AGENCY FOR INTERNATIONAL DEVELOPMENT
WASHINGTON, D. C. 20523

BIBLIOGRAPHIC INPUT SHEET

FOR AID USE ONLY

1. SUBJECT CLASSIFICATION

A. PRIMARY: Agriculture — AE10-0000-0000

B. SECONDARY: Agricultural economics

2. TITLE AND SUBTITLE

A convenient descriptive model of income distribution: the gamma density

3. AUTHOR(S)

Salem,A.B.Z.; Mount,T.D.

4. DOCUMENT DATE: 1973

5. NUMBER OF PAGES: 35p.

6. ARC NUMBER: ARC

7. REFERENCE ORGANIZATION NAME AND ADDRESS

Cornell

8. SUPPLEMENTARY NOTES *(Sponsoring Organization, Publishers, Availability)*

(In A.E.Res.73-8)

9. ABSTRACT

10. CONTROL NUMBER

PN-RAA-375

11. PRICE OF DOCUMENT

12. DESCRIPTORS

Gamma function
Income distribution

13. PROJECT NUMBER

14. CONTRACT NUMBER

CSD-2823 211(d)

15. TYPE OF DOCUMENT

AID 590-1 (4-74)

May 1973

A.E. Res. 73-8

A CONVENIENT DESCRIPTIVE MODEL OF INCOME DISTRIBUTION: THE GAMMA DENSITY

A.B.Z. Salem

T.D. Mount

Department of Agricultural Economics
Cornell University Agricultural Experiment Station
New York State College of Agriculture and Life Sciences
A Statutory College of the State University
Cornell University, Ithaca, New York 14850

A CONVENIENT DESCRIPTIVE MODEL OF INCOME DISTRIBUTION: THE GAMMA DENSITY

by

A. B. Z. Salem and T. D. Mount

Abstract

The distribution of personal income is approximated by a two parameter gamma density function (Pearson Type III). The two parameters may be considered as indicators of scale and of inequality, respectively. Maximum likelihood estimates of the parameters are derived from a random sample using graphical techniques, and a likelihood ratio test for the hypothesis that the inequality parameter remains constant over time or that it is the same for different cross-section distributions is presented. The derivation of both the estimates and the test statistic requires computing only the arithmetic and geometric means from the sample. An empirical application, including a comparison of the gamma and lognormal distributions to demonstrate the better fit of the gamma, is made to personal income data for the U.S. Using the gamma density, the estimated skewness parameters for the years 1960 to 1969 are shown to be significantly related to measures of unemployment and inflation, but not to real growth. In addition, an analysis of the distributional effects of taxation implies that federal income tax reduces the skewness but that state and local taxes are slightly regressive.

1. Introduction[1/]

A number of alternative probability density functions have been proposed as models of income distribution. The requirements of such a function are that it should provide a reasonably close approximation to the true distribution, and that its parameters should be simple to estimate and also to interpret in an economically meaningful way. In practice, these two criteria often compete with each other. Distributions that have been suggested include the Pareto, the lognormal, the Yule and the beta.[2/] Both the Pareto and lognormal distributions have parameters that are easy to estimate and can be directly related to some inequality measure, but neither fits the full range of income data very satisfactorily. On the other hand, parameters of the Yule and beta distributions are more difficult to interpret as measures of income inequality.

The gamma density, which was applied to income data by Amoroso [3] in 1925, provides an alternative model that fits the data reasonably well. The two parameters of this function can be conveniently identified as measures of scale and of skewness, and consequently, are directly related to the concept of proportionate growth (Gibrat's Law) and to inequality, respectively. Estimators of these parameters may be derived from a random sample of observations using maximum likelihood techniques. In addition, the interesting hypothesis that inequality (skewness) does not change over time or is the same for different cross-section categories may be tested using a likelihood ratio test.

1/ The assistance of Professors D. L. Solomon and D. I. Padberg in the development of this research paper is gratefully acknowledged, although they are not responsible for any errors that remain.

2/ Selected references concerning the derivation of these distributions and their application to income data are Quensel [14], Champernowne [4], and Aitchison and Brown [2] for the Pareto and lognormal, Simon and Bonini [15, 16] for the Yule, and Thurow [19] for the beta.

In the sections that follow, the mathematical properties of the gamma distribution are discussed (Section 2), maximum likelihood estimators are derived together with the likelihood ratio test statistic (Section 3), and finally, an empirical application is made to personal income data for U.S. families for the years between 1960 and 1969 (Section 4). First, the goodness of fit of the gamma and the lognormal are compared. Secondly, an attempt is made to relate the observed changes of the gamma parameters to three economic indicators; unemployment, inflation and real output. Finally, the impact of personal taxes on the distribution of income is investigated.

2. The Gamma Distribution

A gamma distribution from Pearson's Type III family [11, p. 71] may be defined[1/]

$$(1) \qquad f_\gamma(X;\alpha,\lambda) = \frac{\lambda^\alpha}{\Gamma(\alpha)} X^{\alpha-1} e^{-\lambda X} \qquad 0 < X < \infty$$

where α and λ are positive parameters, and

$$\Gamma(\alpha) = \int_0^\infty e^{-u} u^{\alpha-1} du \text{ is the gamma function.}$$

The standard parameters of location, dispersion and skewness may be derived in terms of α and λ to be [11, p. 71]

a) arithmetic mean $= \alpha/\lambda$

b) mode $= (\alpha-1)/\lambda$

c) median[2/] $= (3\alpha-1)/3\lambda$

1/ The familiar chi-squared distribution is a special case with $\lambda = 1/2$ and $\alpha = \nu/2 - 1$ where ν is the number of degrees of freedom.

2/ The median is based on Doodson's [22, p. 15] empirical formula: 3(Mean-Median) = Mean-Mode.

d) standard deviation $= \sqrt{\alpha}/\lambda$

e) coefficient of momental skewness[1/] $= 1/\sqrt{\alpha}$

Although the first four parameters depend on both α and λ, skewness is dependent on α but not on λ. Hence, only α is directly related to standard inequality measures. It should also be noted that if the mean is held constant, the standard deviation increases as the skewness increases. The second parameter λ is a scale parameter. If all observations of X are multiplied by a constant to give $Z = aX$, as would happen under Gibrat' Law of proportionate growth, the density function of Z is equivalent to replacing λ in (1) by λ/a.

$$(2) \quad f_\gamma(X;\alpha,\lambda)dX = \left[\frac{\lambda^\alpha}{\Gamma(\alpha)}\left[\frac{Z}{a}\right]^{\alpha-1} e^{-\left[\frac{\lambda Z}{a}\right]}\right]\frac{dZ}{a}$$

$$= \left[\frac{\left[\frac{\lambda}{a}\right]^\alpha}{\Gamma(\alpha)} Z^{\alpha-1} e^{-\left[\frac{\lambda}{a}\right]Z}\right]dZ$$

$$= f_\gamma(Z;\alpha,\lambda/a)dZ$$

In summary, changes of λ with α constant are related to proportionate growth effects, and changes of α influence the skewness of the distribution. An almost identical type of distinction may be made between the two parameters of the lognormal distribution. However, it is shown in Section 4 that the gamma fits income data better than the lognormal.

1/ This equals the third central moment divided by twice the cube of the standard deviation.

One of the most widely accepted measures of the inequality of income distribution is the Lorenz concentration ratio.[1] If income per family is represented by X in (1), then, using steps similar to those followed by Aitchison and Brown [2, p. 111] for the lognormal distribution, the concentration ratio may be written (See Appendix A[2]).

$$(3) \qquad L_\gamma = 2B_{.5}(\alpha,\alpha+1) - 1$$

where $B_{.5}(\alpha,\alpha+1)$ is the incomplete beta function evaluated at .5 for parameter values α and $\alpha + 1$. Tabulated values of this function may be found in [12].

An alternative measure of inequality has been proposed by Theil [11, p. 91]. This measure is based on the entropy of the distribution which may be written

$$(4) \qquad H = \sum_{i=1}^{N}\left[\frac{X_i}{X_.}\right]\log\left[\frac{X_.}{X_i}\right]$$

where $i = 1,2,\ldots,N$ represent individuals

X_i is the income of individual i

$X_. = \sum_i X_i$ is total income

If all individuals have the same income $X_./N$ (perfect equality), then $H = \log N$ is a maximum. Conversely, if one person has all the income (perfect inequality), then $H = 0$ is a minimum. The measure of inequality proposed by Theil is to subtract H from the maximum value to give

$$(5) \qquad I = \log N - H$$

$$= \sum_{i=1}^{N}\left[\frac{X_i}{X_.}\right]\log\left[\frac{NX_i}{X_.}\right]$$

1/ The concentration ratio is the ratio of the area between the Lorenz curve and the diagonal to the complete triangular area under the diagonal [2, p.112].

2/ Appendix A also contains the derivation of an expression for the Gini coefficient of mean difference.

If the distribution of income is represented by (1) which is a continuous function, then (5) may be written

$$(6) \qquad I_\gamma = \int_0^\infty \left[\frac{Nx}{X_s}\right] \log\left[\frac{Nx}{X_s}\right] f_\gamma(x;\alpha,\lambda)dx$$

Remembering that the mean $X_s/N = \alpha/\lambda$ and substituting (1) into (6) gives

$$(7) \qquad I_\gamma = \int_0^\infty \left[\frac{\lambda x}{\alpha}\right] \log\left[\frac{\lambda x}{\alpha}\right] \frac{\lambda^\alpha}{\Gamma(\alpha)} x^\alpha - 1 e^{-\lambda x} dx$$

Redefine λx as t and simplify to

$$(8) \qquad I_\gamma = \frac{1}{\alpha\Gamma(\alpha)}\left[\int_0^\infty t^\alpha e^{-t} \log(t)dt - \log(\alpha)\int_0^\infty t^\alpha e^{-t} dt\right]$$

$$= \frac{1}{\Gamma(\alpha+1)}\left[\Gamma'(\alpha+1) - \log(\alpha)\Gamma(\alpha+1)\right]$$

$$= \frac{1}{\alpha} + \psi(\alpha) - \log\alpha$$

as $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$ and $\psi(\alpha+1) = \frac{1}{\alpha} + \psi(\alpha)$

where $\psi(\alpha) = \frac{\Gamma'(\alpha)}{\Gamma(\alpha)} = \frac{d\log\Gamma(\alpha)}{d\alpha}$ is the digamma function [6, p. 179 and pp. 277-78]

The digamma function is tabulated in [6], and an alternative computing procedure is derived in the next section when maximum likelihood estimators of α and λ are used to compute (8).

It should be noted that both the Lorenz concentration ratio (3) and the Theil inequality measure (8) are monotonically decreasing functions of α, and that neither expression depends on λ. In fact, if α is sufficiently large, the gamma is approximately symmetric and approaches a normal distribution [11, p. 72].

3. Maximum Likelihood Estimation

If a random sample of N observations (for example, family income per year) are drawn from a population represented by the gamma density defined in (1), the likelihood function may be written.

$$(9) \qquad L = \prod_{i=1}^{N}\left[\frac{\lambda^{\alpha}}{\Gamma(\alpha)} X_i^{\alpha-1} e^{-\lambda X_i}\right]$$

and it follows that

$$(10) \qquad \log L = N\left[\alpha\log\lambda - \log\Gamma(\alpha)\right] + (\alpha-1)\Sigma_i \log X_i - \lambda\Sigma_i X_i$$

The maximum likelihood estimators of λ and α may be derived by solving the following two equations

$$(11) \qquad \frac{d\log L}{d\lambda} = \frac{N\hat{\alpha}}{\hat{\lambda}} - \Sigma_i X_i = 0$$

$$(12) \qquad \frac{d\log L}{d\alpha} = N\log\hat{\lambda} - N\psi(\hat{\alpha}) + \Sigma_i \log X_i = 0$$

Simplifying (11) and (12) gives

$$(13) \qquad \hat{\lambda} = \hat{\alpha}/\bar{X}$$

$$(14) \qquad \log\hat{\alpha} - \psi(\hat{\alpha}) = \log\bar{X} - \log\tilde{X} = \log(\bar{X}/\tilde{X})$$

where $\bar{X} = 1/N\Sigma_i X_i$ is the arithmetic mean

$\tilde{X} = (\prod_i X_i)^{1/N}$ is the geometric mean

Expression (14) is nonlinear in $\hat{\alpha}$, and no explicit solution can be derived. However, $[\log\hat{\alpha} - \psi(\hat{\alpha})]$ has been tabulated for different values of $\hat{\alpha}$ [1, pp. 272-73]1/

1/ A more complete tabulation of the digamma alone is given in [6].

and an estimate may be determined using graphical or interpolation procedures after first computing the ratio of the arithmetic to the geometric mean.[1/] In Figure 1, the values of $\hat{\alpha}$ are plotted against different ratios of the two observed means. As this ratio gets larger, $\hat{\alpha}$ decreases, and consequently, the skewness of the distribution increases. The corresponding value of $\hat{\lambda}$ may be computed directly from (13). Hence, the only statistics that are required from the sample to compute $\hat{\alpha}$ and $\hat{\lambda}$ are the arithmetic and the geometric means.

As $\hat{\lambda}$ and $\hat{\alpha}$ are maximum likelihood estimators, the asymptotic variance-covariance matrix may be derived by taking the negative inverse of the matrix of the second order partial derivatives. Asymptotically, these estimators are binormally distributed

$$(15) \qquad \begin{bmatrix} \hat{\lambda} \\ \hat{\alpha} \end{bmatrix} \text{ are Normal } \left[\begin{bmatrix} \lambda \\ \alpha \end{bmatrix}, \begin{bmatrix} a\theta(\alpha)\lambda^2 & a\lambda \\ a\lambda & a\alpha \end{bmatrix} \right]$$

where $a = [N(\alpha\theta(\alpha) - 1)]^{-1}$

$\theta(\alpha) = \dfrac{d^2 \log \Gamma(\alpha)}{d\alpha^2}$ is the trigamma function which has been tabulated [7].

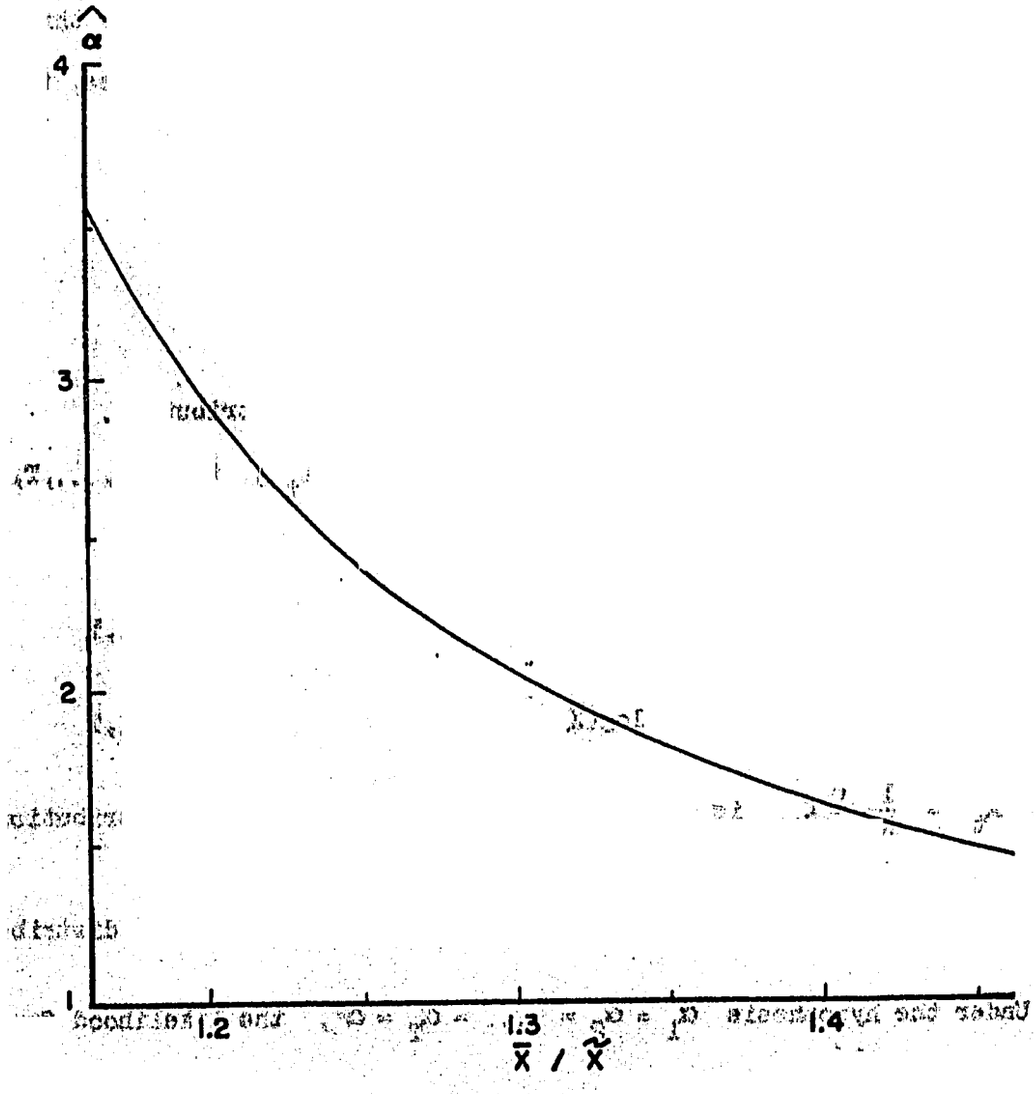
The estimated variance-covariance matrix may be computed by substituting $\hat{\lambda}$ and $\hat{\alpha}$ from (13) and (14) into (15).

As the Lorenz concentration ratio and the Theil inequality measure depend only on α, the maximum likelihood estimators of these inequality

---

1/ Choi and Wette [5] suggest various iterative procedures for obtaining solutions for $\hat{\alpha}$ in (14).

FIGURE 1. RELATIONSHIP BETWEEN THE RATIO OF MEANS AND THE MAXIMUM LIKELIHOOD ESTIMATOR $\hat{\alpha}$ 1/



1/ Derived from the tabulated values of $\psi(\alpha)$ in [6]

measures may be obtained by using $\hat{\alpha}$ instead of α in (3) and (8), respectively. Hence, both measures may be related directly to the ratio of the arithmetic to the geometric mean. These relationships are shown in Figure 2.

In the remainder of this section, the likelihood ratio test statistic is derived for testing the hypothesis that the skewness parameter α is the same for different distributions. This test can be performed to compare different cross-sectional distributions or to test Gibrat's Law of proportional growth by considering distributions at different points through time.

If random samples are drawn from T different gamma densities, the likelihood function may be written

$$(16) \qquad L = \prod_{t=1}^{T} \prod_{i=1}^{N_t} \left[\frac{\lambda_t^{\alpha_t}}{\Gamma(\alpha_t)} X_{it}^{\alpha_t - 1} e^{-\lambda_t X_{it}} \right]$$

where $t = 1,2,\ldots,T$ are different distributions

$i = 1,2,\ldots,N_t$ are observations from the tth distribution.

The maximum likelihood estimators of α_t and λ_t; $t = 1,2,\ldots,T$, which may be derived directly from (13) and (14), are

$$(17) \qquad \hat{\lambda}_t = \hat{\alpha}_t / \bar{X}_t \qquad t=1,2,\ldots,T$$

$$(18) \qquad \log\hat{\alpha}_t - \psi(\hat{\alpha}_t) = \log(\bar{X}_t/\tilde{X}_t) \qquad t=1,2,\ldots,T$$

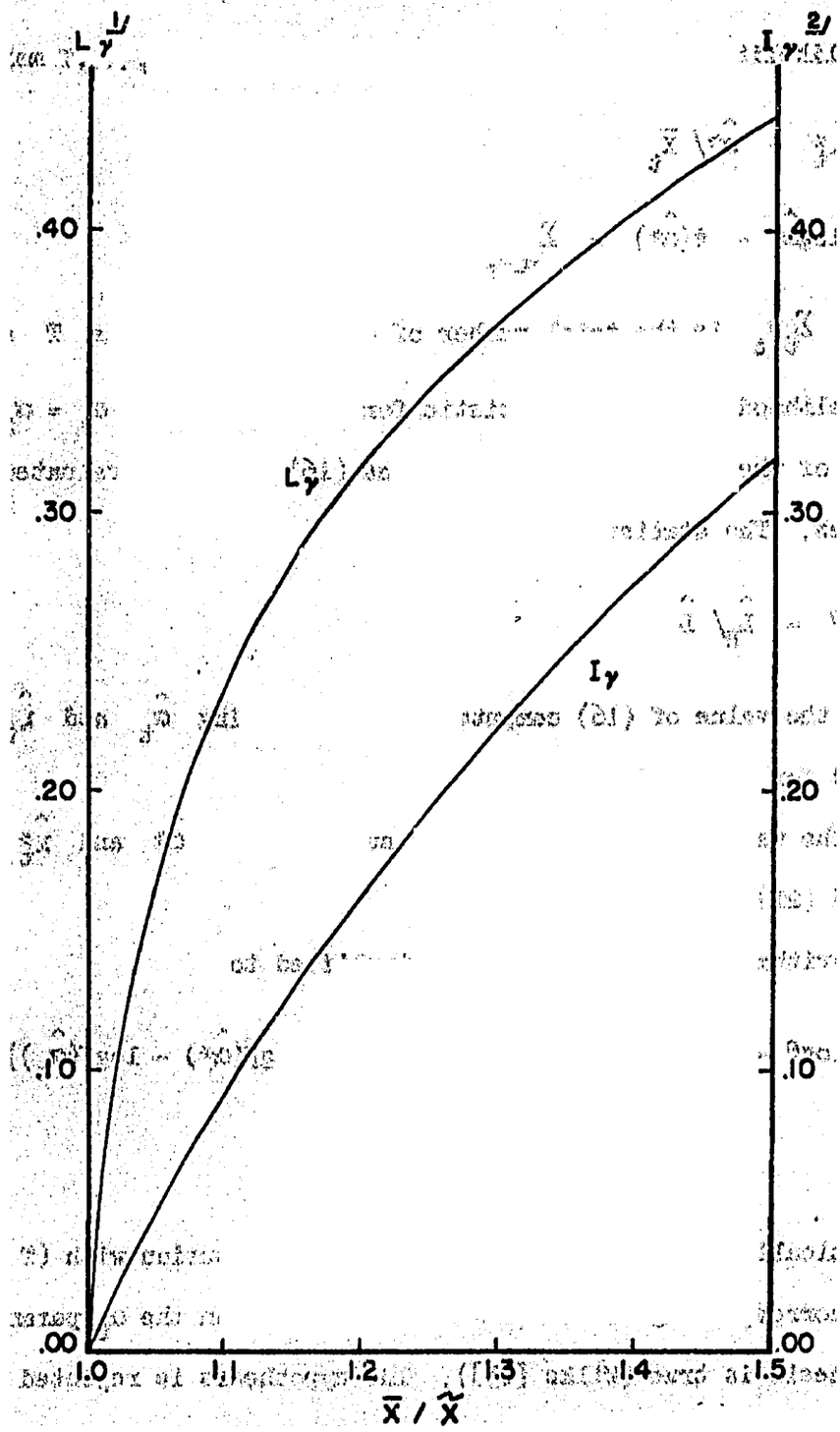
where $\bar{X}_t = \frac{1}{N_t} \sum_i X_{it}$ is the arithmetic mean for the tth distribution

$\tilde{X}_t = (\prod_i X_{it})^{1/N_t}$ is the geometric mean for the tth distribution.

Under the hypothesis $\alpha_1 = \alpha_2 = \ldots = \alpha_T = \alpha^*$, the likelihood function may be rewritten[1/]

1/ If the restrictions $\lambda_1 = \lambda_2 = \ldots = \lambda_T$ are imposed on (16), the maximum likelihood estimators are difficult to compute. However, the economic implications of this hypothesis are not very interesting.

FIGURE 2. RELATIONSHIP BETWEEN THE RATIO OF MEANS AND TWO INEQUALITY MEASURES



1/ Defined in (3)

2/ Defined in (8)

$$(19) \qquad L_H = \prod_{t=1}^{T} \prod_{i=1}^{N_t} \left[\frac{\lambda_t^{*\alpha^*}}{\Gamma(\alpha^*)} x_{it}^{\alpha^*-1} e^{-\lambda_t^* x_{it}} \right]$$

The maximum likelihood estimators of α^* and λ^*_t; $t=1,2,\ldots,T$ may be found by solving

$$(20) \qquad \hat{\lambda}_t^* = \hat{\alpha}^* / \bar{x}_t \qquad t=1,2,\ldots,T$$

$$(21) \qquad \log\hat{\alpha}^* - \psi(\hat{\alpha}^*) = \sum_t \left[\frac{N_t}{N_.} \log(\bar{x}_t / \tilde{x}_t) \right]$$

where $N_. = \sum_t N_t$ is the total number of observations in the T samples.

The likelihood ratio test statistic for the hypothesis $\alpha_1 = \alpha_2 = \ldots = \alpha_T = \alpha^*$ is the ratio of the two likelihood functions (16) and (19) evaluated at their maximum values. The statistic is

$$(22) \qquad \theta = \hat{L}_H / \hat{L}$$

where $\hat{L}$ is the value of (16) computed by substituting $\hat{\alpha}_t$ and $\hat{\lambda}_t$ from (17) and (18) for α_t and λ_t.

$\hat{L}_H$ is the value of (19) computed by substituting $\hat{\alpha}^*$ and $\hat{\lambda}_t^*$ from (20) and (21) for α^* and λ_t^*.

The logarithm of this ratio may be simplified to

$$(23) \qquad \log\theta = \sum_t N_t \Big[(\hat{\alpha}^* \log\hat{\alpha}^* - \hat{\alpha}_t \log\hat{\alpha}_t) - (\log\Gamma(\hat{\alpha}^*) - \log\Gamma(\hat{\alpha}_t)) - (\hat{\alpha}^* - \hat{\alpha}_t)(\log(\bar{x}_t / \tilde{x}_t) - 1) \Big]$$

Asymptotically, $-2(\log\theta)$ has a chi-squared distribution with $(T - 1)$ degrees of freedom, corresponding to the $(T - 1)$ constraints on the α_t parameters, if the hypothesis is true (Wilks [23]). The hypothesis is rejected if the observed level of the statistic is greater than the critical value.

4. Empirical Analysis of Personal Income Data

This section has three main objectives. The first is to determine how well the gamma density function fits the distribution of personal income for families in the U.S. This includes a comparison with the fit of the lognormal. The second is to relate changes in the distribution of income between 1960 and 1969 with three economic measures, and the third is to assess the distributional effects of taxation. A complete description of the data is given in Appendix B.

To accomplish the first objective, gamma and lognormal densities were fitted to income data for 1960 and 1969.[1/] The maximum likelihood estimates of α and λ in (1) were computed using (13) and (14) (Figure 1) for both years. Maximum likelihood estimates of the two lognormal parameters were also computed.[2/] These estimates are summarized in Table 1.

To compare the observed frequencies for each income class with the predicted values, the cumulative density functions must be derived. For the lognormal, this may be done by standardizing the logarithm of the income at

---

1/ Only grouped data were available for family income, and consequently, the assumption that every member of a particular income class receives the same income, measured by the midpoint of that class, had to be made. It should be noted, however, that this assumption would be unecessary if the geometric as well as the arithmetic mean had been reported by the agency conducting the survey.

2/ The maximum likelihood estimates for the lognormal are

$$\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} \log X_i$$

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N} (\log X_i \cdot$$

Table 1. Observed and Predicted Probabilities for Ten Income Classes in 1960 and 1969

Income classes Dollars/family	1960 Observed[1/] probability	1960 Predicted Probability lognormal	1960 Predicted Probability gamma	1969 Observed[1/] probability	1969 Predicted Probability lognormal	1969 Predicted Probability gamma
1. <1000	.050	.0174	.0385	.016	.0002	.0094
2. 1000 to 1999	.080	.1016	.0842	.031	.0089	.0280
3. 2000 to 2999	.087	.1332	.1158	.046	.0364	.0485
4. 3000 to 3999	.098	.1414	.1155	.053	.0676	.0605
5. 4000 to 4999	.105	.1184	.1090	.054	.0873	.0696
6. 5000 to 5999	.129	.0945	.0978	.059	.0942	.0730
7. 6000 to 6999	.108	.0780	.0847	.064	.0913	.0734
8. 7000 to 9999	.200	.1445	.1779	.217	.2320	.2017
9. 10000 to 14999	.106	.1029	.1275	.267	.2185	.2306
10. ≥ 15000	.037	.0681	.0491	.192	.1635	.2054
Estimated Parameters	-	$\hat{\mu} = 8.4946$	$\hat{\lambda} = 3.2418 \times 10^{-4}$	-	$\hat{\mu} = 9.0264$	$\hat{\lambda} = 2.3455 \times 10$
	-	$\hat{\sigma} = 0.7513$	$\hat{\alpha} = 2.06$	-	$\hat{\sigma} = 0.6041$	$\hat{\alpha} = 2.43$
Σ (Observed - Predicted)^2	-	.011869	.003913	-	.007518	.002378

1/ Data Source: Reference [21]. For details see Appendix B.

each class boundary to a Normal(0,1) variable and then using standard normal tables. However, the procedure for the cumulative gamma is more complicated unless α is an integer. In more general situations, the cumulative gamma may be redefined as follows:

$$F_\gamma(X;\alpha,\lambda) = \frac{1}{\Gamma(\alpha)}\int_0^X \lambda^\alpha x^{\alpha-1} e^{-\lambda x} dx$$

$$= I^*(U;\alpha-1)$$

where $U = X\lambda/\sqrt{\alpha}$ (number of standard deviations from the origin)

I^* is the ratio of the incomplete to the complete gamma function.[1/]

The values of I^* have been tabulated [13]. Consequently, if the income levels corresponding to each class boundary are divided by the standard deviation $(\sqrt{\alpha}/\lambda)$ to give the corresponding U value, the cumulative probability may be determined.

1/ If $Z = \lambda X = U\sqrt{\alpha}$, then the cumulative density function for Z may be derived from

$$F_\gamma(X;\alpha,\lambda) = \frac{1}{\Gamma(\alpha)}\int_0^{Z/\lambda} \lambda^\alpha (z/\lambda)^{\alpha-1} e^{-z} \frac{dz}{\lambda}$$

$$= \frac{1}{\Gamma(\alpha)}\int_0^{Z/\lambda} z^{\alpha-1} e^{-z} dz$$

$$= \frac{1}{\Gamma(\alpha)}\int_0^{U\sqrt{\alpha}/\lambda} z^{\alpha-1} e^{-z} dz$$

$$= \Gamma_X(\alpha)/\Gamma(\alpha)$$

$$= I^*(U;\alpha-1)$$

The observed and predicted probabilities of belonging to the 10 income classes are summarized in Table 1 for 1960 and 1969. The observed and predicted income distributions are illustrated in Figure 3 for these two years. It is clear from this figure that the gamma density fits the data better than the lognormal for both 1960 and 1969.[1/] In fact, the sum of squared deviations between the observed and predicted probabilities is over three times greater for the lognormal in both cases. The gamma density does exaggerate the skewness of the distribution, but this tendency is even more marked in the fit of the lognormal.

To investigate how income distribution has changed between 1960 and 1969 for U.S. families, a gamma distribution was fitted to undeflated personal income data before taxes for each year. The estimated parameters and associated inequality measures are summarized in Table 2. It appears from these results that α has increased over time (less inequality) and that λ has decreased. The implication that income has become more equally distributed is also reflected by decreases in the two inequality measures, although it should be noted that the lowest values are actually for 1966. In contrast, the standard deviation has increased over time. This, however, is not inconsistent with a reduction in skewness. It implies that the increase of the mean income is only partially offset by the decrease in skewness, as the standard deviation = arithmetic mean x skewness for a gamma density.

To test the hypothesis that the inequality parameter α has remained constant over time, a likelihood ratio test statistic (23) was computed.[2/]

1/ Metcalf [10] has proposed using a displaced lognormal to improve the fit. It is shown in Appendix C that the economic interpretation of the parameters as well as their statistical properties after this modification are far less tractable than before.

2/ The number of families in the sample, and the relevant sample statistics are listed in Table 2.

FIGURE 3. OBSERVED AND PREDICTED PROBABILITIES OF U.S. FAMILIES IN TEN INCOME CLASSES: 1960 AND 1969

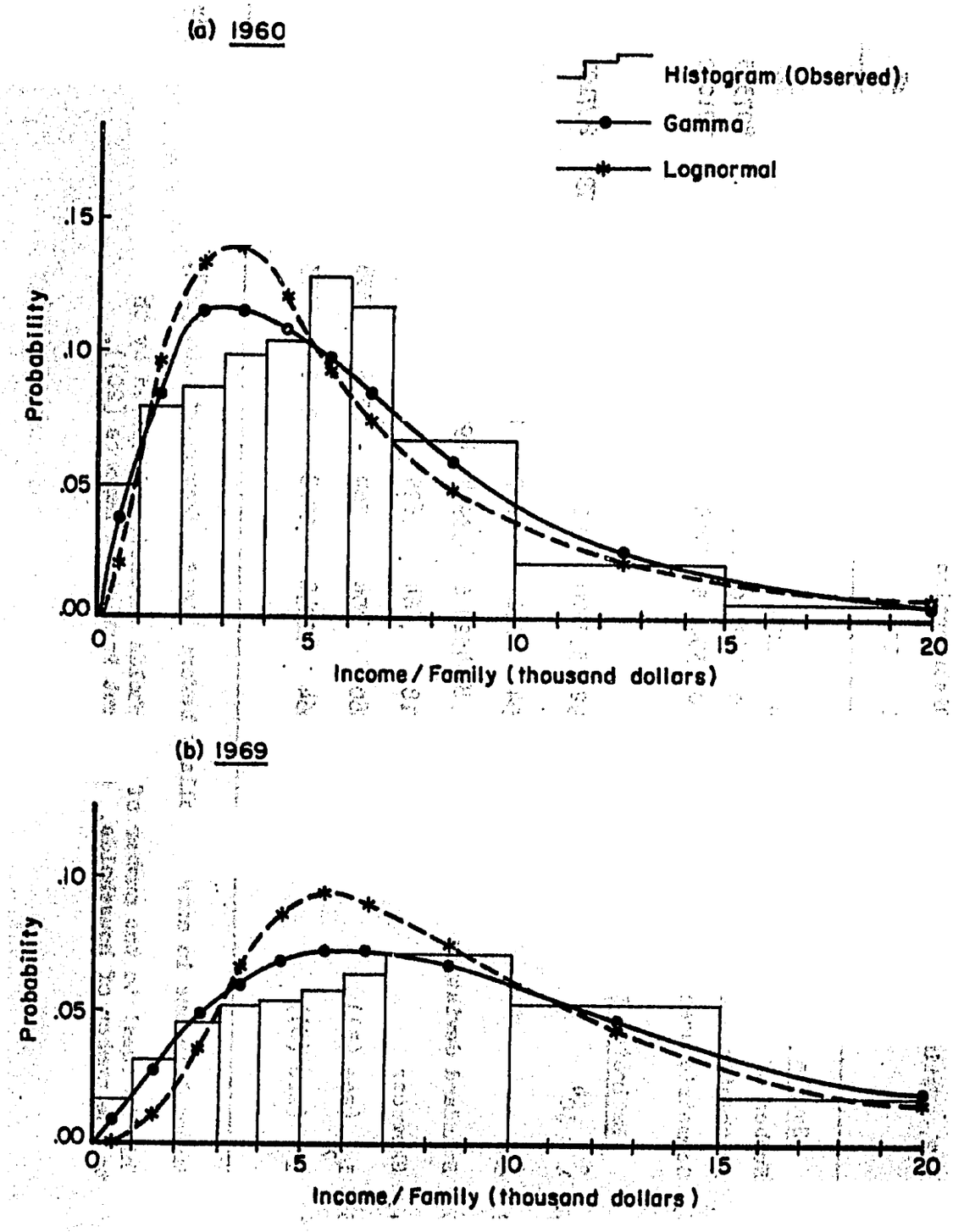


Table 2. Estimated Parameters for the Distribution of Income for Families in the U.S.: 1960 to 1969[1/]

	1960	1961	1962	1963	1964	1965	1966	1967	1968	1969
Sample Statistics										
1. Number of families [2/]	21167	21199	21290	21179	20889	29088	41568	41222	40863	40746
2. Arithmetic mean	6354.5	6578.0	6823.0	7106.5	7439.0	7828.0	8424.5	8973.5	9598.5	10360.5
3. Geometric mean	4888.0	4976.0	5390.5	5478.0	5856.5	6138.0	6809.0	7210.0	7746.0	8320.0
Estimated Parameters										
1. $\hat{\lambda} \times 10^4$	3.2418	2.9492	3.3270	2.9128	3.0112	2.8232	2.9794	2.7191	2.5837	2.3454
2. $\hat{\alpha}$	2.06	1.94	2.27	2.07	2.24	2.21	2.51	2.44	2.48	2.43
3. Standard deviation	4427.4	4722.8	4528.6	4939.4	4970.4	5265.7	5317.5	5744.7	6095.1	6646.3
4. Skewness	.697	.718	.664	.695	.668	.673	.631	.640	.635	.641
5. L_γ (see (3))	.369	.380	.354	.369	.356	.358	.334	.342	.340	.344
6. I_γ (see (8))	.2231	.2364	.2048	.2228	.2072	.2093	.1856	.1910	.1888	.1922

1/ Personal income in current dollars before taxes. Source of data [21], see Appendix B for details.

2/ This is equal to the number of households sampled, multiplied by the ratio of the number of families to the number of households. (Current Population Reports [20]).

The observed value of this statistic is 1153.60. This is much larger than the critical value of the chi-squared statistic (χ^2_9 = 21.67 for a 1% level of significance). Hence, the hypothesis that the α parameters are the same for each year is rejected.

The two estimated parameters do not change consistently over time, and in fact, both exhibit cyclical behavior around a trend line with a period of approximately two years. An attempt was made to relate the estimated parameter values to three economic indicators; the percentage unemployment, consumer price index (CPI), and gross national product (GNP) per capita in real dollars. A similar type of analysis was conducted by Thurow [19] using estimated parameters for the beta density. The estimated ordinary least squares coefficients for log-linear regression models of both parameters are summarized in Table 3.

The signs of the coefficients in the α-model imply that an increase of unemployment or of inflation (CPI), and a decrease of real GNP per capita are associated with a more unequal distribution of income. The fit of the model is good with an R^2 of 0.88. However, none of the coefficients are significantly different from zero at the 5% level of significance. In the λ-model, the estimated coefficients for all three explanatory variables are negative. The R^2 of 0.91 indicates that the fit is good, and only the t ratio for GNP per capita is less than two.

Certain restrictions may be imposed on the models in Table 3 to improve the accuracy of the estimated coefficients. It seems reasonable to expect that the average income per family in current dollars is directly proportional to the product of real GNP per capita and the CPI. Mean family income is α/λ for a gamma density, and consequently, the following log-linear model is proposed.

Table 3. Estimated Coefficients for Predicting the Values of Income Distribution Parameters[1/]

Dependent Variable	Estimated Coefficients	
	Log $\hat{\alpha}$	Log($\hat{\lambda}$ x 10^4)
Independent Variable[2/]		
1. Log Unemployment	-0.3976 (1.80)[3/]	-0.4718 (2.12)[3/]
2. Log Consumer Price Index	-0.5115 (0.98)	-1.7801 (3.38)
3. Log GNP/Capita	+0.2276 (0.38)	-0.6770 (1.11)
4. Constant	+1.1469 (1.16)	+2.4727 (2.46)
1. Number of Observations	10	10
2. Sum of Squared Residuals	0.008189	0.008395
3. Unadjusted R^2	0.88	0.91

1/ The estimated values of α and λ are given in Table 2.

2/ Data Source: Economic Indicators [9].

3/ The numbers in parentheses are the estimated t ratios.

$$(24) \qquad \log\hat{\alpha} = \log(\hat{\lambda} \times 10^4) = a_1 + \log C + \log G + e_1$$

$$\log\alpha = a_2 + b\log U + c\log C + d\log G + e_2$$

where U is the percentage unemployment

C is the CPI

G is the real GNP per capita

The reduced form of this model is

$$(25) \qquad (\log(\hat{\lambda} \times 10^4) + \log C + \log G) = a_2 - a_1 + b\log U + c\log C + d\log G + e_2 - e_1$$

$$(26) \qquad \log\hat{\alpha} = a_2 + b\log U + c\log C + d\log G + e_2$$

The slope coefficients b, c and d in (25) and (26) are identical, and consequently, the ordinary least squares estimators were computed by combining the models together.[1/] The estimate of d in the reduced form is positive but the corresponding t ratio is less than one, and the final model is estimated with the additional constraint $d = 0$. The four estimated coefficients and corresponding t ratios are[2/]

$\hat{a}_1 = -1.3312$ (89.25)

$\hat{b} = -0.5174$ (6.45)

$\hat{c} = -0.5003$ (1.90)

$\hat{a}_2 = 1.5902$ (13.97)

1/ No correction was made for the difference between the residual variances in (25) and (26).

2/ The computed value of the F statistic for testing the hypothesis that (24) is not significantly different from the models in Table 3 is 0.2195. This is less than the critical value (F = 2.48 at the 10% level of significance), and the hypothesis is not rejected.

The estimated t ratios for $\hat{b}$ and $\hat{c}$ are larger than the corresponding values for the α-model in Table 3. The results imply that both increasing unemployment and inflation are associated with a more unequal distribution of income, and that the elasticities of $\hat{\alpha}$ with respect to these two variables have similar magnitudes.[1] However, real growth of GNP per capita has no significant influence on the skewness. In conclusion, the reduction of unemployment between 1960 and 1969 is large enough to offset increasing inflation, and the net effect is a tendency for income to be more equally distributed.

The final part of the empirical section is to investigate the impact of taxation on the distribution of income. The incidence of both federal income tax and state and local taxes on different income classes has been computed by the Tax Foundation Incorporated for 1961 and 1965. A more complete description of this data source is given in Appendix B. A gamma density was fitted to the income data before taxes, after federal taxes, and after all taxes, for both years. The results are summarized in Table 4.

Although state and local taxes account for 32 percent of all taxes in 1961 and also in 1965, the redistributional effects of taxes are entirely attributable to federal taxes.[2] In fact, the skewness and associated inequality measures are slightly larger in the case where all taxes are deducted than in the case where only federal taxes are deducted. These results suggest that state and local taxes are actually regressive.

1/ The elasticities for skewness = $1/\sqrt{\alpha}$ are close to unity in both cases.

2/ The likelihood ratio statistic for testing that the skewness is unchanged by taxes was not computed, as the assumption that the observations before and after taxes are independent is not valid.

Table 4. Estimated Parameters for the Distribution of Income Before and After Taxes for Families in the U.S.: 1961 and 1965[1/]

	1961			1965		
	Before Taxes	After Federal Taxes	After All Taxes	Before Taxes	After Federal Taxes	After All Taxes
Sample Statistics						
1. Arithmetic mean	6578.0	5171.23	4520.26	7828.0	6106.24	5312.72
2. Geometric mean	4976.0	4077.0	3562.5	6138.0	4978.5	4322.6
Estimated Parameters						
1. $\hat{\lambda} \times 10^4$	2.94920	4.37033	4.97759	2.82320	4.25794	4.85627
2. $\hat{\alpha}$	1.94	2.26	2.25	2.21	2.60	2.58
3. Standard deviation	4722.86	3439.85	3013.50	5265.68	3786.93	3307.55
4. Skewness	.718	.665	.666	.673	.620	.623
5. L_γ (see (3))	.3800	.3540	.3545	.358	.3340	.3345
6. I_γ (see (8))	.2364	.2047	.2064	.2093	.1804	.1813

1/ See Appendix B for a description of the taxes included and the assumptions made for generating the data.

5. Conclusions

Any model of income distribution involves some compromise between how well the model fits real data and how easily the parameters can be estimated and interpreted. The gamma density provides a reasonable balance between these two conflicting objectives. The two gamma parameters can be directly related to measures of inequality and of proportionate growth. Both of these concepts have obvious relevance for economists. Maximum likelihood estimates of the parameters may be computed, knowing the arithmetic and geometric means of a sample, using either numerical or graphical procedures. Similar claims for both the estimation and interpretation of parameters may be made for the lognormal density. However, it is shown in Section 4 that the gamma fits family income data much better than the lognormal. In addition, the displaced lognormal is rejected as an alternative due to the difficulty of relating the parameters to an inequality measure (Appendix C).

After fitting gamma densities to family income data for the years 1960 to 1969, inspection of the estimated inequality parameters suggests that income has become more equally distributed over time. A simple regression model relating the ten estimated inequality parameters for 1960-1969 to unemployment, inflation and real GNP per capita suggests that both increasing unemployment and inflation are associated with a more unequal distribution of income. Real growth, however, does not influence the skewness. The fact that income has become more equally distributed during the period analyzed is due to a reduction of unemployment that was sufficient to offset increased inflation.

An analysis of the distribution of family income before and after taxes in 1961 and 1965 was also conducted in Section 4. The estimated gamma

parameters suggest that federal taxes are progressive, but that state and local taxes are slightly regressive.

In summary, the gamma density provides a reasonably close approximation to income distribution data. The skewness parameter is a convenient indicator of inequality that could easily be incorporated as an endogenous variable in an annual macro-economic model. The impact of policy changes on income distribution could then be predicted. The advantage of using the inequality parameter α, instead of an empirical measure such as the entropy (5), is that the proportion of families in different income ranges could be predicted (assuming that the scale parameter λ was also determined by the model). Hence, quite detailed information on income distribution could be extracted from the model without adding unduly to its complexity.

Appendix A: Derivation of the Lorenz Concentration Ratio and Gini Coefficient for a Gamma Density

If X is distributed as the gamma density $f_\gamma(X;\alpha,\lambda)$, the $J^{\underline{th}}$ moment distribution function is defined with $\Gamma(X;\alpha,\lambda) = \int^{X} f_\gamma(u;\alpha,\lambda)du$ as follows

$$\text{(A1)} \qquad \Gamma_J(X;\alpha,\lambda) = \int_0^X u^J d\Gamma(u;\alpha,\lambda) \Big/ \int_0^\infty u^J d\Gamma(u;\alpha,\lambda)$$

$$= \left[\lambda^J \Big/ \prod_{j=1}^{J}(\alpha+j-1)\right] \int_0^X u^J \frac{\lambda^\alpha}{\Gamma(\alpha)} u^{\alpha-1} e^{-\lambda u} du$$

$$= \frac{\lambda^{\alpha+J}}{\Gamma(\alpha+J)} \int_0^X u^{\alpha+J-1} e^{-\lambda u} du$$

$$= \Gamma(X;\alpha+J,\lambda)$$

If X_1 and X_2 are independently distributed as $\Gamma(X_1;\alpha_1,\lambda_1)$ and $\Gamma(X_2;\alpha_2,\lambda_2)$ respectively, then:

$$\text{(A2)} \qquad P(X_1 \leq X_2) = \int_0^\infty \left[\int_0^{X_2} d\Gamma(X_1;\alpha_1,\lambda_1)\right] d\Gamma(X_2;\alpha_2,\lambda_2)$$

$$= \int_0^\infty \Gamma(X_2;\alpha_1,\lambda_1) d\Gamma(X_2;\alpha_2,\lambda_2)$$

but: $$P(X_1 \leq X_2) = P\left(\frac{X_1}{X_1 + X_2} \leq \frac{1}{2}\right)$$

Since X_1 and X_2 are independent gamma densities, the ratio $\frac{X_1}{X_1 + X_2}$ is distributed as the following beta density [22, p. 153]

$$\text{(A3)} \qquad P\left(\frac{X_1}{X_1 + X_2} \leq \frac{1}{2}\right) = \int_0^\infty \Gamma(X_2;\alpha_1,\lambda_1)\, d\,\Gamma(X_2;\alpha_2,\lambda_2)$$

$$= B_{.5}(\alpha_1,\alpha_2)$$

The Lorenz curve and the corresponding measure of income concentration are extensively discussed in [2, pp. 111-115]. The Lorenz concentration ratio can be defined in terms of the distribution function $F(X)$ and the first moment distribution function $F_1(X)$

$$\text{(A4)} \qquad L = 1-2\int_0^\infty F_1(X)\,dF(X)$$

When $F(X) = \Gamma(X;\alpha,\lambda)$, the Lorenz concentration ratio is

$$L = 1-2\int_0^\infty \Gamma_1(X;\alpha,\lambda)\, d\,\Gamma(X;\alpha,\lambda)$$

From (A1) this implies that

$$L = 1-2\int_0^\infty \Gamma(X;\alpha+1,\lambda)\, d\,\Gamma(X;\alpha,\lambda)$$

and using (A3) with $\alpha_1 = \alpha+1$ and $\alpha_2 = \alpha$, gives

$$\text{(A5)} \qquad L = 1-2B_{.5}(\alpha+1,\alpha)$$

L is also related to the Gini coefficient of mean difference which on writing $\Gamma(X) = \Gamma(X;\alpha,\lambda)$ may be defined [2, p. 13]

$$G = \int_0^\infty \int_0^\infty |u-v|\, d\,\Gamma(u)\, d\,\Gamma(v)$$

$$= \int_0^\infty \left[\int_0^U (u-v)\, d\,\Gamma(v)\right] d\,\Gamma(u) + \int_0^\infty \left[\int_U^\infty (v-u)\, d\,\Gamma(v)\right] d\,\Gamma(u)$$

$$= 2\int_0^\infty \left[\int_0^u (u-v)\,d\Gamma(v)\right] d\Gamma(u)$$

$$= 2\int_0^\infty u\left[\int_0^u d\Gamma(v)\right] d\Gamma(u) - 2\int_0^\infty \left[\int_0^u v\,d\Gamma(v)\right] d\Gamma(u)$$

$$= 2\int_0^\infty u\,\Gamma(u)\,d\Gamma(u) - 2\int_0^\infty \left[\int_0^u v\,d\Gamma(v)\right] d\Gamma(u)$$

Using the definition of $\Gamma_1(X;\alpha,\lambda)$ in (A1), this may be rewritten

$$G = 2\frac{\alpha}{\lambda}\left[\int_0^\infty \Gamma(u;\alpha,\lambda)\,d\Gamma(u;\alpha+1,\lambda) - \int_0^\infty \Gamma(u;\alpha+1,\lambda)\,d\Gamma(u;\alpha,\lambda)\right]$$

where $\frac{\alpha}{\lambda} = \int_0^\infty u\,d\Gamma(u)$ is the mean of X

Using (A2) and then (A3) with $\Gamma(X_1;\alpha,\lambda)$ and $\Gamma(X_2;\alpha+1,\lambda)$ gives

$$G = 2\frac{\alpha}{\lambda}\left[P(X_1 \le X_2) - P(X_2 \le X_1)\right]$$

$$= 2\frac{\alpha}{\lambda}\left[2P(X_1 \le X_2) - 1\right]$$

$$= 2\frac{\alpha}{\lambda}\left[2B_{.5}(\alpha,\alpha+1) - 1\right]$$

But by the symmetry of the beta density and using the definition of the Lorenz concentration ratio L in (A4)

$$\text{(A6)} \qquad G = 2\frac{\alpha}{\lambda}\left[1-2B_{.5}(\alpha+1,\alpha)\right]$$

$$= 2\frac{\alpha}{\lambda}L$$

Appendix B: A Description of the Data Sources

Personal Income Before Taxes. The data used for the distribution of family income between 1960 and 1969 were taken from [21]. This series provides in each issue a table entitled, "Money Income - Percent Distribution of Families by Income Level ....". The exact definitions of income and of a family are given in [21] and need not to be repeated here in detail. Essentially, income refers to the "consumer money income for the calendar year." It is personal income before taxes and does not cover nonmoney items of income and is not deflated. In [21], the income scale is divided into ten classes. Although a more detailed classification with more classes, such as the one used in [8], would have been preferred, [21] was chosen for its continuity and consistency.

An important assumption is made when computing the gamma parameters. The midpoint of each income class is considered as representative of that class. In general, the midpoint is not equal to either the arithmetic or geometric mean of the class, and in addition, the fact that the geometric mean is less than or equal to the arithmetic mean is ignored. For the highest income class, a value of 20,000 dollars was used instead of the midpoint.

The Incidence of Taxes. A source for the allocation of taxes, consistent with the family income data in [21], was not readily available. However, another source has been published by Tax Foundation, Inc. [17] for 1961 and 1965. Similar information for 1962 and 1968 may be found in [8], but the data in [8] is not suitable for our purposes. The derivation of the data in [17] involves many assumptions that are not repeated here.

Taxes are reported on the percentage of income before taxes for each income class in [17, Tables 2, 4]. The midpoints of each income class before

taxes are transformed by subtracting the percentage of income going to taxes. This procedure is followed for federal taxes and also for total taxes (federal, state and local).

Appendix C: The Displaced Lognormal Density

Metcalf [10] has proposed using a displaced lognormal density to model income distribution. This implies that $\log(X + C)$ is Normal(μ,σ^2), where X is the income level and C is the displacement parameter. Metcalf suggests an estimation procedure based on the observed values of X at the first, fifth and ninth decile. This procedure, which is described fully in [10], was applied to U.S. family income data for the years 1960 and 1969. The estimated parameters are:

$$1960;\quad \hat{C} = 4990,\quad \hat{\mu} = 9.2695,\quad \hat{\sigma}^2 = .1366$$
$$1969;\quad \hat{C} = 6660,\quad \hat{\mu} = 9.6862,\quad \hat{\sigma}^2 = .1507$$

The positive values of $\hat{C}$ imply that income is distributed between $-\hat{C}$ and ∞, and consequently, the cumulative probability at $X = 0$ is positive.[1/] Either this probability must be added to the probability for the first income class ($0 \leq X < 1000$), or the whole distribution must be truncated at $X = 0$. The first procedure inflated the cumulative probability for the first income class (< 1000) and the fit was worse than with the second procedure. Hence the second procedure was adopted, and the computed probabilities for each income class were rescaled so that the sum of all probabilities equalled one. This rescaling implies dividing the computed probabilities for each interval over the positive range of incomes by one minus the cumulative probability at $X = 0$. In both years, the displaced lognormal fitted better than the gamma.

1/ The cumulative probabilities at $X = 0$ are .0207 and .0116 for 1960 and 1969, respectively. These values are in the same order of magnitude as the observed probabilities of being in the first income class.

The sum over all income classes of squared deviations between the observed and predicted probabilities for the displaced lognormal was 54 percent and 77 percent of the corresponding value for the gamma in 1960 and 1969, respectively (Table 1, p. 14).

Even though the displaced lognormal provides a good fit to the data, there are two serious drawbacks that reduce its usefulness as a model of income distribution. First, the statistical properties of the estimators proposed by Metcalf [10] are unknown, and consequently, testing hypotheses about the parameters is difficult. Secondly, skewness depends on both σ^2 and C.[1/] Hence, the economic interpretation of the parameters is no longer straightforward. In addition, estimates of the Lorenz or Theil inequality measures must be computed empirically as expressions have not been derived for the truncated distribution ($C > 0$).

Table C1. Estimated Theil Inequality Measures

Distribution	Year 1960	Year 1969	Percentage Change
1. Gamma	.2327	.1705	-27
2. Lognormal	.2526	.1488	-41
3. Displaced lognormal	.2113	.1659	-21
4. Observed	.2117	.1647	-22

1/ A change in C changes both the standard deviation and the median, and may result either from proportionate growth or a change in the skewness (see [10, pp. 658-59]).

If the probabilities for each income class are given, the Theil inequality measure may be computed by modifying (5).[1/] The estimates for 1960 and 1969 are summarized in Table C1 using the predicted probabilities from the gamma, lognormal and displaced lognormal, and also the observed probabilities.[2/] Although the estimates from the displaced lognormal are closest to the estimates from the observed probabilities, the gamma does not perform badly in this respect. It should be remembered that both Theil and Lorenz inequality measures may be computed directly from the gamma and lognormal parameters, but this is not the case for the displaced lognormal.

1/ The modified expression for the Theil inequality measure is

$$I = \Sigma_j P_j (X_j/\bar{X}.)\log(X_j/\bar{X}.)$$

where P_j is the probability of being in the jth income class

X_j is the midpoint of the jth income class

$\bar{X}.$ is the mean income.

2/ Neither (8) nor the corresponding expression for the lognormal [18, p. 97] are used to compute these estimates. This ensures that all models are compared under the same restrictions imposed by the use of grouped data.

References

[1] Abramowitz, M. and Stegum, I. A., Handbook of Mathematical Functions, Washington, U.S. Govt. Print. Off. (1964).

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

[3] Amoroso, L., "Ricerche intorno alla curva dei redditi," Annali di Mathematica Pura ed Applicata, Ser. 4-21, p. 123 (1925).

[4] Champernowne, D. G., "A Model of Income Distribution," Economic Journal 63, p. 318 (1953).

[5] Choi, S. C. and Wette, R., "Maximum Likelihood Estimation of the Parameters of the Gamma Distribution and their Bias," Technometrics V. II, No. 4, pp. 683- 90 (1969).

[6] Davis, H. T., Tables of the Higher Mathematical Functions, Principia Press, Inc. VI. (1933).

[7] _____, Tables of the Higher Mathematical Functions, Principia Press, Inc. V. II. (1935).

[8] Herriot, R. A. and Miller, H. P., "Tax Changes Among Income Groups 1962-1968," Business Horizons, Vol. XV, No. 1, pp. 41-50 (Feb. 1972).

[9] Joint Economic Committee, The Council of Economic Advisors, Economic Indicators, U. S. Govt. Print. Off., Washington, D. C.

[10] Metcalf, C. E., "The Size Distribution of Personal Income During the Business Cycle," American Economic Review 59, pp. 657-69 (1969).

[11] Pearson, K., Pearson's Early Statistical Papers, Cambridge University Press (1948).

[12] _____, Tables of the Incomplete Beta Function, The "Biometrika" Office, University College, London (1934).

[13] _____, Tables of the Incomplete Gamma Function, His Majesty's Stationery Office (1922).

[14] Quensel, C. E., Inkomstfördelning och Skattetryck, Stockholm, Serviges Industriforbund (1944).

[15] Simon, H. A. and Bonini , Ch. P., "The Size Distribution of Business Firms," American Economic Review 54, (Sept. 1958).

[16] Simon, H. A., "On a Class of Skew Distribution Functions," Biometrika 52, pp. 425- 40 (1955).

[17] Tax Foundation, Inc., "Allocating Tax Burdens and Government Benefits by Income Class," Gov. Finance Brief No. 8, (Jan. 1967).

[18] Theil, H., Economics of Information Theory, Rand McNally (1967).

[19] Thurow, L. C., "Analyzing the American Income Distribution," American Economic Review, Papers and Proceedings Issue, pp. 261-69 (1970).

[20] U.S. Bureau of the Census, Current Population Reports, Ser. P. 60 and P. 20.

[21] U.S. Department of Commerce, Bureau of the Census, Statistical Abstract of the United States (1960-71).

[22] Weatherburn, C. E., A First Course in Mathematical Statistics, Cambridge, The University Press (1947).

[23] Wilks, S. S., "The Large-Sample Distribution of the Likelihood Ratio for Testing Composite Hypotheses," Annals of Mathematical Statistics 9, (March 1938).