra demografía en los inmediatos decenios será, lógicamente, la que marque de forma muy sensible la tendencia general a seguir por la población en un horizonte temporal amplio.

BIBLIOGRAFIA

ALCAIDE, A.: "Nueva determinación de la curva logística de la población de España". *Revista de Economía Política;* septiembre–diciembre, 1955; p.p. 141–157.

ALCAIDE, A.: "La población de España en el período 1970–2000". *Información Comercial Española;* diciembre, 1974; p.p. 11–21.

LEGUINA, J.: "Fundamentos de Demografía". Ed. Siglo XXI, 3.ª edición, junio 1981.

SUMMARY

NEW PROCEDURE FOR THE ADJUSTMENT OF THE LOGISTICAL CURVE APPLICATION TO THE SPANISH POPULATION

This article presents an adjustment procedure of the logistical curve, based on the previous knowledge of the inflection point. This procedure provides as an estimation the higher and lower asyntoths of the curve. The article ends with its application to the Spanish population.

Key words: Adjustment; logistical curve; inflection point; intercensal increase rate.

AMS, 1980. Subject classification: 62P10.