Income

FACULTY OF ACTUARIAL SCIENCE & ECONOMETRICS

A & E REPORT

REPORT AE 6/80

ERRORS AND DISTURBANCES AND THE PERFORMANCE OF SOME COMMON INCOME DISTRIBUTION FUNCTIONS

J.S. Cramer

University of Amsterdam

FACULTY OF ACTUARIAL SCIENCE & ECONOMETRICS

Report AE 6/80

ERRORS AND DISTURBANCES AND THE PERFORMANCE OF SOME COMMON INCOME DISTRIBUTION FUNCTIONS

J.S. Cramer

Free copies can be obtained from:
Faculty of Actuarial Science & Econometrics
Jodenbreestraat 23
1011 NH Amsterdam, the Netherlands
telephone number: 020-5254217/87

University of Amsterdam

Errors and disturbances and the performance of some common income distribution functions.

Abstract

Recent papers by Kloek & van Dijk and by McDonald & Ransom report maximum likelihood estimates of various income distribution functions and provide test statistics of goodness of fit. According to the latter virtually all models should be rejected, yet this conclusion is avoided, and rightly so for the test is too strict. It allows for sampling variation only, while in fact income distribution functions like any other econometric model are not expected to hold exactly. If this is so one should allow for the presence of errors and disturbances in analysing income distribution data.

The simplest approach is to treat observed income as the sum of a systematic component which follows some specific distribution and an independent normal error. We apply this model to three traditional two-parameter functions, viz. the Pareto, Gamma and Lognormal distribution. Since maximum likelihood estimation is impracticable we proceed by the method of moments.

The evidence of a few data sets suggests that the Pareto nor the Gamma distribution are redeemed by the introduction of an error term. The verdict on the Lognormal must be postponed; as a by-product of the analysis we find that it does rather better than was suggested some years ago by Salem & Mount. A major difficulty is that estimation by moments turns out to be a delicate operation when higher moments are involved, especially if income class frequency data must be used.

Errors and disturbances and the performance of some common income distribution functions.

Acknowledgements. I am heavily indebted to Bernard van Praag for kindly providing me with full statistics of the sample of incomes of Belgian economists, to my students W.S. de Vries and J.G.M. Boos for diligently executing most of the analyses, and to T. Mourik for programming the simulations of page 10 . I take full responsibility for all errors and mistakes.

1. Introduction

As a spate of recent papers on the subject shows, the quest for a moderately simple function that will fit the income distribution continues with nineteenth century zest. Old functions are being revived, as by Salem and Mount [11] , and new functions are put forward, as by Singh and Maddala [12]; moreover the advent of the computer has permitted a marked improvement in the technique of estimation. Both Kloek and van Dijk [5],[6],[7] and McDonald and Ransom [8] have reported maximum likelihood estimates of a wide range of functional forms for a variety of samples. These analyses all yield values of the chi-square test statistic of goodness of fit, and in the majority of cases these indicate that the model under review must be rejected.

Kloek and van Dijk show a good deal of concern over this matter. They typically consider samples of up to a thousand incomes, classified by about ten income classes, that are homogeneous in respect of such variables as occupation, age, family size or the level of education. Among this wealth of empirical material the goodness of fit varies considerably between samples, yet all models must be rejected more often than would follow from the significance level adopted, and the favourable results must be discounted because they tend to refer to the smaller samples. The authors conclude already in their first paper that "one might conjecture that very large sample sizes will lead to a rejection of (almost) any family of distributions if the five percent rule is adopted irrespective of sample size" ([6],p.72). Their later analyses, however valiant, do not materially alter the situation.

McDonald and Ransom use the frequency distributions by 10 income classes of the U.S.Bureau of the Census [13] which is based on a vast sample. They are mainly interested in a comparison of estimation methods and models, and they comment only in passing on the absolute level of the chi-square values which they report. In the case of scoring or maximum likelihood estimation these provide valid test statistics of goodness of fit with seven or eight degrees of freedom, depending on the number of parameters in the model considered. Over four functions and eight annual samples the least chi-square is 43, and values of well over a hundred are recorded even for the model that does best. ([8], p.1518-1519). There is no doubt that all models must be unambiguously rejected on this score, as the authors recognize in a footnote.

If these conclusions were taken seriously there would be little sense in a further discussion of the relative performance of the various models, and a case could be made for abandoning the subject altogether. The reason that this course is seldom advocated is that the chi-square test is clearly much too strict; it allows for sampling variation only, and thus tests the hypothesis that the income distribution function under review holds exactly. If this were the appropriate hypothesis any observed income distribution with cell frequencies of a few hundred incomes or more should conform very closely indeed with the theoretical model, and all the proposed functions could be rejected out of hand against the evidence of nationwide data. But this is not what most practitioners of income distribution studies have in mind. As Kloek and van Dijk put it, "it will (almost certainly) be impossible to find a family of distributions without specification error" ([7], p.442). Income distribution functions, in other words, are put forward as an approximate descriptive device, and just like other econometric models such functions cannot be expected to hold exactly.

If this view is accepted the direct application of chi-square tests to specific income distribution functions is invalid since the formulation of the null hypothesis is much too strict. If we wish to capture what is intended we must revise the income distribution model so as to allow for the approximate nature of the proposed functional forms. The simplest approach is to introduce discrepancies between observed incomes and their systematic component, which should conform strictly to some given function. These discrepancies cover shocks and disturbances, such as institutional rigidities, as well as the straightforward reporting errors that have recently been brought to the fore by van Praag et al. [10], but for brevity we shall refer to them as errors.

This change in the statistical model will of course affect the parameter estimates as well as the assessment of the goodness of fit of the various income distribution functions considered. It may even perhaps improve the standing of some of the older two-parameter income distributions that have been seriously discredited by recent research. We explore this possibility in the present paper by attaching error terms to the Pareto, Gamma and Lognormal distributions.

2. The method of analysis

We use simple models. In the <u>additive error</u> model observed income y is the sum of two independent variates,

$$y = x + z; \tag{1}$$

here x is the systematic component, which follows one of three alternative two-parameter distributions - the Pareto, Gamma and Lognormal - and z is the error with a normal $(0, \sigma(z))$ distribution. The distribution of the observed y is therefore determined by three parameters.

The proper way to estimate these parameters would be to derive y's theoretical distribution and then to apply the appropriate maximum likelihood technique to observed income class frequencies or to individual income data. Unfortunately, however, analytical expressions for the density of y cannot be given, nor can we proceed by means of the characteristic function, as has been done by van Dijk and Kloek [5], since this is not available for the Pareto or the Lognormal distribution. We shall therefore use the <u>method of moments</u>, which is particularly well suited to the problem in hand. The parameter estimates are derived by equating three observed moments of y to their theoretical counterparts, and solving for the parameters. The required expressions are obtained from the definitions

$$\mu_j'(y) = E(y^j), \tag{2}$$

$$\mu_j(y) = E\{y - \mu_1'(y)\}^j \tag{3}$$

As x and z are independent, we have

$$\mu_1'(y) = \mu_1'(x) + \mu_1'(z), \tag{4}$$

$$\mu_2(y) = \mu_2(x) + \mu_2(z), \tag{5}$$

$$\mu_3(y) = \mu_3(x) + \mu_3(z), \tag{6}$$

$$\mu_4(y) = \mu_4(x) + \mu_4(z) + 6\mu_2(x).\mu_2(z). \tag{7}$$

The moments of z follow from the assumption that this error is normally distributed with zero mean. We substitute these values and at the same time change over to conventional transformations of the moments, adapting the notation correspondingly. This yields

$$\text{(mean)} \qquad \mu(y) \overset{d}{=} \mu_1'(y) \;, \qquad \mu(y) = \mu(x) \;; \tag{8}$$

$$\text{(variance)} \qquad \sigma^2(y) \overset{d}{=} \mu_2(y) \;, \qquad \sigma^2(y) = \sigma^2(x) + \sigma^2(z) \;; \tag{9}$$

$$\mu_3(y) = \mu_3(x) \;, \tag{10}$$

$$\text{(skewness)} \qquad \gamma_1(y) \overset{d}{=} \frac{\mu_3(y)}{\sigma^3(y)} \;, \qquad \gamma_1(y) = \left\{\frac{\sigma(x)}{\sigma(y)}\right\}^3 \gamma_1(x) \;; \tag{11}$$

$$\mu_4(y) = \mu_4(x) + 3\sigma^4(z) \;, \tag{12}$$

$$\text{(kurtosis)} \qquad \gamma_2(y) \overset{d}{=} \frac{\mu_4(y)}{\sigma^4(y)} - 3, \; \gamma_2(y) = \left\{\frac{\sigma(x)}{\sigma(y)}\right\}^4 \gamma_2(x) \;. \tag{13}$$

In order to complete the expressions on the right in terms of the parameters of the specific distributions of x that will be considered we require the latters' first four moments or the related characteristics. We shall not go into the algebra involved and give the necessary formulae in table 1. Please note that in order to avoid confusion we have adopted a rather unusual notation for the Pareto distribution parameters: the lower limit of its range is denoted by κ, and the slope parameter , better known as 'Pareto's alpha', as π. For the Gamma distribution we retain the usual (α,λ), and the parameters of the Lognormal are given as $\tilde{\mu}(x)$, $\tilde{\sigma}(x)$ where the tilde indicates that they refer to (natural) logarithms.

Table 1. Some formulae from the distributions of x

	Pareto	Gamma[o)]	Lognormal[*)]
$f(x)$	$\dfrac{\pi\kappa^{\pi}}{x^{(\pi+1)}}$, $x>\kappa$	$\dfrac{\lambda^{\alpha}}{\Gamma(\alpha)}\, x^{\alpha-1}e^{-\lambda x}$, $x>0$	$\dfrac{1}{x\tilde{\sigma}\sqrt{2\pi}}\ \exp -\tfrac{1}{2}\left(\dfrac{\ln x-\tilde{\mu}}{\tilde{\sigma}}\right)^2$
μ	$\kappa\cdot\dfrac{\pi}{\pi-1}$	α/λ	$\exp(\tilde{\mu}+\tfrac{1}{2}\tilde{\sigma}^2)$
σ^2	$\left\{\dfrac{\kappa}{\pi-1}\right\}^2\cdot\dfrac{\pi}{\pi-2}$	α/λ^2	$\tilde{\mu}^2\eta^2$
γ_1	$2\,\dfrac{\pi+1}{\pi-3}\sqrt{\dfrac{\pi-2}{\pi}}$	$2/\sqrt{\alpha}$	$\eta^3+3\eta$
γ_2	$6\,\dfrac{\pi^3+\pi^2-6\pi-2}{\pi(\pi-3)(\pi-4)}$	$6/\alpha$	$\eta^8+6\eta^6+15\eta^4+16\eta^2$
$\tilde{\mu}$	$1/\pi+\ln\kappa$	$\psi(\alpha)-\ln\lambda$	not required
$\tilde{\sigma}^2$	$1/\pi^2$	$\psi'(\alpha)$	
$\tilde{\gamma}_1$	2	$\psi''(\alpha)/\{\psi'(\alpha)\}^{3/2}$	
$\tilde{\gamma}_2$	6	$\psi'''(\alpha)/\{\psi'(\alpha)\}^2$	

*) with $\eta^2 = e^{\tilde{\sigma}^2}-1$. o) with $\psi(\alpha)=\Gamma'(\alpha)/\Gamma(\alpha) = (\log\Gamma(\alpha))'$.

The expressions are largely our own work with some help from handbooks such as Mood et al. [9] and, for the lognormal, from Aitchison & Brown [2].

The <u>multiplicative</u> <u>error</u> model starts from the basic assumption

$$\log y = \log x + \log z \tag{12}$$

and proceeds exactly along the same lines with the proviso that all symbols are given tildes. We now assume that the error has a lognormal distribution, or $\log z \sim N(\tilde{\mu}(z),\tilde{\sigma}(z))$, and all subsequent equations (2) to (13) hold without changes but with the addition of tildes throughout. It will be clear, however, that the expressions for the moments of logx in terms of the income distribution parameters change a great deal. The Lognormal must be discarded altogether, since in this case logy is the sum of two normal variates, and the parameters of these distributions are no longer identified; for the Pareto and the Gamma we have had to work out the moments of logx, so that e.g. $\tilde{\mu}(x)$ stands for the logarithm of the geometric mean. This is once more a matter of algebra; the results are again shown in table 1.

In the applications we substitute the expressions from table 1 for each distribution in turn into some of the equations (8) to (13), equate the result to the observed sample statistics, and solve for the parameters by straightforward manipulation or, if necessary, by numerical approximation. All reported estimates are based on the first three moments, and the fourth moment will hardly appear in the sequel at all. The measures of skewness and kurtosis, which provide equivalent representations of the third and fourth moment, have some added intuitive appeal, and we shall use them in our discussion.

Parameter estimates by the method of moments are consistent, and in view of the large samples we use this is an attractive property. But without going much further we have no standard errors to indicate their precision, nor do we have ready statistics of the goodness of fit. These defects underline the tentative and exploratory character of the present analysis.

We may judge the succes of our approach to some extent by referring to the idea that prompted it. This idea was that some of the simple two-parameter distributions might be retrieved from the disfavour they have incurred by allowing for the presence of errors and disturbances in the observed distributions. This may to some extent be judged by examining the plausibility of the parameter estimates for x, the systematic part of observed income. It is also clear that the idea fails if the error component dominates the observed distribution. Since

$$\sigma^2(y) = \sigma^2(x) + \sigma^2(z) \qquad (14)$$

(and similarly for logarithms), the ratio

$$\phi = \sigma^2(z)/\sigma^2(y) \qquad (15)$$

will indicate what part of the observed income variation is due to errors and shocks. If the model is of any use this ratio should be of moderate value, and offhand we should hope that it does not exceed say .25 .

3. U.S. Family income distributions, 1960 and 1969.

The first set of data that we have considered consists of the frequency distributions of family incomes for 1960 and 1969, as reported by the U.S. Bureau of the Census ([13], table 500). These are exactly the same data that were used by Salem and Mount [11]; McDonald and Ransom use tabulations from the same source, but with a different array of income classes. The shape of these distributions is shown in figure 1, and the full data are given in the appendix.

The trouble with income class frequency data is that they do not permit accurate calculation of the higher moments. This is basically the same problem as Salem and Mount encountered when they had to estimate the geometric mean from these grouped data. The difficulties arise in the treatment of the open-ended class at the upper end of the distribution, and to a lesser extent in the allowance for intra-class variation.

The last of the 10 income classes covers annual incomes of $15.000 and over; this represents only 4% of the sample in 1960, but 20% in 1969. Salem and Mount put the class mid-point at $20.000 ([11],p.1120, footnote), but this is clearly much too low. Since frequencies are presumably concentrated towards the left, the class mean would be even lower; if the upper tail were described by a Pareto distribution this would imply a Pareto alpha of over 4, which is an unlikely figure (for the argument involved see [4],p.145, footnote). For an alpha of 2.5 we would find a class mean of $25.000 and a mid-point even farther to the right. We adopt values of $27.500 for 1960 and of $32.500 for 1969. In the sequel we denote Salem and Mount's treatment of the upper income class as case 1, and our own as case 2.

It can be argued that wide income classes are poorly represented by <u>any</u> single value since this always neglects the frequency distribution within the class. Class point values that lead to the correct overall mean will under-estimate the variance, and vice versa. We have therefore subdivided all three wider classes into new classes of the uniform width of $1000, and allotted frequencies to the latter by artful interpolation. This is case 3. It prepares the way for an application of Sheppard's corrections for intra-class variation in classes of equal width (see [3],p.361). This gives case 4.

Table 2. Moments of U.S. family income distributions, 1960 and 1969*

	1960				1969			
	1	2	3	4	1	2	3	4
mean	6.35	6.63	6.34	6.34	10.37	13.27	10.67	10.67
standard error	4.24	5.25	4.63	4.62	5.78	10.19	7.24	7.23
coefficient of variation	.62	.79	.73	.73	.56	.77	.68	.68
skewness, γ_1	1.23	2.26	2.34	2.35	.39	1.15	1.72	1.72
kurtosis, γ_2	1.86	6.64	9.84	9.92	-.83	-.08	4.43	4.44

* mean and standard error in $1.000 per annum

Table 3. Estimates of additive error models, U.S. data

distribution of x	1960	1969
Pareto	κ 4.538 π 3.524 $\sigma(z)$ 3.725 ϕ .650	κ 7.812 π 3.732 $\sigma(z)$ 5.892 ϕ .663
Gamma	α 1.480 λ .234 $\sigma^2(z)$ negative	α 1.931 λ .181 $\sigma^2(z)$ negative
Lognormal	$\tilde{\mu}$ 1.640 $\tilde{\sigma}$ .642 $\sigma(z)$ .940 ϕ .041	$\tilde{\mu}$ 2.200 $\tilde{\sigma}$ .578 $\sigma(z)$ 2.660 ϕ .135

Data adjustments of this rather messy type are not usually very prominently displayed, although they may materially affect results*).

*) We have also repeated Salem and Mount's comparison of the lognormal and the Gamma distribution for case 4 data, fitting both models by moments. The fit of the lognormal is much better than before, and that of the Gamma much worse, and Salem and Mount's conclusion that the latter provides a superior fit is much weakened, though not reversed. By their RSS criterion the two distributions perform equally well in 1969, but in 1960 the Gamma is still twice as effective as the lognormal.

For the sample moment statistics this is shown in table 2. While Sheppard's corrections turn out to be negligible - cases 3 and 4 are nearly identical - the other changes make for considerable differences, especially in the higher moments. It is clear that case 1 underestimates both mean and variance by putting a much too low value on the top interval, and that case 2 errs in the opposite direction.

These results do not inspire a great deal of confidence in the higher moments, and this is the main reason why we have eschewed four parameter models that would make use of the fourth moment in estimation. We also desist from the further processing of the grouped data that would be needed to calculate the moments of logincome as required by the multiplicative error model.

The estimates of the additive model, based on case 4 sample statistics, are shown in table 3. They are reasonable for the lognormal but awful for the Pareto and the Gamma distribution.

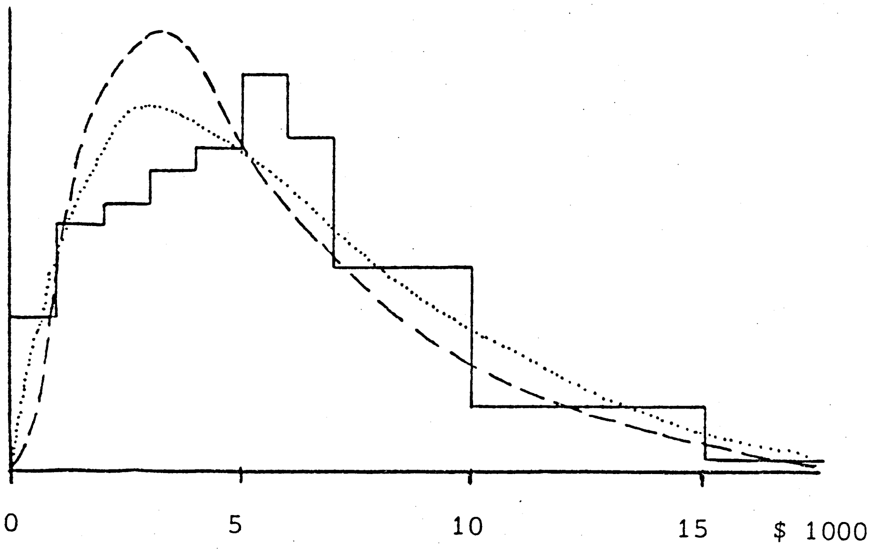
(See here Table 3)

For the Pareto model we find a π of over three, as was to be expected; for otherwise the third moment would not exist, and we have assumed and indeed used its existence. The error term remedies the well-known failure of the pure Pareto to describe the lower income frequencies, but the Pareto limit κ of the systematic income component has scarcely slipped below the modal income, and the major price we have to pay for describing the lower half of the income distribution is that the normal error dominates very strongly indeed. The values of ϕ show that the Pareto part of the model accounts for only a third of income variation, and this is not worthwile.

In the case of the Gamma distribution the observed moments suggest a negative error variance. This of course makes nonsense of our model, but it also reflects on the Gamma model of the income distribution as such. The unacceptable result arises because the sample variance is too small, as it were, relatively to the first and third moment; and since these depend on the distribution of x alone, and not on z, the anomalous outcome has very little to do with the error model but brings out that the Gamma fails to describe the income distribution at all accurately. If we go by the first and third moment, the implied Gamma variance is too large, and if we use the mean and variance for estimation the theoretical value of the third moment falls short of the observed value. This confirms that while the Gamma may be a convenient function it fails to describe the upper reaches of the income distribution: its right-hand tail is much too thin.

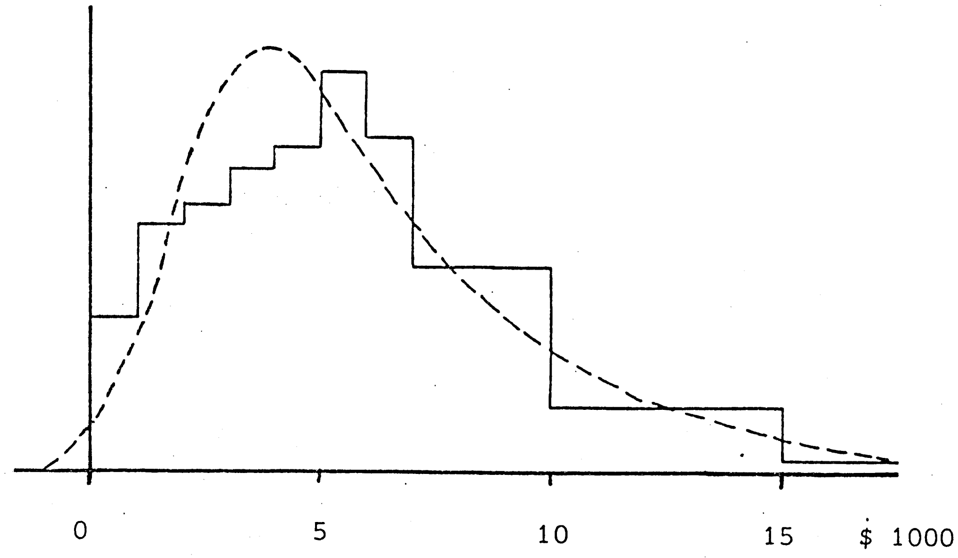
(See here Figures 1a & 1b)

The Lognormal is the only one to present acceptable results, and even then this holds for 1969 rather than for 1960 when the error term is almost negligible. The shape of the theoretical distribution has been determined by simulating 30.000 random drawings of observed income for a pair of



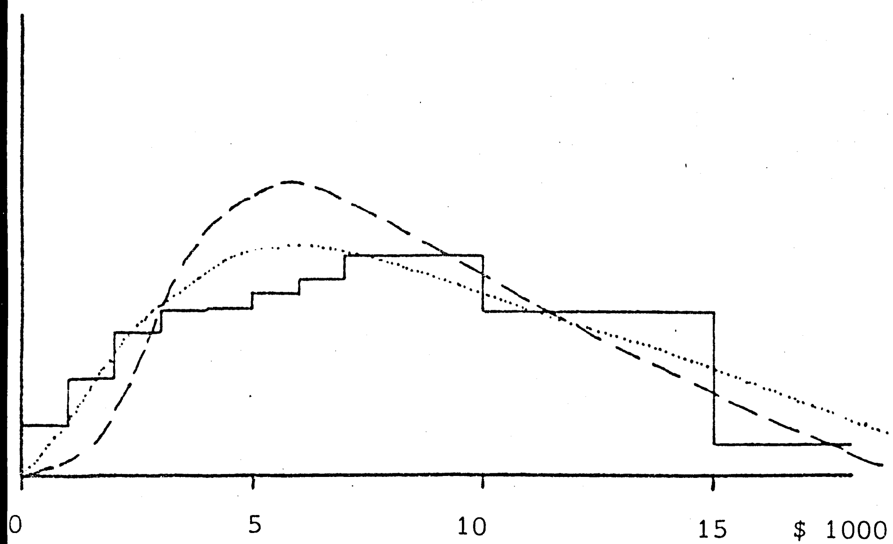
1960

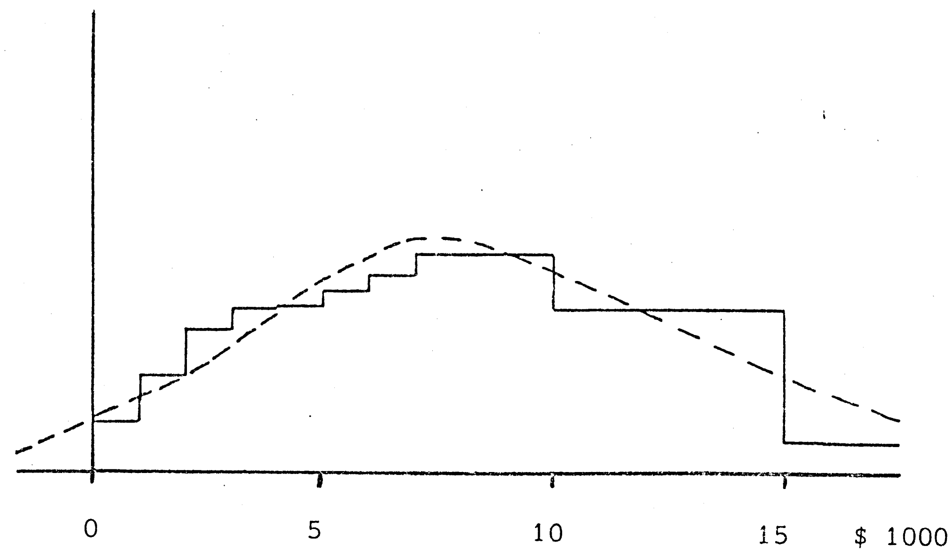
Figure 1a. U.S. Income Distribution as analysed by Salem & Mount: -- Lognormal, ... Gamma



1960

Figure 1b. U.S. Income Distribution and fitted Lognormal with additive Error





lognormal and normal variates, and the result is shown in figure 1b*).

---

*) With a sample size of 30.000 and theoretical probabilities not exceeding .25, the 95% or two standard deviations margin of the computed frequencies is .005, so that they are precise to the second digit.

It would seem that our results are a considerable improvement on those of Salem & Mount, shown in figure 1a; but part of this is simply due to the introduction of a third parameter, and moreover a major improvement in the fit of the lognormal as such can already be obtained by merely adjusting the data set and the top income class in particular, without introducing an error term at all.

We conclude that the analysis of the present paragraph has shed more light on the defects of grouped data and on the relative merits of Pareto, Gamma and Lognormal models than on the additive error model.

4. Belgian economists, 1979

Some of the difficulties we have encountered are avoided if the observed moments can be accurately calculated from individual income returns. Such data are rare, for in many cases incomes are not only reported but actually recorded by income classes. The 1979 survey of Belgian economists is an exception. It is a postal survey among the would-be participants of the 1979 Flemish Economic Congress, and from the results we take the data on the income of the (male) head of the household, after tax. After screening there are 850 returns; the frequency distribution of incomes is shown in figure 2, and some summary statistics are given in the Appendix.

There is nothing to prevent us from calculating the moments of both income and logincome from these data, and we may therefore apply the multiplicative as well as the additive error model. The results that we obtain from this excellent data set are however uniformly bad, and we need only briefly comment on table 4.

(See here table 4)

For the Pareto distribution we find, as before, that unlikely high slope coefficients are coupled with preponderance of the error term. In the multiplicative version the Pareto component comes close to a (multiplicative) constant, and the (log)normal error distribution reigns supreme. Yet if we specify the Lognormal for the systematic part, the observed variance is completely absorbed and there is no room for the additional error variance. This case has already earlier arisen, and we have argued that it indicates a right-hand tail that is too thick to be accomodated by the model.

If the observed tail is too thick for the lognormal it is certainly too thick for the Gamma - and so it is; the additive error model fails in the variance, and the multiplicative model can not be estimated at all. This is so because $\tilde{m}_3(y) = \tilde{m}_3(x)$, and for the latter we find

$$\tilde{m}_3(x) = \psi''(\alpha) \;;$$

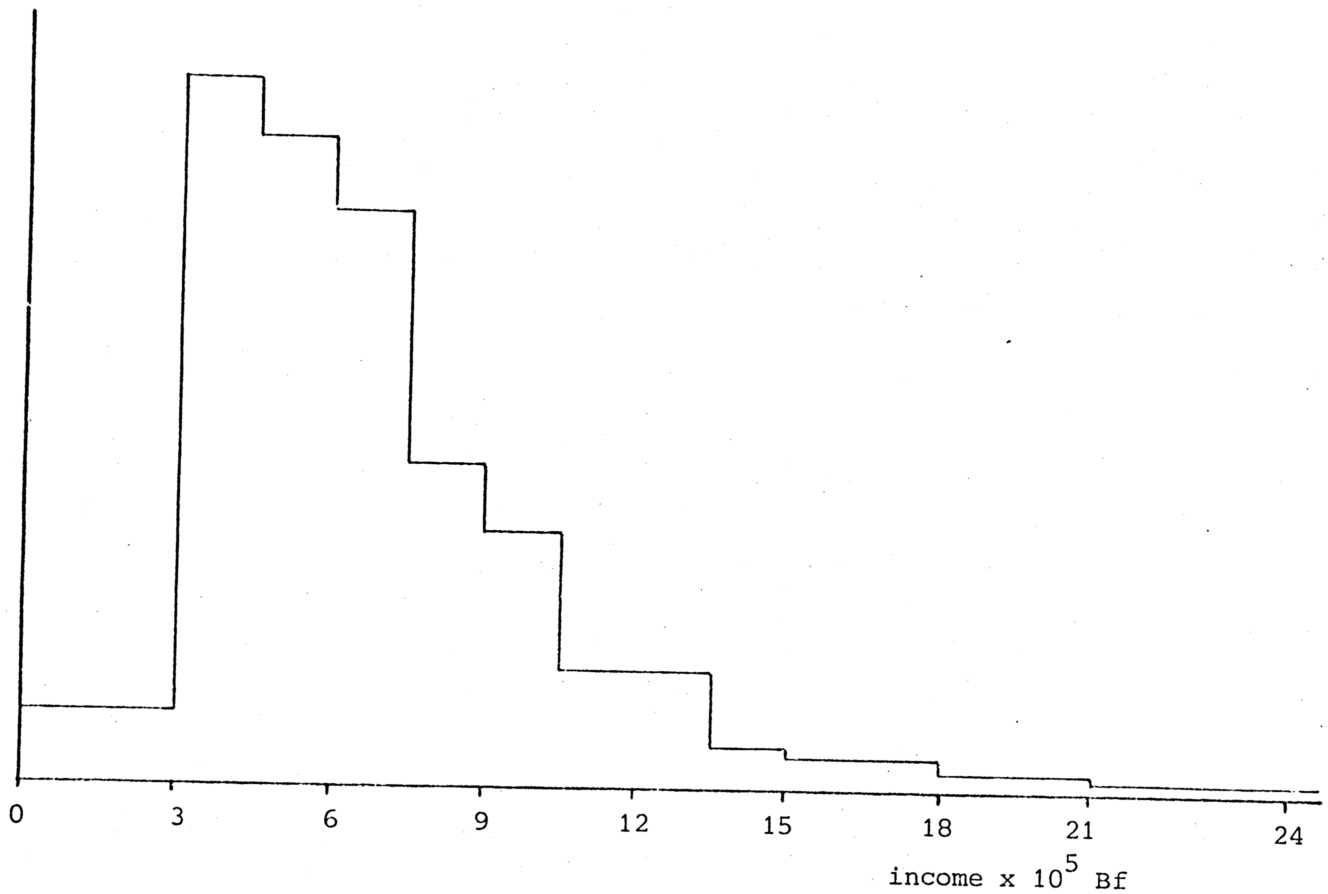


Figure 2. Income distribution of Belgian economists

Table 4. Estimates of error models, Belgian economists, 1979.

distribution of x	additive error	multiplicative error
Pareto	κ 5.034 π 3.651 $\sigma(z)$ 2.824 ϕ .568	κ 5.351 π 8.660 $\tilde{\sigma}(z)$ .515 ϕ .952
Gamma	α 1.755 λ .253 $\sigma^2(z)$ negative	third moment positive: inadmissible
Lognormal	$\tilde{\mu}$ 1.756 $\tilde{\sigma}$ .601 $\sigma^2(z)$ negative	not identified

this trigamma function is negative for all α (see, e.g., [1], p.260, 6.4.1), yet the observed third moment is positive. This may be interpreted as yet another sign that the right-hand tail is much too thick for the model.

We must conclude that all the models considered fail to fit the present data.

5. Conclusions

The limited scope of the analysis does not permit sweeping conclusions. Yet we have gained the distinct impression that the introduction of an error term does not save the Pareto or the Gamma distribution as an adequate description of the income distribution. The verdict on the Lognormal must be postponed.

A major methodological difficulty that has turned up is that the error model must be fitted by the method of moments, and that this is a delicate operation since it involves the third moment and is quite sensitive to small deviations in its observed value. This is particularly painful when we have class frequency data only and must attribute a single representative value to the upper open-ended income class.

References

[1] Abramovitz, M., and I.A. Stegun: Handbook of Mathematical Functions. New York, Dover, 1965.

[2] Aitchison, J., and J.A.C. Brown: The Lognormal Distribution. Cambridge, Cambridge University Press, 1969.

[3] Cramèr, H.: Mathematical Methods of Statistics. Princeton, Princeton University Press, 1945.

[4] Cramer, J.S.: Empirical Econometrics. Amsterdam, North Holland, 1969.

[5] van Dijk, H.K., and T. Kloek: Inferential Procedures in Stable Distributions for Class Frequency Data on Incomes. Econometrica, forthcoming.

[6] Kloek, T., and H.K. van Dijk: Efficient Estimation of Income Distribution Parameters. Journal of Econometrics, vol.8 (1978), p.61-74.

[7] Kloek, T., and H.K. van Dijk: Further Results on Efficient Estimation of Income Distribution Parameters. Economie Appliquée, vol. 30 (1977), p.439-459.

[8] McDonald, J.B., and M.R. Ransom: Functional Forms, Estimation Techniques and the Distribution of Income. Econometrica, vol.47 (1979), p.1513-1526.

[9] Mood, A.M., F.A. Graybill and D.C. Boes: Introduction to the Theory of Statistics. Tokyo, McGraw Hill, 1974.

[10] van Praag, B.M.S., A. Hagenaars and N. Bouma: Income Inequality, Measurement and Reality. Unpublished paper, Leyden, 1980.

[11] Salem, A.B.Z., and T.D. Mount: A Convenient Descriptive Model of Income Distribution: the Gamma Density. Econometrica, vol.42 (1974), p.1115-1128.

[12] Singh, S.K., and G.S. Maddala: A Function for Size Distribution of Incomes. Econometrica, vol.44 (1976), p.963-970.

[13] U.S. Bureau of the Census: Statistical Abstract of the United States, 1970. Washington, Government Printing Office, 1971.

Appendix. The data

1. U.S. data, 1960 and 1969.

Source: Statistical Abstract of the United States, 1970, table 500.

income level of family, x \$1.000 per annum	1960	Fractions	1969
< 1	.050		.016
1 - < 2	.080		.031
2 - < 3	.087		.046
3 - < 4	.098		.053
4 - < 5	.105		.054
5 - < 6	.129		.059
6 - < 7	.108		.064
7 - < 10	.200		.217
10 - < 15	.106		.267
15 and over	.037		.192

2. Belgian economists, 1979

Income of male head of household, sample size 850, in 10^5 Belgian francs per annum, or log of 10^3 Belgian francs p.a.

moments	income	logincome
m_1	6.93254118	1.792837
m_2	18.48036244	0.277982
m_3	216.40057963	0.003079
m_4	4981.65891781	0.349496

Frequency distribution Belgian economists 1979

Income x 10^5 Belgian francs per annum:	fraction
< 3	.049
3 - < 4.5	.229
4.5 - < 6	.210
6 - < 7.5	.187
7.5 - < 9	.105
9 - < 10.5	.083
10.5 - < 12	.038
12 - < 13.5	.038
13.5 - < 15	.014
15 - < 18	.021
18 - < 21	.012
21 and over	.014