BRADFORD'S LAW: AN INDEX APPROACH

YE-SHO CHEN,* F. F. LEIMKUHLER**

*Department of Quantitative Business Analysis, Louisiana State University,
Baton Rouge, LA 70803 (USA)
**School of Industrial Engineering, Purdue University, W. Lafayette, IN 47907 (USA)

(Received June 17, 1986)

A rigorous analysis of Bradford's law is made using an index for the observed values of
the variables. Three important properties relating size and frequency are identified. Using,
this approach, the shape of Bradford-type curves can be described in terms of three distinct
regions and two shape parameters.

1. Introduction

The most commonly used tool for studying journal productivity is Bradford's method of plotting the cumulative number of journals against the cumulative number of papers they contain, where the journals are in decreasing order of productivity and are plotted on a logarithmic scale. Bradford argued that the apparent linearity of this relationship supported his observation that linear increases in yield required geometric increases in the number of journals, which is known as Bradford's law of scattering. Vickery showed that Bradford's law implies a J-shaped curve with an initial concave segment leading to a longer linear segment and that such a curve gives a better fit to the available data. Subsequently, Groos, O'Neill, and others pointed out that data on large collections of journals tended to have more of an S-shape with a pronounced convex “drop” in the upper section of the curve.

Most efforts to explain the shape of Bradford-type curves have relied on a priori models of the relationship between journal size, rank, and frequency, such as those proposed by Brookes, Haspers, Asai, and others. In this paper, we take a different approach and study the shape of Bradford-type curves in terms of the empirical characteristics of typical journal data. This approach leads to a more precise and realistic understanding of the factors affecting the shape of the Bradford-type curves. This result has considerable practical significance, since it provides rigorous foundations for the applications of Bradford's law, e.g., in modeling informational data and in library systems management.

As a prelude to examining the paper, Section 2 examines the traditional approaches
YE-SHO CHEN, F. LEIMKÜHLER: BRADFORD’S LAW

and the associated problems. Section 3 discusses the index approach and Section 4 identifies six classes of Bradford-type curves. Section 5 derives the slopes of a Bradford-type curve and the shapes of the curves with respect to three important regions are discussed in Sections 6, 7, and 8. Finally, Section 9 presents the conclusion.

2. Problems of traditional approaches

Bradford-type data on journal productivity is based on observations of journal size, i.e., the number of papers of a certain kind contained in a journal; the number or frequency of journals of each size; and journal rank, i.e., the cumulative frequency of journals of the same or greater size. Note that, when several journals have the same size, they are assumed to have the same rank, which is the largest possible rank.

Typically, the observed values of size and rank, beyond the first smaller values, will “jump” to larger values in progressively larger steps; that is, they contain “gaps” and do not run consequentially from one to some maximum number, as is often assumed in the literature. The nature of the gaps has a significant impact on the shapes of the Bradford-type curves and cannot be ignored. Besides ignoring the gaps, there are some other problems associated with traditional approaches.

The two traditional approaches are (1) continuous variable of rank (size) and (2) discrete variable of rank (size). The first approach assumes that rank is a continuous and differentiable variable, e.g., Egghe. This approach only displays the asymptotic behavior of the Bradford law for large rank and overlooks the discrete nature of smaller ranks (Morse and Leimkuhler).

The discrete approach assumes that rank goes from one, two, three, ..., to some maximum number. One common problem with this approach is the goodness-of-fit test in modeling. For example, consider the yield formula of Bradford’s distribution proposed by Haspers:

\[ R(n) = h \log \left( \frac{n}{u} + 1 \right) + R(0), \quad n = 1, 2, \ldots, \]

where \( R(n) \) is the cumulative yield of the first \( n \) journals; \( h, u, \) and \( R(0) \) are constants. In Table 2 of his paper, \( R(174) = 3352, R(242) = 3420, \) and \( R(n) = 0 \) for \( n = 175, \ldots, 241 \). However, the yield formula “predicts” \( R(n) > 0 \) for \( n = 175, \ldots, 241 \).

A more realistic approach is to associate an index, \( i = 1, 2, \ldots, m \), with each step or observed value of size and rank. The discussion follows.