CENTRE FOR DEVELOPMENT STUDIES

WORKING PAPER No.18

SOME ANALYTICAL CHARACTERISTICS OF THE DISTRIBUTION OF LAND: AN EXERCISE IN GRADUATION BY THE LOGNORMAL, GAMMA AND LOGGAMMA LAWS

by

N. Krishnaji

Centre for Development Studies
Aakulam Road, Ulloor,
Trivandrum 695011

February 1974

SOME ANALYTICAL CHARACTERISTICS OF THE DISTRIBUTION OF LAND: AN EXERCISE IN GRADUATION BY THE LOGNORMAL, GAMMA AND LOGGAMMA LAWS

1. INTRODUCTION

For studying the analytical characteristics of a distribution, ideally one should have a random sample of observations from the corresponding population. Such a sample makes it possible for us to discriminate between alternative forms of distribution, test various parametric hypotheses in a rigorous manner, estimate inequality measures and in general make probability statements relating to the population. Published data on the distribution of land do not permit such an analysis. The National Sample Survey (NSS) give estimates (see e.g. [9]) of the number of households with ownership holdings belonging to different size-classes; these estimates are themselves based on sample surveys. Thus what we have are estimates (for the different States of India) of the population distribution function $F(x)$ corresponding to a fixed set of values $x = x_i$, $i = 1, 2 \ldots k$. Although 'fitting' theoretical forms like the lognormal law to such data can be done in a conventional manner it is obvious that statistical tests of significance are possible only if the underlying sampling distributions — which involve the complicated survey designs — are derived. However, one can 'approximate' the published estimates of $F(x)$ by different theoretical forms and rely on judgment based on some form of a distance measure rather than rigorous tests to discriminate between alternative forms.

In this paper we study some aspects of inequality in the distribution of land in the different States of India in 1961-62. For this purpose we have approximated the NSS estimates of the population distribution by three alternative families of distributions viz. the lognormal, gamma and loggamma laws. Our analysis shows that both the gamma and loggamma distributions approximate the NSS distributions uniformly better than the lognormal distribution. As a consequence it is possible to obtain a number of interesting analytical characterisations of the inequality in the distribution. Finally we have also considered the problem of estimating the concentration ratio on the basis of NSS data.

We may add that this paper forms part of a study on the structure of leasing and the distribution of land which is primarily based on the following two empirical observations:

(a) Leased in area appears to form a constant proportion of owned area, and

(b) the distribution of owned area and operated area (which is owned area minus leased out area plus leased in area) appear to be nearly identical.

For an examination of these hypotheses and their implications we may refer to Raj [11]. In a forthcoming paper we have attempted to make a precise and testable formulation of (a) and (b) above; we have shown, in particular, that the form of the distributions of owned and operated area determine the structure of leasing and *vice versa*; two alternative sets of assumptions on the structure of leasing are shown to imply that the distribution of land must be of the gamma or loggamma type; the underlying probability models also explain the phenomenon that characterises (b) above. This is the background of the present paper.

2. THE THREE DISTRIBUTIONS

2.1. Definitions: The lognormal distribution, trusted old friend of the economist, needs no introduction. A random variable (r.v.) X has a gamma distribution with parameters (α, λ), $\alpha > 0$, $\lambda > 0$, if it has a density

$$f(x) = [\Gamma(\alpha)]^{-1} \exp(-\lambda x)(\lambda x)^{\alpha-1}\lambda, \quad x \geq 0 \qquad \ldots (1)$$
$$= 0 \qquad x < 0$$

and a r.v.Y has a loggamma distribution with parameters (x_0, β, ν), $x_0 > 0$, $\beta > 0$, $\nu > 0$, if $\log(Y/x_0)$ has a gamma distribution with parameters (β, ν). A simple transformation of (1) shows that the density of Y is

$$g(y) = [\Gamma(\beta)]^{-1} (x_0/y)^{\nu+1} [\nu \log(y/x_0)]^{\beta-1} (\nu/x_0), \quad y \geq x_0 \quad .. (2)$$
$$= 0 \qquad y < x_0$$

For $\beta = 1$ (2) reduces to $(x_0/y)^{\nu+1}(\nu/x_0)$ which is a Pareto density. Thus the Pareto distribution can be termed as the log exponential and the loggamma is a simple generalisation of the Pareto law.

2.2. Moments: We immediately note that while for the gamma distribution all moments are finite it is not so far the loggamma. In fact, if Y has the density g(.) since $X = \log(Y/x_0)$ has a gamma law, we have

$$E(Y^n) = E(x_0^n e^{nX}) = x_0^n \int_0^\infty [\Gamma(\beta)]^{-1} \exp[x(n-\nu)] (\nu x)^{\beta-1} dx$$

which is not finite when $n \geq \nu$ for any value of $\beta (>0)$. It is easy to verify that

$$E(Y^n) = x_0^n (1 + \frac{n}{\lambda})^{-\beta} \quad \text{for} \quad n < \lambda \qquad \ldots (3)$$

where n is not necessarily integral valued.

We also note the following reproductive property of the loggamma which is similar to the one enjoyed by the lognormal distribution: If X and Y have independent loggamma distributions with parameters (x_0, α, λ) and (y_0, β, λ) respectively then the product XY has a loggamma distribution with parameters $(x_0 y_0, \alpha + \beta, \lambda)$. This follows from a well-known property of the gamma law. (Note that the parameter λ has to be the same for the two distributions.)

2.3. <u>Lorenz Measure of Concentration</u>: For any non-negative r.v.X with a finite first moment if we write F(.) for the cumulative distribution function (c.d.f.) a much used measure of concentration, the Lorenz measure, is given by (see e.g. Aitchison and Brown [1])

$$L = 1 - 2\int_0^\infty F_1(x)\, dF(x) \qquad \ldots (4)$$

where

$$F_1(x) = \int_0^x t\, dF(t) \Big/ \int_0^\infty t\, dF(t) \qquad \ldots (5)$$

is also a c.d.f. which gives the 'share' of the population below x in the total. When X has a lognormal distribution with parameters (μ, σ) i.e. when log X is distributed normally (μ, σ) it is well-known [1] that

$$L = 2\,\Phi(\frac{\sigma}{\sqrt{2}}) - 1 \qquad \ldots (6)$$

where $\Phi(.)$ is the c.d.f. corresponding to the standard normal distribution.

distribution.

When $F(.)$ has a density given by (1), substitution of (1) in (5) shows that $F_1(.)$ corresponds to the c.d.f. of a gamma variate with parameters $(\alpha + 1, \lambda)$. Now (4) can be rewritten as

$$L = 1 - 2P(U \leq V) \qquad \ldots (7)$$

where U and V are an arbitrary pair of independent r.v.s with c.d.f.s $F_1(.)$ and $F(.)$ respectively. Thus when X is gamma (α, λ) for computing L we set in (7) a pair of independent gamma variates with parameters $(\alpha + 1, \lambda)$ and (α, λ) for U and V respectively. But in this case $U/(U+V)$ has a beta distribution with parameters $(\alpha + 1, \alpha)$ and hence

$$\begin{aligned} P(U \leq V) &= P(U/(U+V) \leq 1/2) \\ &= B(\alpha+1, \alpha)^{-1} \int_0^{1/2} x^{\alpha} (1-x)^{\alpha-1} dx \end{aligned} \qquad \ldots (8)$$

where $B(.,.)$ is the beta function. Now (7) and (8) yield after simplification

$$L = \frac{\Gamma(2\alpha + 1)}{[\Gamma(\alpha+1)]^2} (1/2)^{2\alpha} \qquad \ldots (9)$$

which gives the Lorenz measure corresponding to a gamma distribution given by (1).

For computing L corresponding to the loggamma distribution we use a similar procedure. When $F(.)$ has a density given by (2) we note that the first moment is finite and equals $x_0(1+ \nu^{-1})^{-\beta}$ only when $\nu > 1$; the Lorenz measure is thus defined only when $\nu > 1$. A substitution of $g(t)dt$ (given in (2)) for $dF(t)$ in (5) then shows that $F_1(.)$ is the c.d.f. of some r.v. distributed as a loggamma with parameters $(x_0, \beta, \nu+1)$. We then must compute the probability in equation (7) inserting for U and V a pair of independent loggamma variates with parameters $(x_0, \beta, \nu+1)$ and (x_0, β, ν) respectively. In this case $U' = (\nu+1) \log (U/x_0)$ and $V' = \nu \log (V/x_0)$ are independent and identical gamma variates with parameters $(\beta, 1)$ and hence $U'/(U'+V')$ has a beta distribution (β, β). It is easily verified that $P(U/V \leq 1) = P\left[U'/(U'+V') \leq (\nu+1)/(2\nu+1)\right]$ and hence from (7)

$$L = 1 - 2B_{(\nu+1)/(2\nu+1)}(\beta, \beta) \qquad \ldots (10)$$

where $B_x(m,n)$ stands for the incomplete beta integral

$$\frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)} \int_0^x t^{m-1} (1-t)^{n-1} \, dt$$

Equations (6), (9) and (10) give the Lorenz measures of concentration corresponding to the lognormal, gamma and loggamma laws respectively. In the case of the lognormal (μ, σ) it is well known [1] that $\sigma^2 = \log (1 + \eta^2)$ where η is the coefficient of variation (c.v.) of the distribution and hence (6) implies that ranking (of different States, say) by L is equivalent to ranking by the c.v. Similarly the (c.v) corresponding to the gamma distribution (1) is given by $\alpha^{-1/2}$ and it is easy to check that L given by (9) decreases with increasing α. Hence even when the underlying distributions are of the gamma type ranking by L is equivalent to ranking by the (c.v.). We have not been able to verify whether this property holds for the loggamma law.

3. GRADUATION

3.1. Landless households: The three distributions considered here are all absolutely continuous with respect to Lebesgue measure over their respective supports. This introduces a complication because none of these distributions can explain the mass at zero i.e. the presence of landless households.* With respect to the loggamma there is the further difficulty that it assigns a zero mass to $[0, x_0]$; in principle this can be taken care of by setting a very small $x_0 > 0$ but the problem of the landless remains. For the purposes of this paper we therefore restrict our approximations to the ranges in which the respective densities are positive although in computing the concentration ratios we attempt to make some adjustments that take into account the positive mass at zero.

3.2. Method of Graduation: Our basic data relate to the seventeenth round of the NSS (1961-62) and give estimates for all the States of India of the number of households belonging to different size groups of acreage of household ownership holdings [9]; the corresponding estimated average size of holding in each size group is also given. Thus while estimates of the population mean are already available estimates of variances are not. Accordingly for the purpose of graduation we have used the following procedure:

---

* In fact, there is practically no theory that simultaneously explains the formation of the 'landless' category and the distribution of land among the 'landed' in an economy characterised by private property. The present author has made one feeble attempt [6] to explore the process of formation of landless households under conditions of semi-feudalism.

(a) Lognormal distribution: The following are computed for each State separately:

$$\hat{\mu} = N^{-1} \sum f_i \bar{x}_i \quad \text{and} \quad \hat{\sigma}^2 = N^{-1} \sum f_i \bar{x}_i^2 - \hat{\mu}^2 \qquad \ldots (11)$$

where f_i = the NSS estimate of number of households belonging to the i-th size group, $\sum f_i = N$, and $\bar{x}_i = \log \bar{X}_i$ where $\bar{X}_i$ = the NSS estimate of the average size of holding of the i-th group, all the summations (hereafter also) being from 1 to k (k = 13 for all the States).

A lognormal distribution with parameters $(\hat{\mu}, \hat{\sigma})$ is then regarded as an approximation to the distribution estimated by the NSS. (We may add that it is possible to derive an approximation by calculating the first two moments of the untransformed data (i.e. X and not log X) and equating them to the moments of a lognormal distribution. We have actually carried out this procedure and found that the resulting approximations are uniformly poorer than those reported in the paper. One also intuitively feels that if instead of $\bar{X}_i$ the corresponding geometric means are used in (11) better approximations would result; however the latter are not available).

(b) Gamma distribution: Here again we ape the relevant sampling theory. Let $v = \log \bar{X} - \hat{\mu}$ where $\bar{X} = N^{-1} \sum f_i \bar{X}_i$, the NSS estimate of the population mean and $\hat{\mu}$ is the same as in (11). Then $\hat{\alpha}$ is derived by making use of Tables for the maximum likelihood estimators of the parameter α in (1) corresponding to the value v (pp.304-305 of [10]). Then we set $\bar{X} = \hat{\alpha} / \hat{\lambda}$ and regard the gamma distribution

$(\hat{\alpha}, \hat{\lambda})$ as an approximation. Finally we have used simple linear interpolation based on the Khamis Tables [5] for obtaining the gamma approximations to the number of households in the NSS size groups.

(c) Loggamma distribution: This distribution can explain the frequency of only those households that belong to (x_o, ∞) but x_o is unknown. To circumvent this problem we tried calculations with $x_o = 0.005$ acres which is in fact the smallest size that is included in the NSS estimates. However, the results were not satisfactory and accordingly we chose $x_o = 1.0$ acre for the purpose of this paper. Since the hypothesis that X has a log gamma distribution on $(1, \infty)$ is equivalent to log X having a gamma distribution, a procedure similar to (b) above is used after making the relevant logarithmic transformations for deriving the loggamma approximations.

3.3. Some results: For judging the closeness of approximation the following measure is employed for each of the three distributions:

$$d = \sum \left[(f_i - p_i)^2 / p_i \right] \qquad \ldots (12)$$

where f_i and p_i are the number of households in the i-th size class estimated by the NSS and the number graduated by the approximating law respectively. Although the actual computations were made for the 13 size classes of the NSS data, Table 1 that follows gives the figures for a collapsed set of five groups; the corresponding d values (reproduced in Table 2) are also computed on the basis of these five groups for the lognormal and gamma laws and the last four groups for the loggamma distribution.

Table 1

The estimated distribution of household ownership holdings and some theoretical approximations

Size Group (Acres)	Number of households (in thousands): Estimated by NSS	Approximated by: LN	G	LG
Andhra Pradesh				
0- 1	2759	3340	2540	-
1- 5	1952	1639	1893	2162
5-10	704	470	861	628
10-20	447	331	608	335
> 20	324	406	284	303
Bihar				
0- 1	3611	4111	3336	-
1- 5	2790	2416	2880	2909
5-10	846	613	1018	768
10-20	391	348	458	314
> 20	131	292	77	165
Jammu & Kashmir				
0- 1	118	166	139	-
1- 5	317	262	289	337
5-10	82	66	36	61
10-20 } > 20 }	21	43	21	22
Madhya Pradesh				
0- 1	1090	1573	1235	-
1- 5	1469	1525	1485	1690
5-10	1063	616	867	1013
10-20	867	477	794	626
> 20	489	789	596	559

Size Group (Acres)	Number of households (in thousands): Estimated by NSS	Approximated by: LN	G	LG
Assam				
0- 1	442	498	546	-
1- 5	739	678	616	767
5-10	226	168	214	181
10-20	52	80	78	54
> 20	3	37	8	17
Gujarat				
0- 1	723	1024	792	-
1- 5	694	722	743	761
5-10	492	268	415	556
10-20	451	227	386	350
> 20	318	438	342	287
Kerala				
0- 1	1034	1024	345	-
1- 5	552	589	730	574
5-10	96	76	127	79
10-20 } > 20 }	40	32	20	35
Madras				
0- 1	2537	3128	2520	-
1- 5	1877	1510	1865	2008
5-10	489	282	536	398
10-20	168	131	185	131
> 20	54	75	19	50

Table 1 (Continued)

Number of households (in thousands)

Maharashtra

Size Group (Acres)	Estimated by NSS	Approximated by LN	Approximated by G	Approximated by LG
0- 1	1266	1701	1331	-
1- 5	1154	1217	1207	1349
5-10	741	457	673	781
10-20	706	382	639	510
>20	585	695	601	545

Mysore

Size Group (Acres)	Estimated by NSS	Approximated by LN	Approximated by G	Approximated by LG
0- 1	603	758	641	-
1- 5	853	1025	917	1045
5-10	754	409	557	601
10-20	422	311	482	356
>20	270	400	306	297

Orissa

Size Group (Acres)	Estimated by NSS	Approximated by LN	Approximated by G	Approximated by LG
0- 1	1393	1765	1284	-
1- 5	1346	1064	1211	1450
5-10	490	296	542	402
10-20	202	188	348	168
>20	71	188	118	89

Punjab

Size Group (Acres)	Estimated by NSS	Approximated by LN	Approximated by G	Approximated by LG
0- 1	1018	1179	930	-
1- 5	436	510	591	585
5-10	341	164	281	224
10-20	246	127	229	148
>20	145	204	155	211

Rajasthan

Size Group (Acres)	Estimated by NSS	Approximated by LN	Approximated by G	Approximated by LG
0- 1	441	653	541	-
1- 5	772	844	735	880
5-10	621	420	481	625
10-20	531	335	511	432
>20	482	594	578	468

U.P.

Size Group (Acres)	Estimated by NSS	Approximated by LN	Approximated by G	Approximated by LG
0- 1	4534	5879	4790	-
1- 5	5639	4367	5110	5934
5-10	1836	1258	1984	1594
10-20	763	784	946	630
>20	229	712	171	309

West Bengal

Size Group (Acres)	Estimated by NSS	Approximated by LN	Approximated by G	Approximated by LG
0- 1	1896	2285	1832	-
1- 5	1716	1350	1745	1794
5-10	482	333	515	422
10-20	164	189	165	132
>20	39	138	40	52

Notation:

LN = lognormal
G = gamma
LG = loggamma

Table 2

Measures of distance between NSS estimates and theoretical approximations

State	Value of d corresponding to the approximation by the		
	Lognormal distribution	Gamma distribution	Loggamma distribution
Andhra Pradesh	335	98	67
Assam	73	58	25
Bihar	291	103	41
Gujarat	529	35	46
Jammu and Kashmir	41	5	8
Kerala	10	113	5
Madhya Pradesh	907	88	133
Madras	368	70	41
Maharashtra	582	20	109
Mysore	433	87	88
Orissa	355	108	37
Punjab	354	64	185
Rajasthan	308	79	37
Uttar Pradesh	1274	135	101
West Bengal	308	4	24

Table 3

Parameters of the distributions used for approximation

State	Lognormal		Gamma		Loggamma ($x_0=1$)	
	μ	τ	α	λ	β	ν
Andhra Pradesh	-0.2209	2.1303	0.4076	0.0896	2.1046	1.4025
Assam	0.5231	1.2611	0.6730	0.2174	3.6762	2.9485
Bihar	-0.1281	1.7507	0.5106	0.1680	2.8026	2.1035
Gujarat	0.7023	2.3389	0.4492	0.0528	3.4479	1.7295
Jammu & Kashmir	0.6037	1.2130	1.0590	0.3369	3.3044	2.9116
Kerala	-0.3021	1.2520	0.6736	0.3697	2.2287	2.2115
Madhya Pradesh	0.9760	2.0150	0.5452	0.0651	3.6362	1.8759
Maharashtra	0.6969	2.2800	0.4378	0.0496	3.2240	1.6057
Madras	-0.4504	1.5800	0.4880	0.2086	2.8847	2.4592
Mysore	1.1084	1.7370	0.6265	0.0788	3.6477	1.9380
Orissa	-0.0260	1.8800	0.4977	0.1124	2.8774	2.1162
Punjab	-0.2546	2.4715	0.3434	0.0621	1.9910	1.0046
Rajasthan	1.4286	1.9270	0.5756	0.0468	3.7174	1.7414
Uttar Pradesh	0.2058	1.7485	0.5866	0.1659	2.9722	2.2296
West Bengal	-0.0769	1.6602	0.5896	0.2227	3.3903	2.7149

Since the NSS estimates are given in figures rounded off to thousands the d values (see (12)) will also be in the same units. Finally Table 3 gives the parameters of the distributions used for approximation.

Now, Tables 1 and 2 clearly show that both the gamma and loggamma distributions approximate the NSS distribution uniformly better than the lognormal distribution. (We must add that the d values corresponding to the loggamma are, strictly speaking, not comparable to the other two sets because of the difference in the range of approximation but the visual impression of the closeness as revealed in Table 1 does lead to the stated conclusion.) The only exception appears to be Kerala for which the lognormal yields a better approximation than the gamma but even in this case the loggamma appears to perform best. No clear verdict is possible in respect of the comparison between the gamma and loggamma laws. In an earlier study Mukherji [8] has also found the superiority of the gamma over the lognormal distribution. Table 1 also shows that the lognormal overestimates (with respect to the NSS data) the tail frequencies at the expense of the middle groups while the gamma distribution underestimates the upper tail; the loggamma appears to be free of such systematic biases in graduation.

Turning to Table 3 we find that, again with the exception of Kerala, the α parameter corresponding to the gamma distribution is less than unity in all the States. This implies that all these distributions have a decreasing failure rate; the implications of this are examined in Section 5 below. Another interesting fact is that α is close to one-half in a number of cases; this implies that a square-root transformation (instead of the logarithmic one) may make the distribution approximate a normal distribution. Thus the 'square-root normal' must also be regarded as a close competitor of the lognormal.

Before concluding this section we must note that the gamma approximations have been derived here by using simple interpolation between tabulated values; the approximations are likely to improve if better methods are used.

4. ESTIMATION OF LORENZ RATIO

4.1. Some methods and results: A commonly used formula for estimating the Lorenz measure on the basis of grouped data of the NSS type is

$$L = \sum_{i=1}^{k-1} \left[\hat{F}(x_i)\, \hat{F}_1(x_{i+1}) - \hat{F}(x_{i+1})\, \hat{F}_1(x_i)\right] \qquad \ldots (13)$$

where (x_i, x_{i+1}) defines the i-th size class and $\hat{F}(x)$ and $\hat{F}_1(x)$ stand for the estimated proportions of households below x and their corresponding share in total respectively. Formula (13) is easily seen to be an approximation to the integral (4); this is derived on the basis of a two-point quadrature formula for each size group. It is difficult to predict the sign of the bias produced by (13) but it is obvious that the approximation is likely to be good only when the width of the size classes is narrow.

Alternatively one can employ equation (4) taking for $F(.)$ and $F_1(.)$ the approximate theoretical forms. The resulting approximations to L will be only as good as the approximation of $F(.)$ to the original data.

Estimates of L based on formulae (6) and (9) are given below. Since concentration ratios based on (10) corresponding to the loggamma law ignore the range $[0, x_0]$ we have not computed them.

Table 4

Estimates of Lorenz Ratio

State	Estimate based on		State	Estimate based on	
	Lognormal	Gamma		Lognormal	Gamma
Andhra Pradesh	0.8680	0.6765	Madras	0.7346	0.6413
Assam	0.6276	0.5779	Mysore	0.7806	0.5934
Bihar	0.7844	0.6325	Orissa	0.8160	0.6349
Gujarat	0.9018	0.6574	Punjab	0.9195	0.7078
Jammu & Kashmir	0.6092	0.4890	Rajasthan	0.8272	0.6089
Kerala	0.6238	0.5777	Uttar Pradesh	0.7836	0.6052
Madhya Pradesh	0.8458	0.6196	West Bengal	0.7596	0.6042
Maharashtra	0.8930	0.6623			

Since the lognormal overestimates the tails and the gamma approximations are on the whole better we must conclude that the estimates based on the lognormal have a strong upward bias. On the other hand one might argue that even the gamma approximations are not good enough for estimating a summary measure like L. However, we find that in respect of Jammu and Kashmir, Maharashtra and West Bengal the gamma approximations are fairly close and we may accept the corresponding estimates of L as the best possible. By the same token we find the lognormal approximation to Kerala good enough for estimating L. In what follows we discuss the concentration in these four States.

4.2. Adjustment for the landless category: The Lorenz ratios given in Table 4 refer to concentration among the households owning some land. For certain purposes, however, it is necessary to compute the inequalities among the entire population including the landless households. If $F(.)$ is the c.d.f. corresponding to the distribution of land among the landed then

$$G(x) = p + (1-p)F(x) \qquad \ldots (14)$$

where $G(0) = p > 0$ is the proportion of landless households in the total, gives the c.d.f. of the distribution of land among all households. The corresponding 'share' distribution $G_1(x)$ is easily seen to be equal to $F_1(x)$ and a simple computation of (4) leads to

$$L_G = 1 - (1-p)(1-L_F) \qquad \ldots (15)$$

where L_F and L_G are the Lorenz ratios corresponding to $F(.)$ and $G(.)$ respectively. It is interesting to note that (15) is valid for the approximate ratios computed by formula (13); this has been proved by V.M.Rao [12].

It can be similarly shown that if

$$G(x) = pH(x) + (1-p)F(x), \quad p > 0 \qquad \ldots (16)$$

where $H(x)$ and $F(x)$ are c.d.f.s with supports $[0, x_0]$ and (x_0, ∞) respectively and p is the proportion of households below x_0 in the total then

$$L_G = 1 - pa(1-L_H) - qb(1-L_F) \qquad \ldots (17)$$

where $q = 1-p$, $a = \mu_H / \mu_G$ and $b = \mu_F / \mu_G$, the μ's standing for the corresponding means.

Formula (17) can be employed for computing the Lorenz ratio of the whole distribution when in (16) F(.) corresponds to a loggamma but then we would need knowledge of H(.), the distribution of land in the range $[0, x_0]$. We have not attempted to approximate H(.) in this paper.

The concentration measures adjusted for the landless category in respect of the four States mentioned above are given below (Table 5).

Table 5

Estimates of Lorenz Ratio

State	Approximation used	Proportion of landless households (%)	Lorenz Ratio for distribution among	
			Landed households	All households
(1)	(2)	(3)	(4)	(5)
1. Jammu & Kashmir	Gamma	10.93	0.4890	0.5449
2. Kerala	Lognormal	30.90	0.6238	0.7400
3. Maharashtra	Gamma	16.03	0.6623	0.7164
4. West Bengal	Gamma	12.56	0.6042	0.6599

Source for Col.(3): Reference [9].

The crudely estimated values of L using formula (13) comparable to Col.(4) above, are 0.4749, 0.6645, 6513 and 6205 respectively for the four States 1 to 4 listed in the above Table. The lesson to draw is that if ranking is the objective one must be extremely cautious with crude methods and fitted distributions alike; inference based on summary measures acquire some validity only when the underlying approximations are good.

5. SOME ANALYTICAL ASPECTS OF INEQUALITY

5.1. Failure rate analysis: As we have seen one generalisation that emerges from this exercise is that the land distributions are approximated well by gamma distributions with the α parameter less than unity. Now it is well known [2] that this implies that these distributions have a decreasing failure rate (DFR). This means that the c.d.f. F(.) is such that

$$\frac{F(t+x) - F(t)}{U(t)} \quad \ldots (18)$$

where $U(t) = 1-F(t)$, is decreasing in t for $x > 0$ and $t \geq 0$. Since the gamma distribution has the density f(t) given by (1) this is equivalent to saying that

$$r(t) = \frac{f(t)}{U(t)} \quad \ldots (19)$$

is decreasing in t (≥ 0). r(t) is called the failure rate of F(t).

This has interesting implications for the skewness in the distribution F(.). For, suppose the households are arranged in classes of equal width x; then we know that the tail U(t) — the proportion of households beyond t — is a decreasing function of t. But (18) implies that the frequency in each size group declines at a faster rate than the corresponding tail U(t) which is a precise description of the relative abundance of the small holdings all along.

At the same time as we have already noted the share distribution $F_1(.)$ is gamma $(\alpha+1, \lambda)$ and the first parameter being greater than unity implies that $F_1(.)$ has an increasing failure rate (IFR) [2]. This means that $[F_1(t+x) - F_1(t)] / [1-F_1(t)]$

is increasing in t for $x > 0$ and $t \geq 0$. So while the share in the tail declines as we go along the acreage axis the share of the groups (of width x) increases at a rate faster than the rate of decline of the corresponding tail share. This describes the relative 'richness' of the large holders in terms of acreage. We may paraphrase the situation by saying that the strength of the small holders lies in their number while the strength of the large holders lies in their wealth.

For the lognormal distribution the DFR property holds only in the large-size range and not throughout $(0, \infty)$ irrespective of the values of the parameters [2]. Since the share distribution is also lognormal it follows that the corresponding failure rate declines ultimately (and does not increase as in the case of the gamma).

We shall now investigate the behaviour of the loggamma distribution. Suppose Y has a loggamma distribution with parameters $(1, \beta, \nu)$. Then log Y is gamma (β, ν) and if we write s(t) and r(t) for the failure rates of Y and X respectively it is easy to establish that

$$s(t) = \frac{r(\log t)}{t} \qquad t > 1 \qquad \ldots (20)$$

Let us recall that Table 3 tells us that for the distributions we are interested in $\beta > 1$. Let us explicitly assume that $\beta > 1$. It follows that r(t) is an increasing function of t (≥ 0). From (20) we get

$$s'(t) < 0 \quad \text{iff} \quad r'(\log t) < r(\log t) \text{ for } t > 1 \qquad \ldots (21)$$

By using the explicit expression for the gamma density (β, ν) it is easy to show that

$$r'(t) < r(t) \quad \text{iff} \quad r(t) < 1+\nu - \frac{\beta-1}{t} \quad \text{for } t>0 \qquad \ldots (22)$$

But it is known that when $\beta>1$ the failure rate $r(t)$ of the corresponding gamma (β, ν) is bounded above by ν [2, p.16] and therefore the second inequality in (22) is automatically satisfied when $1 - (\beta-1)/t > 0$ or when $t > \beta - 1$. The equivalence (21) then implies that

$$s'(t) < 0 \quad \text{for} \quad t > \exp(\beta - 1)$$

and thus the loggamma with $\beta > 1$, like the lognormal, has the DFR property in the large-size range; since the share distribution is also a loggamma with the same β parameter it has a similar property.

5.2. A property of the loggamma law

Manish Bhattacharjee has recently proved the following interesting theorem [3].

If $F(.)$ is a non-discrete distribution with a monotone failure rate (i.e. either DFR or IFR) and support $[0,\infty]$ then it has a unique representation

$$F(t) = 1 - e^{-\rho t} L(e^t) \qquad t > 0 \qquad \ldots (23)$$

where $0 \leq \rho < \infty$ and $L(.)$ is slowly varying.

We may refer to Feller [4] for the concept of slowly varying functions.

Now suppose that X has a loggamma distribution with $x_0 = 1$. If we write $G(t)$ and $F(t)$ for the c.d.f.s of X and log X respectively then F(.) corresponds to a gamma distribution which has a monotone failure rate and hence F(.) satisfies (23). And since G(t) = F(log t) we then get

$$G(t) = 1 - t^{-\alpha} L(t) \qquad \ldots (24)$$

Now it can be shown (see Feller [4] p.279, problem 27) that if U and V have a common distribution H(.) satisfying

$$H(x) \sim 1 - x^{-\alpha} L(x) \qquad \ldots (25)$$

with L(.) slowly varying then

$$P(U > t \mid U+V > t) \rightarrow 1/2 \quad \text{as} \quad t \rightarrow \infty \qquad \ldots (26)$$

From equation (24) we see that G(t) the c.d.f. of X satisfies something stronger than (25). Hence if X and Y have a common loggamma distribution then (26) is valid with X and Y replacing U and V respectively which can best be paraphrased in Feller's words: "Roughly speaking, a large value for the sum is likely to be due to the contribution of one of the two variables".

This is an interesting aspect of skewness which was first noticed by Mandelbrot [7] in the context of income distributions; he shows that certain stable distributions which are considered to be appropriate in income distribution analysis satisfy the property (26).

ACKNOWLEDGEMENT

Thanks are due to M.K.Sukumaran Nair and M.Purushothaman Nair for participating in some early computational experiments that have led to the results reported here. Inspiration for the last section was provided by MCB.

REFERENCES

[1] AITCHINSON, J. and BROWN, J.A.C.(1969): The lognormal distribution, Cambridge University Press.

[2] BARLOW, RICHARD E., et.al.(1967): Mathematical Theory of reliability, John Wiley.

[3] BHATTACHARJEE, MANISH C.(1974): A representation of distributions with monotone failure rate and their tail behaviour, paper to be contributed to the 37th meeting of IMS, Edmonton, Canada.

[4] FELLER, W.(1966): An introduction to probability theory and its applications, Vol.II, John Wiley.

[5] KHAMIS, SALEM H. (1965): Tables of the incomplete gamma function ratio, Justus von Liebeg Verlag, Darmstadt, Germany.

[6] KRISHNAJI, N.(1973): Semi-feudalism and perpetual indebtedness: A Comment, Working Paper, Centre for Development Studies, Trivandrum.

[7] MANDELBROT, BENOIT (1960): The Pareto-Levy law and the distribution of income, International Economic Review, 1, pp.79-106.

[8] MUKERJI, V.(1967): Type III distribution and its stochastic evolution in the context of distribution of income, landholdings and other economic variables, Sankhya, Ser A, 29, pp.405-416.

[9] THE NATIONAL SAMPLE SURVEY (1968): Number 144, Tables with notes on some aspects of landholdings in rural areas, Government of India.

[10] PEARSON, E. and HARTLEY, H.O.(Eds) (1972): Biometrika tables for Statisticians, Cambridge University Press.

[11] RAJ, K.N.(1970): Ownership and distribution of land, Indian Economic Review, Vol.V, No.1.

[12] RAO, V.M.(1970): Adjustment of measure of inequality in rural land ownership for landless categories, Indian Journal of Agricultural Economics, Vol.XXV No.2, pp.59-64.