Information Processing & Management Vol. 26, No. 5, pp. 655–672, 1990
Printed in Great Britain.

0306-4573/90 $3.00 + .00

AN EMPIRICAL EXAMINATION OF THE EXISTING MODELS FOR BRADFORD'S LAW

LIWEN QIU*
School of Library and Information Science, The University of Western Ontario, London, Ontario, Canada N6G 1H1

(*Received* 16 *October* 1989; *accepted in final form* 23 *January* 1990)

Abstract—All the existing models for Bradford's law were summarized and classified into different categories (e.g., rank-frequency cumulative, rank-frequency noncumulative, size-frequency, and other forms). The relationships between some models were established by mathematical deduction. Nineteen data sets were used to estimate the parameters of the models and then goodness of fit tests were conducted to identify empirically the model in each category which can best describe the phenomenon of journal productivity.

1. INTRODUCTION

Bradford's law of scatter describes a quantitative relaltionship between journals and the papers they publish. At present there are two widely recognized formulations of the so-called Bradford's law: the verbal formulation, which is derived from the verbal statement of Bradford's conclusions, and the graphical formulation, which is an empirical expression derived from the graph of a distribution of journals over periodicals.

Bradford's law as originally stated reads:

> If scientific periodicals were arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several groups or zones containing the same number of articles as the nucleus, when the number of periodicals of the nucleus and the succeeding zones will be as 1: $\alpha : \alpha^2 \ldots$ [1]

The constant α is a property of the collection of periodicals called the Bradford multiplier. This statement is called the verbal formulation of Bradford's law.

The graphical formulation is obtained by plotting a curve in a plane whose coordinates are the cumulative number of articles (in the y axis) and the logarithm of the cumulative number of journals of the collection (in the x axis), where journals are cumulated from most to least productive. This curve has invariably an ascending shape which, after a certain point, approaches to a straight line.

Bradford did not provide a mathematical equation to support either the verbal or the graphical description of his law. However, it is generally accepted that the graphical formulation is expressed by

$$Y = A + B * \log X$$

where Y is the proportion of the cumulative number of papers contained in the proportion X of the first most productive journals.

Many different models have been provided by later scholars. For example, Brookes developed the graphical formulation [2]

$$R(r) = \begin{cases} \alpha * r^{\beta} & (1 \leq r \leq C) \\ K * \log r/s & (C \leq r \leq N) \end{cases}$$

where $R(r)$ is the cumulative number of papers in the first r journals when journals are ranked from most to least productive and N is the total number of journals.

Leimkuhler developed a model based on Bradford's verbal formulation [3]

$$F(x) = \frac{\log(1 + B * X)}{\log(1 + B)}$$

where $F(x)$ is the relative total number of papers contained in the proportion X of the most productive journals.

Vickery [4] was the first to point out that the verbal and graphical formulations of Bradford's theory were not mathematically equivalent. Later, Wilkinson [5] discussed the ambiguity of Bradford's law in terms of the two different approaches. She claimed that the verbal formulation expressed Bradford's theory while the graphical formulation expressed his observation.

In this study, a different approach, neither graphical nor verbal but one of empirical validation, was taken to examine the existing forms of Bradford's law. The purpose of this study was to find out, by statistical testing, which mathematical model can best describe the phenomenon of journal productivity.

2. LITERATURE REVIEW

A literature search was carried out to collect all the existing forms of Bradford's law. The search included:

1. ERIC CD-ROM search. The time span was 1981–Mar. 1987.
2. A manual search on *Library Literature*. The time span was 1972–1988 (from 1972, *Library Literature* has the entry "Bradford's Law").
3. A manual search on *Scientometrics* from Vol. 1 until the most recent issue, since *Library Literature* does not cover *Scientometrics*.
4. Because of the time delay of *Library Literature*, papers published after 1988 may not be reported. Therefore, a manual search of JASIS, *Journal of Documentation* from 1988 to the most recent issue was conducted.
5. Papers published before 1972 were found through the references of later papers (the references of every searched paper were examined).

Papers were summarized into the following four categories:

1. Providing new forms of mathematical models (e.g., reference 3, 5, and 7).
2. Discussion of the meaning or characteristics of parameters and curves (e.g., reference 13, 20, 22, and 23).
3. Fitting data set(s) to model(s) (e.g., reference 29, 30, and 35).
4. Summarizing different models and determining their relationships (e.g., reference 8 and 14).

It is surprising to find that so many papers have been published on Bradford's law but that no comprehensive summarizing of the models has been carried out. As Hubert [7] pointed out, "a set of data that conforms to a particular formula tells little about the underlying phenomena." Although so many data sets have been collected and fitted to different models, few goodness of fit tests have been conducted. The study reported here, therefore, summarizes all the existing models and examines their relationships and their validity and applicability by means of statistical tests.

3. EXISTING MODELS

A summary of the 22 previously published models of Bradford's law is presented in Appendix 1. The basic variables of these models are the number of journals and the number of papers. Using the independent variables, we can classify these models into two basic types: rank-frequency and size-frequency. If $g(p)$ is the number of journals having p papers, we speak of a size-frequency relationship. If $f(r)$ is the number of papers in a journal of rank r, we speak of a rank-frequency relationship. In the former we can estimate the number of journals, given the number of papers; in the latter we can estimate the number of papers given the rank of the journals. Hubert [8] has presented a statistical relationship linking these two types of productivity distributions. Within the basic categories of rank-frequency and size-frequency, these models can be further classified into cumulative and noncumulative types. In Appendix 1, rank noncumulative models are represented by $f(r)$, the number of papers contained in a journal of rank r. The rank cumulative category can be further divided into fractional and nonfractional types. If $R(r)$ is the cumulative frequency of papers in the r most productive journals, r is the rank of journal, N is the total number of journals, and $R(N)$ is the total number of paprs, then $F(x) = R(r)/R(N)$ is the rank-cumulative-fractional model, where $X = r/n$. That is, the relationship between models in part 2 of Appendix 1 and those in part 1 is $X = r/N$, $F(x) = R(r)/R(N)$.

The original notation of parameters in some models is altered in order to avoid confusion and to make the comparison of different formulae easier. Models of stochastic process type (e.g., Burrell's model [9]) are not included in Appendix 1 because these models have a time parameter which is impossible to test using the present data sets (all the present data sets have no time variable).

Models which have a Bradford multiplier (models 2–7, 5–1, and 5–2) were not tested in this study because the value of the Bradford multiplier depends on the number of Bradford zones. Different divisions of zones will result in different values of the Bradford multiplier. Although some authors have discussed the appropriate choice of Bradford multiplier [31], there is not yet a standard and generally accepted way of deciding this parameter.

The 22 models may seem overwhelming, but after we unify the original notation, it is clear that some forms of models are only disguised forms of others. For example, model 2–2, 2–3, 2–4 and 2–5 are similar because all are in the form of $A * \log(1 + B * r) + C$, where A, B and C are constants. Model 2–2 is a specific case of model 2–3, when the constant $R(0) = 0$. Model 2–2 is also a specific case of model 2–5, when $j_1 = j_2$ and $a_1 = a_2$. Model 2–4 and model 1–2 are equivalent when constant j in model 2–4 equals the total number of papers $R(N)$ (remembering that $F(x) = R(r)/R(N)$). Because j may not be equal to $R(N)$, both models were tested to see which one fits better.

Model 1–2 can be derived from model 2–2 [5]:

$$R(r) = j * \log(r/a + 1) \longrightarrow \text{model 2–2}$$

$$F(x) = R(r)/R(N)$$

$$= j * \log(r/a + 1)/j * \log(N/a + 1)$$

$$\text{Let } B = N/a \quad \text{and} \quad X = r/N$$

$$\text{then } F(x) = \log(1 + B * x)/\log(1 + B) \longrightarrow \text{model 1–2}$$

Because these two models are substantively the same, only one of them (model 1–2) was tested.

Similarly, model 1–2 can be derived from model 2–4:

$$R(r) = j * \log(1 + a * r)/\log(1 + a) \longrightarrow \text{model 2–4}$$

$$F(x) = R(r)/R(N) = \frac{j * \log(1 + a * r)/\log(1 + a)}{j * \log(1 + a * N)/\log(1 + a)} = \frac{\log(1 + a * r)}{\log(1 + a * N)}$$

Let $B = a * N$ and $x = r/N$

$$\text{then } F(x) = \frac{\log(1 + a * x * N)}{\log(1 + a * N)} = \frac{\log(1 + B * x)}{\log(1 + B)} \longrightarrow \text{model 1–2}$$

Therefore, model 2–4 does not need to be tested since model 1–2 is tested.

Model 3–1 can be derived from model 2–2 if n is assumed to be continuous. In this case, we may derive the noncumulative distribution from the cumulative one by taking the derivative:

$$R(r) = j * \log(1 + r/A) \longrightarrow \text{model 2–2}$$

$$\frac{dR(r)}{dr} = \frac{j/A}{1 + r/A}$$

Let $j/A = a$ and $d = 1/A$

$$\frac{dR(r)}{dr} = \frac{a}{1 + d * r} \tag{1}$$

In eqn (1), $dR(r)/dr$ is virtually $f(r)$ in model 3–1. Therefore, eqn (1) is equivalent to

$$f(r) = \frac{a}{1 + d * r} \longrightarrow \text{model 3–1}$$

So, model 3–1 was not tested.

Model 2–6 is equivalent to model 2–2:

$$R(r) = \log_b\left(\frac{a + r}{a}\right) \longrightarrow \text{model 2–6}$$

Substituting $b = (a + N)/a$

$$R(r) = \frac{\log(1 + r/a)}{\log(1 + N/a)}$$

$$\text{let } j = \frac{1}{\log(1 + N/a)}$$

$$R(r) = j * \log(1 + r/a) \longrightarrow \text{model 2–2}$$

Therefore, model 2–6 was not tested.

Model 5–1 and model 1–2 are equivalent [7, p 455]: Suppose, in model 5–1, we do not group the journals. In other words, each group contains only one journal. Then we could take $S = 1$ in model 5-1 and S_k is equivalent to X, the cumulated proportion of journals, in model 1-2. The Bradford multiplier k in model 5-1 is equivalent to $F(X)$, the cumulated proportion of papers contained in the proportion of journals X, in model 1-2. Then we can write model 5-1 as

$$X = \frac{n^{F(X)} - 1}{n - 1} \qquad 0 \leq F(X) \leq 1$$

which yields

$$F(X) = \frac{\log[X(n-1)+1]}{\log n} \qquad 0 \leq X \leq 1$$

Let $n - 1 = B$

$$F(X) = \frac{\log(B * X + 1)}{\log(B + 1)} \longrightarrow \text{model 1-2}$$

Model 5-2 is a changed form of model 2-2. Because $X = r/N$, model 5-2 can be changed into

$$R(r) = j * \log(1 + C * X)$$
$$= j * \log(1 + C * r/N) = j * \log(1 + r/a)$$

Therefore only model 2-2 need be tested. Because it has been shown that model 2-2 is equivalent to model 1-2 (see above), only model 1-2 will be tested.

To summarize, models 1-1, 1-2, 1-3, 1-4, 1-5, 2-1, 2-3, 2-5, 3-2, 3-3, 4-1, 4-2, 4-3 and 5-4 will be tested.

4. DATA SETS

More than 30 sets of previously published data were found in the literature search. The sources of these data sets are quite varied. Apart from journal-paper types of data, the standard sources of Bradford distributions, there are monograph-publisher types of data and even library records of journal usage data. Because library usage data represent a different application, they were discarded in the first round of data filtration. As a result, 27 sets of data remained.

For strict conformity with Bradford's law, certain conditions have to be imposed on the data set. Bradford himself did not give a standard for data collection. However, Brookes' [2] three conditions are generally accepted as the standard. They are:

1. the subject of the bibliography must be well defined;
2. the bibliography must be complete, that is, all relevant papers and periodicals must be listed;
3. the bibliography must be of limited time span so that all contributing periodicals have the same opportunity to contribute papers.

If these strict conditions are followed, very few present data sets can be used. Nevertheless, it is unnecessary to be so strict because "it is found that the form of the graph is surprisingly stable even when these conditions are not fully satisfied" [2]. Therefore, for the sake of practicality, eight sets of incomplete data (data that do not have a complete list from the most productive to the least productive journals) were deleted in the second round of data filtration. Those data sets which do not have an ideal time span but have a reasonable number of data points (larger than 10 data points) were kept. Finally, 19 sets of data remained with a wide range of subject topic and type of materials (see Appendix 2).

5. ESTIMATION AND TESTING

In order to estimate the parameters in the models, the method of nonlinear least square estimation was used (maximum likelihood estimation is not appropriate for a cumulative distribution). A microcomputer software program was used to do the estimating.

After estimating the parameters, a goodness of fit test was conducted to determine which model fitted the data sets best, that is, which model can best describe Bradford phenomena. The two most commonly used goodness of fit tests are the chi-square and Kolmogorov-Smirnov (K-S) tests. In this study, the K-S test was used in dealing with cumulated data while the chi-square test was applied to uncumulated data.

Theoretically speaking, statistical tests such as the chi-square test and K-S test are not suitable for testing the hypothesis of this kind [39]. The assumptions of independence and randomness which underlie the validity of the chi-square and K-S tests may not be satisfied in the data being used here. However, there is no other statistical test which is more suitable for the purpose of this study. In fact, bibliometricians have used the chi-square and K-S tests to test this kind of model in practice.

Models 2-1 and 2-5 require the decision of the transition point that separates the nuclear zone and periphery zone. There is no standard way of deciding this critical point. Although Rousseau suggested a p-nucleus formula for calculating the data points that should be included in the nuclear zone [38], it is not used in this study to decide the nuclear zone since it necessitates the estimation of a parameter which requires an extra fitting. Rather, the data points were plotted on the screen and transition points were picked out by looking at the shape of the curve.

The results of parameter estimation and the goodness of fit test for each model tested is listed from Table 1 to Table 14. All the logarithms in this study are natural logarithms and all the statistical tests in this study are at the significance level of 5%.

6. ANALYSIS OF TEST RESULTS

A. *Rank-frequency cumulative models*

(1) **Model 1-1** $F(X) = 1 + B * \log X$ **and model 1-5** $F(X) = A + B * \log X$. They are simplest in form and therefore permit the easiest estimation of the parameters. It is surprising to see that their goodness of fit is so poor—both models passed only 1 K-S test out of 19 (see Table 1 and Table 2. Each table contains the values of estimated parameters and the goodness of fit test results). Model 1-1 was developed by Cole, but the model did not fit even his data (data set 1)!

Model 1-1 is a specific case of model 1-5 when $A = 1$. The fitting results of model 1-5 in Table 2 show that the value of A is very close to 1. In order to find out if it is necessary to have parameter A in model 1-5, a paired T test of 19 K-S test values for the two models was conducted and showed no significant difference between these two models in terms of goodness of fit. It can be concluded that parameter A in model 1-5 is unnecessary. So model 1-1 is preferred over model 1-5 as it has fewer parameters.

(2) **Model 1-2** $F(X) = \dfrac{\log(1 + B * X)}{\log(1 + B)}$, **model 1-4** $F(X) = A * \log(X + C) + B$,

model 2-3 $R(r) = j * \log(1 + r/a) + R(0)$,

and model 2-5 $R(r) = \begin{cases} j_1 * \log(1 + r_1/a_1) \\ j_2 * \log(1 + r_2/a_2). \end{cases}$

These four models all passed the K-S test for eight out of 19 data sets (see Table 3-6).

The justification that Asai gave his model (1-4) over other models, including model 1-2 and model 2-2, was that model 1-4 has a smaller minimal value of root-weighted square error, but no statistical test was conducted to see if the difference was significant. To compare the four models in terms of goodness of fit, a within subject F test was con-

Table 1. Test results for model 1-1

$F(X) = 1 + B * \log X$

Data set	B	K-S*	C. V.**	Pass
1	0.18	0.086	0.045	N
2	0.19	0.101	0.054	N
3	0.14	0.045	0.023	N
4	0.15	0.073	0.016	N
5	0.15	0.035	0.032	N
6	0.13	0.151	0.035	N
7	0.16	0.059	0.020	N
8	0.17	0.135	0.015	N
9	0.18	0.111	0.037	N
10	0.22	0.178	0.068	N
11	0.21	0.164	0.051	N
12	0.15	0.160	0.014	N
13	0.17	0.07	0.040	N
14	0.184	0.041	0.053	Y
15	0.207	0.357	0.020	N
16	0.195	0.302	0.034	N
17	0.172	0.129	0.021	N
18	0.187	0.222	0.028	N
19	0.189	0.089	0.047	N

*K-S stands for the K-S test result.
**C. V. is the critical value for the K-S test.

ducted on the K-S values for the four models. The result of an F test ($F = 1.496, p > 0.05$) showed no significant difference among the four models. That is, Asai's model is not significantly better than the other models. Since model 1-2 has the least number of parameters, it is preferred over the other three models.

Leimkuhler [3], when presenting model 1-2, claimed "parameter B is related to the subject field and completeness of the collection." Examining the B values and their corresponding subject fields, the claimed relationship was not found. For example, data set 4 and data set 11 refer to the same subject: information science. But their B values are very different ($B = 736.68$ for data set 4, $B = 56.93$ for data set 11). Drott and Griffith [23] used

Table 2. Test results for model 1-5

$F(X) = A + B * \log X$

Data set	A	B	K-S*	C. V.**	PASS
1	0.980	0.180	0.093	0.045	N
2	0.950	0.180	0.089	0.054	N
3	0.990	0.140	0.057	0.023	N
4	1.080	0.170	0.126	0.016	N
5	0.980	0.150	0.041	0.032	N
6	1.130	0.180	0.134	0.035	N
7	1.050	0.180	0.124	0.020	N
8	1.040	0.180	0.181	0.015	N
9	0.950	0.170	0.092	0.037	N
10	0.900	0.190	0.121	0.068	N
11	0.970	0.210	0.145	0.051	N
12	0.990	0.150	0.152	0.014	N
13	0.980	0.170	0.058	0.040	N
14	1.000	0.180	0.043	0.053	P
15	0.890	0.180	0.266	0.020	N
16	0.750	0.140	0.251	0.034	N
17	1.030	0.180	0.158	0.021	N
18	0.950	0.170	0.186	0.028	N
19	1.080	0.220	0.081	0.047	N

*K-S stands for the K-S test result.
**C. V. is the critical value for the K-S test.

Table 3. Test results for model 1-2

$$F(X) = \frac{\text{Log}\ (1 + B * X)}{\text{Log}\ (1 + B)}$$

Data set	B	K-S*	C. V.**	Pass
1	160.320	0.032	0.045	Y
2	112.790	0.026	0.054	Y
3	913.000	0.032	0.023	N
4	736.680	0.078	0.016	N
5	513.000	0.017	0.032	Y
6	2097.000	0.157	0.035	N
7	387.000	0.062	0.020	N
8	164.200	0.103	0.015	N
9	140.400	0.020	0.037	Y
10	36.790	0.006	0.068	Y
11	56.930	0.055	0.051	N
12	556.060	0.035	0.014	N
13	254.450	0.016	0.040	Y
14	170.430	0.041	0.053	Y
15	40.120	0.049	0.020	N
16	14.090	0.012	0.034	Y
17	251.700	0.060	0.021	N
18	113.710	0.048	0.028	N
19	162.170	0.110	0.047	N

*K-S stands for the K-S test result.
**C. V. is the critical value for the K-S test.

23 data sets to fit the Bradford curve and they did not find a relationship between the parameters of Bradford curve and subject fields.

(3) **Model 2-1** $R(r) = \begin{cases} \alpha * r^{\beta} & (1 \le r \le C) \\ K * \log(r/s) & (C < r \le N). \end{cases}$

Model 2-1 passed K-S tests for nine out of 19 data sets (see Table 7).

Brookes claimed that parameter s could be an objective measure of subject breadth.

Table 4. Test results for model 1-4

$$F(X) = A * \text{Log}\ (X + C) + B$$

Data set	A	B	C	K-S*	C. V.**	PASS
1	0.210	1.030	0.010	0.034	0.045	P
2	0.220	1.010	0.010	0.140	0.054	P
3	0.147	1.007	0.001	0.028	0.023	N
4	0.181	1.107	0.001	0.107	0.016	N
5	0.160	1.004	0.002	0.018	0.032	P
6	0.162	1.108	−0.003	0.107	0.035	N
7	0.208	1.080	0.007	0.082	0.020	N
8	0.239	1.110	0.008	0.112	0.015	N
9	0.210	1.010	0.009	0.019	0.037	P
10	0.280	0.990	0.030	0.005	0.068	P
11	0.270	1.050	0.020	0.058	0.051	N
12	0.171	1.050	0.002	0.048	0.014	N
13	0.182	1.008	0.003	0.009	0.040	P
14	0.200	1.020	0.003	0.020	0.053	P
15	0.304	1.040	0.034	0.052	0.020	N
16	0.380	0.970	0.080	0.009	0.034	P
17	0.210	1.086	0.004	0.087	0.021	N
18	0.230	1.050	0.009	0.055	0.028	N
19	0.230	1.100	0.003	0.097	0.047	N

*K-S stands for the K-S test result.
**C. V. is the critical value for the K-S test.

Table 5. Test results for model 2-3

$R(r) = j * \text{Log}(r/a + 1) + R(0)$

Data set	j	a	$R(0)$	K-S*	C. V.**	pass
1	193.330	1.850	29.270	0.034	0.045	Y
2	141.980	2.400	37.760	0.014	0.054	Y
3	508.580	0.453	−290.330	0.028	0.023	N
4	1335.260	1.100	−953.630	0.107	0.016	N
5	285.910	0.810	19.290	0.018	0.032	Y
6	260.650	0.006	−974.600	0.134	0.035	N
7	917.220	1.830	275.140	0.082	0.020	N
8	2065.130	5.650	−344.710	0.112	0.015	N
9	279.810	3.070	39.630	0.019	0.037	Y
10	108.670	4.370	−1.390	0.005	0.068	Y
11	191.460	3.190	−19.890	0.058	0.051	N
12	1695.600	3.140	−320.210	0.048	0.014	N
13	203.980	0.940	−28.580	0.009	0.040	Y
14	127.670	0.440	−98.540	0.020	0.053	Y
15	1475.790	21.940	63.230	0.052	0.020	N
16	612.890	59.110	19.740	0.009	0.034	Y
17	842.740	2.230	−250.210	0.087	0.021	N
18	545.200	5.150	−77.400	0.055	0.028	N
19	491.090	26.730	207.090	0.346	0.047	N

*K-S stands for the K-S test result.
**C. V. is the critical value for the K-S test.

He reported "$s = 1$ for only narrow scientific subjects: as the subject widens, so does s" [2, p 953]. Examining the s values of the 19 data sets and their subject fields (see Table 7 and Appendix 2), it does not seem to be true that the s value stands for the subject breadth. For example, data set 14 (international research in social science), a very broad area, has an s value of 0.84 while data set 12 (schistosomiasis), a very narrow subject, has an s value of 3.08. Data set 2 and 7 have the same subject area (library science), but the s values are somewhat different, the former is 1.36 while the latter is 1. The s values for data set 6 and 19

Table 6. Test results for model 2-5

$$R(r) = \begin{cases} j_1 * \text{Log}(1 + r_1/a_1) & (r_1 = 1,2,\ldots n) \\ j_2 * \text{Log}(1 + r_2/a_2) & (r_2 = 1,2,\ldots p) \end{cases}$$

Data set	j_1	a_1	j_2	a_2	n	K-S*	C. V.**	pass
1	146.87	0.87	188.38	1.43	7	0.029	0.045	Y
2	114.66	1.05	136.44	1.56	11	0.013	0.054	Y
3	705.00	1.83	477.25	0.59	10	0.027	0.023	N
4	2785.09	9.80	1320.49	2.19	10	0.100	0.016	N
5	355.17	1.10	289.82	0.80	6	0.020	0.032	Y
6	363.79	0.85	173.52	0.03	8	0.039	0.035	N
7	704.65	0.68	874.94	1.10	14	0.073	0.020	N
8	2064.93	9.10	1865.90	4.47	6	0.091	0.015	N
9	231.88	1.67	279.49	2.60	16	0.015	0.037	Y
10	104.94	4.32	108.15	4.37	6	0.005	0.068	Y
11	170.41	3.43	191.97	3.60	6	0.057	0.051	N
12	1940.18	5.45	1697.43	3.84	13	0.047	0.014	N
13	211.40	1.20	204.42	1.09	15	0.011	0.040	Y
14	193.82	2.45	128.57	1.00	5	0.021	0.053	Y
15	1141.49	13.71	1391.49	17.57	25	0.041	0.020	N
16	307.28	20.28	611.06	56.47	10	0.008	0.034	Y
17	1009.97	5.06	829.32	2.88	12	0.079	0.021	N
18	732.07	10.51	539.19	5.79	16	0.049	0.028	N
19	196.73	2.67	191.67	1.28	3	0.093	0.047	N

*K-S stands for the K-S test result.
**C. V. is the critical value for the K-S test.

Table 7. Test results for model 2-1

$$R(r) = \begin{cases} \alpha * r^{\beta} & (1 \le r \le C) \\ K * \text{Log } (r/s) & (C \le r \le N) \end{cases}$$

Data set	α	β	K	S	C^*	K-S**	C. V.***	Pass
1	123.92	0.49	181.30	1.19	7	0.025	0.045	Y
2	95.24	0.45	132.63	1.36	11	0.016	0.054	Y
3	361.14	0.57	473.47	0.56	10	0.014	0.023	Y
4	305.46	0.81	1282.99	1.89	10	0.094	0.016	N
5	246.20	0.56	282.12	0.70	6	0.025	0.032	Y
6	315.57	0.48	173.49	0.03	8	0.039	0.035	N
7	757.46	0.40	859.79	1.00	14	0.072	0.020	N
8	271.78	0.74	1830.84	4.03	6	0.091	0.015	N
9	152.39	0.45	272.82	2.30	16	0.045	0.037	N
10	24.18	0.73	96.17	2.83	6	0.01	0.068	Y
11	47.37	0.73	173.86	2.47	6	0.045	0.051	Y
12	401.23	0.70	1619.90	3.08	13	0.035	0.014	N
13	163.35	0.45	201.24	1.00	15	0.035	0.040	Y
14	69.63	0.71	124.00	0.84	5	0.017	0.053	Y
15	153.85	0.62	1280.49	12.76	2	0.036	0.020	N
16	22.68	0.73	485.24	27.90	10	0.021	0.034	Y
17	222.16	0.70	794.71	2.36	12	0.072	0.021	N
18	83.68	0.76	500.72	4.26	16	0.037	0.028	N
19	63.58	0.77	184.80	1.07	3	0.092	0.047	N

*C is the transition point. That is, C = 7 means that the firsts 7 data points fit the first part of the model.
**K-S is the K-S test result.
***C. V. is the critical value for the K-S test.

(both are statistical methodology) are even more different, with 0.03 for data set 6 and 1.07 for data set 19.

In model 2-1, parameter K is the slope of the log-linear curve. Brookes conjectured that $K = N$ (N is the total number of journals in the data set) [2, p. 953]. When the fitted K values are compared with their corresponding N values, the relationship $K = N$ does not seem to hold. But a correlation test of K values and N values showed a significant correlation between the two values ($r = 0.756$; the critical value of Pearson r with $df = 17$ is 0.456).

(4) **Model 1-3 $F(X) = A * \log[B + C * X + D * \log(1 + C * X)]$.** Model 1-3 passed the K-S tests for 11 out of 19 data sets (see Table 8), the best record among all rank frequency cumulative models. It is not surprising that this model turned out to be the best one, because (a) it contains more parameters; (b) it it the result of a direct mathematical deduction from the verbal formulation of Bradford's law.

An interesting feature to notice is that in parameter estimation, different original values can result in different values of the parameters. For example, for data set 2, when the original values were $A = 0.3$, $B = 1.4$, $C = 2.4$, $D = 19$, then, after 80 iterations, the converged results were $A = 0.25$, $B = 1.35$, $C = 0.73$, $D = 93.99$; when the original values were $A = 0.3$, $B = 1.4$, $C = 10$, $D = 1$, then after 14 iterations, the converged results were $A = 0.34$, $B = 1.29$, $C = 6.5$, $D = 5.39$. Putting these two sets of parameters into model 1-3, we get similar results for the expected value of $F(x)$ and the K-S test values (the K-S value is 0.011 for the first set of parameters and 0.012 for the second set of parameters. The critical value is 0.054). When the two sets of parameters were compared, it was found that there was little difference in the values of A and B, but the values of C and D changed inversely. The first set of parameters is smaller in C and greater in D while the second set of parameters is greater in C and smaller in D. Examining model 1-3, it is obvious that C and D have an inverse relationship. This explains why model 1-3 can have different converged estimation results for the same data sets. Because the estimation results of parameters C and D can be so different for model 1-3, it is legitimate to claim that the parameters in this model have no realistic meaning. That is, they are not related to subject field, completeness of data, etc., as may be the case with other models.

Table 8. Test results for model 1-3

$F(X) = A * \text{Log}\,[B + C * X + D * \text{Log}(1 + C * X)]$

Data set	*A*	*B*	*C*	*D*	K-S*	C. V.**	Pass
1	0.29	1.27	3.18	19.14	0.009	0.045	Y
2	0.25	1.35	0.73	93.99	0.011	0.054	Y
3	0.23	1.05	41.94	7.77	0.018	0.023	Y
4	0.25	0.83	9.87	22.49	0.031	0.016	N
5	0.17	0.58	338.32	1.02	0.014	0.032	Y
6	0.24	0.22	8.42	36.10	0.060	0.035	N
7	0.29	1.48	3.46	23.30	0.052	0.020	N
8	0.32	0.94	3.67	15.86	0.079	0.015	N
9	0.23	1.03	89.07	0.53	0.019	0.037	Y
10	0.39	1.01	6.56	2.56	0.003	0.068	Y
11	0.23	1.47	95.67	−1.36	0.039	0.051	Y
12	0.26	1.06	21.76	8.37	0.013	0.014	Y
13	0.18	0.93	265.67	−0.13	0.008	0.040	Y
14	0.18	1.65	302.44	−1.67	0.012	0.053	Y
15	0.28	1.11	42.79	−0.42	0.044	0.020	N
16	0.39	1.03	12.01	0.06	0.009	0.034	Y
17	0.18	1.75	344.69	−1.80	0.060	0.021	N
18	0.20	1.17	170.64	−0.97	0.033	0.028	N
19	0.18	5.17	372.00	−4.70	0.065	0.047	N

*K-S is the K-S test result.
**C. V. is the critical value for the K-S test.

B. *Rank frequency noncumulative models*

(1) **Model 3-2** $f(r) = a * r^{-c}$. The model passed the chi-square tests for 10 out of 19 data sets (see Table 9). Hubert stated that *a* and *c* are parameters of the subject area [7, p 465]. But comparing the *a* and *c* value of similar subjects in Table 9 does not seem to prove this statement.

(2) **Model 3-3** $f(r) = a * (r + b)^c$. This model passed chi-square tests for 14 out of 19 data sets (see Table 10). This model can be seen as a generalized form of model 3-2 since it has a shift parameter *b* and because of this it fits the data better. It is the best model among rank-frequency noncumulative models.

Table 9. Test results for model 3-2

$f(r) = a * r^{-c}$

Data set	*a*	*c*	*df**	Chi.V.**	C. V.***	Pass
1	116.47	0.86	17	6.52	27.59	Y
2	84.41	0.87	14	1.73	23.68	Y
3	372.33	0.83	25	105.92	37.65	N
4	350.51	0.52	64	466.20	83.68	N
5	242.02	0.91	22	22.13	33.92	Y
6	292.71	0.88	23	83.02	35.17	N
7	713.09	1.06	48	358.36	65.17	N
8	253.43	0.40	15	130.83	25.00	N
9	105.62	0.66	21	22.09	32.67	Y
10	23.66	0.47	11	2.60	19.68	Y
11	48.40	0.53	20	8.52	31.41	Y
12	391.80	0.55	70	291.93	90.53	N
13	127.36	0.78	21	11.52	32.67	Y
14	69.19	0.65	19	20.71	30.14	Y
15	109.29	0.41	42	49.23	58.12	Y
16	19.39	0.32	11	1.42	19.68	Y
17	222.61	0.56	53	107.85	70.99	N
18	81.85	0.42	38	53.94	53.38	N
19	69.53	0.49	21	63.19	32.67	N

*df is the degree of freedom in the chi-square test.
**Chi.V. stands for the chi-square test result.
***C. V. is the critical value for the chi-square test.

Table 10. Test results for model 3-3

$f(r) = a * (r + b)^c$

Data set	a	b	c	df*	Chi.V**	C. V.***	Pass
1	68.64	−0.60	−0.62	18	6.07	28.87	Y
2	70.89	−0.60	−1.72	4	1054.00	9.49	N
3	219.77	−0.63	−0.60	26	160.94	38.88	N
4	39522.98	16.94	−1.70	60	44.54	79.08	Y
5	304.50	0.27	−1.00	22	19.81	33.92	Y
6	1475.65	2.02	−1.50	22	35.12	33.92	N
7	273.31	−0.82	−0.60	52	131.11	69.83	N
8	1330.03	7.76	−0.83	15	32.58	25.00	N
9	511.09	2.82	−1.23	18	21.29	28.87	Y
10	70.58	3.05	−0.84	10	0.16	18.31	Y
11	58.66	0.45	−0.60	19	6.71	30.14	Y
12	2596.76	5.97	−1.08	67	10.10	87.11	Y
13	261.44	1.03	−1.07	19	5.75	30.14	Y
14	1055.34	4.93	−1.58	16	7.49	26.30	Y
15	254.27	3.58	−0.65	41	16.46	56.94	Y
16	22.80	0.76	−0.37	11	0.78	19.68	Y
17	617.88	2.85	−0.86	52	26.70	69.83	Y
18	1528.27	12.92	−1.19	35	2.10	49.80	Y
19	59447.12	20.28	−2.26	19	13.64	30.14	Y

*df is the degree of freedom in the chi-square test.
**Chi.V. stands for the chi-square test result.
***C. V. is the critical value for the chi-square test.

C. *Size frequency model*

(1) **Model 4-1** $P(U) = (C/U) - D$ where $P(U)$ is the proportion of journals having a yield not less than U. The model passed K-S tests for 14 out of 19 data sets (see Table 11).

(2) **Model 4-2** $J_p = \dfrac{1}{P * (P + 1)}$ where J_p is the relative number of journals having p references each. The model passed the chi-square tests for 2 out of 19 data sets (see Table 12). This model is unique in terms of simplicity since it has no parameter at all. It is the easiest to use but the goodness of fit is too poor.

Table 11. Test results for model 4-1

$P(U) = (C/U) - D$

Data set	C	D	K-S*	C. V.**	Pass
1	1.01	−3.670	0.027	0.096	Y
2	1.02	0.020	0.016	0.104	Y
3	0.97	0.020	0.055	0.049	N
4	0.99	−0.010	0.036	0.043	Y
5	0.99	0.010	0.031	0.070	Y
6	0.93	−0.030	0.048	0.105	Y
7	1.17	−0.110	0.281	0.086	N
8	1.00	−0.080	0.175	0.052	N
9	1.02	0.010	0.020	0.075	Y
10	1.02	0.070	0.063	0.105	Y
11	1.02	−0.010	0.041	0.105	Y
12	0.99	0.001	0.019	0.032	Y
13	0.99	0.020	0.045	0.082	Y
14	0.97	0.020	0.056	0.102	Y
15	1.23	−0.001	0.274	0.053	N
16	1.09	0.090	0.038	0.050	Y
17	1.07	−0.030	0.095	0.059	N
18	0.98	0.010	0.042	0.056	Y
19	0.97	−0.030	0.040	0.112	Y

*K-S is the K-S test result.
**C. V. is the critical value for the K-S test.

Table 12. Test results for model 4-2

$$J_P = \frac{1}{P * (P + 1)}$$

Data set	df*	Chi.V.**	C. V.***	Pass
1	6	26.80	12.59	N
2	6	8.23	12.59	Y
3	12	33.83	21.03	N
4	19	211.45	30.14	N
5	10	34.08	18.31	N
6	6	85.78	12.59	N
7	9	466.47	16.92	N
8	5	4767.88	11.07	N
9	9	18.83	16.92	N
10	6	11.05	12.59	Y
11	7	25.91	14.07	N
12	26	142.65	38.88	N
13	8	24.93	15.51	N
14	6	68.56	12.59	N
15	15	520.07	25.00	N
16	11	41.94	19.68	N
17	14	163.66	23.68	N
18	14	40.54	23.68	N
19	5	41.21	11.07	N

*df is the degree of freedom in the chi-square test.
**Chi.V. stands for the chi-square test result.
***C. V. is the critical value for the chi-square test.

(3) **Model 4-3** $f(X) = K * X^{-\alpha}$ where $f(X)$ is the number of journals contributing X articles. The model passed chi-square tests for 5 out of 19 data sets (see Table 13).

D. *Other forms*

(1) **Model 5-4** $F(n) = (B/n)^D - C$ where $F(n)$ is the relative rank of a journal with productivity n. Leimkuhler derived this model by showing that Bradford's law is a special case of the Zipf-Mandelbrot law. In the Zipf-Mandelbrot law ($n = B * (F_n + C)^{-1/D}$),

Table 13. Test results for model 4-3

$$f(X) = K * X^{-\alpha}$$

Data set	K	α	*df**	Chi.V.**	C. V.***	Pass
1	99.26	1.66	7	22.50	14.07	N
2	84.98	1.62	7	4.13	14.07	Y
3	424.27	1.73	13	27.75	22.36	N
4	532.63	1.85	19	218.46	30.14	N
5	202.50	1.82	10	33.26	18.31	N
6	90.02	1.92	5	88.99	11.07	N
7	61.27	1.09	18	33.45	28.87	N
8	273.67	1.52	6	3117.41	12.59	N
9	168.40	1.71	10	13.43	18.31	Y
10	101.95	2.00	5	9.52	11.07	Y
11	81.08	1.73	6	21.18	12.59	N
12	907.56	1.75	28	89.20	41.34	N
13	154.08	1.92	7	30.23	14.07	N
14	24.42	1.15	6	4.65	12.59	Y
15	139.19	0.86	38	136.90	53.38	N
16	387.00	1.52	13	67.45	22.36	N
17	212.13	1.42	22	32.67	33.92	Y
18	327.96	1.90	13	46.42	22.36	N
19	72.80	1.79	5	44.30	11.07	N

**df* is the degree of freedom in the chi-square test.
**Chi.V. stands for the chi-square test result.
***C. V. is the critical value for the chi-square test.

Table 14. Test results for model 5-4

$F(n) = (B/n)^D - C$

Data set	B	D	C	K-S*	C. V.**	Pass
1	1.01	0.93	0.010	0.023	0.096	Y
2	1.02	1.01	0.020	0.016	0.104	Y
3	1.00	1.24	−0.004	0.019	0.049	Y
4	0.97	0.92	−0.004	0.045	0.043	N
5	1.00	1.07	0.001	0.021	0.070	Y
6	0.94	1.06	−0.040	0.039	0.105	Y
7	1.75	0.38	0.190	0.091	0.086	N
8	1.06	0.46	0.080	0.075	0.052	N
9	1.02	0.98	0.020	0.018	0.075	Y
10	1.00	1.35	0.001	0.021	0.105	Y
11	1.05	0.79	0.050	0.046	0.105	Y
12	0.99	1.00	0.001	0.019	0.032	Y
13	0.99	1.12	−0.001	0.028	0.082	Y
14	0.98	1.22	−0.020	0.024	0.102	Y
15	1.85	0.45	0.230	0.091	0.053	N
16	1.07	1.10	0.070	0.047	0.050	Y
17	1.04	0.73	0.030	0.017	0.059	Y
18	1.00	1.04	0.003	0.037	0.056	Y
19	0.96	0.92	−0.170	0.037	0.112	Y

*K-S stands for the K-S test result.
**C. V. is the critical value for the K-S test.

parameter $B > 0$, $C \geq 0$, and $D \geq 1$. However, the fitting results of model 4-4 showed that C can be smaller than 0 and D can be less than 1 (see Table 14). It is a little surprising that this model passed K-S tests for 15 out of 19 data sets, the best record in all the models tested.

7. CONCLUSIONS

When using rank-frequency cumulative models, Egghe's model $F(X) = A * \log[B + C * X + D * \log(1 + C * X)]$ should be considered first since it fits better than others. In choosing rank-frequency noncumulative models, Chen and Leimkuhler's model $f(r) = a * (r + b)^c$ should have the first consideration. Fairthorne's model $p(U) = (C/U) - D$ is the best among size-frequency models. Leimkuhler's model $F(n) = (B/n)^D - C$ is the only model tested among models with other forms and it passed goodness of fit tests more times than any other models. Therefore, it is a good choice in general.

Overall, rank-frequency cumulative models (including those which can be deducted from other models and therefore have not been tested directly) passed 78 of the goodness of fit tests, i.e., 37.3% of the tests (each model was tested against 19 data sets). Rank-frequency noncumulative models passed 32 (56.1%) of the goodness of fit tests. Size-frequency models passed 21 (36.8%) of the goodness of fit tests. Models with other forms passed 23 (60.5%) of the goodness of fit tests. Because two types of goodness of fit tests, e.g., K-S test and chi-square test, were used, no direct comparison can be made as to which approach (rank-frequency or size-frequency) is better.

It is interesting to note that all the rank-frequency cumulative models, including the best model $F(x) = A * \log[B + C * X + D * \log(1 + C * X)]$, failed K-S tests with data sets 4, 6, 7, 8, 15, 17, 18, and 19. Further examination of the curve shapes of these data sets showed that they all have the so-called "Groos droop," and that the rest of the data sets have no Groos droops. Exploring the fitting residuals (the difference between expected values and observed values) for these data sets, it is found that the greatest residuals occurred at the tail end of the curves, i.e., the Groos droop part of the curves. It can be concluded, therefore, that none of these models can describe the Groos droop. When Egghe put forward his model, he claimed that this model can describe the Groos droop, while Leimkuhler's model $F(x) = \log(1 + B * X)/\log(1 + B)$ cannot [11]. However, the results of this study showed that Egghe's model failed to do so – it is exactly in the data sets which have Groos droop that the model failed the K-S tests. Based on the above analysis, it is obvious that a direction for further research in Bradford's law is to find a model which can describe the Groos droop part.

Brookes states the necessity of expressing the Bradford curve in two parts: "No single general form of the Bradford law can thus be expected to fit the whole bibliographical data" [37, p. 81]. In fact, Brookes' models, which express the curves in two mathematical formulae, do not fit better than Egghe's single formula model and take double effort to fit. However, since the existing models cannot describe the Groos droop, it might be necessary to have two parts, one formula to describe the Groos droop in the tail, and the other formula to describe the rest of the distribution.

Acknowledgement – The author gratefully acknowledges Dr. Jean Tague for her suggestions and comments concerning this study.

REFERENCES

1. Bradford, S.C. Documentation. London: Crosby Lockwood and Son Ltd.; 1948.
2. Brookes, B.C. Bradford's Law and the Bibliography of Science. Nature, 224(6): 953-956; 1969.
3. Leimkuhler, F.F. The Bradford Distribution. Journal of Documentation, 23(3): 197-207; 1967.
4. Vickery, B.C. Bradford's Law of Scattering. Journal of Documentation, 4(3): 198-203; 1948.
5. Wilkinson, E.A. The Ambiguity of Bradford's Law. Journal of Documentation, 28(2): 122-130; 1972.
6. Maia, M.J.F.; Maia, M.D. On the Unity of Bradford's Law. Journal of Documentation, 40(3): 206-216; 1984.
7. Hubert, J.J. Bibliometric Models for Journal Productivity. Social Indicators Research, 4(2): 441-473; 1977.
8. Hubert, J.J. A Relationship Between Two Forms of Bradford's Law. Journal of the American Society for Information Science, 29(3): 159-161; 1978.
9. Burrell, Q.L. Modelling the Bradford Phenomenon. Journal of Documentation, 44(1): 1-18; 1988.
10. Cole, P.F. A New Look at Reference Scattering. Journal of Documentation, 18(2): 58-64; 1962.
11. Egghe, L. Consequences of Lotka's Law for the Law of Bradford. Journal of Documentation, 41(3): 173-189; 1985.
12. Egghe, L. The Dual of Bradford's Law. Journal of the American Society for Information Science, 37(4): 246-255; 1986.
13. Naranan, S. Bradford's Law of Bibliography of Science: an Interpretation. Nature, 227: 631-632; 1970.
14. Asai, I. A General Formulation of Bradford's Distribution: The Graph-Oriented Approach. Journal of the American Society for Information Science, 32(2): 113-119; 1981.
15. Haspers, J.H. The Yield Formula and Bradford's Law. Journal of the American Society for Information Science, 27(4): 281-287; 1976.
16. Leimkuhler, F.F. Operational Analysis of Library Systems. Information Processing and Management, 13(1): 79-93; 1977.
17. Brookes, B.C. The Haitun Dichotomy and the Relevance of Bradford's Law. Journal of Information Science, 8(1): 19-24; 1984.
18. Fairthorne, R.A. Empirical Hyperbolic Distributions (Bradford-Zipf-Mandelbrot) for Bibliometric Description and Prediction. Journal of Documentation, 25(4): 319-343; 1969.
19. Brookes, B.C.; Griffiths, J.M. Frequency-Rank Distributions. Journal of the American Society for Information Science, 29(1): 5-13; 1978.
20. Chen, Y.; Leimkuhler, F.F. Bradford's Law: An Index Approach. Scientometrics, 11(3-4): 183-198; 1987.
21. Leimkuhler, F.F. An Exact Formulation of Bradford's Law. Journal of Documentation, 36(4): 285-292; 1980.
22. Egghe, L.; Rousseau, R. Causes of the Groos Droop. Scientometrics, 14(5-6): 493-511; 1988.
23. Drott, M.C.; Griffith, B.C. An Empirical Examination of Bradford's Law and the Scattering of Scientific Literature. Journal of the American Society for Information Science, 29(4): 238-246; 1978.
24. DePew, J.N.; Basu, S. The Application of Bradford's Law in Selecting Periodicals on Conservation and Preservation of Library Materials. Collection Management, 8(1): 55-64; 1986.
25. Sivers, R. Partitioned Bradford Ranking and the Serials Problem in Academic Research Libraries. Collection Building, 8(2): 12-19; 1987.
26. Pope, A. Bradford's Law and the Periodical Literature. Journal of the American Society for Information Science, 26(3): 207-213; 1975.
27. Kendall, M.G. The Bibliography of Operational Research. Operational Research Quarterly, 11(1-2): 31-36; 1960.
28. Saracevic, T.; Perk, L.J. Ascertaining Activities in a Subject Area Through Bibliometric Analysis. Journal of the American Society for Information Science, 24(2): 120-134; 1973.
29. Lawani, S.M. Periodical Literature of Tropical and Subtropical Agriculture. UNESCO Bulletin of Libraries, 26(2): 88-93; 1972.
30. Saracevic, T. Five Years, Five Volumes and 2345 Pages of the Annual Review of Information Science and Technology. Information Storage and Retrieval, 7(3): 127-139; 1971.
31. Rousseau, R. Documentation Note: The Nuclear Zone of a Leimkuhler Curve. Journal of Documentation, 43(4): 322-333; 1987.
32. Seetharama, S. Documents on Survey Analysis. Library Science with a Slant to Documentation, 9(3): 384-395; 1972.

33. Lancaster, F.W. Evaluation of the MEDLARS Demand Search Service. Report presented to the U.S. Department of Health, Education, and Welfare Public Health Service; 1968.
34. Freeman C. Bradford Bibliography and the Literature of Marine Science. Australian Academic and Research Libraries, 5(2): 65–71; 1974.
35. Goffman, W.; Warren, K.S. Dispersion of Papers among Journals based on a Mathematical Analysis of Two Diverse Medical Literatures. Nature, 221: 1205–1207; 1969.
36. Goffman, W.; Morris, T.G. Bradford's Law and Library Acquisitions. Nature, 226: 922–923; 1974.
37. Brookes, B.C. A Critical Commentary on Leimkuhler's "Exact" Formulation of the Bradford Law. Journal of Documentation, 37(2): 77–88; 1981.
38. Brooks, T.A. Perfect Bradford Multipliers: A Definition and Empirical Investigation. Paper presented at the 2nd International Conference on Bibliometrics, Scientometrics and Informatics, London, Canada; 1989.
39. Ijiri, Y.; Simon, H.A. Skew Distributions and the Sizes of Business Firms (p. 109). Amsterdam: North Holland Publishing Company. 1977.

APPENDIX 1. A List of the Existing Models

Part 1. *Rank-cumulative (fractional)*

1-1 *Cole—1962*

$$F(X) = 1 + B * \log X$$

1-2 *Leimkuhler—1967*

$$F(X) = \frac{\log(1 + B * X)}{\log(1 + B)}$$

1-3 *Egghe—1985*

$$F(X) = A * \log[B + C * X + D * \log(1 + C * X)]$$

1-4 *Asai—1980*

$$F(X) = A * \log(X + C) + B$$

1-5 *Bradford—1948*

$$F(X) = A + B * \log(X)$$

Part 2. *Rank-cumulative (non-fractional)*

2-1 *Brookes—1969*

$$R(r) = \begin{cases} \alpha * r^{\beta} & (1 \le r \le C) \\ K * \log(r/s) & (C \le r \le N) \end{cases}$$

2-2 *Wilkinson—1972*

$$R(r) = j * \log(r/a + 1)$$

2-3 *Hasper—1976*

$$R(r) = j * \log(r/a + 1) + R(0)$$

2-4 *Leimkuhler—1977*

$$R(r) = j * \log(1 + a * r)/\log(1 + a)$$

2-5 *Brookes—1984*

$$R(r) = \begin{cases} j_1 * \log(1 + r_1/a_1) & r_1 = 1,2,\ldots n \quad \text{nucleus} \\ j_2 * \log(1 + r_2/a_2) & r_2 = 1,2,\ldots p \quad \text{periphery} \end{cases}$$

2-6 *Brookes—1978*

$$R(r) = \log_b(1 + r/a) \qquad \text{where } r = 1,2,\ldots N \text{ and } b = (a + N)/a$$

2-7 *Maia—1984*

$$R(n_k) = j * \log(n_k - b_k) \qquad \text{where} \quad b_k = j * \log(\alpha_k/n_k)$$

n_k—cumulative sum of all periodicals in the collection up to the class of order k
α—Bradford multiplier
$R(n_k)$—cumulative sum of all papers in the collection up to the class of order k

Part 3. *Rank-non cumulative*

3-1 *Fairthorne—1969*

$$f(r) = a/(1 + d * r)$$

3-2 *Hubert—1977*

$$f(r) = a * r^{-c}$$

3-3 *Chen—1987*

$$f(r) = a * (r + b)^c$$

Part 4. *Size-frequency*

4-1 *Fairthorne—1969*

$$P(U) = (C/U) - D$$

$P(u)$ is the proportion of journals having a yield not less than u.

4-2 *Kendall—1960*

$$J_p = \frac{1}{P * (P + 1)}$$

J_P is the relative number of journals having P references each. $P = 1,2,\ldots N$

4-3 *Naranan—1970*

$$f(X) = K * X^{-\alpha}$$

$f(X)$ is the number of journals contributing X articles.

Part 5. *Other Forms*

5-1 *Vickery—1948*

$$S_k = S(n^k - 1) = S_1 \frac{n^k - 1}{n - 1}$$

S_k is the cumulative number of journals in the most productive k groups ($k = 1,2\ldots$).

5-2 *Fairthorne—1969*

$$R(X) = j * \log(1 + C * X)$$

$R(X)$ is the total yield from the more productive fraction of X of all journals.

5-3 *Egghe—1986*

$$m_p = K_p/e^E = 0.56K^p$$

where K is the Bradford multiplier
E is the number of Euler
m_p is the maximal number of papers in a journal in group p.

5-4 *Leimkuhler—1980*

$$F(n) = (B/n)^D - C$$

$F(n)$ is the relative rank of a journal with frequency (productivity) n.

APPENDIX 2. List of Data Sets

Data set	Subject Topic	Reference Source	Time Span
1	petroleum industry	Cole—1962	unknown
2	library science	DePew—1986	3 years
3	remote sense of earth resources	Sivers—1987	3 years
4	information science	Pope—1975	6 years
5	operational research	Kendall—1960	unknown
6	statistical methodology	Kendall—1960	unknown
7	library science	Saracevic—1973	1 year
8	tropical & subtropical agriculture	Lawani—1972	4 years
9	applied geophysics	Bradford—1948	3 years
10	lubrication	Bradford—1948	3 years
11	information science	Saracevic—1971	1 year
12	schistosomiasis	Goffman—1969	>10 years
13	transplantation-immunology	Goffman—1970	3 years
14	international research in social science	Seetharama—1972	unknown
15	medicine	Lancaster—1968	unknown
16	medicine	Lancaster—1968	unknown
17	fishery	Freeman—1974	1 year
18	mast cell	Coffman—1969	>10 years
19	statistical methods	Egghe—1988	unknown