Faculty of Economic and Social Studies
University of Manchester/Manchester/M13 9PL
Telephone: 061-273 7121
Ext. No.
Department of Econometrics and Social Statistics

19 May 1987

The Cowles Librarian
Department of Economics
Yale University
New Haven
Connecticut 06520-2125
USA

Dear Ms Gifford

I enclose a copy of my paper on Harold T Davis which I have submitted for publication in Oxford Economic Papers.

I should be most gratefu for any comments.

Best wishes

R W Farebrother

R W Farebrother

HAROLD THAYER DAVIS AND THE LEAST SQUARES ESTIMATION OF NONLINEAR EQUATIONS

R.W. Farebrother

Department of Econometrics and Social Statistics
University of Manchester

May 1987

Published in 1988 by Marcel Dekker
Linear Least Squares Computation

SUMMARY

In this paper we give brief details of the life of H.T. Davis (1892-1974) and discuss the traditional least squares curve fitting approach to the estimation of nonlinear economic models as outlined in his (1941a,b) textbooks.

1. INTRODUCTION

Carl Christ's (1952) history of the first twenty years of the Cowles Commission opens with the following passage:

"The Cowles Commission for Research in Economics was founded in 1932. Alfred Cowles, president of Cowles and Company, an investment counseling firm in Colorado Springs, Colorado, initiated some inquiries into the accuracy of professional stock market forecasters over the period 1928-32. This aroused his interest in fundamental economic research, which led him to offer his financial support toward the establishment of the Cowles Commission and to bear a significant share of the burden each year. Fortunately, at the outset, he encountered Harold T. Davis, a professor of mathematics at Indiana University whose interests included mathematical economics and statistics. Davis was to become an important figure in the founding of the Cowles Commission and on its progressive development through the years...."

And yet the name of Harold Thayer Davis is little known to present-day econometricians despite his many contributions to their subject in its early years. Indeed, it is unlikely that the author would have undertaken this study of Davis's work had he not been intrigued by the extensive citation of Davis's (1941a,b) books in the texts published by his Cowles Commission colleagues,[1] Gerhard Tintner (1952) and Oskar Lange (1959).

2. BIOGRAPHICAL DETAILS

Harold Thayer Davis was born in Beatrice, Nebraska, on 5 October 1892 and died in Bloomington, Indiana, on 14 November 1974. He was educated at Colorado College[2] (A.B. 1915, hon. LL.D. 1949), Harvard College (A.M. 1919) and the University of Wisconsin (Ph.D. 1926). He was employed as an instructor in mathematics at the University of Wisconsin from 1920 to 1923, then he was successively assistant professor, associate professor, and full professor of mathematics at Indiana University from 1923 to 1937. After a brief one semester stay as acting professor of econometrics at Colorado College, he was appointed professor of mathematics at Northwestern University in 1937, becoming Chairman of the Department in 1940. Following his retirement from Northwestern University in 1955 he served in a similar capacity at Trinity University until

While at Indiana University, Davis established the Principia Press, "his purpose being to stimulate [the] publication of limited edition treatises. Later at Trinity University, he was likewise associated with enlargement of the Trinity Press" [Colorado College, 1974].

Davis was a prolific writer in many areas and a forthright advocate of the principles he believed in. On one occasion "he expressed a vigorous opinion on a subject widely removed from his primary field of knowledge which called forth a challenge [from a member of his hiking group]....'ne sutor ultra crepidam' ('let the shoemaker stick to his last')" [Colorado College, 1974]. This jibe gave Davis the title for his autobiography, The Adventures of an Ultra-Crepidarian (1962). "The 468 page volume is filled with fascinating fragments of Dr. Davis' life, with penetrating insights into areas which intrigued him....It reflects the author's warm human qualities, many amusing sidelights in his ever zestful life, his rich fertility of mind, his intense application to work and duty, and his abiding confidence in the ultimate victory of reason over ignorance and of good over evil" [Colorado College, 1974].

3. PUBLICATIONS

Davis's publication list includes "some 60 volumes [of full length]....along with at least an equal number of treatises and dissertations of lesser size. His major writings range in content from a 20 volume 'Encyclopedia of Mathematical Functions and their Tables' to one 'On the Fine Art of Punning' and a series of imaginative Christmas tales [published[3] as 'The Spectral Santa Claus'] sent annually over a 20 year period to his wide circle of friends" [Colorado College, 1974].

In this paper we will restrict our attention to two of his more serious works The Analysis of Economic Time Series (1941a) and The Theory of Econometrics (1941b) which were both published by the Principia Press. These books were written in parallel as Davis explains in his preface to the latter (1941b, pp. xii-xiii). "It soon became clear to the author that some extensive coordination of the material on time series was necessary before the second part of the book could be completed satisfactorily. It was necessary, therefore, to interrupt the writing of the present volume [The Theory of Econometrics] and to prepare a second work on the subject of economic dynamics. This work, The Analysis of Economic Time Series, is now completed....The author has borrowed heavily from the material in this treatise. Much of the analysis and many of the charts have been reproduced in the present work. Since The Analysis of Economic Time Series, however, was written as a treatise, and the present volume as a text, the material in the latter has been treated less completely...."

Thus, The Analysis of Economic Time Series (1941a) was a treatise on time series analysis covering such topics as harmonic analysis, serial correlation analysis, the analysis of trends and periodogram analysis, whilst The Theory of Econometrics (1941b) is an unusual blend of mathematical economics, statistical economics, economic statistics, actuarial mathematics,[3a] and time series analysis. This blend would not now be accepted as the subject matter of econometrics but Davis (1941b,

p. xi) had a clear view of what he understood by the term: "Since the newly-coined term 'econometrics' connotes better than either 'mathematical economics' or 'statistical economics' the measure of [Davis's] task, this word has been adopted as a descriptive title. As one may infer, it implies that the phenomena of economics are to be investigated and such patterns as may be observed are to be described in mathematical terms and by means of mathematical equations. Although this may remove the book from the reach of some who do not know the language of mathematics, there seems to be no other way in which so large an area can be exactly surveyed."

Both books are written at a high mathematical level and assume a sound knowledge of elementary statistics.[4] But they seem to be devoid of the concept of a statistical model[4a] as we shall illustrate in the following sections which examine Davis's approach to the fitting of nonlinear equations by the method of least squares.

4. INCOME DISTRIBUTION

4.1 Pareto's Law

For our first example we shall consider Davis's (1941a, pp. 394-418; 1941b, pp. 23-50) discussion of Pareto's (1897, Vol. 2, pp. 299-345) law of the distribution of income. According to this law, the number of persons y with income in excess of x is given by

$$y = ax^{-\nu} \qquad (4.1)$$

or $$\log y = \log a - \nu \log x \qquad (4.2)$$

Given data on the number of persons in each of a sequence of income classes, Davis cumulates this data to form series for x and y, takes logarithms, and estimates ν and log a in Equation (4.2) by the method of least squares.

Despite the fact that this procedure ignores the nature of the errors in the data, it provides an excellent fit to the data for the higher income classes.[5] Davis argues that the Pareto Law in the upper tail of the income distribution should be reflected in the upper tails of the distributions of measurements of intelligence and skill. He criticises Binet's Intelligence Quotient for not having such a distribution[6] in its upper tail and provides data on his colleagues' proficiency at billiards which tends to support his thesis. He also cites Lotka's (1926) analysis of publications in chemistry and physics to the same effect.

4.2 Generalisations

Davis mentions several generalisations of Equation (4.1) including those due to Pareto (1897)

$$y = Ae^{-\beta x}(x-c)^{-\nu} \tag{4.3}$$

Champernowne[7] (1936)

$$y = A[\cosh(Bw-C) - D]^{-1} \tag{4.4}$$

and Gibrat (1931)

$$y = N(2\pi)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}g^2) \tag{4.5}$$

where N is the number of income recipients, w = log x is income power and

$$g = a\log(x-k) + b \tag{4.6}$$

But he also developed the frequency function

$$\phi(z) = a_* z^{-(\mu+1)}[e^{b/z} - 1]^{-1} \tag{4.7}$$

as a generalisation of the Pareto function

$$\phi(x) = a\nu\, x^{-\mu} \qquad (4.8)$$

where $z = x-c$ and $\mu = \nu+1$.

Davis gives two methods for estimating the parameters of Equation (4.7). The first is relatively straightforward but involves the use of the gamma and zeta functions:

For a suitable approximation to the 'wolf point' c and for sufficiently large values of x, we have

$$\phi(z) \approx (a_*/b) z^{-\mu} \qquad (4.9)$$

which takes the same form as Equation (4.7) and may therefore be estimated by the method of Section 4.1.

Given this value for $\mu = \nu+1$, the total number of income recipients N, and the total income I, we may solve[7a]

$$N = \int_0^\infty \phi(z)dz = a_* b^{-\mu}\, \Gamma(\mu)\zeta(\mu) \qquad (4.10)$$

and[8]

$$I = \int_0^\infty \phi(z)z\ dz = a_* b^{-\nu}\, \Gamma(\nu)\zeta(\nu) \qquad (4.11)$$

for the remaining parameters a_* and b.

Davis's (1941a, pp. 414-415) second method is rather more complicated but it avoids using dubious values for N and I by using the modal income z_0 and the modal frequency ϕ_0 instead. Differentiating Equation (4.7) with respect to z, we have

$$d\phi/dz = z^{\mu}(\phi^2/a_*)\ [e^{b/z}(b/z - n) + n] \qquad (4.12)$$

so that

$$z_0 = b/(n-p) \qquad (4.13)$$

where $n = \mu+1$ and p is the nontrivial solution of

$$pe^{-p} = ne^{-n} \tag{4.14}$$

and
$$\phi_o = a_* pb^{-1} z_o^{-\mu} \tag{4.15}$$

For the value of μ determined by the method of Section 4.1, Davis obtains approximations N(z) to the frequencies ϕ(z) by numerical differentiation of the integral of Equation (4.8). Rewriting this equation in the form

$$a_*/b = N(z)z^{\mu} \tag{4.16}$$

he estimates a_*/b by the geometric mean of the values on the right. Then he uses the observed values of $n = \mu+1$ and ϕ_o to solve Equations (4.14), (4.15) and (4.13) for p, z_o and b and finally he estimates $c = x_o - z_o$ by comparing the observed modal income x_o with the calculated value of z_o.

4.3 Applications

Davis illustrated all three estimation methods by applying them to data on personal income recipients in the United States in 1918.

5. ENGEL CURVES AND LOGISTIC TREND LINES

For our second example we will examine Davis's (1941a, pp. 247-271; 1941b, pp. 161-165 and pp. 215-239) methods for estimating Engel curves of the form[9]

$$x = a - be^{-\mu I} \tag{5.1}$$

and logistic trend lines of the form[10]

$$y = k/(1 + be^{at}) \tag{5.2}$$

5.1 A Simple Method

The Engel curve in Equation (5.1) suggests that as a family's income I grows its expenditure x on a particular commodity (food, clothing or housing) will increase at a decreasing rate. Given data on I and x, Davis approximates Equation (5.1) by[11]

$$x = (a - b) + \mu b I \qquad (5.3)$$

from which he obtains estimates of a - b and μb by the method of least squares. These estimates are then combined with an estimate of the value of a, obtained by visual inspection of the graph of x against I, to yield estimates of b and μ.

5.2 Pearl and Reed's Method

Davis discusses three methods for estimating the logistic trend line (5.2), to which we may add a fourth based on the method of the previous section applied to the transformed equation

$$1/y = (1/k) + (b/k)e^{at} \qquad (5.4)$$

Davis's first method is due to Pearl and Reed (1920). In the first stage of this method we have to select three equally spaced points[12] $t = t_0$, $t = t_0 + t_1$ and $t = t_0 + 2t_1$ and define $a_0 = \log b + at_0$, $a_1 = a$ and $t_* = t - t_0$ so that Equation (5.2) may be rewritten as

$$y = k/(1 + e^{a_0 + a_1 t_*}) \qquad (5.5)$$

or as

$$a_0 + a_1 t_* = \log[(k-y)/y] \qquad (5.6)$$

Substituting $t_* = 0$, $t_* = t_1$ and $t_* = 2t_1$ in Equation (5.6) we have three equations in three unknowns from which we may obtain two expressions for a_1 in terms of k. These two expressions may be identified to yield a single expression for k from which we may obtain suitable expressions for a_1 and a_0.

In the second stage of this method, the preliminary estimates of k and a_0 are discarded but we define $c = 1/b$, $d = k/b$ and $h = a_1 - a$ so that Equation (5.2) may be rewritten as

$$y = d/(e^{a_1 t} e^{-ht} + c) \tag{5.7}$$

Approximating e^{-ht} by $1 - ht$ we have

$$y = d/[e^{a_1 t}(1 - ht) + c] \tag{5.8}$$

and thus

$$d - cy + ht.ye^{a_1 t} = ye^{a_1 t} \tag{5.9}$$

which may be estimated for a given value of a_1 by the method of least squares.

5.3 Schultz's Method

Davis (1941b, pp. 220-221) also notes that "it has been pointed out by Henry Schultz [1930] that the adjustment of the parameters of the logistic as given above does not lead to the true least squares solution which is obtained by minimizing the sum of the squares of the differences between the data and the functional values. He suggests the following alternative procedure:

'We first determine approximate values of the parameters, which we may designate by a_0, b_0 and k_0. The desired corrections may then be denoted by Δa, Δb and Δk . The function [5.2] is then expanded to a first approximation in the corrections by means of Taylor's series....The computation of the corrections is then effected by the method of least squares'."

This passage clearly describes the method now known as the Gauss-Newton method;[13] however "the principal objection to this adjustment is found in the fact that Δk may be so large that differences of second order cannot be neglected. Hence the correction must be made successively several times before an approximation better than the Pearl-Reed method is attained".

5.4 Hotelling's Method

Davis's third method for estimating the parameters of a logistic trend line is a variant of the method proposed by Hotelling (1927). Let $c = -a$ then we may rewrite Equation (5.2) as

$$y = k/(1 + be^{-ct}) \tag{5.10}$$

and its derivative as

$$dy/dt = cy - (c/k)y^2 \tag{5.11}$$

Replacing dy/dt by $\Delta y/\Delta t$ where $\Delta t = 1$, we have the equation

$$\Delta y = cy - (c/k)y^2 \tag{5.12}$$

but Davis prefers to estimate c and c/k by applying the method of least squares to the expression[14]

$$\Delta y/y = c - (c/k)y \tag{5.13}$$

which is linear in y.

This procedure yields estimates of c and k but not b. Davis offers three different methods for estimating b:

(i) If t_0 is the value of t for which $y = \frac{1}{2}k$, then b is given by

$$\log b = ct_0.$$

(ii) If (5.10) is rewritten as

$$k/y_t = 1 + be^{-ct} \tag{5.14}$$

then we have

$$k \sum_{t=m}^{n} 1/y_t = n-m+1 + b \sum_{t=m}^{n} e^{-ct} \tag{5.15}$$

which may be solved for b.

(iii) Alternatively, if (5.10) is rewritten as

$$b = [(k - y_t)/y_t]e^{ct} \tag{5.16}$$

then we have

$$b = (n-m+1)^{-1} \sum_{t=m}^{n} [(k-y_t)/y_t]e^{ct} \tag{5.17}$$

5.5 Applications

Davis applies one or other of the methods of the last three sections to data on the production of pig iron, the production of wheat, the index of industrial production and the population of the United States. In fact, he applies all three methods in the last case and finds that they yield similar predictions: 183-184 millions for the population of the US in the year 2000, and 195-196 millions in the year 2100. Both sets of figures had already been exceeded by 1980.

Davis also suggests that the logistic curve may be fitted to data on the proportion of the population of a country concentrated in large towns and to data on the growth of cities.[15]

6. THE MODIFIED EXPONENTIAL, LOGISTIC AND GOMPERTZ CURVES

6.1 The Gompertz Curve

Davis (1941a, p. 249; 1941b, p. 240) notes that the theory of the logistic function is closely related to that of the Gompertz curve

$$y_t = pq^{r^t} \tag{6.1}$$

but he does not explain how the parameters of this curve are to be estimated. For such an explanation, we turn to Appendix D of the contemporary textbook by Mills (1938, pp. 667-680):

Taking logarithms in Equation (6.1), we have

$$z_t = a + br^t \tag{6.2}$$

where $z_t = \log y_t$, $a = \log p$ and $b = \log q$. Now, suppose that we have 3n observations on z_t indexed by $t = 0,1,\ldots,3n-1$, then we may divide them into three equal groups and obtain the expressions

$$S_1 = na + b(r^n - 1)/(r - 1) \tag{6.3}$$

$$S_2 = na + br^n(r^n - 1)/(r - 1) \tag{6.4}$$

and $$S_3 = na + br^{2n}(r^n - 1)/(r - 1) \tag{6.5}$$

where

$$S_{j+1} = \sum_{t=1}^{n} z_{nj+t} \qquad j = 0,1,2 \tag{6.6}$$

Further $$D_1 = S_2 - S_1 = b(r^n - 1)^2/(r - 1) \tag{6.7}$$

and $$D_2 = S_3 - S_2 = br^n(r^n - 1)^2/(r - 1) \tag{6.8}$$

so that we have

$$r = (D_2/D_1)^{1/n} \tag{6.9}$$

$$b = D_1(r - 1)/(r^n - 1)^2 \tag{6.10}$$

and $$a = [S_1 - D_1/(r^n - 1)]/n \tag{6.11}$$

These are the expressions used by Mills to estimate the parameters $p = \exp(a)$, $q = \exp(b)$ and r of the Gompertz curve (6.1).

6.2 Modified Exponential Curve

Mills also discusses the estimation of Equation (6.4) when $z_t = y_t$ or $z_t = 1/y_t$. If this equation is rewritten as

$$z_t = a + be^{-ct} \tag{6.12}$$

where $c = -\log r$ then it can readily be identified with our Equations (5.1) and (5.4) in these two cases. In the first case, Mills calls Equation (6.4) "the modified exponential curve" and in the second case he calls it "the logistic curve, sometimes termed the Pearl-Reed growth curve".

Mills recommends the method of Section 6.1 for estimating the parameters of Equation (6.4) when $z_t = 1/y_t$ but he suggests a different method when $z_t = y_t$. Substituting Equations (6.7), (6.8) and (6.9) into Equation (6.11) we have

$$a = (M_1 M_3 - M_2^2)/(M_1 + M_3 - 2M_2) \tag{6.13}$$

where $M_j = S_j/n$. Having thus determined a value for a, Mills subtracts it from y_t and obtains values for $\log b$ and $\log r$ by fitting the equation

$$\log(y_t - a) = \log b + t \log r \tag{6.14}$$

by the method of least squares.

6.3 Special Case

Peters and van Voorhis (1940, pp. 435-437) examined the special case of the Gompertz curve when $p = 1$, $0 < q < 1$ and $0 < r < 1$. In this case Equation (6.3) takes the form

$$\log y_t = - dr^t \tag{6.15}$$

where $d = - \log q$. Negating[16] this expression and taking logarithms we have[17]

$$\log[-\log (y_t)] = \log d + t \log r \tag{6.16}$$

which may be estimated by the method of least squares.

6.4 Applications

Mills (1938) illustrated his methods for fitting the modified exponential curve and the Gompertz curve using data on the production of rayon filament yarn in the United States between 1920 and 1937. He illustrated his method for fitting the logistic curve using data on the length of railways in operation in the United States between 1850 and 1935.

Peters and van Voorhis (1940) used Burt's data on the proportion of boys passing test 20 (Fingers) of the Binet test to illustrate their method for fitting their special case of the Gompertz curve.

6.5 Demand Curve

Finally, we must record that Davis (1941b, pp. 100-101), following Moore (1914), suggested that classical demand curves of the form

$$\frac{\Delta p}{p} = b \frac{\Delta u}{u} - c \tag{6.17}$$

should be estimated by applying the method of least squares to Equation (6.17) where p represents price and u production.

Expressions of this form may be derived by the methods of Section 5.4 from expressions of the form

$$p = au^b e^{-ct} \tag{6.18}$$

but Davis does not do so. On the other hand he does (1941a, p. 438; 1941b, p. 154) obtain the expression

$$\frac{dP}{P} = p\,\frac{dL}{L} + q\,\frac{dC}{C} \tag{6.19}$$

from the Cobb-Douglas production function

$$P = AL^pC^q \tag{6.20}$$

Christ (1985) has made a detailed study of the contributions of Moore and Schultz in this area.

7. CONCLUSION

In the last three sections we have examined Davis's treatment of the traditional curve fitting approach to econometric estimation. This approach was standard before 1941, see Croxten and Cowden (1939), Mills (1938) or Peters and van Voorhis (1940), but it was swept away by the Probabilistic Revolution[18] initiated by Haavelmo (1944) and completed by Hood and Koopmans (1953).

The traditional approach was characterised by a lack of concern for the statistical properties of the economic variables being studied, and thus by the use of arbitrary estimation procedures, whereas the new approach stressed the probabilistic foundations of the model and recommended the use of maximum likelihood estimation procedures.

Although Davis remained a member of the Cowles Commission and an associate editor of Econometrica,[19] he seems to have had little influence on the direction taken by mainstream econometrics after 1941. Indeed none of the papers published in the volumes[20] edited by Koopmans (1950) or Hood and Koopmans (1953) refer to him by name. On the other hand, his influence is clearly visible in the papers published by Hayakawa (1951) and Champernowne (1952) and in the books published by Tintner[21] (1952) and Lange (1959). Nor did his influence come to an end in the 1950s as the text by Lange survived into a fourth edition (1978) and, as late as 1980, Badger[22] (1980) fitted a generalised Pareto curve to data on income recipients in the United States in 1935-6 by traditional curve fitting methods.

Thus, Davis had the misfortune to publish his books on The Theory of Econometrics and The Analysis of Economic Time Series just three years before Haavelmo formally introduced[23] The Probability Approach to Econometrics which replaced Davis's traditional approach by one based on the use of statistical models.

Nevertheless, Davis's books deserve to be studied by historians of econometrics as they give a clear indication of the nature of the subject before the Probabilistic Revolution of 1944. Further, it should be noted that these books also contain a considerable body of material which is of interest to the historians of mathematical economics and time series analysis but which we have not had space to discuss in this paper.

APPENDIX

Kelvin's Dictum

In his history of the Cowles Commission, Christ (1952, p. 61) notes that its motto[24] from 1932 to 1952, "Science is Measurement....was originally suggested by one of Davis' favourite quotations from the British physicist Lord Kelvin to the effect that when you can measure what you are speaking about, you know something about it, but when you cannot measure it, your knowledge is of a meager and unsatisfactory kind". See Davis (1941b, p. 3).

Christ (1952, p. 20) also notes that from 1939 the Cowles Commission was lodged in a suite of four offices on the fourth floor of the Social Science Research Building at the University of Chicago. But he fails to mention that a longer version of Kelvin's dictum is carved on the facade of this building. The origins of this longer version are traced by Merton, Sills and Stigler (1984).

Acknowledgement

I am indebted to the archivists of the Cowles Commission, Colorado College, Northwestern University and Trinity University for locating the material discussed in the first three sections of this paper.

I am indebted to Christopher Gilbert, David Hendry and Mary Morgan for many useful comments.

Notes

1. Tintner, Davis and Lange were the authors of Cowles Commission monographs 5, 6 and 8 respectively.

2. Davis's wife, Agnes Holm, and his sister, Marjorie Davis, were also graduates of Colorado College.

3. The Book Editor of the Colorado Springs Gazette notes (8 December 1967) that "The stories are the kind with which academic friends may greet each other in their lighter communications".

3a. Seven tables are printed at the end of Davis's The Theory of Econometrics (1941b, pp. 465-471). Six of these tables are actuarial in nature, the seventh tabulates values of $1/(2p+1)$ and $3/[p(p+1)(2p+1)]$ to aid the fitting of a linear trend.

4. For example (1941a, p. 399; 1941b, p. 26) assumes that the reader is familiar with the moments of a distribution and Pearson's measures of skewness and kurtosis. Both texts contain several references to the textbook by Davis and Nelson (1937) which Davis wrote with another colleague from the Cowles Commission.

4a. This judgement may seem a little harsh in view of Davis's use of the terms 'frequency function' and 'cumulated frequency function' to describe $\phi(x)$ and $y(x)$ in his discussion of income distributions. But his use of these terms has no impact on his estimation procedures which implicitly assume that Equations (4.1) and (4.7) are deterministic.

5. A procedure which takes account of the different numbers in the various income classes would, of course, provide a close fit to the more numerous lower income classes.

6. Recent work suggests that the Intelligence Quotient may have heavier than normal tails for male subjects.

7. The misplaced B in Davis's statement of Champernowne's formula is reproduced from the abstract published in Econometrica in 1937. Further, the left side of this equation should be replaced by $\phi(w)$ as this expression represents a frequency function rather than an accumulated frequency function. See Champernowne (1936, Appendix 1; 1952).

8. The left side of Equation (4.10) should presumably be replaced by I-cN.

9. Davis does not justify this choice of a functional form other than to note (1941b, p. 161) that "it is probable that a function of the form [5.1] would furnish in general a better approximation to Engel curves....than does the straight line....".

10. Davis also discusses the general logistic function defined by

$$y = k/(1 + be^{\phi(t)})$$

where $\phi(t)$ is a polynomial function of t. He suggests that $\phi(t)$ should be a cubic function if y represents the numbers in a population under the direction of a central mechanism.

11. Davis would have been able to estimate a, b and μ directly if he had approximated (5.1) by

$$x = (a-b) + \mu bI - \frac{1}{2}\mu^2 bI^2$$

rather than by (5.3).

12. Davis implicitly assumes that $t_0 = 0$.

13. Schultz's Taylor series expansion of Equation (5.2) would seem to be incorrect.

14. Davis notes that $y^{-1}dy/dt$ may be approximated more closely by the formula

$$\Delta F(t) - \frac{1}{2}\Delta^2 F(t) + \frac{1}{3}\Delta^3 F(t) + \ldots.$$

where $F(t) = \log y(t)$. "For most statistical data the first two terms of the series are sufficient." Equation (5.13) retains the first term only.

15. Singer (1936) has proposed the Pareto distribution as a model for the size distribution of urban settlements.

16. Peters and van Voorhis are somewhat mysterious concerning this negation step.

17. If we define $\alpha = -\log d$ and $\beta = -\log r$ then (6.16) may be rewritten as

$$y = \exp[-\exp(-\alpha - \beta t)]$$

which is the standard form for a Gumbel distribution.

18. See Morgan (1987).

19. Davis's own contributions to <u>Econometrica</u> comprise sixteen papers published by abstract and one by title on such topics as the distribution of income and time series analysis. He also published an obituary notice for Charles Roos, the first Director of the Cowles Commission.

20. These volumes were issued as Cowles Commission Monograph 10 and Cowles Foundation Monograph 14 respectively.

21. Tintner's book represents an intermediate phase of the Probabilistic Revolution as it discusses both the old curve fitting methods and the new maximum likelihood methods.

22. In a second contribution to the volume edited by West (1980), Budgor and West (1980) illustrated the value of mathematical modelling for problems in the social sciences by fitting a Gumbel curve (see Note 17) to data on the minimum height of the Nile

flood.

23. Haavelmo's (1944) paper had been circulated in manuscript in 1941.

24. In 1952 it was changed to _Theory and Measurement_ at the suggestion of Clifford Hildreth.

References

W.W. Badger (1980), "An Entropy-Utility Model for the Size Distribution of Income", in West (1980, pp. 87-120).

A. Budgor and B.J. West (1980), "Natural Forces and Extreme Events", in West (1980, pp. 35-45).

D.G. Champernowne (1936), The Distribution of Incomes Between Persons. Fellowship Prize Dissertation, King's College, Cambridge. Reprinted with additional material, Cambridge University Press, Cambridge, 1973.

D.G. Champernowne (1952), "The Graduation of Income Distributions", Econometrica, 20, pp. 591-615.

C.F. Christ (1952), "History of the Cowles Commission 1932-1952", in Economic Theory and Measurement: A Twenty Year Research Report, Cowles Commission for Research in Economics, Chicago, 1952, pp. 3-65.

C.F. Christ (1985), "Early Progess in Estimating Quantative Economic Relationships in America", American Economic Review, 75, pp. 39-52.

Colorado College (1974), "News Release re Dr. Harold Thayer Davis", Colorado College, Colorado Springs.

F.E. Croxton and D.J. Cowden (1939), Applied General Statistics, Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

H.T. Davis (1941a), The Analysis of Economic Time Series, Principia Press, Bloomington, Indiana. Reprinted by Trinity University Press.

H.T. Davis (1941b), The Theory of Econometrics, Principia Press, Bloomington, Indiana.

H.T. Davis (1962), The Adventures of an Ultra-Crepidarian, San Antonio, Texas.

H.T. Davis and W.F.C. Nelson (1937), Elements of Statistics (Second Edition), Principia Press, Bloomington, Indiana.

R. Gibrat (1931), Les Inegalites Economiques, Recueil Sirey, Paris.

T. Haavelmo (1944), "The Probability Approach in Econometrics", Econometrica, 12, Supplement.

M. Hayakawa (1951), "The Application of Pareto's Law of Income to Japanese Data", Econometrica, 19, pp. 174-183.

W.C. Hood and T.C. Koopmans, Eds. (1953), Studies in Econometric Methods, Yale University Press, New Haven.

H. Hotelling (1927), "Differential Equations Subject to Error and Population Estimates", Journal of the American Statistical

Association, 22, pp. 283-314.

T.C. Koopmans, Ed. (1950), *Statistical Inference in Dynamic Economic Models*, John Wiley and Sons, Inc., New York.

O.R. Lange (1959), *Introduction to Econometrics*, Pergamon Press Ltd., Oxford and Polish Scientific Publishers, Warsaw [Fourth Edition, 1978].

A.J. Lotka (1926), "The Frequency Distribution of Scientific Productivity", *Journal of the Washington Academy of Sciences*, 16, pp. 317-323.

R.K. Merton, D.L. Sills and S.M. Stigler (1984), "The Kelvin Dictum and Social Science: An Excursion into the History of an Idea", *Journal of the History of the Behavioural Sciences*, 20, pp. 319-331.

F.C. Mills (1938), *Statistical Methods Applied to Economics and Business* (Revised Edition), Henry Holt and Company, New York.

H.L. Moore (1914), *Economic Cycles: Their Law and Cause*, Macmillan, New York.

M.S. Morgan (1987), "The Probabilistic Revolution in Econometrics", in L. Kruger, G. Gigerenza and M.S. Morgan (Eds.), *The Probabilistic Revolution: Ideas in the Sciences*, MIT Press, Cambridge, Massachusetts.

V. Pareto (1897), *Cours d'Economie Politique*, F. Pichon, Paris.

R. Pearl and L.J. Reed (1920), "On the Rate of Growth of the Population of the United States since 1790 and its Mathematical Representation", *Proceedings of the National Academy of Sciences*, 6, pp. 275-288.

C.C. Peters and W.R. van Voorhis (1940), *Statistical Procedures and Their Mathematical Bases*, McGraw-Hill Book Company, Inc., New York.

H. Schultz (1930), "The Standard Error of Forecast from a Curve", *Journal of the American Statistical Association*, 25, pp. 139-185.

H.W. Singer (1936), "Courbes des Populations: A Parallel to Pareto's Law", *Economic Journal*, 46, pp. 254-263.

G. Tintner (1952), *Econometrics*, John Wiley and Sons, Inc., New York.

B.J. West, Ed. (1980), *Mathematical Models as a Tool for Social Sciences*, Gordon and Breach Inc., New York.

SUPPLEMENT

THE IDENTIFICATION PROBLEM

Davis resumes his analysis of Moore's study by noting that he had obtained a positive value for the slope parameter, b, from data on corn prices and output but a negative value from data on pig iron. "Since the methods employed in the two cases were the same Moore was led to the conclusion that he had obtained the demand curve for pig iron....[However] if one assumes that agricultural commodities are associated with their price in essentially the same way as manufactured goods are associated with their price, then it is much more sensible to interpret the two contradictory results in the light of the hypothesis advanced by Working [1927]. In the case of corn the demand curve was fixed and the supply curve varied, while in the case of pig iron, the supply curve remained stable and the demand curve varied. Hence, we obtained in the first instance an approximation to the classical curve of demand, and in the second an approximation to the curve of supply" [Davis (1941b, p. 102)].

Davis continues his discussion of the identification problem by considering the cobweb model in which price is a function of lagged quantity and quantity is a function of current price

$$p_t = f(q_{t-1})$$

$$q_t = g(p_t)$$

Successive iterations of this model yield points (p_1,q_1), (p_2,q_2),.... on the supply curve and points (p_1,q_0), (p_2,q_1),.... on the demand curve.

Davis (1941a, p. 141; 1941b, p. 105) quotes a passage from Schultz (1938, pp. 78-79):

"Thus far we have assumed that the two unknown curves (of demand and supply) remain fixed and have shown that when an interval elapses between changes in price and corresponding changes in supply, it is possible to deduce both curves statistically....The importance of such a demand-supply relationship lies in that it admits of a straightforward statistical 'verification'. If by correlating prices and output (consumption) for synchronous years (or other intervals) we obtain a high negative correlation; and if by correlating the same data but with out-put lagged by, say, one year we get a high positive correlation; <u>and if these correlations have meaning in terms of the industry or commodity under consideration</u>, the statistical demand and supply curves thus obtained are probably very close approximations to the theoretical curves."

But Davis (1941b, p. 105) does not seem to believe this result: "The situation thus commented upon is not unlike that of the stranger in Aesop's fable who aroused the suspicions of his host because he blew hot to warm his hands and then blew cold to cool his porridge".

Additional References

A.S. Goldberger (1972), "Structural Equation Methods in the Social Sciences", <u>Econometrica</u>, 40, pp. 979-1001.

N. Kaergaard (1984), "The Earliest History of Econometrics: Some Neglected Danish Contributions", <u>History of Political Economy</u>, 16, pp. 437-444.

H. Schultz (1938), <u>The Theory and Measurement of Demand</u>, The University of Chicago Press, Chicago.

E.J. Working (1927), "What do Statistical Demand Curves Show?", <u>Quarterly Journal of Economics</u>, 41, pp. 212-235.