ON THE NUMERICAL PROGRAMMING AND USE OF ASYMPTOTICALLY PARETIAN DISTRIBUTIONS.

by

Michel Biart and Jean Pierre Lemaître*

Working Paper 7108

August, 1971

* The authors are "Aspirant du Fonds National de la Recherche Scientifique" and "Assistant à l'Université Catholique de Louvain" respectively.

1. INTRODUCTION

Assymptotically paretian distributions are well known to economists interested in industrial organization or in income theory and measurement. (IV)(IX)(XIII)

For large values of X, they are approximated by the following formula of cumulated frequencies :

$$F(X) = 1 - \left(\frac{X}{X_o}\right)^{-\alpha} \qquad X \geq X_o \qquad (1)$$

from which one might estimate, by ordinary least-squares, the regression :

$$\ln \left[1 - F(X)\right] = \alpha \ln X_o - \alpha \ln X \qquad (2)(XII)$$

Unfortunately, such a procedure would overlook several difficulties:

1°) The estimation of X_o ought to be carried out by non linear procedures.

2°) The grouping of data and the cumulative character of the "greater than frequencies" violates the usual assumption of independence of observations and accordingly is bringing about a definite scheme of autocorrelation of residuals.

3°) The logarithmic transformation itself is bound to affect the distribution of the error term.

4°) Last but not least, the fact that empirical distributions are well-fitted by a Pareto curve in their upper part only would lead to a misspecification error if equation (2) were to be tested on the whole range of a variable.

Points 2°) and 3°) are explicitely dealt with by Aigner and Goldberger in their article on which this paper relies.(VI) They get rid of the first difficulty by using a device that we shall justify in part 3. That will lead us to a mildly optimistic view with respect to specification. Part 4 will be devoted to the use of α for descriptive purposes. In the last two sections all relevant information will be given on the program and the way one may handle it.

2. AIGNER AND GOLDBERGER GENERALIZED LEAST SQUARE ESTIMATOR

The principal originality of Aigner and Goldberger's paper is perhaps that the autocorrelation of residuals is entirely explained by properties of sampling. The authors start from the fact that if ϕ_i is the probability that, according to a Pareto law, an observation were to be found in the interval (X_i, X_{i+1}), the number of observations in each interval may be considered as drawn at random from a multinomial distribution (V) Accordingly one can consider that a sample has a probability mass function $M(N; \phi_o, \ldots, \phi_i, \ldots, \phi_T)$ where M stands for multinomial and N for the size of the sample.

If the absolute frequencies have the multinomial distribution, the means, variances and covariances of the relative ones are the following :

$$E(f_t) = \phi_t \qquad Var(f_t) = \frac{1}{N}\phi_t(1-\phi_t) \qquad Covar(f_t,f_s) = -\frac{1}{N}\phi_t\phi_s \ ^{1}$$

Aigner and Goldberger consider that the observed relative frequencies are related to the true ones by the expression :

$$f = \phi + \varepsilon$$

where the multinomial hypothesis defines the properties of the disturbance vector :

$$E(\varepsilon) = 0, \qquad E(\varepsilon\varepsilon') = \frac{1}{N}(\Phi - \phi\phi')$$

with Φ the T X T diagonal matrix built from the ϕ_i. To obtain a regression model linear in the parameters, a transformation of variables is required. Aigner and Goldberger perform it in two steps.

First, they introduce the "greater than frequencies" :

$$\pi_t = Prob(X_t < X) = \sum_{s=t}^{T} \phi_s = x_t^{-\alpha} \qquad (t=1,\ldots,T)$$

$$Y_t = \sum_{s=t}^{T} f_s$$

and set up the regression model :

$$Y = \pi + u \qquad (4)$$

---

[1] Those results are straightforward if one remembers that, for a multinomial, $E(X)=np$ $Var(X)=np(1-p)$ and $Cov(X_i,X_j)=-np_ip_j$ while any discrete distribution is such that

$$E(\frac{X}{N})=\frac{1}{N}E(X), Var(\frac{X}{N})=\frac{1}{N}2\ Var(X), Cov(\frac{X_i}{N},\frac{X_j}{N})=\frac{1}{N}2\ Cov(X_i,X_j)$$

Defining a T X T matrix A with elements :

$$a_{ts} = \begin{cases} 1 \text{ if } t \geq s \\ 0 \text{ if } t < s \end{cases}$$

they relate (4) to (3) by the expressions

$$Y = A'f = A'\phi \qquad u = Y - \pi = A'(f-\phi) = A'\varepsilon$$

and therefore $E(u) = 0 \qquad E(uu') = A'E(\varepsilon\varepsilon')A = \frac{1}{N}\,\Omega$

where $\Omega = A'\sum A = A'\Phi A - A'\phi\phi'A = A'\Phi A - \pi\pi'$

Afterwards they apply a logarithmic transformation

$$\ln Y = \ln\pi + v \tag{5}$$

and as one may write from (4)

$$Y_t = \pi_t \left(1 + \frac{u_t}{\pi_t}\right)$$

$$v_t = \ln\left(1 + \frac{u_t}{\pi_t}\right) = \ln\left(1 + \frac{u_t}{x_t^{-\alpha}}\right) \tag{6}$$

according to the Pareto law.

Classical Mac Laurin expansion enables to write :

$$v_t = \frac{u_t}{\pi_t} + R_t$$

And by retaining the first term only we get the following regression model :

$$Y^* = -\alpha X^* + v \quad \text{where } v = \pi^{-1}u$$

π being the T X T diagonal matrix built from the π_t

$$E(V) = 0 \qquad E(vv') = \frac{1}{N}\,\pi^{-1}\,A'(\Phi-\phi\phi')A\pi^{-1}$$

Application of the generalized least-squares procedure to minimize $v'\hat{\Lambda}^{-1}v$, where $\hat{\Lambda}^{-1}$ is a consistent estimator of $E(vv')$, yields :

$$\hat{\alpha} = -(X^{*\prime}\hat{\Lambda}^{-1}X^*)^{-1}(X^{*\prime}\hat{\Lambda}^{-1}Y^*)$$

with $\hat{\sigma}^2_{\hat{\alpha}} = \hat{v}'\hat{\Lambda}^{-1}\hat{v}\ (x^{*\prime}\hat{\Lambda}^{-1}x^{*})\ (T-1)^{-1}$,

the regression passing through the origin.

This estimator is consistent and thus assymptotically unbiased, properties it shares with the O.L.S. estimator :

$$\alpha^{*} = -\sum_{t=1}^{T} X_t^{*}\, Y_t^{*} \Big/ \sum_{t=1}^{T} X_t^{*2}$$

In addition, it is assymptotically efficient i.e. the assymptotic variance of the Generalized least-squares estimator as defined here above is smaller than the one of any other estimator because $\hat{\Lambda}$, three-diagonal matrix of the form :

$$\lambda_{ii} = y_i^2\,(f_{i-1}^{-1} + f_i^{-1}),\ \lambda_{i,i+1} = \lambda_{i,i-1} = -y_{i+1}y_i f_i^{-1}\ ,$$

is a consistent estimator of Λ :

$$\lambda_{ii} = \pi_1^2\,(\phi_{i-1}^{-1} + \phi_i^{-1}),\ \lambda_{i,i+1} = \lambda_{i,i-1} = -\pi_{i+1}\pi_i\phi_i^{-1} \qquad [1]$$

Amongst economists, it is common faith that the assymptotic properties of the variance-covariance matrix of an estimator hold for "large" samples (see VI, p.128). Sceptics may accept this viewpoint after having checked the closeness of estimates obtained e.g. by generalized and ordinary least-squares. They might then conjecture that the joint likelihood of these estimates is maximum for a true value of the parameter in their range.

---

[1] Consistency is obvious as an infinite sample would be characterized by $f_{i's}$ and $y_{i's}$ equal to the $\phi_{i's}$ and the $\pi_{i's}$ respectively.

3. THE TRUNCATED PARETO LAW.

In the former section was introduced the regression model :

$$\ln Y = - \alpha \ln X + v. \qquad (7)$$

This involves the estimation of one parameter only and allows to avoid the non linearity problem appearant in the traditional regression model for a Pareto :

$$\ln Y = \alpha \ln X_o - \alpha \ln X$$

Aigner and Goldberger set up the model (7) thanks to the change of variable $x = \frac{Z}{Z_o}$, were Z is the variable under study and Z_o the value it takes at the lower limit of the second class interval, from the origin. The use of the Jacobian method shows indeed straight away that :

$$g(x) = f[h(x)] \frac{dZ}{dX} = \alpha Z_o^{\alpha}(Z_o X)^{-\alpha-1} = \alpha X^{-\alpha-1}$$

but we must know whether the choice of an a priori value for Z_o imposes a restriction on the model, especially on the estimation of α.

Discarding observations below some point ξ is equivalent to considering a Pareto truncated to the right. The cumulative frequencies of this distribution are easily obtained from the general formula of truncated distributions :

$$G(X) = \frac{F(X)-F(\xi)}{1-F(\xi)} ,$$

i.e. the ratio of non discarded observations smaller than X to all non discarded observations.

For a Pareto, one has :

$$G(X) = \frac{1-(\frac{X}{X_o})^{-\alpha} - [1-(\frac{\xi}{X_o})^{-\alpha}]}{1-[1-(\frac{\xi}{X_o})^{-\alpha}]} = 1-(\frac{X}{\xi})^{-\alpha}$$

This shows that truncating a Pareto distribution does not affect the value of α and replaces X_o by ξ, the truncation point. Therefore, we may accept the Aigner-Goldberger simplification as fully valid.

Starting from this result, we wrote a program that truncates the Pareto law at each lower limit of class interval. It enables to decide whether the approximation of empirical distributions by Pareto ones may be done soon enough to yield a succession of $\hat{\alpha}$ having very similar values. In such a case, the smallest of the relevant lower limits is taken as the base income of the paretian part of the distribution.

A comparison of estimates by G.L.S. and O.L.S. was performed on income data for spanish households in 1964. G.L.S. results are to be found in (VIII). They were grouped as follows :

INCOME CLASSES 10^3 pesetas	HOUSEHOLDS	INCOME CLASSES 10^3 pesetas	HOUSEHOLDS
0 - 21.6	1,828	60. - 72	2,224
21.6 - 24.	364	72. - 96.	3,084
24. - 30.	1,093	96. - 120.	1,639
30. - 36.	1,416	120. - 144.	911
36. - 42.	1,485	144. - 180.	693
42. - 48.	1,573	180. - 240.	478
48. - 54.	1,461	240. - 500.	375
54. - 60.	1,461	500. - +	74
		TOTAL	20,002

Source : Encuestas de presupuestos familiares,
Madrid, Instituto Nacional de Estatistica, 1965

O.L.S. and G.L.S. results (7) are the following. (The α^* and $\sigma^*_{\alpha^*}$ are the O.L.S. estimates, the $\hat{\alpha}$ and the $\hat{\sigma}_{\hat{\alpha}}$ are the G.L.S. ones obtained as above.)

TRUNCATION POINT	α^*	$\sigma^*_{\alpha^*}$	$\hat{\alpha}$	$\hat{\sigma}_{\hat{\alpha}}$
21.6	1.351	.107	.500	.124
24.0	1.433	.108	.604	.127
30.0	1.620	.0991	.877	.130
36.0	1.778	.0896	1.112	.127
42.0	1.913	.0792	1.352	.116
48.0	2.022	.0704	1.541	.107
54.0	2.114	.0624	1.716	.0941
60.0	2.188	.0562	1.861	.0833
72.0	2.309	.0397	2.129	.0515
96.0	2.429	.0163	2.367	.0260
120.0	2.478	.00477	2.478	.00635
144.0	2.484	.00827	2.498	.00466
180.0	2.478	.0119	2.495	.00621

These results show that the Paretian base income may be fixed at 120,000 pesetas with a corresponding $\alpha = 2.48$. For the Paretian range both estimators give similar α. For the truncated Pareto chosen,, with base income 120,000 pesetas, the variance obtained by G.L.S. is higher than the O.L.S. one. This is not surprising because, in case of + autocorrelation, estimates of variances by O.L.S. tend to be biased downwards Once the distribution is truncated at lower bounds, α^* and $\hat{\alpha}$ widely diverge because the estimated variance-covariance matrix $\hat{\Lambda}^{-1}$ weights heavily the first observations.

A last comparison of both estimates refers to their predictive value. We compared F(X) obtained by O.L.S. and G.L.S. to evaluate their relative accuracy in fitting the distribution truncated at 120,000 pesetas, the base income.

Economists studying industrial concentration are essentially interested by the tail of the distribution, others do not want to privilege one part of it. We have therefore computed absolute differences between actual cumulative frequencies (the Kolmogorov-Smirnov procedure) and the ratio between the two corresponding frequencies that was likely to take large value at the tail as a consequence of the small number of individuals.

Results are as follows :

CLASS INTERVAL	Fact.	Fg.1.s.	Fo.1.s.	Fg.1.s. - Fact.	Fo.1.s. - Fact.	Fg.1.s. / Fact.	Fo.1.s. / Fact.
120.0 - 144.0	.360	.363	.364	.003	.004	1.008	1.011
144.0 - 180.0	.634	.633	.634	.001	.000	.987	.985
180.0 - 240.0	.823	.820	.821	.003	.003	.989	.989
240.0 - 500.0	.971	.971	.971	.000	.000	1.015	1.014
over 500.0	1.000	1.000	1.000	.000	.000	.996	.996

Those results do not allow any discrimination between the two estimators.

4. USE OF α FOR DISPERSION AND CENTRAL MEASURES

α has long been known as a dispersion index. In particular for $\alpha > 1$ the ratio of the open interval mean to its lower bound is :

$$\frac{\bar{X}}{\xi} = \frac{\alpha}{\alpha - 1}$$

When α is smaller than 1, $\bar{X}$ is infinite.

Another index may be worked out for discrete empirical distributions well approximated by Pareto ones : the well known concentration ratio.

We know that $F(X) = 1 - \left(\frac{X}{\xi}\right)^{-\alpha}$

Let us write $F(X) = q$, the corresponding quantile X_q is derived

$$1 - q = \left(\frac{X_q}{\xi}\right)^{-\alpha}$$

$$X_q = \xi \left(\frac{1}{1-q}\right)^{\frac{1}{\alpha}}$$

The concentration ratio is defined as the share of the N largest individuals in the sample. Its computation is straightforward provided one has an estimated size for each of these individuals. Let us write $(1 - q) = \frac{1}{N}$. Our $\hat{\alpha}$ gives us, by (8), an estimate of the quantile such that i individuals are larger than it. If we make the assumption that the lowest of those has the size of the quantile, we may generate an estimated size for each of the top observations $X_i = \xi \left(\frac{N}{i}\right)^{\frac{1}{\alpha}}$, where N is the number of individuals above the truncation point. In the truncated distribution, the concentration ratio is expressed :

$$CR_n = \frac{1}{\sum_{i=1}^{N} X_n} \xi \left[N^{\frac{1}{\alpha}} + \frac{N}{2}^{\frac{1}{\alpha}} + \ldots + \frac{N}{n}^{\frac{1}{\alpha}} \right] \qquad (9)$$

To use (9) for distribution paretian in their upper part only, we have to consider that N is the total number of individuals, provided that the Pareto law applies to the n

largest individuals at least[1] . Another dispersion measure, proposed by H. Lyddal (IX, p.139), the percentiles, and defined as $100 \frac{X_q}{X_M}$, where X_M is the median, is immediately computed :

$$P_{1-q} = 100 \left(\frac{.5}{1-q}\right)^{\frac{1}{\alpha}}$$

$$\text{as } X_M = \left(\frac{1}{.5}\right)^{\frac{1}{\alpha}}$$

for strictly paretian distributions.

In the asymptotic case, the formula becomes

$$P_{1-q} = \frac{100}{X_M} \xi \left(\frac{PW}{1-q}\right)^{\frac{1}{\alpha}}$$

where X_M is graphically determined

ξ is the truncation point or class frontier at the base, and

PW the Pareto Weight (cfr.p.29).

---

[1] This procedure may be extended to any measure of dispersion, like the entropy, and adapted to any distribution. For those having a first moment, it is better to use it, as did P. Pashygian (XI), unless data are grouped according to their real value and not to arbitrary limits of classes.
When total effectives are not available, one may rely on the formula derived, along the same line of reasoning as ours, by B. Mandelbrot and D. Orr (X).
The above expression verifies that, in a Pareto distribution, two observations of rank i and j are related by the formula

$$\frac{x_i}{x_j} = \left(\frac{j}{i}\right)^{1/\alpha}$$

cfr. Y. Ijiri and H.A. Simon (VI). An empirical mean might be generated by this quantile method also if α were smaller than 1. Eventually let us notice that our assumption about the size of individuals is consistent with the fact that the smallest would be equal to the truncation point.

The last symbol is introduced because

$$X_q^o = X^o_{1-(1-q)} = X^{Pa}_{1-(\frac{1-q}{PW})} = \xi \left(\frac{PW}{1-q}\right)^{\frac{1}{\alpha}}$$

where o refers to the overall distribution and Pa to its Paretian part.

From our samples, we obtained :

Percentiles	Spanish Incomes	Belgian Incomes
P_1	585	707
P_2	443	504
P_5	306	321
P_{10}	231	229

With respect to those of other occidental countries, these distributions look quite unequalitarian (cfr.the Appendix of (IX)). Comparing them, we notice that the smaller belgian α does not affect P_{10}. This is due to the wider paretian range in the belgian case; for the same α, the smaller is ξ, the more incomes are equally distributed (cfr. a double-log graph of cumulated income frequencies). As P_{10} roughly takes the same value in both countries, it is straightforward that higher percentiles must take larger values in Belgium, the distribution of relevant incomes being determined by α's only.

The greater inequality found in Belgium might look somewhat surprising. Part of the explanation could be that people enjoying top incomes might be reluctant to fill the forms required for budget enquiries. One must also take into account the fact that, unlike spanish data, the belgian ones do not include social benefits that are more equally distributed than labour or capital income.

For central tendency indexes, like the mean, difficulties arise from the non paretian character of empirical distributions, in most of their range. To overcome it, we used a simple device. The mean was defined :

$$\bar{X} = \frac{N_1}{N}\left(\frac{U_1+L_1}{2}\right) + \ldots + \frac{N_{p-1}}{N}\left(\frac{U_{p-1}+L_{p-1}}{2}\right) + \sum_{i=p}^{T} N_i \bar{X}_p \quad \text{(see VI,p.57)}$$

where N_i is the number of individuals belonging to class i;
N, the total number of individuals;
p, the index of the first paretian class;
$\bar{X}$, the paretian mean of the distribution truncated at the base income[1].

For Spanish incomes in 1964, we obtained $\bar{X}$ = 73,638 pesetas to be compared with 75,139 reported after direct computation. This result shows that the approximation is reasonably reliable. Better estimates might be obtained by describing

---

[1] If some classes contain no element, $\hat{\Lambda}^{-1}$ cannot be computed. Therefore we censored the cumulative frequency distributions to discard those classes but they were taken into account in the computation of the mean.

by one statistical law the non-paretian part of the distribution. A good candidate might be the truncated lognormal. Fitting a lognormal on the whole distribution of spanish incomes by the so-called graphical method (II, p.31), we obtained $\overline{X}$ = 74,796 pesetas, result that is worth of special notice as the ratios of the lognormal frequencies to the actual ones were as follows.

CLASS INTERVAL	$F_{act.}$ $f(x)$	$F_{log\text{-}normal}$	$F_{log\text{-}normal} / F_{act.}$
0 - 21.6	.0911	.0872	.957
21.6 - 24.0	.0181	.0255	1.405
24.0 - 30.0	.0545	.0705	1.293
30.0 - 36.0	.0706	.0744	1.054
36.0 - 42.0	.0740	.0733	.991
42.0 - 48.0	.0784	.0694	.885
48.0 - 54.0	.0728	.0641	.879
54.0 - 60.0	.0680	.0581	.855
60.0 - 72.0	.1109	.0988	.891
72.0 - 96.0	.1537	.1390	.904
96.0 - 120.0	.0817	.0850	1.040
120.0 - 144.0	.0454	.0525	1.156
144.0 - 180.0	.0354	.0450	1.302
180.0 - 240.0	.0238	.0333	1.398
240.0 - 500.0	.0187	.0227	1.215
over 500.0	(.0037) $1 - \Sigma F(act)$	.0014	.382

Those figures clearly show thet the lognormal poorly fits the paretian part of the spanish income distribution. Therefore, it might pay to estimate the mean by weighted average of a paretian and a truncated lognormal one.

5. COMPOSITION AND USE OF THE PROGRAM "NANETTE"

The program "NANETTE" is written in FØRTRAN for the I.B.M.-360/44 System. It consists of a main routine and of seven subroutines, "TR1N", "GLSP1R", "ADJ1", T1BLE", "DW1T".

The main program reads the information of the user who follows steps a-b-c.

a) It records

"T" : the number of points in the regression (i.e. the number of observations minus two);

"IFSCAN" : a code which if different from zero causes a scanning on truncations ranging from TRØN + 1 (i.e. first truncation chosen, to T);

"TRØN" : the index of the observation at the base of the truncation if it is higher than zero;

"T", "IFSCAN" and "TRØN" are read on a format card (313).

b) The program then records

- a format card of the class frontiers ("FXØ": format 20A4);
- a format card of the frequency distributions ("FNF" : format 20A4);
- the set of lower class frontiers (T+2 data) according to the above "FXØ" format.

c) For each distribution which he wants to process the user introduces

- a title card ("NAME" : format 20A4);
- the set of frequency distributions (T+2) according to the above "FNF" format).

The card following the last set of frequency distributions,

- if punched with the word "NEXT" from column one to four causes the return to a;
- if punched with the word "ØUT" from column one to three causes Job cancellation.

The main program - carries out the transformation of of the data unaffected by the truncation point and organizes the scanning if indicated by the users (with "IFSCAN");

- computes the weights of the elements of the mean for the non-paretian part of the distribution;

- rearranges the data should frequencies in the middle of the series be null. Series with begin or end leg with null frequencies should be read from the first non-zero frequency to the last one. (Information for this should be given in a-b-c cards.)

The sub-routine "TRIN" performs the transformation of the data that were not run on the main program (see point 2 and 3 of this paper).

The sub-routine "GLP1R" computes the parameters of the regression by the generalized least-squares method on data computed in "TR1N". This is done following point 2 of this paper.

It computes also the raw estimated frequencies and the means of both the paretian and the non-paretian part of the distribution as well as the overall mean.

Sub-routine "GLP1R" calls sub-routines "ADJ1" and "DW1T".

"ADJ1" gives the R-squared and the R-squared ajusted for the degrees of freedom according to the formula $R^2-(\frac{1}{N+1})\ (1-R^2)$.

"DW1T" computes the Durbin-Watson statistic.

Sub-routine GLSP1R" also calls "T1BLE" at the end of the scanning in case the users asked for (see : "IFSCAN"). The sub-routine prints out a summary of results according to the different truncations.

6. PRINT-OUT OF THE PROGRAM, OF THE INPUT TO BE PUNCHED BY THE USER AND THE CORRESPONDING OUTPUT.

The print-out hereafter refers to belgian data on households incomes, before tax, in 1963. Comments are to be found in (III). The chosen truncation point is 80,000 B.F. We obtained an α of 2.04 with a standard deviation of 0.03. 39 % of all households belong to the paretian part of the whole distribution. Estimated mean income amounts to 91,910 B.F. per annum and the corresponding paretian mean is 157,000 B.F. per annum.

```
C PRØGRAM NANETTE
      REAL ※ 8 XØ(30),X(30),Y(30),V(30),L(30,30),A(30),W(32),MXØ(30),
     1LXØ(30),L1XØ(30),LW(32)
      REAL※4    N(30)          ,NF(30),TSN,LNF(30)
      INTEGER※4 ØUT, NEXT
      INTEGER※2 I,J,H,HI(30),T,NN,LNN,K,P,R,TRØN,NNN
      CØMMØN SN
      DATA IØ,IP,IN/3,2,1/
      DATA ØUT/'ØUT'/
      DATA NEXT/'NEXT'/
      DIMENSION NAME (30),FXØ(20),FNF(20)
    4 READ(IN,99)T,IFSCAN,TRØN
   99 FØRMAT(513)
      LNN=T+2
      READ(IN,98)FXØ
      READ(IN,98)FNF
      READ(IN,FXØ)(L1XØ(I),I=1,LNN)
    2 P=TRØN+3
      K=P-1
      R=P-2
      DØ 10 I=1,LNN
   10 LXØ(I)=L1XØ(I)
      READ(IN,98)(NAME(I),I=1,20)
   98 FØRMAT(20A4)
      WRITE(IØ,198)(NAME(I),I=1,20)
  198 FØRMAT(1H1,20A4/)
      IF(NAME(1).EQ.ØUT)CALL EXIT
      IF(NAME(1).EQ.NEXT)GØ TØ 4
      READ(IN,FNF)     (LNF(I),I=1,LNN)
      IF(LNF(LNN).EQ.0.0)WRITE(IØ,199)
  199 FØRMAT(1H1,4X,27('※')//1H ,'※ ERROR. YOUR LAST FREQ.=0 ※'/
     F1H ,4X,27('※'))
      TSN=0.0
      DØ 22 I=1,LNN
   22 TSN=TSN+LNF(I)
      WRITE(IØ,202)
  202 FØRMAT(1HO,'ABSØLUTE FREQUENCY IN THE I-TH INTERVAL',
     F9X,'LØWER LIMIT ØF THE I-TH INTERVAL'/);
      LW(1)=((LXØ(1)+LXØ(2))/2.D0)※(LNF(1)/TSN)
      DØ 12 I=2,LNN
   12 LW(I)=LW(I-1)+((LXØ(I)+LXØ(I+1))/2.D0)※(LNF(I)/TSN)
      DØ 13 I=1,LNN
      H=I-2
   13 WRITE(IØ,203)(H,LNF(I),LXØ(I))
  203 FØRMAT(1H ,I2,I2,9X,F10.3,11X))
      NN=LNN
      DØ 32   I=1,NN
      NF(I)=LNF(I)
      XØ(I)=LXØ(I)
      W(I)=LW(I)
      IF(LNF(I).NE.0.0) GØ TØ 30
```

```
      NNN=NN-I
      H=1
      DØ 33 J=1,NNN
      IF(LNF(I+J).NE.0.0.)GØ TØ 34
   33 H=H+1
   34 NN=NN-H
      XØ(I)=LXØ(I+H)
      NF(I)=LNF(I+H)
      W(I)=LW(I+H)
      DØ 31 J=I,NN
      LNF(J)=LNF(J+H)
      LW(J)=LW(J+H)
   31 LXØ(J)=LXØ(J+H)
   30 CØNTINUE
   32 CØNTINUE
      IF(NN.EQ.LNN)GØ TØ 40
      WRITE(IØ,204)
  204 FØRMAT(1HØ,'TAKING ØUT NULL FREQUENCIES'/)
      WRITE(IØ,202)
      DØ 41 I=1,NN
      H=I-2
   41 WRITE(IØ,203)(H,NF(I),XØ(I))
   40 CØNTINUE
    1 CØNTINUE
      WRITE (IØ,195)R
  195 FØRMAT(1H1,'SET NUMBER',I3/)
      DØ 3 I=K,NN
    3 X(I)=XØ(I)/XØ(K)
      WRITE(IØ,197)XØ(K)
  197 FØRMAT(1HO,'CLASS FRØNTIER AT THE BASE ØF THE TRUNCATIØN=',D15.9)
      CALL TRIN(X,NF,N,V,Y,   P,NN)
      CALL GLSP1R(X,N,V,Y,  A,P,NN,T,XØ,NF,W,TSN,TRØN,NAME)
      P=P+1
      K=P-1
      R=P-2
      IF(P.GT.(NN-1)) GØ TØ 2
      IF(IFSCAN.EQ.0) GØ TØ 2
      GØ TØ  1
      END



      SUBROUTINE TRIN(X,NF,N,V,Y,   P,NN)
      CØMMØN SN
      REAL*8 X(1),Y(1),V(1)
      REAL*4 N(1),    NF(1)
      INTEGER*2 I,J,H,HI(30),NN,P,K
      DATA IØ,IP,IN/3,2,1/
      K=P-1
      WRITE(IØ,201)
```

```
201 FØRMAT(1HØ,5X,'SAMPLE SIZE',22X,'RELATIVE FREQUENCY',3X,
   F3X,'GREATER THAN REL.FREQ.',3X,'RATIØ ØF LØWER CLASS FRØNT',/
   F1H,91X,'TO BASE CLASS FRØNT'/)
    SN=NF(K)
    DØ 1 I=P,NN
  1 SN=SN+NF(I)
    DØ 2 I=K,NN
    HI(I)=I-K
    V(I)=0.0
  2 N(I)=NF(I)/SN
    DØ 3 I=K,NN
    DØ 3 J=I,NN
  3 V(I)=V(I)+N(J)
    WRITE(IØ,202)SN,(HI(I),N(I),V(I),X(I),I=K,NN)
202 FØRMAT(1HØ,4X,E12.5/(1H,31X,I2,8X,E12.5,12X,D12.5,17X,D12.5)))
    WRITE(IØ,205)
205 FØRMAT(1HO,'VARIABLES USED FØR REGRESSIØN',3X,'LØG Y',9X,'LØG X'/)
    DØ 4 I=P,NN
    H=I-P+1
    Y(I)=DLØG(V(I))
    X(I)=DLØG(X(I))
  4 WRITE(IØ,206)H,Y(I),X(I)
206 FØRMAT(1H ,15X,I2,9X,2(D15.9,2X))
    RETURN
    END



    SUBRØUTINE GLSP1R(X,N,V,Y,  A,P,NN,T,XØ,NF,W,TSN,TRØN,NAME)
    CØMMØN SN
    REAL*8 X(1),Y(1),V(1),R(30),YEST(31),VLV,SDIF,DIFX2(30),DIFXA(30)
    REAL*8 XØ(1),W(1),MM,M(30),MU(30),RC(30),WE(30),SHA(30),A(30),R2C
    REAL*4 N(1)    ,NF(1),TSN
    DIMENSIØN NAME(1)
    INTEGER*2 I,J,H,HI(30),NN,T,K,P,TRØN
    DØUBLE PRECISIØN L(30,30),SHSA,XL(30),XLX,XLY,VL(30),E
    DATA IØ,IP,IN/3,2,1/
    DØ 1 I=1,NN
    DØ 1 J=1,NN
  1 L(I,J)=0.DO
    DØ 2 I=P,NN
    HI(I)=I-P+1
    L(I,I)=V(I)**2*(1.DO/N(I-1)+1.DO/N(I))
    L(I-1,I)=-V(I)*V(I-1)*(1.DO/N(I-1))
  2 L(I,I-1)=L(I-1,I)
    WRITE(IØ,201)
201 FØRMAT(1HO,'INVERTED VARIANCE-COVARIANCE MATRIX OF RESIDUALS'/)
    DØ 3 I=P,NN
    H=I-P+1
```

```
  3 WRITE(IØ,202)H,(HI(J),L(I,J),J=P,NN)
202 FØRMAT(1HØ,I3,6(5(2X,I3,2X,E15?9),/4X))
    DØ 11 I=P,NN
    XL(I)=0.DO
    DØ 11 J=P,NN
 11 XL(I)=XL(I)+X(J)*L(J,I)
    XLX=0.DO
    XLY=0.DO
    DØ 22 I=P,NN
    XLX=XLX+XL(I)*X(I)
 22 XLY=XLY+XL(I)*Y(I)
    WRITE(IØ,210)
210 FØRMAT(1HO, 4X , '   LOG Y       ',3X , 'EST LOG Y      ',
   F3X,'RESIDUALS'/)
    A(P)=-(1.DO/XLX)*XLY
    DO 33 I=P,NN
    H=I-P+1
    R(I)=Y(I)+A(P)*X(I)
    YEST(I)=-A(P)*X(I)
 33 WRITE(IØ,207)H,Y(I),YEST(I),R(I)
    CALL ADJ1(YEST,Y      ,P,NN,R2C)
    CALL DW1T(R,P,NN)
    RC(P)=R2C
    E=0.271828D01
    DØ 44 I=P,NN
    VL(I)=0.DO
    YEST(I)=E**YEST(I)
    DØ 44 J=P,NN
 44 VL(I)=VL(I)+R(J)*L(J,I)
    VLV=0.DO
    DØ 55 I=P,NN
 55 VLV=VLV+VL(I)*R(I)
    SHSA=VLV*(1.DO/XLX)*(1.DO/(DFLØAT(NN+P-6)))
    SHA(P)=DSQRT(DABS(SHSA))
    K=NN-1
    WRITE(IØ,204)
204 FØRMAT(1HO,5X,'ALFA STAR   SIGMA HAT       SQRT')
    WRITE(IØ,205)A(P),SHSA,SHA(P)
205 FØRMAT(1HØ,2X,5(D12.6,2X))
    WRITE(IØ,206)
206 FØRMAT(1HO,4X,'ACTUAL G.T.FREQUENCY',3X,'EST. G.T. FREQUENCY'/)
    DØ 4 I=P,NN
    H=I-P+1
  4 WRITE(IØ,207)H,V(I),YEST(I)
207 FØRMAT(1HØ,I2,6X,3(E15.9,3X))
    WRITE(IØ,208)
208 FØRMAT(1HO,44,'ACTUAL FREQUENCIES  ',3X,'ESTIMATED FREQUENCIES'/)
    V(NN+1)=0.DO
    YEST(NN+1)=0.DO
    DØ 5 I=P,NN
    DIFX2(I)=(V(I)-V(I+1))*SN
    DIFXA(I)=(YEST(I)-YEST(I+1))*SN
    H=I-2
```

```
  5 WRITE(IØ,207)H,DIFX2(I),DIFXA(I)
    MU(P)=XØ(P-1)**A(P)/(DABS(A(P)-1.DO))
    WE(P)=(SN/TSN)**1.DO
    MM=W(P-2)**(TSN/(TSN-SN))
    M(P)=(MU(P)**WE(P))+(MM**(1.DO-WE(P)))
    WRITE(IØ,209)MM,WE(P),MU(P),M(P)
209 FØRMAT(1HO,'NØN PARETIAN MEAN          =',D19.6/
   F 1HO,'PARETØ WEIGHT                   =',E19.6/
   F 1HO,'MEAN ØF THE ESTIMATED PARETØ    =',D19,6/
   F 1HO,'ØVERALL MEAN                    =',D19.6)
    IF(P.EQ.NN-1)CALL TAB1L(XØ,A,WE,MU,M,RC,SHA,TRØN,NAME,NN)
    RETURN
    END



    SUBROUTINE ADJ1  (YE,YA       ,NI,NN,R2C)
    REAL**8 YE(1),YA(1),R2,R2C,SYE,SYA,SYEDM,SYADM
    INTEGER**2       NI,NN
    IØ=3
    SYE=0.DO
    SYA=0.DO
    SYEDM=0.DO
    SYADM=0.DO
    DO 1 I=NI,NN
    SYE=SYE+YE(I)
  1 SYA=SYA+YA(I)
    SYA=SYA/(NN-NI+1.DO)
    SYE=SYE/(NN-NI+1.DO)
    DO 2 I=NI,NN
    SYEDM=SYEDM+(YE(I)-SYE)**2
  2 SYADM=SYADM+(YA(I)-SYA)**2
    R2=SYEDM/SYADM
    R2C=R2-(1.DO/(NN-NI))* (1.DO-R2)
    WRITE(IØ,200)      R2,R2C
200 FØRMAT (1HO,'R-SQUARE= ',D15.9/1H ,'CØRRECTED= ',D15.9)
    RETURN
    END



    SUBROUTINE DWIT(R,P,NN)
    REAL**8 R(30),SDR2,SR2,DW
    INTEGER**2 P,NN,K
    SR2=0.DO
    SDR2=0.DO
    IØ=3
    K=P+1
    DØ 1 I=K,NN
  1 SDR2=SDR2+((R(I)-R(I-1))**2)
    DØ 2 I=P,NN
```

```
      2 SR2=SR2+(R(I)**2)
        DW=SDR2/SR2
        WRITE(IØ ,200) DW
    200 FØRMAT(1HO,'DURBIN-WATSØN=',2X,D12.4/)
        RETURN
        END



        SUBRØUTINE TAB1L(XØ,A,WE,MU,M,RC,SHA,TRØN,NAME,NN)
        INTEGER*2 TRØN,NN,P,NNN,IN
        REAL*8 XØ(1),A(1),WE(1),MU(1),M(1),RC(1),SHA(1)
        IØ=3
        DIMENSION NAME(1)
        WRITE(IØ,300)(NAME(I),I=1,20)
    300 FØRMAT(1H1,8X,20A4/1HO,3X,
       F                    'CLASS FRØNTIER',5X,'ALFA',7X,'PARETØ WEIGHT'
       1 ,2X,'PARETØ MEAN',4X,'ØVERALL MEAN',3X,'CØRRECTED R-SQUARE'/
       21HØ,4X,'AT THE BASE',2X,'(STAND.DEV.)'/)
        IN=TRØN+3
        NNN=NN-1
        DØ 1 P=IN,NNN
      1 WRITE(IØ,301)XØ(P-1),A(P),WE(P),MU(P),M(P),RC(P),SHA(P)
    301 FØRMAT(1HØ,6(3X,D10.4,2X)/1HØ,17X,'(',D10.4,')'/)
        RETURN
        END
```

INPUT:

```
 22 99
(23F3.0,F4.0)
(13F6.3/13F6.3)
  5 25 30 35 40 45 50 60 70 80 90100110130150175200225250300400500750 1000
BELGIUM:STAT.FISC.DS.REV.GLØB.NETS TAXABLES DS.PERS.PHYS. 1963(CUM.REV.EPØUX)INS
130822 928921221891347521494301636103513933078112506752075081759801325321703 07
102365 81472 51100 45412 24117 30818 31099 14670 14049  4917  6345
ØUT
```

OUTPUT

BELGIUM'STAT.FISC.DS.REV.GLOB.NETS TAXABLES DS.PRES.PHYS.'1963(CUM.REV.EPOUX)INS

	ABSOLUTE FREQUENCY IN THE I-TH INTERVAL	LOWER LIMIT OF THE I-TH INTERVAL
-1	130.822	5.000
0	92.892	25.000
1	122.189	30.000
2	134.752	35.000
3	149.430	40.000
4	163.610	45.000
5	351.393	50.000
6	307.811	60.000
7	250.675	70.000
8	207.508	80.000
9	175.980	90.000
10	132.532	100.000
11	170.307	110.000
12	102.365	130.000
13	81.472	150.000
14	51.100	175.000
15	35.412	200.000
16	24.117	225.000
17	30.818	250.000
18	31.099	300.000
19	14.670	400.000
20	14.049	500.000
21	4.917	750.000
22	6.345	1000.000

SET NUMBER 11

CLASS FRONTIER AT THE BASE OF THE TRUNCATION=0.100000000D 03

SAMPLE SIZE		RELATIVE FREQUENCY	GREATER THAN REL.FREQ.	RATIO OF LOWER CLASS FRONT.TO BASE CLASS FRONT.
0.69920E 03				
	0	0.18955E 00	0.10000D 01	0.10000D 01
	1	0.24357E 00	0.81045D 00	0.11000D 01
	2	0.14640E 00	0.56688D 00	0.13000D 01
	3	0.11652E 00	0.42048D 00	0.15000D 01
	4	0.73083E-01	0.30396D 00	0.17500D 01
	5	0.50646E-01	0.23087D 00	0.20000D 01
	6	0.34492E-01	0.18023D 00	0.22500D 01
	7	0.44076E-01	0.14573D 00	0.25000D 01
	8	0.44478E-01	0.10166D 00	0.30000D 01
	9	0.20981E-01	0.57181D-01	0.40000D 01
	10	0.20093E-01	0.36200D-01	0.50000D 01
	11	0.70323E-02	0.16107D-01	0.75000D 01
	12	0.90746E-02	0.90746D-02	0.10000D 02

VARIABLES USED FOR REGRESSION	LOG Y	LOG X
1	-.210160980D 00	0.953101798D-01
2	-.567606930D 00	0.262364264D 00
3	-.866363586D 00	0.405465108D 00
4	-.119087086D 01	0.559615788D 00
5	-.146588679D 01	0.693147181D 00
6	-.171353883D 01	0.810930216D 00
7	-.192596740D 01	0.916290732D 00
8	-.228613366D 01	0.109861229D 01
9	-.286153535D 01	0.138629436D 01
10	-.331870064D 01	0.160943791D 01
11	-.412850569D 01	0.201490302D 01
12	-.470227285D 01	0.230258509D 01

INVERTED VARIANCE-COVARIANCE MATRIX OF RESIDUALS

1	1	0.616194427D 01	2	-.188620933D 01	3	0.0	4	0.0	5	0.0
	6	0.0	7	0.0	8	0.0	9	0.0	10	0.0
	11	0.0	12	0.0						
2	1	-.188620933D 01	2	0.351432679D 01	3	-.162811747D 01	4	0.0	5	0.0
	6	0.0	7	0.0	8	0.0	9	0.0	10	0.0
	11	0.0	12	0.0						
3	1	0.0	2	-.162811747D 01	3	0.272497164D 01	4	-.109685418D 01	5	0.0
	6	0.0	7	0.0	8	0.0	9	0.0	10	0.0
	11	0.0	12	0.0						
4	1	0.0	2	0.0	3	-.109685418D 01	4	0.205706539D 01	5	-.960211210D 00
	6	0.0	7	0.0	8	0.0	9	0.0	10	0.0
	11	0.0	12	0.0						
5	1	0.0	2	0.0	3	0.0	4	-.960211210D 00	5	0.178178263D 01
	6	-.821571421D 00	7	0.0	8	0.0	9	0.0	10	0.0
	11	0.0	12	0.0						
6	1	0.0	2	0.0	3	0.0	4	0.0	5	-.821571421D 00
	6	0.158305768D 01	7	-.761486263D 00	8	0.0	9	0.0	10	0.0
	11	0.0	12	0.0						
7	1	0.0	2	0.0	3	0.0	4	0.0	5	0.0
	6	-.761486263D 00	7	0.109761530D 01	8	-.336129032D 00	9	0.0	10	0.0
	11	0.0	12	0.0						
8	1	0.0	2	0.0	3	0.0	4	0.0	5	0.0
	6	0.0	7	-.336129032D 00	8	0.466821924D 00	9	-.130692891D 00	10	0.0
	11	0.0	12	0.0						
9	1	0.0	2	0.0	3	0.0	4	0.0	5	0.0
	6	0.0	7	0.0	8	-.130692891D 00	9	0.229350400D 00	10	-.986575082D-01
	11	0.0	12	0.0						
10	1	0.0	2	0.0	3	0.0	4	0.0	5	0.0
	6	0.0	7	0.0	8	0.0	9	-.986575082D-01	10	0.127676117D 00
	11	-.290186086D-01	12	0.0						

```
 11  1 0.0                    2 0.0                    3 0.0                    4 0.0                    5 0.0
     6 0.0                    7 0.0                    8 0.0                    9 0.0                   10 -.290186086D-01
    11 0.498033268D-01   12 -.207847182D-01
 12  1 0.0                    2 0.0                    3 0.0                    4 0.0                    5 0.0
     6 0.0                    7 0.0                    8 0.0                    9 0.0                   10 0.0
    11 -.207847182D-01   12  0.207847182D-01
```

	LOG Y	EST LOG Y	RESIDUALS
1	-.210160980D 00	-.200637004D 00	-.952397546D-02
2	-.567606930D 00	-.552301760D 00	-.153051700D-01
3	-.866363586D 00	-.853542663D 00	-.128209237D-01
4	-.119087086D 01	-.117804452D 01	-.128263375D-01
5	-.146588679D 01	-.145914082D 01	-.674597453D-02
6	-.171353883D 01	-.170708533D 01	-.645350844D-02
7	-.192596740D 01	-.192887924D 01	0.291184433D-02
8	-.228613366D 01	-.231268348D 01	0.265498220D-01
9	-.286153535D 01	-.291828164D 01	0.567462882D-01
10	-.331870064D 01	-.338802006D 01	0.693194251D-01
11	-.412850569D 01	-.424156273D 01	0.113057037D-01
12	-.470227285D 01	-.484716088D 01	0.144888029D 00

R-SQUARE=0.107598759D 01

CORRECTED= 0.108289555D 01

DURBIN-WATSON= 0.1091D 00

ALFA STAR	SIGMA HAT	SORT
0.210510D 01	0.140323D-03	0.118458D-01

	ACTUAL G.T. FREQUENCY	EST. G.T. FREQUENCY
1	0.810453769D 00	0.818209495D 00
2	0.566880401D 00	0.575623551D 00
3	0.420477804D 00	0.425903669D 00
4	0.303956445D 00	0.307880447D 00
5	0.230873164D 00	0.232436122D 00
6	0.180226870D 00	0.181393935D 00
7	0.145734705D 00	0.145311154D 00
8	0.101658750D 00	0.989953969D-01
9	0.571809001D-01	0.540265506D-01
10	0.361998379D-01	0.337755611D-01
11	0.161069296D-01	0.143851354D-01
12	0.907462835D-02	0.785066042D-02

	ACTUAL FREQUENCIES	ESTIMATED FREQUENCIES
11	0.170306962D 03	0.169616553D 03
12	0.102364975D 03	0.104684427D 03
13	0.814719557D 02	0.825220612D 02
14	0.510999696D 02	0.527508161D 02
15	0.354119851D 02	0.356887941D 02
16	0.241169873D 02	0.252291491D 02
17	0.308179915D 02	0.323840656D 02
18	0.310989976D 02	0.314423030D 02
19	0.146699986D 02	0.141595305D 02
20	0.140489997D 02	0.135578226D 02
21	0.491699844D 01	0.456891733D 01
22	0.634499743D 01	0.548919671D 01

NON PARETIAN MEAN = 0.575712D 02

PARETO WEIGHT = 0.250946D 00

MEAN OF THE ESTIMATED PARETO = 0.190490D 03

OVERALL MEAN = 0.909266D 02

BELGIUM'STAT.FISC.DS.REV.GLOB.NETS TAXABLES DS.PERS.PHYS.'1963(CUM;REV.EPOUX)INS

CLASS FRONTIER AT THE BASE	ALFA (STAND.DEV.)	PARETO WEIGHT	PARETO MEAN	OVERALL MEAN	CORRECTED R-SQUARE
0.2500D 02	0.4841D 00 (0.9911D-01)	0.9530D 00	0.2346D 02	0.2306D 02	0.2527D-01
0.3000D 02	0.6821D 00 (0.1069D 00)	0.9197D 00	0.6437D 02	0.6082D 02	0.8772D-01
0.3500D 02	0.8723D 00 (0.1097D 00)	0.8759D 00	0.2390D 03	0.2124D 03	0.1634D 00
0.4000D 02	0.1081D 01 (0.1058D 00)	0.8275D 00	0.5312D 03	0.4444D 03	0.2652D 00
0.4500D 02	0.1291D 01 (0.9537D-01)	0.7739D 00	0.1997D 03	0.1617D 03	0.3851D 00
0.5000D 02	0.1480D 01 (0.8056D-01)	0.7151D 00	0.1543D 03	0.1202D 03	0.5063D 00
0.6000D 02	0.1723D 01 (0.6304D-01)	0.5890D 00	0.1430D 03	0.1011D 03	0.6958D 00
0.7000D 02	0.1899D 01 (0.4643D-01)	0.4785D 00	0.1478D 03	0.9479D 02	0.8536D 00
0.8000D 02	0.2035D 01 (0.2616D-01)	0.3886D 00	0.1573D 03	0.9191D 02	0.9919D 00
0.9000D 02	0.2110D 01 (0.1093D-01)	0.3141D 00	0.1711D 03	0.9086D 02	0.1080D 01
0.1000D 03	0.2105D 01 (0.1185D-01)	0.2509D 00	0.1905D 03	0.9093D 02	0.1083D 01
0.1100D 03	0.2084D 01 (0.9181D-02)	0.2034D 00	0.2115D 03	0.9113D 02	0.1068D 01
0.1300D 03	0.2061D 01 (0.7806D-02)	0.1423D 00	0.2525D 03	0.9137D 02	0.1049D 01
0.1500D 03	0.2052D 01 (0.8361D-02)	0.1055D 00	0.2926D 03	0.9147D 02	0.1046D 01
0.1750D 03	0.2032D 01 (0.7216D-02)	0.7628D-01	0.3445D 03	0.9163D 02	0.1027D 01
0.2000D 03	0.2023D 01 (0.7623D-02)	0.5794D-01	0.3954D 03	0.9170D 02	0.1021D 01
0.2250D 03	0.2002D 01 (0.3645D-02)	0.4523D-01	0.4496D 03	0.9182D 02	0.9952D 00
0.2500D 03	0.1998D 01 (0.3809D-02)	0.3657D-01	0.5004D 03	0.9184D 02	0.9878D 00
0.3000D 03	0.2010D 01 (0.3333D-02)	0.2551D-01	0.5971D 03	0.9182D 02	0.1004D 01
0.4000D 03	0.2019D 01 (0.4094D-02)	0.1435D-01	0.7925D 03	0.9186D 02	0.1035D 01
0.5000D 03	0.1996D 01 (0.1912D-03)	0.9084D-02	0.1002D 04	0.9196D 02	0.1004D 01

APPENDIX : On the estimation bias of the classical least - squares estimator of α

It is well-known that the estimates of classical least-squares variances are biased when the variance-covariance matrix of residuals differs from the unit matrix[1]. In the Paretian case, one may write :

$$E(\varepsilon\varepsilon') = \sigma^2\Omega = \frac{1}{N}\,\Lambda\,X \qquad (1)$$

$$\text{with } \Lambda = \left[\Pi_k^{-1}(1-\Pi_k)\right] \quad (k = \min i, j)$$

$$\sigma^2 = \frac{\sum \varepsilon_{ii}}{T} = \frac{\sum \lambda_{ii}}{NT} \quad \text{for normalisation}$$

$$\text{and } \Omega = \frac{T}{\sum \lambda_{ii}}\,\Lambda = \frac{T}{\sum \Pi_i^{-1}(1-\Pi_i)}\left[\Pi_k^{-1}(1-\Pi_k)\right]$$

The general formula :

$$ES^2\,[X'X]^{-1} = \sum_{bb} + \sigma^2(X'X)^{-1}\,X'(I-\Omega)\,X\,(X'X)^{-1} + \sigma^2\,\frac{tr(X'X)^{-1}\,X'\,(I-\Omega)\,X}{T-K-1}\,(X'X)^{-1}$$

becomes, as X'X is a scalar :

$$ES^2_{\hat{\alpha}} = \sigma^2_{\hat{\alpha}} + \frac{1}{N}\,\frac{\sum \Pi_i^{-1}(1-\Pi_i)}{T-1}\,\left(\sum x_i^{*2}\right)^{-2}\,X'\,(I-\Omega)\,X$$

---

[1] cfr. (V), p.238 onwards.

The sign of the bias is given by $X'(I-\Omega)X =$

$$\sum x_i^{*2} - \frac{T}{\sum \Pi_i^{-1}(1-\Pi_i)} \left[\sum \Pi_i^{-1}(1-\Pi_i)x_i^{*2} + 2 \sum \Pi_i^{-1}(1-\Pi_i)x_i^{*} \sum_{j=i+1}^{T} x_j^{*} \right]$$

As all x_i^* are positive, $\Pi_i^{-1}(1-\Pi_i)$ and x_i^{*2} are monotonously related and the bias must be negative. Indeed :

$$\Pi_1^{-1}(1-\Pi_1)\, X_1^{*2} + \ldots + \Pi_T^{-1}(1-\Pi_T)\, X_T^{*2} > \frac{\sum X_i^{*2}}{T} \sum \Pi_i^{-1}(1-\Pi_i)$$

and

$$\sum X_i^{*2} - \frac{T}{\sum \Pi_i^{-1}(1-\Pi_i)} \left[\frac{\sum X_i^{*2}}{T} \sum \Pi_i^{-1}(1-\Pi_i) + 2 \sum \Pi_i^{-1}(1-\Pi_i)\, X_i^{*} \sum X_j^{*} \right] =$$

$$- \frac{2T}{\sum_i \Pi_i^{-1}(1-\Pi_i)} \sum_i \Pi_i^{-1}(1-\Pi_i)\, X_i^{*} \sum_j X_j^{*} \quad < \quad 0$$

To see whether the bias was likely to affect the O.L.S. estimates seriously, we computed it for the spanish sample, under the assumption that the observed frequencies were identical to those of the population, and obtained : $ES^2_{\hat{\alpha}} - \sigma^2_{\hat{\alpha}} =$ -.00397. If one remembers that the estimates of $\sigma^2_{\hat{\alpha}}$ by O.L.S. and G.L.S. were .0000228 and .0000403 respectively, one must admit that O.L.S. estimates are unreliable.

R E F E R E N C E S

(I) D.J. AIGNER and A.S. GOLDBERGER, "Estimation of Pareto's Law from Grouped Observations", Journal of the American Statistical Association, V LXV, 1970.

(II) J. AITCHISON and J.A.C. BROWN, The Lognormal Distribution, Cambridge University Press, Cambridge, 1957, p.31.

(III) A. CLOSON, Etude comparée de la dispersion des revenus des ouvriers dans les pays de la C.E.E. à partir d'une estimation de la loi de Pareto, Mémoire de Licence, Institut des Sciences Economiques, Université Catholique de Louvain, (forthcoming).

(IV) J.S.CRAMER, Empirical Econometrics, North Holland C°, Amsterdam, 1969, ch.4.

(V) W. FELLER, An Introduction to Probability Theory T.1, Wiley, London, 1950, ch.6, sect.9.

(VI) A.S. GOLDBERGER, Econometric Theory, Wiley, London, 1964

(VII) Y. IJIRI and H.A. SIMON, "Effects of Mergers and Acquisition on Business Firm Concentration", Journal of Political Economy, V LXXIX, 1971.

(VIII) C. LLUCH, "Distribution de la renta en España, por provincias y categorias socio-económicas, según la encuesta de presupuestos familiares de 1964 - 65", (forthcoming).

(IX) H. LYDALL, The Structure of Earnings, Clarendon, Oxford, 1968.

(X) B. MANDELBROT and D. ORR, "Firms Size Distribution "Laws" and Industrial Concentration", Memorandum IBM Research Center, Yorktown Heights, New York, 1966.

(XI) P. PASHYGIAN, The Relation between the Size Distribution of Firms and the Concentration Ratio, mimeographed, 1966.

(XII) R.E. QUANDT, "Old and New Methods of Estimating Pareto Distributions", Metrika, V X, 1966.

(XIII) J. STEINDL, Random Processes and the Growth of Firms - A Study of the Pareto Law, Griffin, London, 1965.

E R R A T A

p.3, 1.8 $f = \phi + \varepsilon$ (3)

1.17 $\sum_{s=t}^{T} \phi_s = X_t^{-\alpha}$

p.4, 1.5 $A'f = A'(\phi + \varepsilon)$; $u = Y - \pi$

1.12 $\ln(1 + \frac{u_t}{\pi_t}) = \ln(1 + \frac{u_t}{X_t^{-\alpha}})$

p.6, 1.11 $X = \frac{Z}{Z_o}$

1.15 $g(X) = f[h(X)]\frac{dZ}{dX} = [\alpha Z_o^{\alpha} (Z_o X)^{-\alpha-1}] Z_o$

p.11, 1.3 $X_q = \xi(\frac{1}{1-q})^{\frac{1}{\alpha}}$ (8)

1.12 $(1-q) = \frac{i}{N}$

1.19 $CR_n = \frac{1}{\sum_{i=1}^{N} X_i} \xi \left[N^{\frac{1}{\alpha}} + \frac{N}{2}^{\frac{1}{\alpha}} + \ldots + \frac{N}{n}^{\frac{1}{\alpha}} \right]$