INFORMETRICS 89/90
L. Egghe and R. Rousseau (Editors)

JOURNAL PRODUCTIVITY IN ECONOMICS

I.K. RAVICHANDRA RAO

Documentation Research and Training Centre
Indian Statistical Institute
8th Mile, Mysore Road, R.V. College Post
Bangalore - 560059 (India)

Abstract

A bibliography in economics consisting of 4130 articles from 744 journals is examined to study the scattering of articles in various journals. It has been observed that the empirical distribution of articles is reverse-J shape with a long tail and distribution is highly positively skewed. Further, it has been observed that 99 % of the journals contain 90 % of the articles. Only one percent of the journals contribute nearly 10 % of the periodical literature in economics.

This study suggests that the negative binomial distribution does not fit the empirical distribution of articles in economics. However, it fits very well to the truncated data (without the data pertaining to the most productive journals). It fits equally well to the most (12 out of 15) of the frequency distributions of articles in homogeneous group of subjects (i.e. in the subdivisions of economics).

This study further suggests that the rank distribution of journals in economics does not confirm to the perfect Bradford's log-linear model. The shape of the Bradford curve for the economics data is S-shape and Groos droop exist. Leimkuhler's model as such doesn't fit the data. However, Leimkuhler's model as expressed by $F(x) = a * \log(1 + bx)$, where a and b are constants and F(x) is the number of articles contained in x top most journals, give rise to a good result when Egghe's algorithm to compute Bradford multiplier (k), and the parameters a and b is used. A simple procedure to compute k (by assuming the number of journals in the nucleus (r_0) is equal to one and considering the number of articles (y_0) it contains) also gives rise to a good approximat-on for the rank distribution of journals.

A few applications of bibliometric studies are discussed in the context of special libraries in India. A core list of journals in economics is compiled.

I. INTRODUCTION

As early as 1950's, Prof. P.C. Mahalanobis, the Founder of the Indian Statistical Institute, suggested that statistics is the 'Key Technology' for all development and forecasting studies. In recent past, it has been applied to many subjects such as perspective planning, industrial development, agricultural development, etc. Statistics and other related tools when applied to a field of activity generates a new field derived out of fusion. For example, application of quantitative methods to library organization and services is known as librametry [18]. Statistical analysis of bibliographical organization and services is generally termed as bibliometrics [17].

Application of statistical methods to organization of science and its productivity analysis give rise to scientometrics, and when it is applied to information system and services, we call it informetrics.

In this paper, I shall discuss only bibliometric aspects with an emphasis to distribution of articles and rank distribution of journals. I shall particularly discuss :

(a) journal productivity in economics;

(b) how far negative binomial distribution is good enough to describe the distribution of articles in economics?

(c) how far Bradford's model, Leimkuhler's model, and Egghe's algorithm to compute Bradford multiplier are adequate enough to explain the scattering phenomena?

(d) a few applications of bibliometrics in the context of special libraries in India.

2. SCATTERING OF ARTICLES IN ECONOMICS

In this study, a bibliography* has been examined to collect the required data. A frequency distribution of articles in economics, which were included in the bibliography is compiled and it is given in table 1. Like in other similar studies, the following has been observed :

(a) the empirical distribution of articles in economics is reverse-J shaped with a long tail, and

(b) the distribution is highly positively skewed.

The corresponding frequency curve is shown in figure 1. A Lorenz curve is drawn to study the concentration of articles in various journals, and is shown in figure 2. The cumulative percentages (of journals and articles) to draw Lorenz curve are computed and are given in table 1.

In table 1 as well as from figure 2, we may observe that :

(a) only a small fraction (0.94 %) of the journals (most productive) contain 9.99 % of the articles in economics, and

(b) 5.38 % of the journals (most productive) contribute 32.35 % of the articles in economics.

In other words, 99 % of the journals (when arranged according to the increasing productivity) contain of 90 % of the articles; i.e. one per cent of the journals contribute 10 % of the articles in economics. This suggests us that the 80-20 rule is hardly applicable to study the journal productivity in economics.

3. DISTRIBUTIONS OF ARTICLES

In recent times, many have argued that negative binomial distribution is most suitable to describe bibliometric data. The negative binomial distribution function is given by

$$p(x) = \frac{(k+x-1)!}{(k-1)!\,x!}\, p^k\, (1-p)^x \qquad (1)$$

$$x = 0,1,2,3,4,\ldots,\ k > 0,\ 0 \le p \le 1$$

---

* International Bibliography of Economics, Vol.29, 1980 (International Bibliography of the Social Sciences), Tavistock Publications, London.

Table 1 : A distribution of articles in the International Bibliography of Economics

Number of articles (x)	Number of journals (y)	Total number of articles (x,y)	Cumulative number of Σ xy (%)	Cumulative number of Σ y (%)
(1)	(2)	(3)	(4)	(5)
1	229	229	229 (5.54)	229 (30.78)
2	138	276	505 (12.23)	367 (49.33)
3	88	264	769 (18.62)	455 (61.16)
4	61	244	1013 (24.53)	516 (69.35)
5	40	200	1213 (29.37)	556 (74.73)
6	29	174	1387 (33.58)	585 (78.63)
7	20	140	1527 (36.97)	605 (81.32)
8	14	112	1639 (39.68)	619 (83.20)
9	10	90	1729 (41.86)	629 (84.54)
10	12	120	1849 (44.87)	649 (86.16)
11	7	77	1926 (46.63)	648 (87.20)
12	9	108	2034 (49.25)	657 (88.31)
13	11	143	2177 (52.71)	668 (89.78)
14	6	84	2261 (54.75)	674 (90.59)
15	7	105	2366 (57.29)	681 (91.53)
16	3	48	2414 (58.45)	684 (91.54)
17	1	17	2431 (58.86)	685 (92.07)
18	7	126	2557 (61.91)	692 (93.01)
19	5	95	2652 (64.21)	697 (93.68)
20	5	100	2752 (66.63)	702 (94.35)
21	2	42	2794 (67.65)	704 (94.62)
22	7	154	2948 (71.38)	711 (95.56)
23	1	23	2971 (71.94)	712 (95.70)
25	4	100	3071 (74.36)	716 (96.24)
26	1	26	3097 (74.99)	717 (96.37)
27	3	81	3178 (76.95)	720 (96.77)
28	3	84	3262 (76.98)	723 (97.18)
29	7	203	3465 (83.90)	730 (98.12)
31	2	62	3527 (85.40)	732 (98.39)
33	1	33	3560 (86.20)	733 (98.52)
36	1	36	3596 (87.07)	734 (98.66)
39	1	39	3635 (88.04)	735 (98.79)
41	1	41	3676 (89.01)	736 (98.92)
42	1	42	3718 (90.02)	737 (99.06)
48	1	48	3766 (91.19)	738 (99.19)
53	1	53	3819 (92.47)	739 (99.33)
55	1	55	3874 (93.80)	740 (99.46)
56	1	56	3930 (95.16)	741 (99.60)
57	1	57	3987 (96.54)	742 (99.73)
66	1	66	4053 (98.14)	743 (99.87)
77	1	77	4130 (100.00)	744 (100.00)
	744	4130		

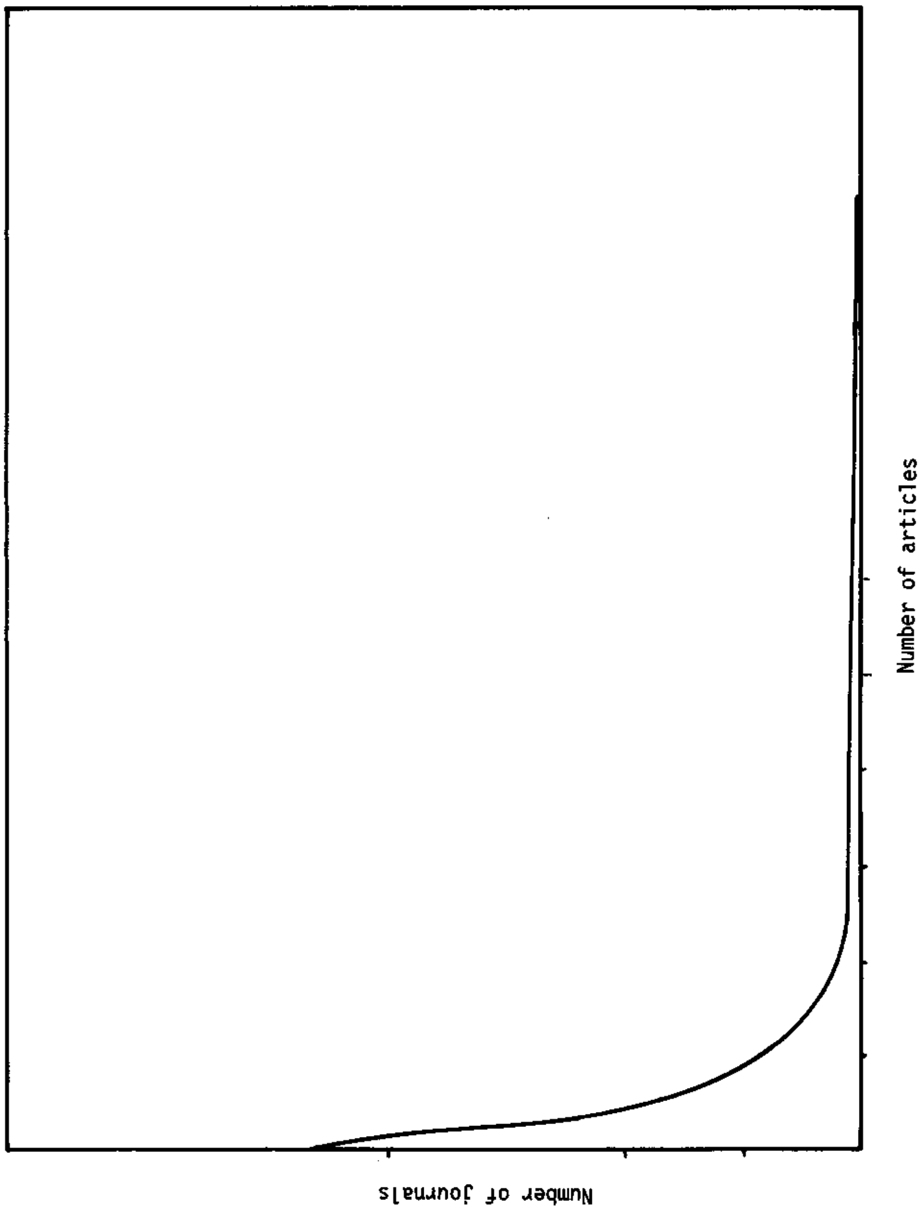

Fig.1 : A frequency curve of distribution of articles in economics

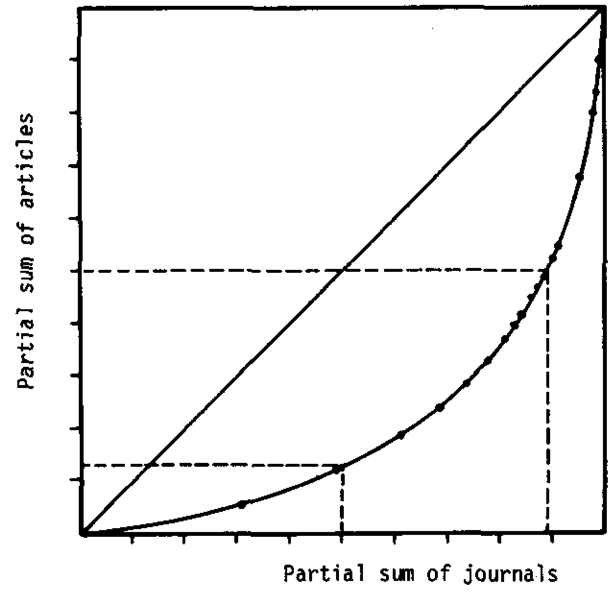

Fig.2 : A Lorenz curve : scattering of articles in economics

p(x) gives the probability that a journal contains x articles. Burrell [4,5,6], Ravichandra Rao [19] and few others have shown that negative binomial distribution fits fairly well to circulation data, scientific productivity data, etc. Burrell [7] and O'Neill [16] have argued that negative binomial distribution fits the journal productivity data.

An attempt has been made here to study how far the negative binomial distribution is suitable to describe the distribution of articles in economics. Theoretical values are computed and given in table 2. Both moment estimators and maximum likelihood estimators (m.l.e.) are used (separately) to compute the theoretical values.

It may be noted from table 2 that theoretical values deviate considerably from the observed frequencies. Chi-square test suggests us to reject the hypothesis that the negative binomial distribution fits the empirical distribution of articles in economics. The poor fit may be due to :

1) incomplete bibliography;
2) heterogeneity nature of data; data might have come from different negative binomial populations.

It is believed that the long tail (as can be seen in figure 1) is primarily due to a mixture of negative binomial populations and the core journals may belong to an entirely different population. By eliminating the core journals (i.e. of truncating the empirical distribution at x = 18, from where the straight line begins in figure 3), theoretical frequencies are computed using the negative binomial distribution. The results (shown in table 3) are much better than the one shown in table 2. The Chi-square test suggests that negative binomial distribution fits very well to the truncated (at right) distribution of articles.

Table 2 : Distribution of articles : observed and theoretical values as computed from Negative Binomial Distribution

Number of articles (x)	Number of journals f(x)	Expected values (Using Moments) (1)	(Using m.l.e.) (2)
1	229	313	220
2	138	93	92
3	88	57	62
4	61	41	47
5	40	32	38
6	29	26	32
7	20	21	27
8	14	18	23
9	10	15	20
10	12	13	18
11	7	12	16
12	9	10	14
13	11	9	12
14	6	8	11
15	7	7	10
16	3	6	9
17	1	6	8
18	7	5	7
19	5	5	7
20	5	4	6
21	2	4	5
22	7	3	5
23	1	3	5
25	4	3	4
26	1	2	3
27	3	2	3
28	3	2	3
29	7	2	3
31	2	1	2
33	1	1	2
36	1	1	1
39	1	1	1
41	1	1	1
42	1	1	1
48	1	0	1
53	1	0	0
55	1	0	0
56	1	0	0
57	1	0	0
61	1	0	0
77	1	0	0

$\bar{x}$ = 5.551100

σ^2 = 69.782300

σ = 8.353600

p = 0.065218

q = 0.934782

k = 0.317521 (Using moments)

k = 0.446533 (m.l.e.)

χ^2 (Computed for (1), d.f. = 17) = 88.69

χ^2 (Computed for (2), d.f. = 20) = 27.16

Critical value :

χ^2 (α = 0.05, d.f. = 17) = 27.6

χ^2 (α = 10.05, d.f. = 20) = 31.4

Table 3 : Truncated distribution of articles : observed and expected values as computed from Negative Binomial Distribution

Number of articles (x)	Number of journals f(x)	Expected frequencies (Using Moments) (1)	(Using m.l.e.) (2)
1	229	230	238
2	138	127	127
3	88	86	85
4	61	61	60
5	40	45	44
6	29	34	33
7	20	25	24
8	14	19	19
9	10	15	14
10	12	11	11
11	7	9	8
12	9	7	6
13	11	5	5
14	6	4	4
15	7	3	3
16	3	2	2
17	1	2	2
18	7	1	1

$\bar{x}$ = 3.6909
σ^2 = 13.1628
σ = 3.628058
p = 0.204750
q = 0.792499
k = 0.693894 (Using moments)
k = 0.671693 (m.l.e.)

χ^2 (Computed for (1), d.f. = 12) = 13.64
χ^2 (Computed for (2), d.f. = 12) = 12.05

Critical value :
χ^2 (α = 0.05, d.f. = 12) = 21.0

Another way of analyzing the data (instead of truncating the distribution) is to eliminate the heterogeniety; i.e. to study only the homogeneous data as far as possible. In this study, to eliminate the heterogeniety, frequency distributions of articles in various subdivisions of econpmics were compiled (using the same bibliography) and these are given in tables 4-11.

For each of the homogeneous groups, an attempt has been made to fit the negative binomial distributions. Except for the three empirical distributions (tables 4(B), 8(J) and 11(O)) the negative binomial fits very well.

Data in tables 4(B), 8(J) and 11(O) refer to the bibliography of methods, money and finance, and international economics. These subjects are mostly interdesciplinary in nature and thus data comes from heterogenous population. Although Sankhya, Journal of Game theory, etc. include articles in econometrics, methods in economics, these journals are not covered in the bibliography. This fact illustrates that the bibliography is incomplete.

From these observations, it is very clear that the negative binomial distribution is unlikely to fit the heterogeneous data and also it is unlikely to fit the data based on incomplete bibliography. Considerable research is needed to estimate the parameters of the distribution in such circumstances.

Table 4 : A frequency distribution of articles in "Preliminaries" and "Methods" of Economics

A. Preliminaries			B. Methods		
Number of articles x	Number of journals f(x)	Theoretical values	Number of articles x	Number of journals f(x)	Theoretical values
1	9	9	1	20	27
2	2		2	16	7
3	1	4	3	5	5
9	1		4	2	
			6	2	9
			9	2	
			19	1	
	13	13		48	48
$\bar{x}$ = 1.9231	p = 0.203655		$\bar{x}$ = 2.5833	p = 0.174445	
σ^2 = 4.5325	q = 0.796345		σ^2 = 9.0764	q = 0.825547	
	k = 0.2360654			k = 0.334569	
			χ^2 (Computed) = 14.4		
			χ^2 (α = 0.05, d.f. = 3) = 7.81		

Table 5 : A frequency distribution of articles in "General and Basic Subjects" and "History and Economic Thought"

C. General and Basic Subjects			D. History of Economic Thought		
Number of articles x	Number of journals f(x)	Theoretical values	Number of articles x	Number of journals f(x)	Theoretical values
1	17	16	1	22	21
2	3	5	2	3	6
3	2		3	5	
4	1	3	4	1	5
5	1		10	1	
	24	24		32	32
$\bar{x}$ = 1.5833	p = 0.502994		$\bar{x}$ = 1.78125	p = 0.345052	
σ^2 = 1.1597	q = 0.497006		σ^2 = 2.2641	q = 0.654949	
	k = 0.590361			k = 0.411594	
χ^2 (Computed) = 1.1958			χ^2 (Computed) = 2.3476		
χ^2 (α = 0.05, d.f. = 2) = 5.99)			χ^2 (α = 0.05, d.f. = 2) = 5.99)		

Table 6 : A frequency distribution of articles in "Economic History" and "Economic Activity"

E. Economic History			F. Economic Activity		
Number of articles x	Number of journals $f(x)$	Theoretical values	Number of articles x	Number of journals $f(x)$	Theoretical values
1	9	9	1	136	131
2	1	2	2	34	42
5	2		3	17	18
			4	12	8
			5	5	4
			6	1	
			7	1	4
			8	1	
	11	11		207	207
$\bar{x}$ = 1.4545		p = 0.339502	$\bar{x}$ = 1.6860		p = 0.469283
σ^2 = 1.3388		q = 0.660498	σ^2 = 1.4618		q = 0.530717
		k = 0.233638			k = 0.606583
			χ^2 (Computed) = 2.2202		
			χ^2 (α = 10.05, d.f. = 5) = 11.1		

Table 7 : A frequency distribution of articles in "Organization of Production" and "Production"

G. Organization of Production			H. Production		
Number of articles x	Number of journals f(x)	Theoretical values	Number of articles x	Number of journals f(x)	Theoretical values
1	143	144	1	150	114
2	38	36	2	66	51
3	17	19	3	38	27
4	12	12	4	9	16
5	12	8	5	10	10
6	3	6	6	1	
7	2		7	2	
8	3		8	7	
9	2		9	1	
10	4	16	12	2	21
11	1		13	1	
12	2		14	1	
13	1		15	1	
15	1				
22	1				
	241	241		289	289

G: $\bar{x} = 2.776$, $\sigma^2 = 7.6790$, $p = 0.179397$, $q = 0.820603$, $k = 0.301165$

χ^2 (Computed) = 3.8286

χ^2 ($\alpha = 0.05$, d.f. = 6) = 12.6

H: $\bar{x} = 2.1903$, $\sigma^2 = 4.5139$, $p = 0.263696$, $q = 0.736304$, $k = 0.426292$

χ^2 (Computed) = 14.3414

χ^2 ($\alpha = 0.01$, d.f. = 5) = 15.1

Table 8 : A frequency distribution of articles in "Prices and Market" and "Money and Finance"

I. Prices and Market			J. Money and Finance		
Number of articles x	Number of journals f(x)	Theoretical values	Number of articles x	Number of journals f(x)	Theoretical values
1	37	44	1	65	80
2	16	9	2	26	20
3	6	5	3	28	11
5	6		4	6	7
8	1	9	5	5	
9	1		6	1	
			7	1	
			8	3	
			10	2	25
			13	1	
			14	1	
			17	1	
			22	1	
			24	1	
			25	1	
	67	67		143	143
$\bar{x}$ = 2.0		p = 0.212025	$\bar{x}$ = 2.951049		p = 0.125413
σ^2 = 4.7164		q = 0.787975	σ^2 = 15.5570		q = 0.874587
		k = 0.269076			k = 0.279773
χ^2 (Computed) = 7.8692			χ^2 (Computed) = 32.9881		
χ^2 (α = 0.05, d.f. = 3) = 9.35			χ^2 (α = 0.05, d.f. = 4) = 11.1		

Table 9 : A frequency distribution of articles in "Incomes and Income Distribution" and "Demand (Use of income)"

K. Income and Income Distribution			L. Demand (Use of Income)		
Number of articles x	Number of journals f(x)	Theoretical values	Number of articles x	Number of journals f(x)	Theoretical values
1	75	63	1	70	66
2	26	34	2	8	16
3	6	13	3	11 }	
4	5 }		4	1 }	10
6	2 }	6	5	2 }	
7	1 }				
12	1 }				
	116	116		92	92

K. Income and Income Distribution	L. Demand (Use of Income)
$\bar{x} = 1.6897$	$\bar{x} = 1.445$
$\sigma^2 = 2.1278$	$\sigma^2 = 0.8123$
$p = 0.794076$	$p = 0.5486554$
$q = 0.205924$	$q = 0.451346$
$k = 2.659429$	$k = 0.541733$
χ^2 (Computed) = 9.4373	χ^2 (Computed) = 5.8424
χ^2 (α = 0.01, d.f. = 3) = 11.3	χ^2 (α = 0.05, d.f. = 2) = 5.95

Table 10 : A frequency distribution of articles in "Economics and Policy" and "Public Economy"

M. Social Economics and Policy			N. Public Economy		
Number of articles x	Number of journals f(x)	Theoretical values	Number of articles x	Number of journals f(x)	Theoretical values
1	74	63	1	133	147
2	17	29	2	46	28
3	4		3	15	14
4	5	12	4	5	8
6	3		5	6	5
7	1		6	2	
			7	2	
			8	2	
			9	1	13
			12	1	
			17	1	
			18	1	
	104	104		215	215
$\bar{x}$ = 1.5577	p = 0.0832689		$\bar{x}$ = 1.9442	p = 0.201924	
σ^2 = 0.669749	q = 0.167311		σ^2 = 4.6759	q = 0.798076	
	k = 2.775576			k = 0.238891	
χ^2 (Computed) = 6.9694			χ^2 (Computed) = 14.9886		
χ^2 (α = 0.025, d.f. = 2) = 7.83			χ^2 (α = 0.01, d.f. = 5) = 15.1		

Table 11 : A frequency distribution of articles in "International Economy"

Number of articles x	Number of journals f(x)	Theoretical values
1	173	223
2	70	32
3	25	17
4	25	11
5	5	8
6	8	6
7	3	26
8	2	
9	3	
10	2	
12	1	
13	1	
14	1	
17	1	
18	1	
35	1	
50	1	
	323	323

$\bar{x} = 2.4815$ $p = 0.0951887$

$\sigma^2 = 25.5636$ $q = 0.9048112$

$k = 0.155856$

χ^2 (Computed) = 82.8250

χ^2 ($\alpha = 0.05$, d.f. = 6) = 12.6

4. THE BRADFORD DISTRIBUTION

A well-known procedure to study journal productivity is Bradford's method of plotting (on a semi-logarithmic paper) the cumulative number of journals against the cumulative number of papers they contain, where the journals are arranged in decreasing order of productivity [23]. Bradford in his study observed that linear increases in the cumulative number of papers required geometric increases in the number of journals. It is now known as Bradford's law. It is expressed in the form of

$$F(x) = a + b \log x$$

where F(x) is the cumulative number of references contained in the first x most productive journals, a and b are constants.

Vickery [21] pointed out the difference between the verbal and graphical interpretation of law of scattering and further showed that Bradford's law implies a J-shape semilogarithmic curve with an initial concave segment leading to a linear segment. Since then many have worked in this area and suggested several models [11,12].

Groos [10], O'Neill [16], Keenan and Atherton [13] and few others have pointed out that for a large set of data, shape of the curve is likely to be a S-shape with a convex segment in the upper portion of the curve. In fact, the convex segment is known as 'Groos droop'. The occurrence of the droop may be due to incomplete bibliographical data. The droop may also occur in case of newly emerging subjects (in which case, there is hardly any core journals) and in case of inter-disciplinary subjects (in which case, there are too many least productive journals). In other words, the droop is likely to occur for the heterogeneous bibliographic data.

In order to study the journal productivity in economics, the cumulative number of journals is thus plotted against the cumulative number of papers they contain, when the journals are arranged in decreasing order. The relevant data (x and F(x)) are given in table 12 (in columns 2 and 4 respectively) and also semi-logarithmic curve is drawn and it is shown in figure 3.

Figure 3 suggests that the curve is not a perfect Bradford curve. It is a S-shape curve and it has Groos droop. It is likely that the data is incomplete and comes from heterogeneous population. It is also likely that the empirical distribution is a mixture of Bradford populations. In fact, the semi-logarithmic curves (figures 4-6) for each of the homogeneous distributions (given in tables 4-11) suggest that the data come from a mixture of Bradford populations. However as such it is difficult to express the distribution function for the mixture of Bradford distributions.

Table 12 : Rank distribution of journals in economics

Number of journals (x)	Cumulative number of Σx	Number of articles (y)	Cumulative number of articles Σx.y = F(x)
(1)	(2)	(3)	(4)
1	1	77	77
1	2	66	143
1	3	57	200
1	4	56	256
1	5	55	311
1	6	53	364
1	7	48	412
1	8	42	454
1	9	41	495
1	10	39	534
1	11	36	570
1	12	33	603
2	13	31 (62)	665
7	21	29 (203)	868
3	24	28 (84)	952
3	27	27 (81)	1033
1	28	26	1059
4	32	25 (100)	1159
1	33	23	1182
7	40	22 (154)	1336
2	42	21 (42)	1378
5	47	20 (100)	1478
5	52	19 (95)	1573
7	59	18 (126)	1699
1	60	17	1716
3	63	16 (48)	1764
7	70	15 (105)	1869
6	76	14 (84)	1953
11	87	13 (143)	2096
9	96	12 (108)	2204
7	103	11 (77)	2281
12	115	10 (120)	2401
10	125	9 (90)	2491
14	139	8 (112)	2603
20	159	7 (140)	2743
29	188	6 (174)	2917
40	228	5 (200)	3117
61	289	4 (244)	3361
88	377	3 (264)	3625
138	515	2 (276)	3901
229	744	1 (229)	4130

NOTE : In column 3, the figures within the brackets refer to x.y.

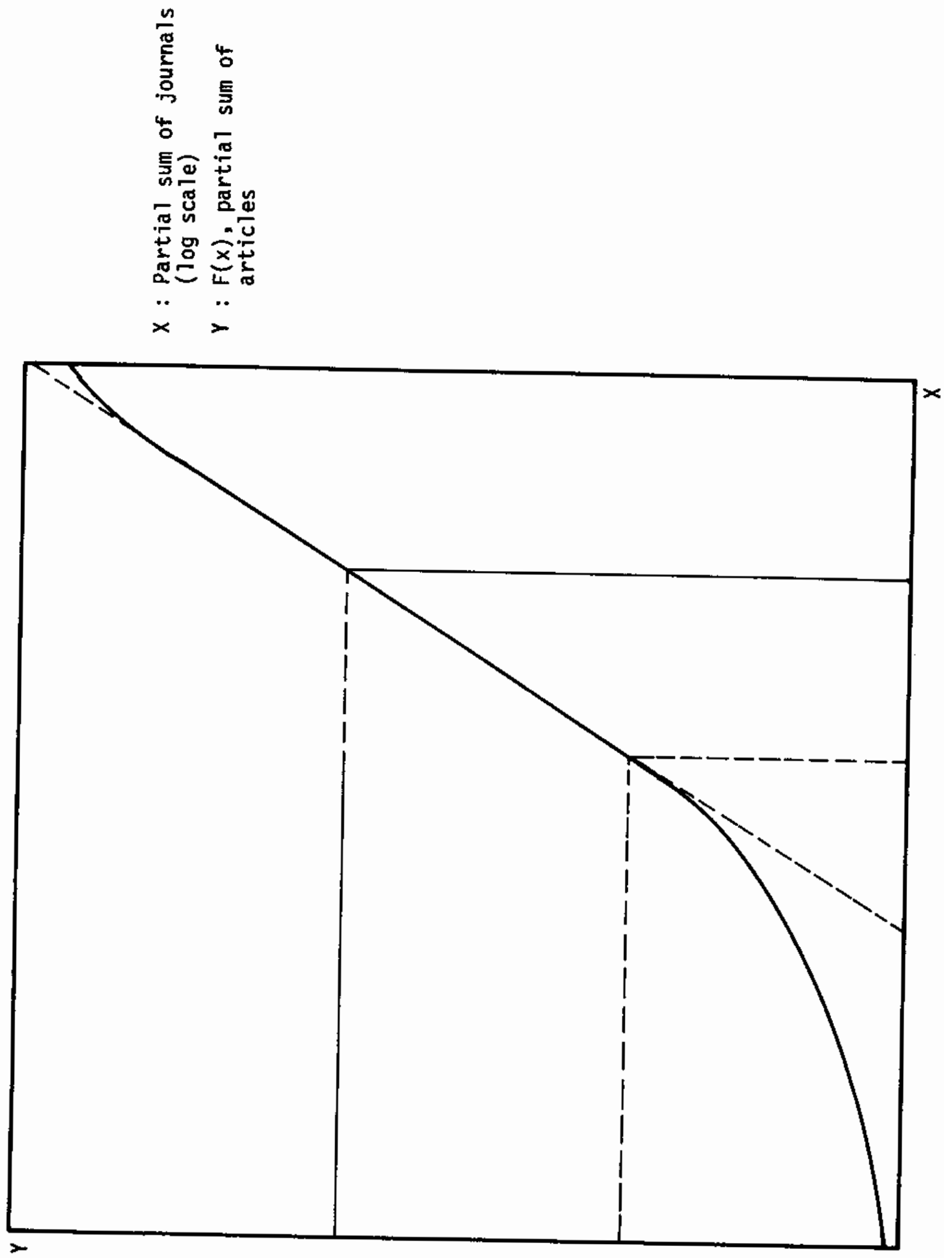

Fig.3 : A Bradford curve

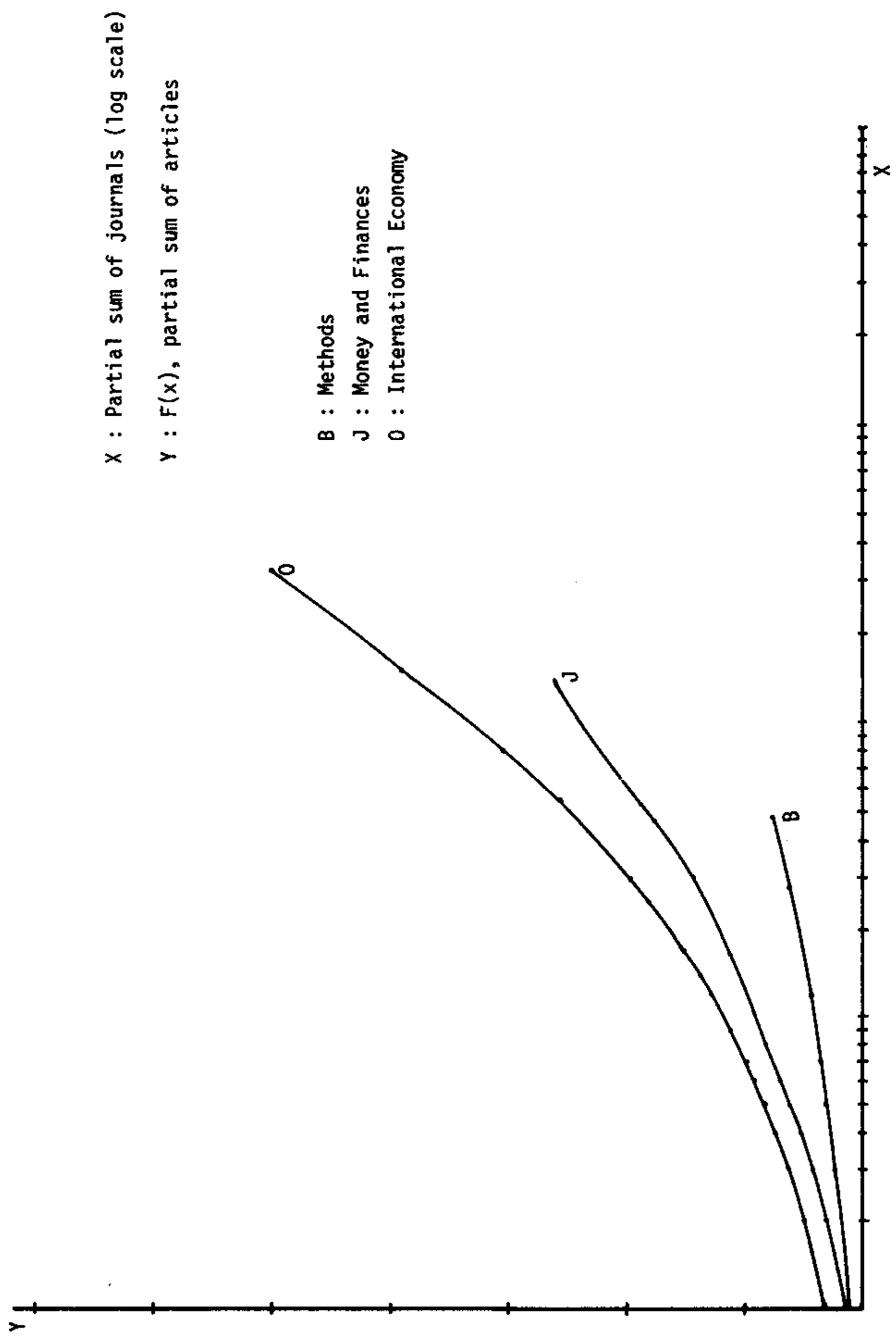

Fig.4 : A Bradford curve

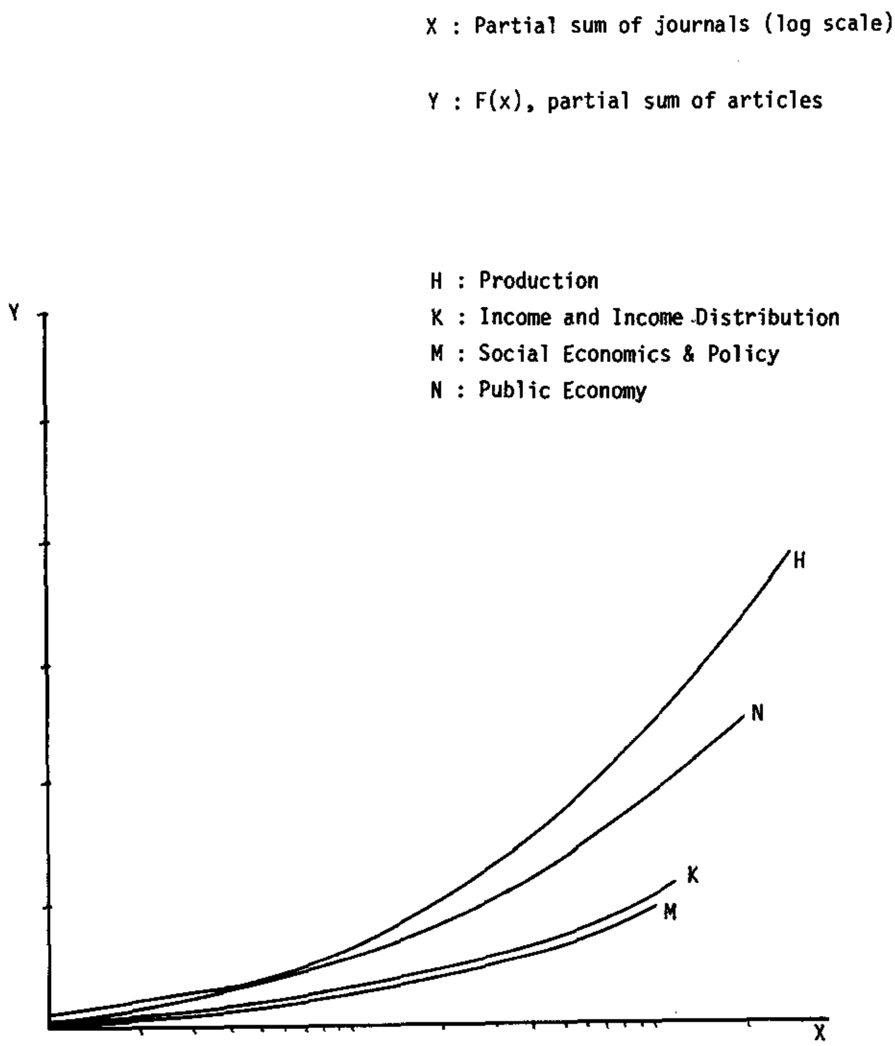

Fig.5 : A Bradford curve

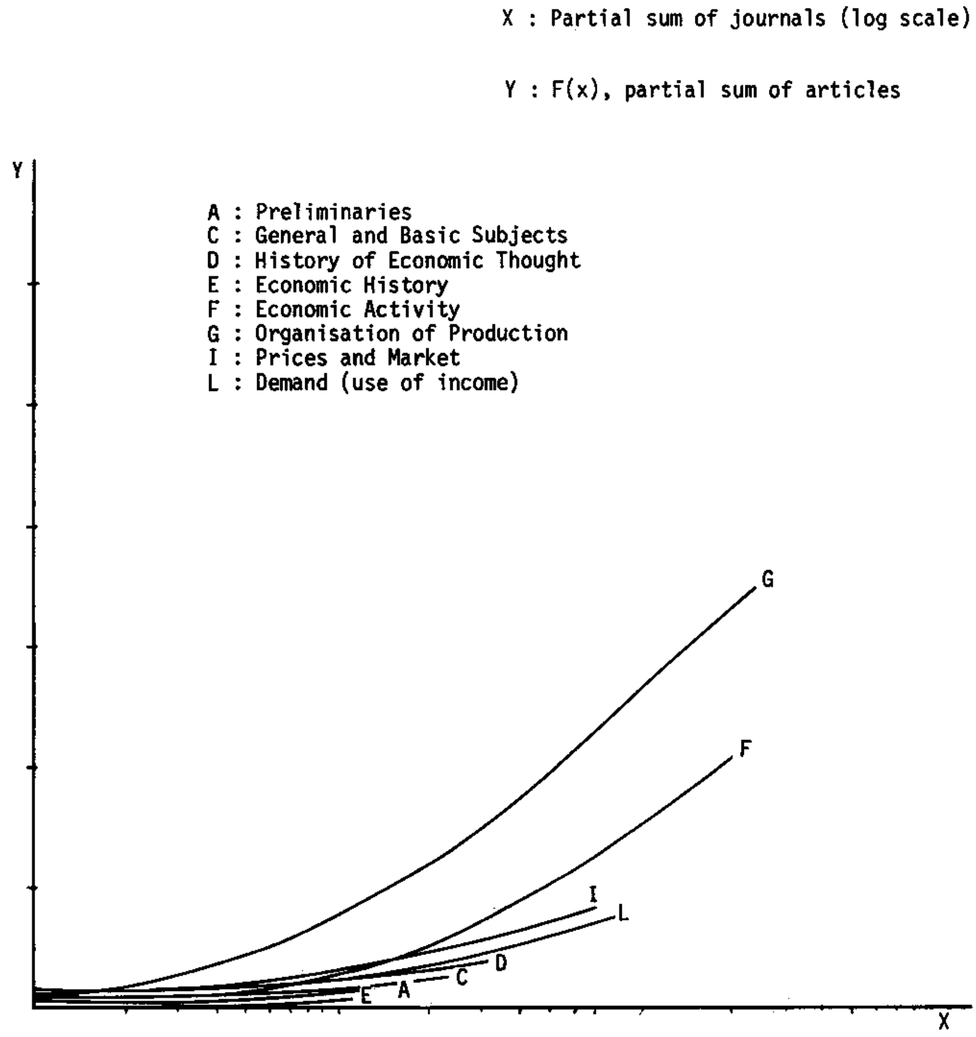

Fig.6 : A Bradford curve

5. LEIMKUHLER'S MODEL

Based on Bradford's law, Leimkuhler in 1967 derived a probability distribution function. It is given by :

$$F(x) = \frac{\log (1+\beta x)}{\log (1+\beta)} \qquad (3)$$

where $\beta > 0$ and $F(x)$ is the relative cumulative frequency of articles contained in x proportion of top most productive journals. β can be computed using the first moment (about zero). It is given by :

$$\mu = \frac{1}{\log (1+\beta)} - \frac{1}{\beta} \qquad (4)$$

This function is likely to fit only for those data which comes from "perfect Bradford population" (the population which confirms to $F(x) = a + b \log x$). An attempt has also been made here to compute $F(x)$, to see how for Leimkuhler's model is adequate enough to study $F(x)$. Table 13 gives x, $F(x)$ (observed value), $F_1(x)$ (computed value) and $D(x)$ (difference between the computed and observed values. $D(x)$ is computed as :

$$D(x) = |F(x) - F_1(x)| / \text{Total number of articles}$$

The maximum of value of $D(x)$ is 0.0399516. The Kolmogorov-Smirnov test therefore suggests us to reject the hypothesis that the Bradford distribution as suggested by Leimkuhler fits the empirical data; the $D_\alpha(x)$ (for $\alpha = 0.01$) is given by 0.0253636. The value of β is 58.75.

Further, if we carefully observe the different values of $D(x)$, we may notice that in 14 cases (out of 41) the values of $D(x)$ goes beyond 0.0253636. These fourteen values refer to the concave and convex portions of the curve.

So, we must have different approach to compute the parameter of Leimkuhler's model. Egghe has suggested a method. This is discussed in the next section.

6. EGGHE'S METHOD

Egghe [9] suggested a theory and an algorithm to compute the parameters of Leimkuhler's model which may be used even for the data based on incomplete bibliography. Leimkuhler's model may be defined as :

$$F(x) = a * \log (1 + bx) \qquad (5)$$

where $F(x)$ is the number of articles contained in x top most journals, a and b are constants.

An attempt has been made here to study whether Egghe's algorithm is good enough to estimate the parameters of the rank distribution of journals in economics.

In his method, Egghe basically suggests to choose a cut-off rank (ρ_0) at which Groos droop becomes apparent and to determine the number of articles per journal, at the cut-off point, say n. Next step is to fix p (the number of zones for the complete bibliography), p must be sufficiently large. Egghe then suggests the following formulae to compute the number of articles (y_0) contained in r_0 most productive journals (for the complete bibliography).

$$r_0 = \frac{\overline{T}(k-1)}{k^{p-[q]-1} - 1} \qquad (6)$$

$$y_0 = \frac{\overline{A}}{p-[q]-1} \qquad (7)$$

Table 13 : Computed values of F(x) for Leimkuhler's model

Number of journals x	Number of articles F(x)	Computed values $F_1(x)$	D(x) $\lvert F(x) - F_1(x)\rvert / F(x=744)$
1	77	77	0.0
2	143	149	0.00145280
3	200	215	0.00363200
4	256	278	0.00532690
5	311	337	0.00629540
6	364	392	0.00677970
7	412	445	0.00788030
8	454	495	0.00992740
9	495	543	0.01162230
10	534	588	0.01307510
11	570	632	0.01501210
12	603	674	0.01719130
14	665	752	0.02106540
21	868	988	0.02905570
24	952	1074	0.02954000
27	1033	1153	0.02905570
28	1059	1178	0.02881360
32	1159	1273	0.02760290
33	1182	1296	0.02760290
40	1336	1440	0.02518160
42	1378	1477	0.02397090
47	1478	1566	0.02130750
52	1573	1647	0.01791770
59	1699	1751	0.01259080
60	1716	1765	0.01186440
63	1764	1805	0.00992740
70	1869	1895	0.00629540
76	1953	1966	0.00314770
87	2096	2084	0.00290560
96	2204	2171	0.00799030
103	2281	2234	0.01138010
115	2401	2334	0.01622280
125	2491	2410	0.01961260
135	2603	2481	0.02954000
159	2743	2633	0.02663440
188	2917	2790	0.03075060
228	3117	2974	0.03462470
289	3361	3202	0.03849880
377	3625	3460	0.03995160
515	3901	3767	0.0324450
744	4130	4131	0.00024210

where : $\bar{T}$ is the number of journals in the truncated bibliography (in $p - [q] - 2$ zones),

$\bar{A}$ is the number of articles in $\bar{T}$,

$[q]$ is the largest whole number smaller than or equal to q (the last Bradford group),

k is the Bradford multiplier and it is given by :

$k = (1.781\ y_m)^{1/p}$, y_m is the number of articles in the most productive journal.

The parameters (a and b) in equation (3) can now be computed using the following functions :

$$a = \frac{y_0}{\log k} \qquad (8)$$

$$b = \frac{k - 1}{r_0} \qquad (9)$$

For the data shown in table 12, these parameters are computed and given below :

$$y_m = 77, \quad p = 10 \quad \text{and} \quad k = 1.6357$$

In figure 3, we may observe that the droop begins at $x = 480$ and the corresponding value of n is 2 (the number of articles per journal at cut-off point). Thus,

$$q = \frac{E + \log n}{\log k}$$

where E is the Euler's constant (= 0.5772). Hence $[q] = 2$ and $q' = 3$. The modified value of n is thus given by

$$n' = \frac{k^n}{1.781} = 2.457237$$

So the final cut-off rank is 9 articles 2 being between the ranks 378 and 515)

$$r' = 515 - 0.457237 \times (515 - 377)$$

$$= 452.36$$

$$\simeq 452$$

Thus, the total number of journals ($\bar{T}$) is 452 and they contain 3779 articles ($\bar{A}$). Substituting the values of $\bar{T}$, $\bar{A}$, k, p and [q] in (6) and (7), we have :

$$r_0 = 9.474455 \quad \text{and} \quad y_0 = 539$$

On substituting the values of r_0 and y_0, in (8) and (9), we have

$$a = 1095.371 \quad \text{and} \quad b = 0.0670962$$

By substituting the values of a and b in equation (3), the theoretical values of F(x) for different values of x are computed and given in table 13 (column 3, $F_1(x)$). In table 14, column 4 gives the difference ($D_1(x)$) between the relative frequencies of observed (F(x)) and theoretical frequencies ($F_1(x)$). It clearly

Table 14 : Theoretical and observed values

x	F(x)	$F_1(x)$	$D_1(x)$	$F_2(x)$	$D_2(x)$
(1)	(2)	(3)	(4)	(5)	(6)
1	77	72	0.00121070	77	0.00000000
2	143	138	0.0012070	149	0.00145280
3	200	201	0.0024210	216	0.00387410
4	256	261	0.00121070	279	0.00556900
5	311	317	0.00145280	338	0.00653750
6	364	371	0.00169490	395	0.00750510
7	412	422	0.00242130	448	0.00871670
8	454	471	0.00411620	499	0.01089590
9	495	518	0.00556900	547	0.01259080
10	534	563	0.00702180	593	0.01428570
11	570	606	0.00871670	638	0.01646490
12	603	647	0.01065380	680	0.01864410
14	665	726	0.01477000	760	0.02300240
21	868	963	0.02324460	1000	0.03196130
24	952	1051	0.02397090	1088	0.03292980
27	1033	1133	0.02421310	1169	0.03292980
28	1059	1159	0.02421310	1195	0.03292980
32	1159	12576	0.02348670	1292	0.03220340
33	1182	1279	0.02348670	1315	0.03220340
40	1336	1429	0.02251820	1463	0.03075060
42	1378	1468	0.02179180	1501	0.02978210
47	1478	1560	0.01985470	1592	0.02760290
52	1573	1645	0.01743340	1675	0.02469730
59	1699	1754	0.01331720	1782	0.02009690
60	1716	1769	0.01283290	1796	0.01937050
63	1764	1812	0.01162230	1838	0.01791770
70	1869	1906	0.00895880	1930	0.01477000
76	1953	1981	0.00677970	2003	0.01210650
87	2096	2106	0.00242130	2124	0.00677970
96	2204	2199	0.00121070	2214	0.00242130
103	2281	2266	0.00363200	2279	0.00048430
115	2401	2372	0.00702180	2382	0.00460050
125	2491	2453	0.00920100	2460	0.00750610
135	2603	2529	0.01791770	2533	0.01694920
159	2743	2692	0.01234870	2690	0.01293290
188	2917	2861	0.01355930	2852	0.01573850
228	3117	3058	0.01428570	3041	0.1840190
289	3361	3303	0.01404360	3276	0.02059110
377	3625	3582	0.01041160	3543	0.01985470
515	3901	3912	0.00266340	3859	0.01016950
744	4130	4130*	0.00000000	4130*	0.00000000

NOTE : (1) x is the number of journals; F(x) is the number of articles contained in x journals; $F_1(x)$ is computed using Leimkuhler's model, where in parameters are computed based on Egghe's algorithm. $D_1(x) = |F(x) - F_1(x)|/F(x=744)$. $F_2(x)$ is computed using Leimkuhler's model, where in parameters are computed for $r_0 = 1$ and $y_0 = 77$. $D_2(x) = |F(x) - F_2(x)|/F(x=744)$
(*) computed such that total number of articles in 4130.

suggests that Leimkuhler's model fits the data extremely well. It of course implies that Egghe's algorithm gives rise to a good estimate of the parameters. Also Kolmogorov-Smirnov test leads us to accept the hypothesis that Leimkuhler's model fits the data at 0.01 level. Computed value of Kolmogorov-Smirnov statistic (D_{max}) is 0.02421310; the critical value $D_{(\alpha = .01)}$ is 0.0253636.

We may note that in Egghe's method, k (the Bradford multiplier) is a function of p (the number of zones) and thus it is likely to give rise to a good estimate. However, it is necessary for us to draw the semi-logarithmic curve and observe carefully to determine the cut-off point (ρ_0) at which Groos droop becomes apparent.

The question is thus that is it possible to compute without looking at the semi-logarithmic curve? The author suggests the following procedure to compute k :

i) assume $r_0 = 1$ (the number of journal in the nucleus zone),

ii) select the corresponding value of y_0 (the number of articles for $r_0 = 1$),

iii) from these two values of r_0 and y_0, and using the Newton-Raphson method, k can be computed from the following equation

$$i + k + k^2 + \ldots + k^{p-1} = T ,$$

where p is the number of zones and T is the total number of journals.

For instance, for the data given in table 12, for $r_0 = 1$, $y_0 = 77$; since T = 744, for $y_0 = 77$, p is 55. The value of k is thus 1.0766334. Substituting the values of k, r_0 and y_0 in (8) and (9), we have a = 1042.808 and b = 0.0766335. By substituting the values of a and b in equation (3), F(x) has been computed the results are given in table 12 (column 6). $D_2(x)$ (in column 6 of table 14) suggests that for $r_0 = 1$ and $y_0 = 77$ give rise to an equally good approximation for k and thus the model fits fairly well to the data. Moreover, the procedures to compute a and b are very simple and straightforward.

We may also observe that values of the parameters computed based on Egghe's algorithm give rise to a good approximation for the convex and concave portions of the curve. The author has also observed further that for increased value of r_0 and smaller value of p, the results are, fairly accurate, but for the lower portions of the semi-logarithmic curve.

Unless one is able to compute the standard error for k, any estimate of parameters is incomplete to a great extent. In order to compute the standard error, we must have a fool-proof mathematical expression. So, research is needed in this direction.

As a part of this study, a list of core journals (consisting of 60 journals) is prepared and is given in Appendix.

7. APPLICATIONS OF BIBLIOMETRICS

In the first Conference on Bibliometrics held at Diepenbeek in 1987, Tague [20] has already discussed two application areas (file design and collection management). In her paper, she emphasized the procedures for estimating the parameters of the Zipf/Lotka and Bradford distribution from samples. In the same conference Bookstein [1] suggested that one way to defend the reality of bibliometric distributions is to demonstrate that one can make reasonable decisions in practical situations by using them. In fact, he suggested a micro-economic approach to allocate resources appropriately, especially for the book collection.

In this conference, I would like to raise again the question of "why bibliometrics?" and try to answer it purely from a view point of my experience in India. In the last fifteen years of my teaching experience, I find students are mostly interested in the direct applications of bibliometrics rather than in the theory itself. Often they raise questions like :
Is it really required to use most complicated or probability or statistical models? Is it not possible for us to get the results with desired precision from simple quantitative analysis of bibliometric data?

In the absence of "quantitative culture" among librarians, it is very difficult to introduce any kind of bibliometric techniques in the day-to-day administration of libraries. In these circumstances, we must perhaps approach the problem with simplest possible tools/techniques, no matter how simple it may be.

With this opinion, let me now briefly discuss a few areas where bibliometric analysis are most useful. Among the many bibliometric techniques/tools, Bradford's law is perhaps the best known bibliometric concepts. Bradford's law has found a wide-ranging application in several areas, not only restricted to journal productivity, but also to study journal use patterns in special libraries. This law can appropriately be applied :

a) to determine the core journals in a subject,

b) to determine the number of journals required to cover a specified fraction of the total periodical literature,

c) to determine the completeness of the bibliography.

Bradford's law can ultimately help us in identifying core, less-productive and least productive journals; it may also help us in identifying frequently and infrequently used journals. Such studies are very useful in special libraries, especially in countries like India.

We find in India that receiving journals on time is a rare event, especially when most of the libraries get most of their journals by surface mail. Because of the budgetary restrictions, very few libraries can afford to get the journals by air-mail. In such cases, one can get only the core journals by air-mail and the less productive journals may be obtained by surface mail.

Also one can subscribe only to hard copies of the core journals and to microfiche/microfilm copies of the less productive journals; articles contained in the least productive journals may be acquired by way of collecting its reprint/pre-print and xerox copies. By adopting such policies, the limited fund may be utilized to a maximum extent.

Special libraries in India face yet another problem of getting required fund for binding journals. In a library where 500 to 2000 journals are subscribed, about Rs.25,000 to 100,000 is required to bind the journals. Since hardly 10 to 20 % of the journals are used very frequently (that too, in about 5-6 years from its acquisition), it is suggested here that it is enough to bind only the frequently used journals. The list of frequently used journals may be obtained by using the appropriate bibliometric techniques.

The analysis of journal usage data also helps us in identifying the optimal retention period for a journal and the optimal number of journals to be maintained in the collection.

As has been discussed by Tague in her paper [2] and Ravichandra Rao [19], bibliometric studies, particularly journal productivity studies, circulation theory, etc. help us in organizing records in databases very effectively. In such studies, usually the problem is to determine the size of the frequently and infrequently borrowed/used or size of core collection. Depending on the purpose and requirement, the size can, however, be determined.

Other bibliometric laws, such as Zipf's law [22] has its main application in information retrieval especially in generation of automatic/semi-automatic indexes to a text. Lotka's law [15] is mainly applicable to determine the size of the elite scientists.

Bibliometric techniques can also be used in identification of trends in a subject and also in the study of scientific communication.

8. CONCLUSION

This study indicates that the negative binomial distribution is not really good enough to describe the distribution of articles in economics, especially when the data is heterogeneous in nature or when the bibliography is incomplete (i.e. when the shape of semi-logarithmic curve is S-shape).

However, Leimkuhler's model, when Egghe's algorithm to compute the parameters is used, fits very well to the empirical data on rank distributions of journals in economics. Also, it is observed that a simple procedure to compute k (the Bradford multiplier) under the assumption r_0 (the number of core journals) = 1 and its corresponding value of y_0 (the number of articles contained in r_0) give rise to good approximations for the parameters of Leimkuhler's model (a and b).

Further it is observed that only 15 % of the journals in economics contribute about 10 % of the total periodical literature in economics.

Research is needed in the direction of estimating the parameters of the negative binomial distribution to fit the empirical distributions of articles in incomplete bibliographies and also research must be carried out further to develop formulae to compute the standard error of k (the Bradford multiplier).

9. BIBLIOGRAPHICAL REFERENCES

[1] Bookstein, A., Applications of the bibliometric distributions. Informetrics 87/88, Select proceedings of the First International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval, Diepenbeek, Belgium, 25-28 August 1987. Edited by Leo Egghe and Ronald Rousseau, Elsevier Science Publishers, Amsterdam, 1988.

[2] Bradford, S.C., Sources of information on specific subject. Engineering : 85-6, Jan.26, 1982.

[3] Bradford, S.C., Documentation. London : Crosby Lockwood, 1948.

[4] Burrell, Q.L., Alternative models for library circulation data. Journal of Documentation, 38(1), 1982, 1-13.

[5] Burrell, Q.L. and Cane, V.R., The analysis of library data. Journal of Royal Statistical Society, Sever A, Part 4, 439-71, 1982.

[6] Burrell, Q.L., Predictive aspects of some bibliometric processes. Informetrics 87/88, Select proceedings of the Theoretical Aspects of Information Retrieval, Diepenbeek, Belgium, 25-28 August 1987. Edited by Leo Egghe and Ronald Rousseau, Elsevier Science Publishers, Amsterdam, 1988.

[7] Burrell, Q.L., Modelling the Bradford Phenomenon. Journal of Documentation, 44(1) 1-18, 1988.

[8] Egghe, L., The dual of Bradford's law. Journal of the American Society for Information Science, 37(4) : 246-55, 1988.

[9] Egghe, L., Applications of the theory of Bradford's law to the calculation of Leimkuhler's law and the completion of bibliography (to appear in Journal of the American Society for Information Science, 1988).

[10] Groos, D.V., Bradford's law and the Keenan-Atherton data. American Documentation, 18(1) : 46, 1967.

[11] Hjerppe, R., Outline of Bibliometrics and citation analysis. Stockholm, Royal Institute of Technology, 1980.

[12] Hubert, John, J., Bibliometric models for journal productivity. Social Indicators Research, 4, 441-73, 1977.

[13] Keenan, S. and Atherton, P., The journal literature of Physics. New York: American Institute of Physics, 1964.

[14] Leimkuhler, F.F., Bradford distribution. American Journal of Documentation, 23(3) : 197-207, 1967.

[15] Lotka, A.J., Frequency distribution of scientific productivity. Journal of Washington Academy of Science, 16 : 317-23, 1926.

[16] O'Neill, E.T., Generalized Bradford Distribution. State University of New York at Buffalo, School of Information and Library Studies, 1972, Research Memorandum No.1.

[17] Pritchard, A., Statistical Bibliography or Bibliometrics. Journal of Documentation, 25(4) : 348-9, 1969.

[18] Ranganathan, S.R., Librametry and its scope. DRTC Annual Seminar, 7, Paper DA, 1969.

[19] Ravichandra Rao, I.K., Document and User Distributions. Library Science with a slant to Documentation, 19(2) : 69-96, 1982.

[20] Tague, J.H., What is the use of bibliometrics? Informetrics 87/88. Select proceedings of the first International Conference on Bibliometrics and Theoretical Aspects of Information Retrieval, Diepenbeek, Belgium, 25-28 August 1987. Edited by Leo Egghe and Ronald Rousseau, Elsevier Science Publishers, Amsterdam, 1988.

[21] Vickery, B.C., Bradford's law of scattering. Journal of Documentation, 4(3) : 198-203, 1948.

[22] Zipf, George, Human behaviour and the principle of least effort. New York, Addison-Wesley, 1949.

APPENDIX : LIST OF CORE JOURNALS IN ECONOMICS

(List prepared based on the quantitative study)

1. Review of Economic Studies
2. Journal of Political Economy
3. Quarterly Journal of Economics
4. Journal of International Economics
5. Econometrika
6. Southern Economic Journal
7. American Economic Review
8. Journal of Finance
9. Integration Latinoamerican
10. Journal of Financial and Quantitative Analysis
11. Review of Economics and Statistics
12. Economic Appliquee
13. Journal of Money, Credit and Banking
14. Planovoe Hozjastvo
15. Applied Economics
16. Economic Development and Cultural Change
17. International Economic Review
18. Journal of Economic Theory
19. Journal of Public Economy
20. Revued Economic Politique
21. Land Economics
22. Economic Journal
23. Journal of Industrial Economics
24. American Journal of Agricultural Economics
25. Ekonomista
26. Journal of Development Economics
27. Oxford Economic Papers
28. Journal of Development Studies
29. Economic Inquiry
30. Journal of Economic Issues
31. Weltwirtschaft
32. Wirtschaftwissenchak
33. Quarterly Review of Economics and Business
34. Canadian Journal of Economics/Revue Canadienne d'Economie
35. Development and Change
36. Economic Record
37. Planovane Hospodarstvi
38. Public Finance Quarterly
39. Revue Economique
40. Acta Occenomica
41. Economska Revija
42. Rassegna Economica (Napoli)
43. Aussenwirtschaft
44. Ekonomicky Casopis
45. Journal of Financial Economics
46. Maandschrift Economics
47. Oxford Bulletin of Economics and Statistics
48. American Journal of Economic and Sociology
49. Revista di Economica Agraria
50. Trimestre Economics
51. Finances et Developent
52. Giornale degli Economisti Annali di Economica
53. Econonich en Social Tiodschrift
54. Economica
55. Review of Income and Wealth
56. Review Intermitall du Travail
57. Revista Internationale di Science Economice Commerciali

58. Revista Internationale di Science Sociali
59. Staff Papers
60. Revista Internationale di Science and Economiche and Economicile

Note : I understand from my colleagues who work in economics that the following journals are equally important in economics and may be considered as a part of the core journals. However, in this study, we may observe that these titles do not appear in the core list!

1. World Development
2. Sankhya
3. Journal of Econometrics
4. European Review of Agricultural Economics
5. Journal of Agricultural Economics
6. Journal of Game Theory
7. Journal of Empirical Economics
8. Bell Journal of Economics