

Article

Temporal Evolution of Bradford Curves in Academic Library Contexts

Haobai Xue

Information and Intelligence Department, University Town Library of Shenzhen, 2239 Lishui Road, Nanshan District, Shenzhen 518055, China; xuehb@utszlib.edu.cn; Tel.: +86-0755-26032370

Abstract: Bradford's law of bibliographic scattering is a fundamental principle in bibliometrics, offering valuable guidance for academic libraries in literature search and procurement. However, Bradford curves can exhibit various shapes over time, and predicting these shapes remains a challenge due to a lack of causal explanation. This paper attributes the deviations from the theoretical J-shape to integer constraints on the number of journals and articles, extending Leimkuhler's function to encompass highly productive core journals, where the theoretical journal number falls below one. Using the Simon–Yule model, key parameters of the extended formulas are identified and analyzed. The paper explains the reasons for the Groos droop and examines the critical points for shape changes. The proposed formulas are validated with empirical data from the literature, demonstrating that this method can effectively predict the evolution of Bradford curves, providing academic libraries with a valuable tool for evaluating journal coverage, optimizing resource allocation, and refining Collection Development Policies (CDP).

Keywords: Bradford's law; bibliometrics; Simon–Yule model; dynamic analysis; academic libraries

1. Introduction

1.1. Background and Significance

Bradford's law, a foundational principle in bibliometrics, holds significant theoretical value in modeling and simulating the dynamics of scientific knowledge production [1–8]. By integrating it into the framework of science dynamics, this research reveals how micro-level behaviors influence macro-level trends in knowledge growth and shifts in core literature. However, generating Bradford curves can be labor-intensive, especially for journals with few relevant papers. Additionally, as the scientific literature in a given discipline often grows exponentially or passes through various developmental stages [9], a Bradford curve created at one point in time may become outdated without adjustments. Some mathematical models predict a J-shaped curve; but, in practice, Bradford curves can take on at least six different shapes, including the S-shaped curve with the so-called Groos droop [10]. There is a lack of causal explanations for these variations, and comprehensive empirical studies are limited [11]. As such, predicting the evolution of Bradford curves remains an open question requiring further investigation.

This paper attributes the different shapes of Bradford curves to the integer constraints of journal and paper numbers. If journal productivity n is high enough that the corresponding theoretical journal number $f_t(n) = C/n^a$ falls below one, the actual journal number $f_e(n)$ can only be zero or one. This discrete nature causes deviations in the core zone from the theoretical predictions of models like Lotka or Simon–Yule. To address this, the paper proposes two distinct formulas for the core and the normal zones, which are analyzed through theoretical methods and Monte Carlo simulations of the Simon–Yule model. The causes of the Groos droop are explained, and the critical points for shape changes are identified. Finally, the proposed formulas are validated using empirical data, demonstrating that this method can predict the evolution of Bradford curves.



Citation: Xue, H. Temporal Evolution of Bradford Curves in Academic Library Contexts. *Publications* **2024**, *12*, 36. <https://doi.org/10.3390/publications12040036>

Academic Editor: Andrew Kirby

Received: 12 July 2024

Revised: 12 September 2024

Accepted: 18 September 2024

Published: 15 October 2024



Copyright: © 2024 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

While Bradford's law may not directly guide acquisition strategies in modern academic libraries, it still provides a valuable framework for evaluating journal coverage and understanding the distribution of core literature within specific fields [12–16]. This suggests its potential for use in ensuring comprehensive and balanced collections, contributing to a more strategic approach to library resource management.

1.2. Literature Review

Bradford's law was first proposed by Bradford in 1934 [17] but did not gain wide recognition until Vickery further developed the theory in 1948 [18]. According to Bradford's law, if journals are arranged in descending order of productivity and divided into p groups with the same number of papers, the number of journals in each group n_i follows the ratio $n_1 : n_2 : \dots : n_p = 1 : k : \dots : k^{p-1}$, where k is the Bradford multiplier. Besides this verbal form, Bradford's law can also be depicted as a J-shaped curve by plotting the accumulated productivity $R(r)$ of the first r journals against the natural logarithm of the journal rank r . Leimkuhler proposed the mathematical formula for Bradford curve in 1967 [19], and Egghe developed a method for determining the parameters of this formula in 1990 [20]. In Leimkuhler's function $R(r) = a \log(1 + br)$, where the key parameters a and b can be calculated from the article number A , journal number T , and the productivity y_m of the most productive journal. Although Leimkuhler's function matches well with many bibliographies, it corresponds to a J-shaped curve, which deviates from those with a Groos droop [20].

Incomplete bibliographies were initially believed to cause the Groos droop, but further research refuted this hypothesis [21]. Egghe demonstrated that if the ranking of each journal r is transformed into $r' = r + r_0$ by adding a large constant $r_0 > 1/b$, then the new curve will concave downwards, showing a Groos droop [22]. The merging of different bibliographies, each with a different maximum journal productivity $y_m^{(i)}$, could explain the large constant r_0 [22]. However, it is also likely that the large core regions (regions with the most productive journals where $f_t(n_i) < 1$) of some bibliographies contribute to the large r_0 [23]. Essentially, y_m in Leimkuhler's function denotes the journal productivity, where $f_t(y_m) = C/y_m^\alpha \approx 1$ [24], rather than the maximum yield X_1 of a journal as claimed by Egghe himself. Thus, if the total number of these journals T_0 exceeds the critical value $r_0 = 1/b$, a Groos droop will emerge. This paper adopts this explanation and extends Leimkuhler's function to predict the evolution of Bradford curves.

In the 1990s, research interest in Bradford's law shifted from the static presentation of data at a particular time to its dynamic and evolutionary aspects [25]. Oluić-Vuković studied how the increase in productivity of core journals affected the shape of the distribution curve over time [26]. By analyzing the research output of Croatian scholars in different subjects, she concluded that the Groos droop or S-shaped curve is caused by an increase in the concentration/dispersal disparity, reflected by the rise in the core/periphery ratio [27]. The dynamic evolution of Bradford curves and the emergence of the Groos droop were presented in her 1992 study [28], and other similar empirical studies partitioning bibliographies over time were conducted by Garg [29], Wagner-Döbler [11] and Sen [30].

Meanwhile, stochastic models like the Simon–Yule model have increasingly been used to study the dynamic characteristics of bibliometric laws [25,31]. Initially introduced by Yule in 1924 for studying the distribution of biological genera by species number, the Simon–Yule model gained recognition when Simon expanded it in 1955 to analyze the frequency distributions of words in writing samples [32]. Besides employing theoretical methods for precisely solving the constant entry rate α of new sources [32], Monte Carlo simulations have been used to explore more complex scenarios, such as declining entry rates α_t [33] and autocorrelated growth rates γ of established journals (also referred to as aging or obsolescence rate) [34]. Chen et al. [35–37] first used the Simon–Yule model to study the evolution of Lotka's and Bradford's laws over time. They found the entry rate α_t and the autocorrelated growth rate γ have significant yet opposite effects on the Bradford curves, offering an explanation for the various types of Bradford curves [36].

Later, Oluić-Vuković also explored the dynamics of Bradford distributions using the Simon–Yule model but found that its steady-state solution was too restrictive to handle time variations, limiting its applicability [25,31]. This paper also utilizes the Simon–Yule model to examine different scenarios’ effects on key parameters (e.g., journal number T_0 , article number A_0 , and maximum productivity X_1 of the core region) of the extended Leimkuhler’s function. However, it is not used directly to forecast the evolution of Bradford curves or to compare them with empirical data. Instead, key parameters are estimated from past empirical data to improve predictions of Bradford curve evolution in the future.

2. Theoretical Study

2.1. Simon-Yule Model

Simon’s generating mechanism for the Bradford distribution is based on the following two assumptions, where $f_t(n, t)$ denotes the number of journals that have published exactly n papers in the first t published papers:

1. There is a constant probability α that the $(t + 1)$ -th paper is published in a new journal—a journal that has not published in the first t papers.
2. The probability that the $(t + 1)$ -th paper is published in a journal that has published n papers is proportional to $nf(n, t)$ —that is, to the total number of papers of all journals that have published exactly n papers.

Therefore, if there are A papers at a given time, then the corresponding journal number T is approximately $T = A\alpha$. Based on Simon’s two assumptions, the steady-state solution of the Bradford distribution can be written as [37]:

$$f_t(n) = \rho B(n, \rho + 1) \approx \rho \Gamma(\rho + 1) n^{-(\rho+1)} \tag{1}$$

where B is the beta function, Γ is the gamma function, and ρ is a function of the entry rate of new journal α , defined as $\rho = 1/(1 - \alpha)$. Equation (1) suggests that the analytical outcome of the Simon–Yule model aligns with Lotka’s law when $\rho \approx 1$.

In addition to the analytical solutions, Monte Carlo simulations are conducted for $\alpha = 0.15$, and the results are compared with the theoretical results of Equation (1), as shown in Figure 1. Detailed procedures for these simulations can be found in [33] and the data and MATLAB code used are available in the Supplementary Materials. To reduce inherent randomness, each case is simulated $N = 10^4$ times, utilizing only the medians of these simulations as the final outputs.

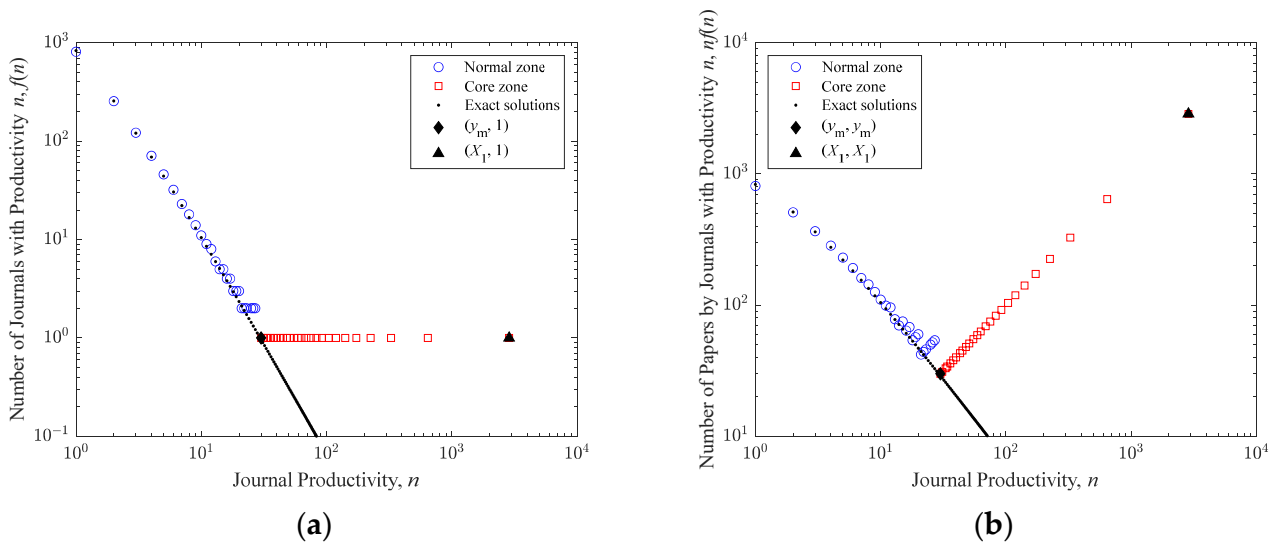


Figure 1. Comparisons of the theoretical and numerical results: (a) number of journals $f(n)$ with productivity n ; (b) number of papers $nf(n)$ produced by journals with productivity n .

Figure 1 illustrates distinct zones in the simulation results: a normal zone (blue circles) and a core zone (red squares). This distinction arises from the necessity for actual journal numbers $f_e(n)$ to be integers, unable to fall below one. Thus, when the journal productivity n is high enough for the theoretical journal number $f_t(n)$ to drop below one, actual journal numbers $f_e(n)$ are constrained to zero or one, deviating from theoretical predictions as shown by the red squares and black dots in Figure 1. Moreover, despite the smaller number of journals in the core region, their contribution to the paper count is significant, as shown in Figure 1b. Hence, accurate prediction of the paper count for each journal X_r in the core zone is crucial for depicting the Bradford curve faithfully.

Estimating journal productivity X_r of the core region involves determining the journal count T_0 and the paper count A_0 first. Since Figure 1 shows a close match between the journal number $f(n)$ and the paper number $nf(n)$ in the normal region, the total journal count T_1 and the total paper count A_1 of the normal region can be directly obtained by summing all journal and paper numbers. These are calculated as $T_1 = \sum_{n=1}^{y_m} f(n)$ and $A_1 = \sum_{n=1}^{y_m} nf(n)$, where y_m is the journal productivity when $f_t(y_m) \approx 1$. In this context, y_m can be seen as the productivity of the most productive journal in the normal region, as indicated by the black diamond in Figure 1. According to Equation (1), the analytical expression for y_m can be derived as:

$$y_m = [A(\rho - 1)\Gamma(\rho + 1)]^{\frac{1}{\rho+1}} \quad (2)$$

Once y_m is calculated, the total journal count T_0 and the paper count A_0 of the core region can be calculated as $T_0 = T - T_1$ and $A_0 = A - A_1$. Alternatively, they can be directly calculated as:

$$T_0 \approx \int_{y_m}^{+\infty} T f(n) dn = \frac{y_m}{\rho} \quad (3)$$

$$A_0 \approx \int_{y_m}^{+\infty} T n f(n) dn = \frac{y_m^2}{\rho - 1} \quad (4)$$

where y_m is calculated using Equation (2).

In the Simon–Yule model with constant entry rate α , the maximum number of papers one journal can have, X_1 , can be estimated using Gumbel's r -th characteristic extreme theory [38–40]:

$$G(X_r) \approx \int_{X_r}^{+\infty} f(i) di = \frac{r}{A} \quad (5)$$

By solving this equation, it can be derived that the productivity of the most productive journal, X_1 , can be expressed as:

$$X_1 = [A\Gamma(\rho + 1)]^{\frac{1}{\rho}} = (\rho - 1)^{-\frac{1}{\rho}} y_m^{\frac{\rho+1}{\rho}} \quad (6)$$

The productivity of the r -th most productive journal, X_r , is related to X_1 by $X_r = X_1 r^{-1/\rho}$. Figure 2 compares Gumbel's r -th characteristic extreme values with the medians of the simulation results. While Gumbel's theory effectively predicts the largest paper number X_1 , it falls short in estimating other paper numbers in the core zone X_r , for $r = 2, 3, \dots, T_0$. Thus, an alternative method is assumed in this paper, where all other X_r , for $r = 2, 3, \dots, T_0$, are related to the largest paper number X_1 through the equation:

$$\frac{X_1}{X_r} = k(r - 1) + 1 \quad (7)$$

where k is the only parameter waiting to be determined. The validity of this equation is supported by Figure 2b, where the blue circles represent the simulation results, and the blue dashed lines show the linear fitting results. Hence, the productivity of the r -th most

productive journal can be derived from Equation (7), and the cumulative productivity of the first r most productive journals can be written as:

$$R_c(r) = \sum_{i=1}^r \frac{X_1}{k(i-1) + 1} \tag{8}$$

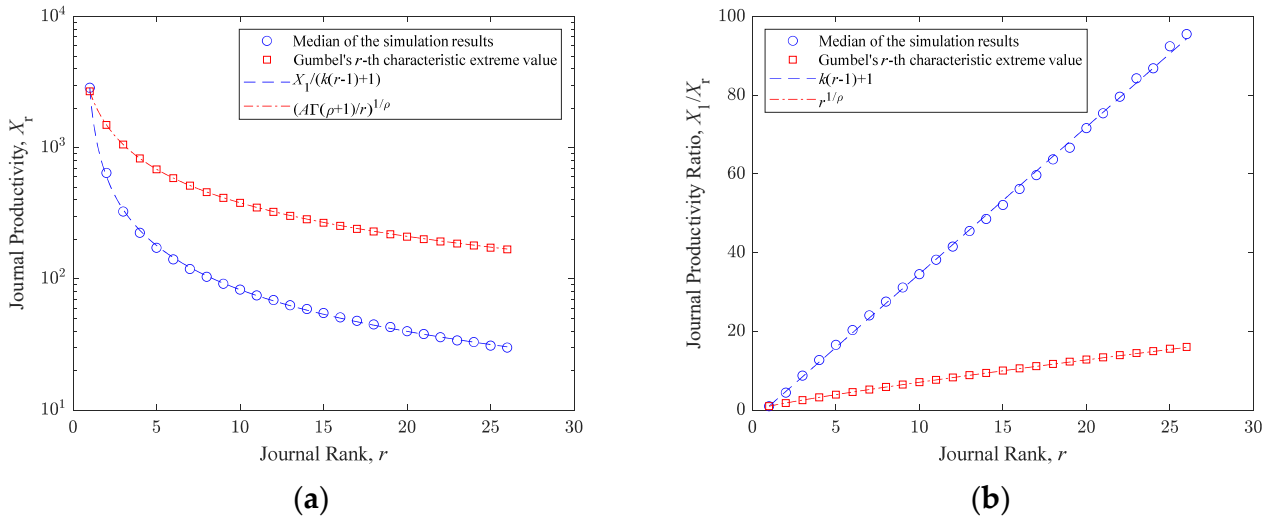


Figure 2. Journal productivity in the core region X_r as a function of journal rank r : (a) journal productivity X_r as a function of r ; (b) journal productivity ratio X_1/X_r as a function of r .

If there are T_0 journals with A_0 papers in the core region and the numbers of T_0 and A_0 are known, then the parameter k can be calculated from the equation $R_c(T_0) = A_0$. Then, Equation (8) can be used to predict the evolution of the core regions ($r \leq T_0$) of Bradford curves.

2.2. Leimkuhler’s Function

After removing the T_0 journals and the A_0 papers of the core region, the remaining T_1 journals and A_1 papers align well with the theoretical results predicted by Equation (1). Consequently, they follow Lotka’s law, and their Bradford curve can be predicted using the revised Leimkuhler’s function [20]:

$$R(r_1) = a \log(1 + br_1) \tag{9}$$

where the key parameters a and b are defined as:

$$a = \frac{A_1}{\log(e^\gamma y_m)} \tag{10}$$

$$b = \frac{e^\gamma y_m - 1}{T_1} \tag{11}$$

where γ is the Euler–Mascheroni constant, $\gamma \approx 0.5772$, and y_m is the journal productivity when the corresponding theoretical journal number $f_t(y_m) \approx 1$. While y_m can be directly calculated from Equation (2), it can also be estimated using the following equation if the values of X_1 , T_0 and A_0 are known:

$$y_m \approx \frac{X_1}{k(T_0 - 1) + 1} \tag{12}$$

Since the core region’s journal productivity is higher than the normal region’s, these significant journals rank lower. Consequently, the Bradford curve for the normal region

starts at the point (T_0, A_0) . Each rank r_1 in the normal region should be transformed to $r = r_1 + T_0$, and the cumulative productivity of the first r journals $R(r)$ should be transformed into $R_n(r) = R(r_1) + A_0$. The revised Leimkuhler’s function for the normal region is then written as:

$$R_n(r) = R(r - T_0) + A_0 = a \log[1 + b(r - T_0)] + A_0 \tag{13}$$

Equation (13) can be used to predict the dynamic evolution of the normal regions $(T_0 < r < T)$ of the Bradford curve. Therefore, Equations (8) and (13) together can be used to predict the dynamic evolution of the Bradford curves.

Figure 3a displays the Bradford curve. The blue circles represent the normal zone, while the red squares indicate the core zone. The blue dashed lines show the prediction results of Equation (13), and the red dotted lines show the prediction results of Equation (8). The black upper triangle, the diamond, and the lower triangle represent the points $(1, X_1)$, (T_0, A_0) , and the (T, A) , respectively. From the discussion above, it is evident that these three points and two lines are crucial for predicting the evolution of the Bradford curve.

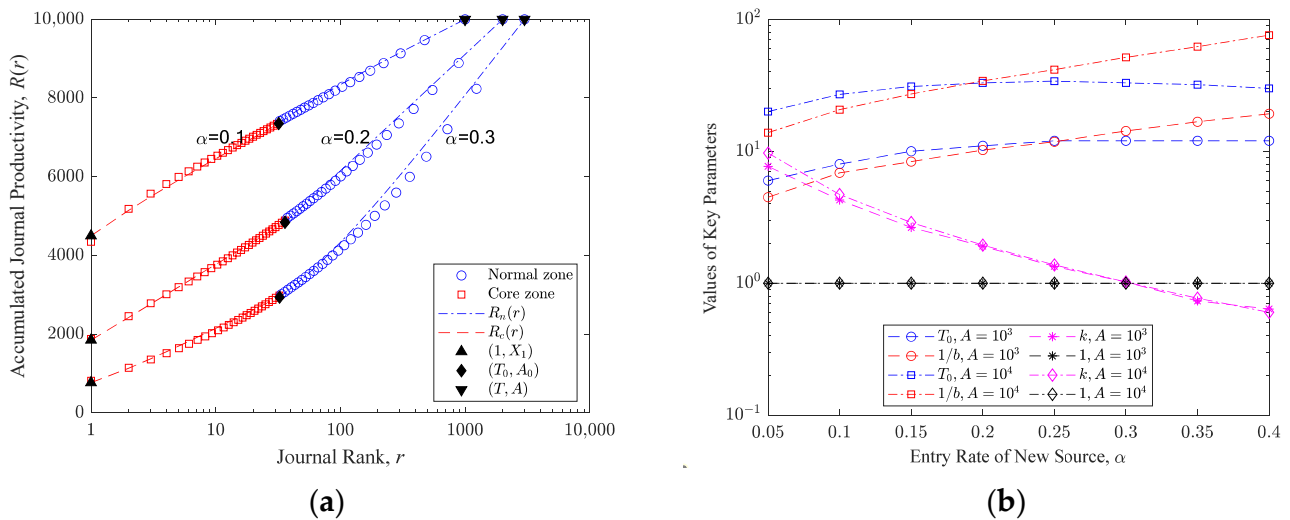


Figure 3. The evolution of the Bradford curves and the cause of the Groos droop: (a) the evolution of the Bradford curves; (b) the cause of the Groos droop.

2.3. Groos Droop

Groos was the first to observe that in some datasets, when the journal productivity is low, the Bradford curve tends to bend downwards [10]. Egghe [22] explained the cause of the Groos droop as merging datasets. However, this section demonstrates that the core region’s existence causes the Groos droop in the normal region.

The first and second derivatives of $R_c(r)$ for the core region can be derived from Equation (8):

$$\frac{\partial R_c(r)}{\partial(\log r)} = \frac{X_1 r}{k(r - 1) + 1} \tag{14}$$

$$\frac{\partial^2 R_c(r)}{\partial(\log r)^2} = \frac{X_1(1 - k)r}{[k(r - 1) + 1]^2} \tag{15}$$

From Equation (15), it can be noted that when $k > 1$, $\frac{\partial^2 R_c(r)}{\partial(\log r)^2} < 0$, the Bradford curve for the core region concaves downwards. Conversely, when $k < 1$, $\frac{\partial^2 R_c(r)}{\partial(\log r)^2} > 0$, the curve concaves upwards. As the entry rate of new journals α increases, the number of journals T rises, and the distribution of articles become more dispersed, leading to a decrease in the largest journal productivity X_1 . From Equation (8) and $R_c(T_0) = A_0$, we know that a lower

X_1 results in a lower k if A_0 is relatively constant. Therefore, as α increases, the Bradford curve for the core region will gradually concave upwards, as shown in Figure 3.

Similarly, the first and second derivatives of $R_n(r)$ for the normal region can be derived from Equation (13):

$$\frac{\partial R_n(r)}{\partial(\log r)} = \frac{abr}{b(r - T_0) + 1} \quad (16)$$

$$\frac{\partial^2 R_n(r)}{\partial(\log r)^2} = \frac{ab(1 - bT_0)r}{[b(r - T_0) + 1]^2} \quad (17)$$

From Equation (17), it can be noted that when $T_0 > 1/b$, $\frac{\partial^2 R_c(r)}{\partial(\log r)^2} < 0$, the Bradford curve for the normal region will concave downwards, showing a Groos droop. When $T_0 < 1/b$, $\frac{\partial^2 R_c(r)}{\partial(\log r)^2} > 0$, the curve will concave upwards, forming a J-shaped curve. As the entry rate of new journals α increases, Figure 3a shows that the T_0 will eventually fall below $1/b$, causing the Bradford curve for the normal region to concave upwards, similar to the core region.

Figure 3b illustrates the variation of key parameters T_0 , $1/b$ and k with the entry rate α . When $A = 10^4$, the normal region will start to concave upwards at critical point $\alpha_n \approx 0.2$, while the core region will do so at $\alpha_c \approx 0.3$. Therefore, when $\alpha < 0.2$, the entire Bradford curve will concave downwards; when $\alpha > 0.3$, it will concave upwards. For $0.2 < \alpha < 0.3$, the Bradford curve will exhibit a reversed S-shape, with the core region concaving downwards and the normal region concaving upwards. Figure 3a shows the three shapes of Bradford curves. In this specific case, since $\alpha_n < \alpha_c$, there is no S-shaped Bradford curve. This is because the aging of journals is not considered here, making the largest journal productivity X_1 relatively large. When aging effects are considered, which will be discussed in detail in Section 3.2, X_1 decreases significantly, which results in a lower k and thus a much lower α_c . If $\alpha_c < \alpha_n$, an S-shaped Bradford curve will appear for $\alpha_c < \alpha < \alpha_n$, with the core region concaving upwards and the normal region concaving downwards.

2.4. Bradford Dynamics

Given the analytical expression of T_0 , A_0 and X_1 (Equations (3), (4) and (6)), the core region of the Bradford curve at any time can be predicted using Equation (8). The parameters for the normal regions can be derived from these factors through $T_1 = T - T_0$, $A_1 = A - A_0$ and Equation (12), allowing the normal region of the Bradford curve to be predicted using Equation (13).

The Bradford curves for the constant entry rate scenario are shown in Figure 4a. Here, the red squares and blue circles represent the simulation results for the core and the normal regions, respectively, while the red dashed lines and the blue dotted lines represent the theoretical results of Equations (8) and (13). The key points $(1, X_1)$, (T_0, A_0) , and (T, A) are also shown as the black upper triangle, the diamond and the lower triangle in Figure 4a. It is notable that although the core region contains far fewer journals than the normal region, its representation is significant due to the x -axis's log scale. Consequently, journals with lower ranks are better represented in Figure 4a.

Figure 3 shows that when $\alpha = 0.15$, the entire Bradford curve concaves downwards for $10^3 < A < 10^4$. Figure 4a depicts the evolution of Bradford curves as the paper number A increases from 10^3 to 10^4 , aligning well with theoretical predictions.

Figure 4b presents the simulation and the analytical results for T_0 , A_0 and X_1 . Hollow symbols represent the simulation results, while solid symbols denote the analytical ones. It can be observed that these three key factors are linear functions of the paper number A , as indicated by Equation (18):

$$\log(Y) = a_\rho + b_\rho \log(A) \quad (18)$$

where the constants a_ρ and b_ρ are functions of ρ . Since ρ is approximately one, a_ρ and b_ρ can be considered constants. The close match between the analytical and the numerical results confirms the validity of Equations (3), (4) and (6). The analytical result for T_0 is slightly lower than the numerical ones because, in the numerical results, all journals with productivity n that satisfy $f(n) = 1$ are considered part of the core region; whereas, in theory, some of them belong to the normal region.

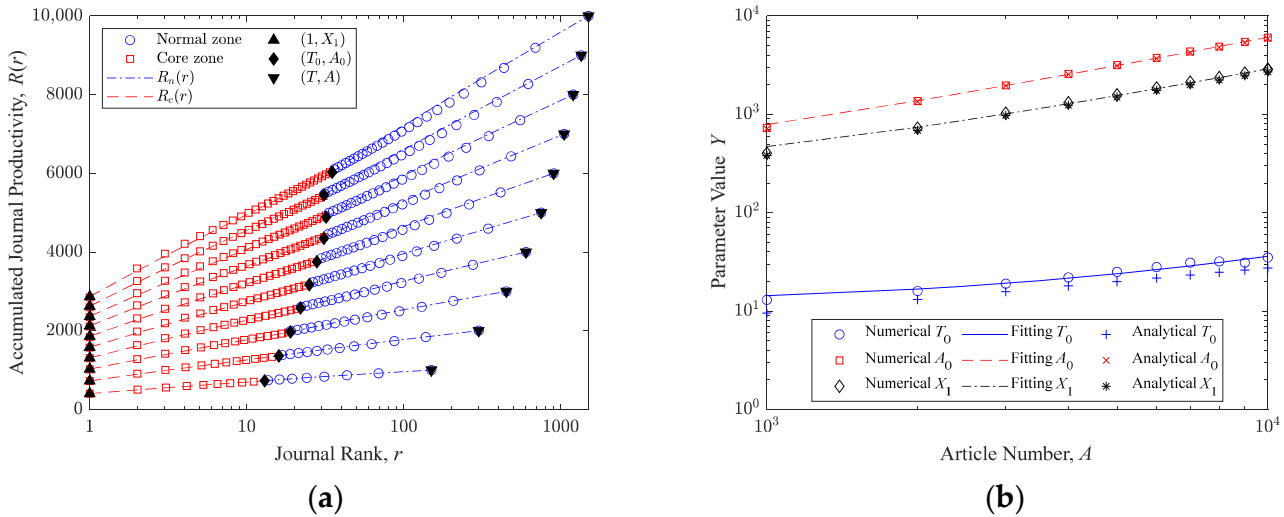


Figure 4. The dynamics of the Bradford curves and the variation of key parameters when $\alpha = 0.15$: (a) the dynamics of the Bradford curves; (b) the variation of key parameters.

3. Numerical Study

3.1. Decreasing Entry Rate

Assume the probability of adding a new journal decreases linearly over the time

$$\alpha(t) = \alpha_s - kt \tag{19}$$

where k is a constant, $k = (\alpha_s - \alpha_f) / A_f$, where α_f and A_f are the entry rate of new journals and the total number of articles in the final state, respectively. The accumulated number of journals can then be expressed as:

$$T = \sum_{t=1}^A \alpha(t) = \alpha_s A - \frac{1}{2} k A^2 \tag{20}$$

Using Equation (20), a quadratic fitting of T and A allows us to determine the values of α_s and α_f . Once these are known, the average entry rate $\bar{\alpha} = (\alpha_s + \alpha_f) / 2$ can be used to calculate the analytical results.

Figure 5 shows the dynamic evolution of Bradford curves and the variations of key parameters when the entry rate decreases linearly from 0.2 to 0.1. The proposed method effectively predicts these variations, with analytical results using $\bar{\alpha}$ matching well with the simulation results. Although the numerical results for A_0 and X_1 are slightly lower than the analytical ones, this suggests that a decreasing entry rate has a slight negative impact on their increase. However, the effect on the overall shape of the Bradford curve and key parameters is relatively insignificant, indicating that the analytical results for a constant entry rate $\bar{\alpha}$ can still be used to predict key parameters without significant errors.

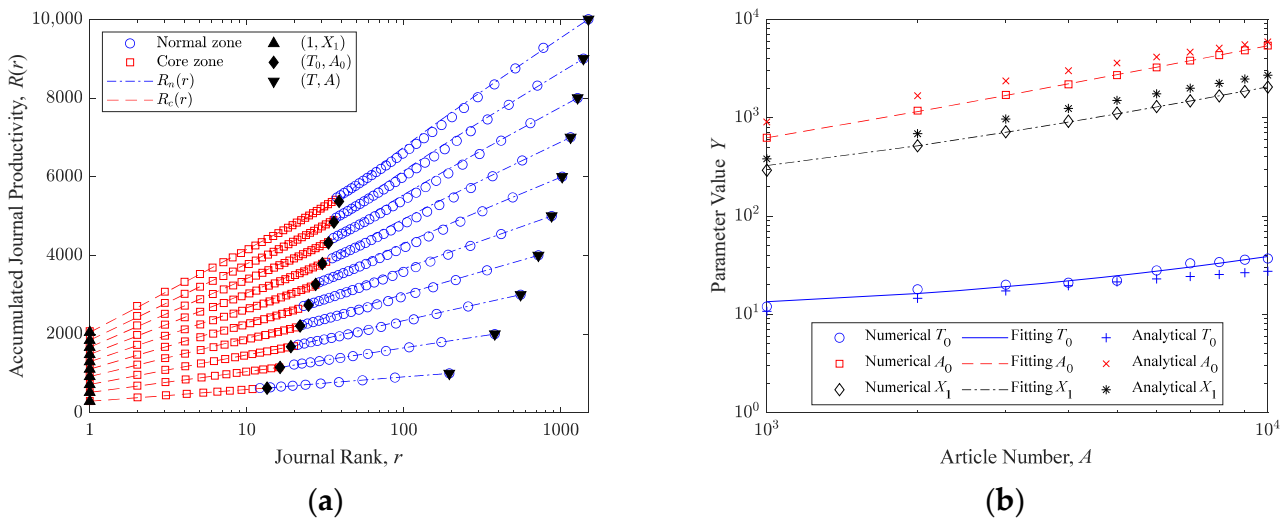


Figure 5. The dynamics of the Bradford curves and the variation of key parameters when α decreases linearly from 0.2 to 0.1: (a) the dynamics of the Bradford curves; (b) the variation of key parameters.

3.2. Aging Rate of Journals

Simon assumes that only one paper gets published in each time period and models the probability of a journal increasing in paper number as proportional to a weighted sum of its past increments. These increments are weighted by a factor that decreases geometrically over time, with the rate of decrease denoted as γ .

Let $y_j(k)$ represent the change in the paper number of the j -th journal during the k -th time interval, where $y_j(k)$ is either 1 (indicating a unit increment) or 0 (indicating no change). The paper number of the j -th journal at the end of the k -th interval is given by $\sum_{\tau=1}^k y_j(\tau)$. The expected increment in the paper number during the $(k + 1)$ -th interval is:

$$p[y_j(k + 1) = 1] = \frac{1}{W_k} \sum_{\tau=1}^k y_j(\tau) \gamma^{k-\tau} \tag{21}$$

where W_k is a time-dependent function consistent across all journals, defined as $W_k = \sum_{j=1}^T w_j(k)$ with $w_j(k) = \sum_{\tau=1}^k y_j(\tau) \gamma^{k-\tau}$. The parameter γ determines how quickly the influence of past growth is diminished and is thus referred to as the aging rate of journals in this paper.

Figure 6 illustrates the impact of the aging rate of journals γ on the dynamics of Bradford curves and key parameters. It shows that the aging factor increases T_0 , making the normal region concave downwards more. The aging effect also markedly reduces X_1 by weakening the Mathew effect, where successful journals attract more papers. As older journals lose their appeal, this “success breeds success” effect diminishes, leading to a substantial decrease in X_1 , as shown in Figure 6a. Consequently, while the number of articles A_0 in the core zone remains relatively unchanged, T_0 must increase to offset the reduction in X_1 . This expansion of the core region increases its share of the Bradford curve, shaping it into a J-shape due to the reduction of k and X_1 , as discussed in Section 2.3. Additionally, with the increase in T_0 , it is more likely that T_0 will exceed $1/b$, further contributing to the concave downwards shape of the normal region. In summary, the aging effect of journals facilitates the Bradford curve to more easily adopt an S-shape.

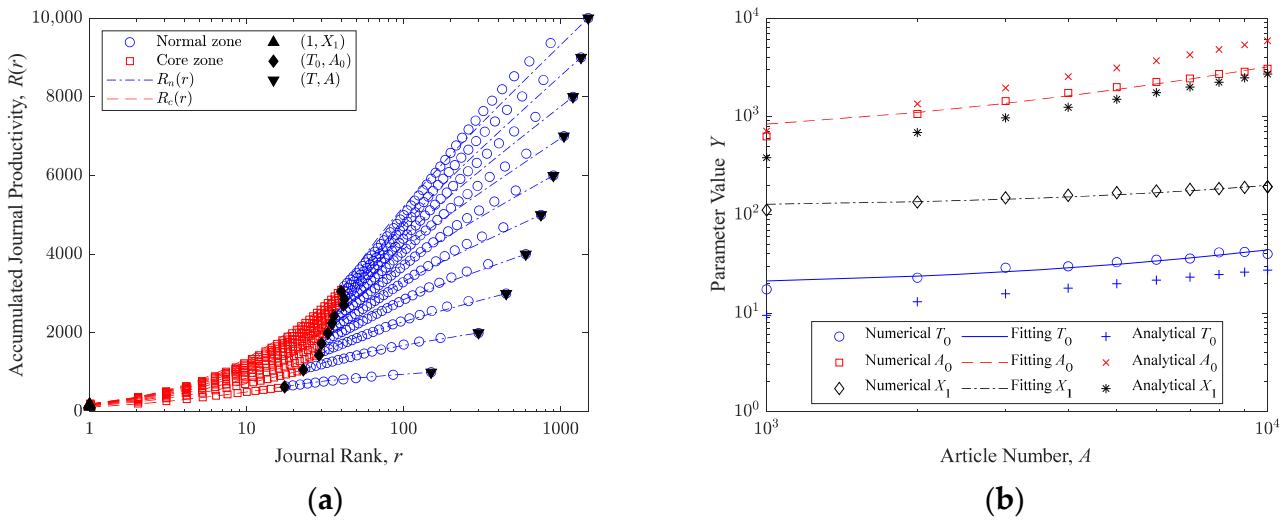


Figure 6. The dynamics of the Bradford curves and the variation of key parameters when $\alpha = 0.15$ and $\gamma = 0.95$: (a) the dynamics of the Bradford curves; (b) the variation of key parameters.

3.3. Varying Entry and Aging Rates

Figure 7 shows the impact of varying entry and aging rates on Bradford curves. In real-world scenarios, both the entry rate and the aging rate often change steadily. For example, the entry rate might decrease linearly from 0.2 to 0.1, while the aging rate increases linearly from 0.95 to 1.0, and the simulation results are shown in Figure 7. Comparing Figures 5–7 reveals that the effects of decreasing entry rate and increasing aging rate are similar to those observed with constant entry and aging rates (Figure 6). However, in Figure 7b, both the article number A_0 and journal number T_0 are even lower compared to Figure 6b, indicating that the decreasing entry rate further exacerbates their decrease. Consequently, the normal region of Bradford curve becomes less concave downwards, and its starting point on the y -axis is notably lower. Importantly, all three key factors continue to show linear relationships with the article number, suggesting that Equation (18) remains useful for predicting them.

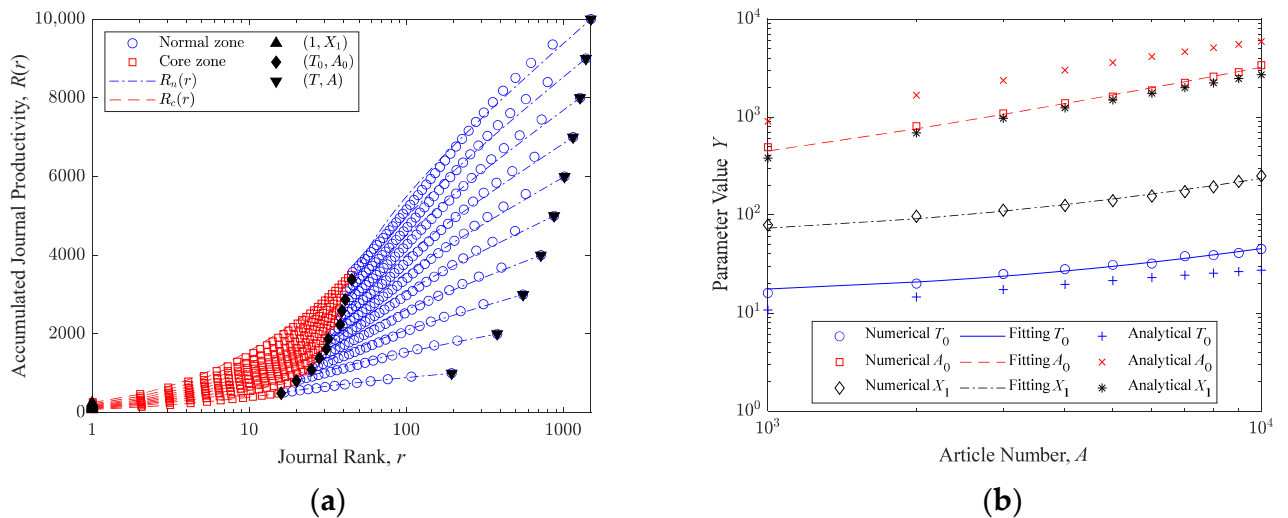


Figure 7. The dynamics of the Bradford curves and the variation of key parameters when α decreases linearly from 0.2 to 0.1 and γ increases linearly from 0.95 to 1.0: (a) the dynamics of the Bradford curves; (b) the variation of key parameters.

4. Empirical Study

4.1. Dataset of Croatian Chemistry Research

Oluić-Vuković used the research output in chemistry by authors from Croatia to prepare full bibliographic references for a ten-year period [28]. This dataset includes only articles published in journals, comprising 2543 papers across 416 journals over a decade. The productivity of the top few (fewer than 10) most prolific journals was taken directly from Figure 1 in [28], while the productivity of other journals was taken from Tables 4 and 6 in [25].

In these tables, the journal productivity n and the number of journals $f(n)$ are recorded for nine cumulative time periods. For journals with higher productivity (where n correspond to $f(n) = 1$), Oluić-Vuković grouped the data into ranges (e.g., $n \geq 77$), which limits the precision of the distribution for high-producing journals. To accurately represent this higher productivity range, we relied on the detailed Bradford curves provided in [28], adjusting the data to match the total number of journals T and articles A as indicated in Table 2 of [28].

As an example, using the data from the ninth interval (also found in Table 1 of reference [27]), we identified the highest productivity level $X_1 = 301$ by observing the last value in the sequence of n . The dividing point between the core and the normal zones was set at $n = 24$ based on the fact that for $n < 24$, most values of $f(n)$ were greater than 1, and n values were continuous; while for $n > 24$, $f(n)$ showed irregular gaps with isolated high n values and lower average journal counts.

The journal count T_0 and the article count A_0 for the core zone were calculated by summing the respective $f(n)$ and $n \cdot f(n)$ for $n \geq 24$. Similarly, the total number of journals T and articles A were obtained by summing the values across all n , ensuring consistency with the overall dataset.

To predict the dynamics of the Bradford curve, the process begins by predicting the variation of the total article number $A(t)$ over time t . Logistic regression analysis [41] was applied to the empirical data to predict the total number of articles $A(t)$ at any given time, as shown in Figure 8a. Logistic regression was applied using MATLAB to model the growth of the article output based on empirical data, providing a reliable prediction of future article numbers.

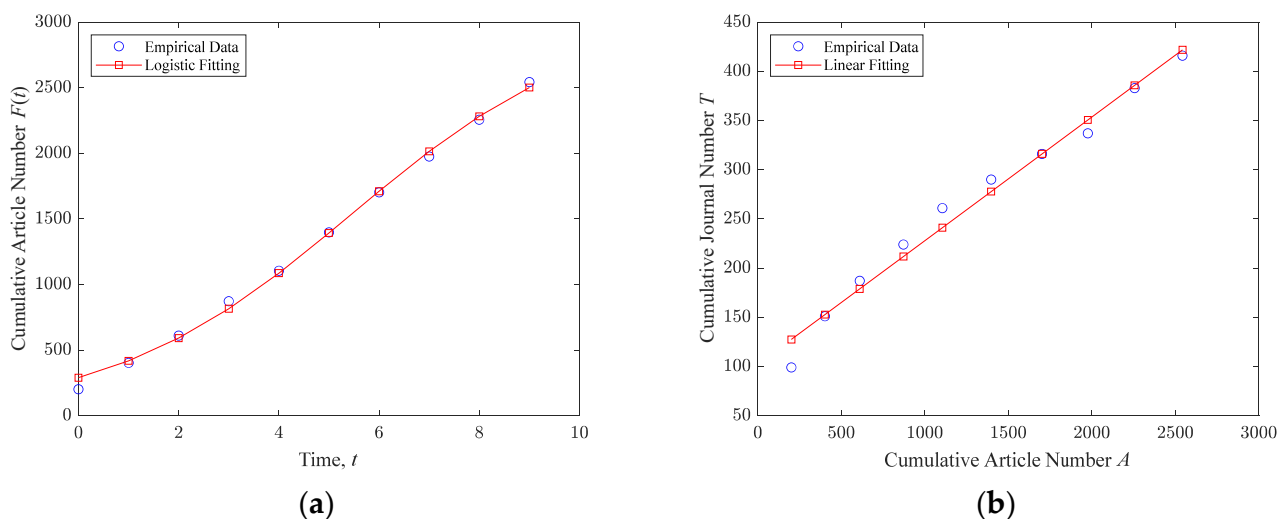


Figure 8. The process of determining the point (T, A) for any given time: (a) the total article number $A(t)$ as a function of time t ; (b) the total journal number T as a function of the article number A .

Next, the total journal number T and the entry rate of new journals α were estimated by plotting the total journal number T against the total article number A and a linear fit of $T = A\alpha$ was applied, as shown in Figure 8b. The linear fit, with an R-squared value of

0.98, confirmed a strong relationship between T and A , providing insights into how new journals enter the system as article output increases.

Once the point (T, A) is determined for any time, linear regression of Equation (18) is used to determine the three key parameters T_0 , A_0 , and X_1 on a log-log axis, with the fitting results shown in Figure 9a. These parameters reveal the structure of the core and the normal regions of the Bradford distribution.

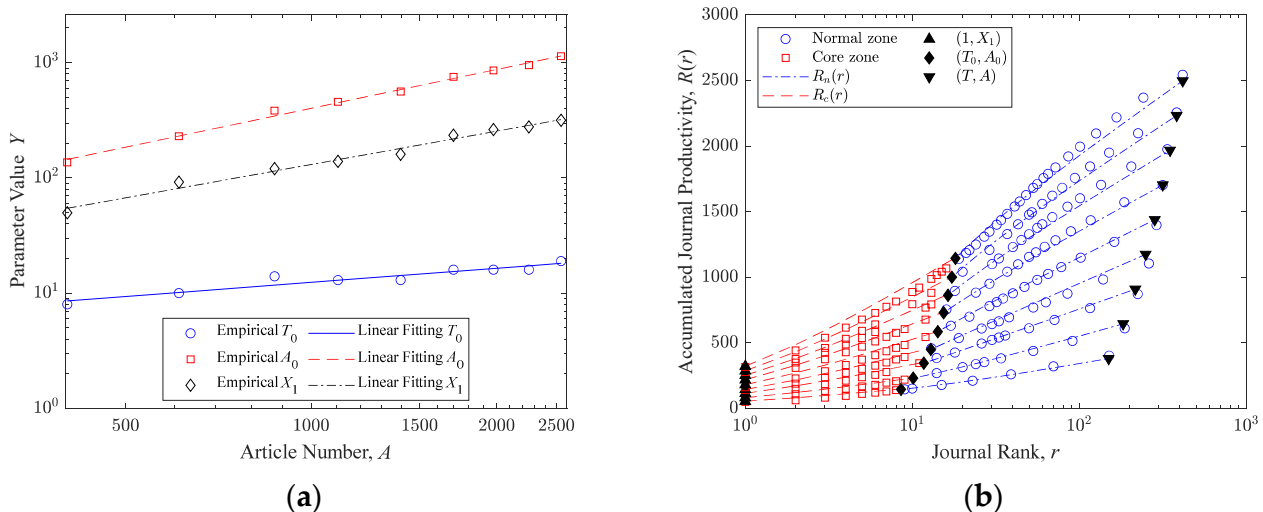


Figure 9. The procedures for predicting the evolution of the Bradford curves: (a) the variation of key parameters T_0 , A_0 , and X_1 with the article number A ; (b) the dynamics of the Bradford curves.

Finally, based on the three key points (T_0, A_0) , $(1, X_1)$, and (T, A) , Equations (8) and (13) were used to model the Bradford curve for the core region and the normal region, respectively. The results, illustrated as red dashed lines and blue dotted lines in Figure 9b, show the gradual transition in the curve's shape. This approach allows librarians and practitioners to model journal productivity trends, helping to optimize collection development strategies and predict journal coverage effectively.

The Bradford curves shown in Figure 9b align closely with the empirical data, demonstrating a strong match between the predicted and the observed outcomes. Notably, the Bradford curve transitions gradually from a J-shape to an S-shape, a transformation that is accurately captured by the analytical predictions.

4.2. Dataset of Solar Power Research

The bibliographies on solar power research for the years 1971, 1974, 1977, 1980, 1983, and 1986 were compiled by Garg et al. [29], encompassing papers published in journals from the Engineering Index. The data for this analysis were directly extracted from Tables 1–7 of the referenced study. Unlike the Croatian Chemistry dataset, Garg's dataset is clearer, as it records every journal's exact productivity n without grouping high-productivity journals.

The dataset is divided into six cumulative periods: 1971, 1971 + 1974, 1971 + 1974 + 1977, 1971 + 1974 + 1977 + 1980, 1971 + 1974 + 1977 + 1980 + 1983, and 1971 + 1974 + 1977 + 1980 + 1983 + 1986. Although the data skips certain years (e.g., 1972 and 1973), this does not affect the Bradford curve calculation. However, when predicting article growth, care must be taken to sum the predicted outputs for each year to match the cumulative totals.

The data processing followed the same four-step method as the Croatian Chemistry dataset. First, logistic regression was applied to predict the cumulative article numbers $A(t)$ over time t , with data from Table 7 used to estimate the article numbers for each desired interval, as shown in Figure 10a. This step involves fitting the empirical data and adjusting for the cumulative nature of the dataset, ensuring accurate predictions for the individual years within each period.

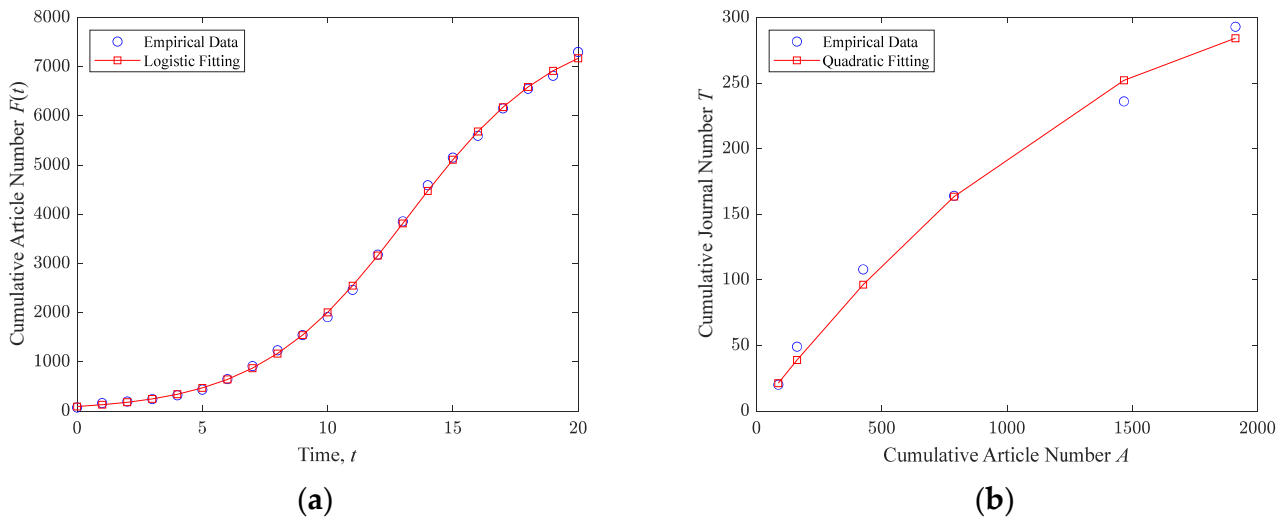


Figure 10. The process of determining the point (T, A) for any given time: (a) the total article number $A(t)$ as a function of time t ; (b) the total journal number T as a function of the article number A .

Next, quadratic fitting (Equation (20)) was applied to the journal and article pairs to estimate the total number of journals T and the entry rate of new journals α , as shown in Figure 10b. Logistic fitting was applied to better capture the effect of a linearly decreasing journal entry rate, providing a more accurate representation of how new journals enter the system as article output increases.

The key parameters T_0 , A_0 , and X_1 were then predicted using linear fitting on the log-log axis, based on the empirical data from the earlier steps. This method, depicted in Figure 11a, was used to determine these parameters at each cumulative time point, allowing the segmentation of the core and the normal regions of the Bradford curve.

Finally, using Equations (8) and (13), Bradford curves for the core and the normal regions were plotted, as shown in Figure 11b. As with the Croatian Chemistry dataset, determining the highest productivity level X_1 , the core zone journal count T_0 , and the core zone article count A_0 was straightforward in Garg’s dataset due to the clear distribution of journals and articles. This simplicity contrasted with the Croatian dataset, where grouping high-productivity journals added complexity.

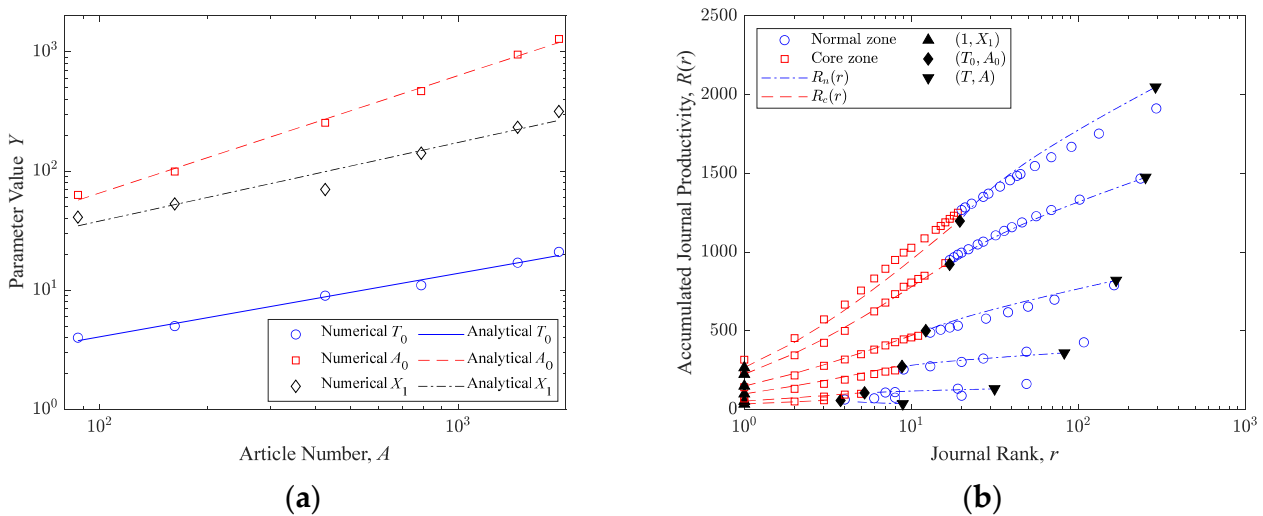


Figure 11. The procedures for predicting the evolution of the Bradford curves: (a) the variation of key parameters T_0 , A_0 , and X_1 with the article number A ; (b) the dynamics of the Bradford curves.

Figures 9b and 11b demonstrate that while the proposed method can predict the general trend of the Bradford curves, there are inherent errors. These errors arise because Bradford's law inherently contains uncertainties. Numerical studies reveal that the article numbers for each journal rank have a large standard deviation, making it practically impossible to predict the precise shape of the Bradford curve. Additionally, the various fitting procedures introduce errors into the process. Therefore, this method can only predict the general trend of Bradford dynamics but cannot accurately predict the article number for each journal at any given time.

Another issue with this method is that the first derivatives of the core region (Equation (14)) and the normal region (Equation (16)) differ at the (T_0, A_0) point, resulting in a non-smooth analytical curve at the intersection point. In contrast, the numerical simulation results are smooth throughout. This problem could be addressed by proposing more complex formulas for the normal region, but this would complicate the overall method. Given the difficulty in accurately predicting Bradford curve dynamics, this aspect is not explored further in this paper.

5. Discussion

5.1. Theoretical Contributions

This study makes contributions in two areas: advancing bibliometric modeling and simulation and addressing extreme value phenomena. Both are critical to understanding the dynamics of scientific knowledge production.

1. Modeling and Simulation of Science or Publications

Modeling and simulations in bibliometrics help uncover micro-level behaviors that drive broader trends, such as publication growth and shifts in core literature [42]. These methods offer controlled environments to reduce data biases and explore extreme scenarios, providing valuable insights for decision-making. While fields like network science, agent-based modeling (ABM), and diffusion models have been widely used in science modeling and simulation [43], the contributions of bibliometric research in these areas have not been fully explored. This study bridges that gap by incorporating models like Bradford's law into the broader science dynamics framework, showing their value in understanding knowledge production over time.

2. Extreme Value and Non-Lotkaian informetrics

This study addresses extreme value phenomena—highly productive authors, highly cited papers, and influential journals—that traditional models like Lotkaian informetrics struggle to capture [44,45]. These extreme cases deviate from typical productivity patterns, necessitating new theoretical approaches. By applying the extreme value theory, order statistics, and the Simon–Yule model, this research provides a more accurate model for the core zone of Bradford curves. These rare but influential entities shape scientific output in ways that Lotkaian informetrics cannot fully explain. To address this, the study introduces a new framework—non-Lotkaian informetrics—which better accounts for extreme values and extends beyond bibliometrics to areas like patent analysis and network studies.

5.2. Practical Applications

Despite bundled licensing agreements, Bradford's law helps academic libraries evaluate their coverage rate in specific disciplines, offering a quantitative framework to estimate the relationship between acquired and available journals for informed decision-making.

1. Journal Coverage Evaluation and Resource Optimization

Bradford's law is a key tool for evaluating journal collection coverage and optimizing resources in academic libraries. By using the Bradford curve, libraries can assess how well their subscriptions cover a specific discipline, focusing on high-impact journals like those in SCI or Scopus while deprioritizing less essential ones such as predatory journals. This method helps libraries understand their coverage rate in particular fields, ensuring

comprehensive collections by focusing on core journals and guiding effective weeding strategies to retire low-usage ones [46–48].

Additionally, Bradford's law supports resource optimization by helping libraries periodically evaluate journal and database subscriptions, even when subscriptions appear stable year-to-year. This assessment not only aids in making informed decisions about renewals and expansions, ensuring continued coverage of core journals as disciplines evolve, but also provides valuable insights for developing and refining Collection Development Policies (CDP). By using Bradford's law as a reference, libraries can better balance budget efficiency with evolving research needs, ensuring that their collections remain comprehensive and aligned with institutional priorities [49,50].

2. Broader Applications in Academic Libraries

Bradford's law provides a versatile framework for assessing academic impact by classifying papers based on citation counts and authors by their h-index. It ranks papers into zones of influence, similar to how top-tier journals are grouped in JCR rankings, helping researchers identify the most influential papers in their field [51]. The same principle applies to authors, where those with higher h-indices are placed in the core zone, offering insights into both productivity and influence [52]. This broader application not only enhances support for impactful research but also strengthens the academic ecosystem by providing valuable insights into scholarly influence.

6. Conclusions

This paper examines how integer constraints on the number of journals T and articles A affect the shape of Bradford curve, dividing it into two distinct zones: the core zone and the normal zone, based on the significance of these integer effects. Using the Simon–Yule model, we derive analytical results for key parameters and distributions under a constant entry rate. Theoretical formulas for each zone are developed, and the reasons behind the various shapes of Bradford curves are analyzed. Monte Carlo simulations are employed to study the impact of decreasing entry rates of new journals and aging rates of journals on the shape of the Bradford curve and the key parameters. Finally, we validate our proposed method using empirical data from the Croatian Chemistry and the Solar Power research datasets. The main conclusions are:

1. Bradford curves should be divided into two separate zones based on the significance of integer constraints on journal and article numbers. Different formulas for each zone should be derived separately;
2. Bradford curves can exhibit four different shapes, determined by the second derivatives of the core and the normal zones;
3. The largest productivity X_1 , the number of journals T_0 , and the number of articles A_0 are key parameters influencing the shapes of Bradford curves. Decreasing entry rates and aging rates of journals affect these parameters.
4. The proposed four-step method can predict general trends in Bradford curves despite some errors.
5. Bradford's law provides a valuable framework for academic libraries to evaluate journal coverage, optimize resource allocation, and refine Collection Development Policies (CDP), ensuring comprehensive and well-balanced collections as research needs evolve.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/publications12040036/s1>, file: Data and Code.

Funding: This research was supported by two projects: the 2024 Shenzhen Library and Information Science Research Project, "Temporal Evolution of Bradford's Law in the Context of Library Professionalization" (Project No. SWTQ2024261), and the 2023 Guangdong Provincial Library Key Research Project, "Joint Analysis and Data Governance of Papers and Patents in the Context of Smart Libraries" (Project No. GDTK23004).

Data Availability Statement: All relevant data and code used in this study are provided in the Supplementary Materials, organized under the “Data and Code” folder. The code, developed in MATLAB, is available to support the reproducibility of the analysis. Due to the complexity and organization of the code, we recommend that researchers utilize tools such as AI-assisted analyzers (e.g., ChatGPT) to facilitate the interpretation of the code. We encourage other researchers to replicate the findings, and for any further inquiries or clarification, the corresponding author can be contacted directly.

Conflicts of Interest: The author declares no conflicts of interest.

References

1. Singh, C.K.; Barme, E.; Ward, R.; Tupikina, L.; Santolini, M. Quantifying the rise and fall of scientific fields. *PLoS ONE* **2022**, *17*, e0270131. [[CrossRef](#)] [[PubMed](#)]
2. Golosovsky, M. *Citation Analysis and Dynamics of Citation Networks*; Springer: Cham, Switzerland, 2019.
3. Sun, X.; Kaur, J.; Milojević, S.; Flammini, A.; Menczer, F. Social dynamics of science. *Sci. Rep.* **2013**, *3*, 1069. [[CrossRef](#)] [[PubMed](#)]
4. Leydesdorff, L. *The Evolutionary Dynamics of Discursive Knowledge: Communication-Theoretical Perspectives on An Empirical Philosophy of Science*; Springer Nature: Cham, Switzerland, 2021.
5. Sjögarde, P. *Mapping the Structure of Science through Clustering in Citation Networks: Granularity, Labeling and Visualization*; Karolinska Institutet: Stockholm, Sweden, 2023.
6. Raimbault, J. Exploration of an interdisciplinary scientific landscape. *Scientometrics* **2019**, *119*, 617–641. [[CrossRef](#)]
7. Schneider, J.W.; Costas, R. Identifying potential “breakthrough” publications using refined citation analyses: Three related explorative approaches. *J. Assoc. Inf. Sci. Technol.* **2017**, *68*, 709–723. [[CrossRef](#)]
8. Manolopoulos, Y.; Vergoulis, T. *Predicting the Dynamics of Research Impact*; Springer: Cham, Switzerland, 2021.
9. Larivière, V.; Archambault, É.; Gingras, Y. Long-term variations in the aging of scientific literature: From exponential growth to steady-state science (1900–2004). *J. Am. Soc. Inf. Sci. Technol.* **2008**, *59*, 288–296. [[CrossRef](#)]
10. Groos, O.V. Brief Communications Bradford’s Law and the Keenan-Atherton Data. *Am. Doc. (Pre-1986)* **1967**, *18*, 46. [[CrossRef](#)]
11. Wagner-Döbler, R. Time dependencies of Bradford distributions: Structures of journal output in 20th-century logic and 19th-century mathematics. *Scientometrics* **1997**, *39*, 231–252. [[CrossRef](#)]
12. Chisaba Pereira, C.A. Data-driven model (DDM) for collection development and management: From library data to institutional value generation. *Qual. Quant. Methods Libr.* **2021**, *10*, 177–187.
13. Crawford, L.S.; Condrey, C.; Avery, E.F.; Enoch, T. Implementing a just-in-time collection development model in an academic library. *J. Acad. Librariansh.* **2020**, *46*, 102101. [[CrossRef](#)]
14. Voorbij, H.; Lemmen, A. Coverage of periodicals in national deposit libraries. *Ser. Rev.* **2007**, *33*, 40–44. [[CrossRef](#)]
15. Zaghmouri, L. The First of Its Kind: Collection Development Techniques for the Vasche Library’s Modern Assyrian Heritage Collection. *Collect. Manag.* **2023**, *48*, 48–55. [[CrossRef](#)]
16. Mwilongo, K.J.; Luambano, I.; Lwehabura, M.J.F. Collection development practices in academic libraries in Tanzania. *J. Librariansh. Inf. Sci.* **2020**, *52*, 1152–1168. [[CrossRef](#)]
17. Bradford, S.C. Sources of information on specific subjects. *Engineering* **1934**, *137*, 85–86.
18. Vickery, B.C. Bradford’s law of scattering. *J. Doc.* **1948**, *4*, 198–203. [[CrossRef](#)]
19. Leimkuhler, F.F. The Bradford distribution. *J. Doc.* **1967**, *23*, 197–207. [[CrossRef](#)]
20. Egghe, L. Applications of the theory of Bradford’s law to the calculation of Leimkuhler’s law and to the completion of bibliographies. *J. Am. Soc. Inf. Sci.* **1990**, *41*, 469–492. [[CrossRef](#)]
21. Qiu, L.; Tague, J. Complete or incomplete data sets. The Groos droop investigated. *Scientometrics* **1990**, *19*, 223–237. [[CrossRef](#)]
22. Egghe, L.; Rousseau, R. Reflections on a deflection: A note on different causes of the Groos droop. *Scientometrics* **1988**, *14*, 493–511. [[CrossRef](#)]
23. Chen, Y.-S.; Leimkuhler, F.F. Bradford’s law: An index approach. *Scientometrics* **1987**, *11*, 183–198. [[CrossRef](#)]
24. Egghe, L. Consequences of Lotka’s Law for the Law of Bradford. *J. Doc.* **1985**, *41*, 173–189. [[CrossRef](#)]
25. Oluić-Vuković, V. Simon’s generating mechanism: Consequences and their correspondence to empirical facts. *J. Am. Soc. Inf. Sci.* **1998**, *49*, 867–880. [[CrossRef](#)]
26. Oluić-Vuković, V. Impact of productivity increase on the distribution pattern of journals. *Scientometrics* **1989**, *17*, 97–109. [[CrossRef](#)]
27. Oluić-Vuković, V. The shape of the distribution curve: An indication of changes in the journal productivity distribution pattern. *J. Inf. Sci.* **1991**, *17*, 281–290. [[CrossRef](#)]
28. Oluić-Vuković, V. Journal productivity distribution: Quantitative study of dynamic behavior. *J. Am. Soc. Inf. Sci.* **1992**, *43*, 412–421. [[CrossRef](#)]
29. Garg, K.; Sharma, P.; Sharma, L. Bradford’s law in relation to the evolution of a field. A case study of solar power research. *Scientometrics* **1993**, *27*, 145–156. [[CrossRef](#)]
30. Sen, S.; Chatterjee, S. Bibliographic scattering and time: An empirical study through temporal partitioning of bibliographies. *Scientometrics* **1998**, *41*, 135–154. [[CrossRef](#)]

31. Oluić-Vuković, V. Bradford's distribution: From the classical bibliometric "law" to the more general stochastic models. *J. Am. Soc. Inf. Sci.* **1997**, *48*, 833–842. [[CrossRef](#)]
32. Simon, H.A. On a class of skew distribution functions. *Biometrika* **1955**, *42*, 425–440. [[CrossRef](#)]
33. Simon, H.A.; Van Wormer, T.A. Some Monte Carlo estimates of the Yule distribution. *Behav. Sci.* **1963**, *8*, 203–210. [[CrossRef](#)]
34. Ijiri, Y.; Simon, H.A. *Skew Distributions and the Sizes of Business Firms*; North-Holland: Amsterdam, The Netherlands, 1977; Volume 24.
35. Chen, Y.-S.; Chong, P.P.; Tong, M.Y. The Simon-Yule approach to bibliometric modeling. *Inf. Process. Manag.* **1994**, *30*, 535–556. [[CrossRef](#)]
36. Chen, Y.S.; Chong, P.P.; Tong, M.Y. Dynamic behavior of Bradford's law. *J. Am. Soc. Inf. Sci.* **1995**, *46*, 370–383. [[CrossRef](#)]
37. Chen, Y.-S. Analysis of Lotka's law: The Simon-Yule approach. *Inf. Process. Manag.* **1989**, *25*, 527–544. [[CrossRef](#)]
38. Glänzel, W. High-end performance or outlier? Evaluating the tail of scientometric distributions. *Scientometrics* **2013**, *97*, 13–23. [[CrossRef](#)]
39. Glänzel, W. The role of the h-index and the characteristic scores and scales in testing the tail properties of scientometric distributions. *Scientometrics* **2010**, *83*, 697–709. [[CrossRef](#)]
40. Gumbel, E.J. *Statistics of Extremes*; Echo Point Books & Media: Brattleboro, VT, USA, 1958.
41. Verhulst, P. Notice on the law that the population follows in its growth. *Corresp. Math. Phys.* **1838**, *10*, 113–126.
42. Gilbert, N. A simulation of the structure of academic science. *Sociol. Res. Online* **1997**, *2*, 91–105. [[CrossRef](#)]
43. Scharnhorst, A.; Börner, K.; Van den Besselaar, P. *Models of Science Dynamics: Encounters between Complexity Theory and Information Sciences*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2012.
44. Egghe, L. *Power Laws in the Information Production Process: Lotkaian Infometrics*; Academic Press: Cambridge, MA, USA, 2005.
45. Burrell, Q.L. Extending Lotkaian infometrics. *Inf. Process. Manag.* **2008**, *44*, 1794–1807. [[CrossRef](#)]
46. Goffman, W.; Morris, T.G. Bradford's law and library acquisitions. *Nature* **1970**, *226*, 922–923. [[CrossRef](#)]
47. Alabi, G. Bradford's law and its application. *Int. Libr. Rev.* **1979**, *11*, 151–158. [[CrossRef](#)]
48. Drott, M.C.; Mancall, J.C.; Griffith, B.C. Bradford's law and libraries: Present applications—Potential promise. *Aslib Proc.* **1979**, *31*, 296–304. [[CrossRef](#)]
49. Tague, J. What's the Use of Bibliometrics? 1988. Available online: <https://documentserver.uhasselt.be/bitstream/1942/846/1/tague271.pdf> (accessed on 12 October 2024).
50. Xu, F. A standard procedure for Bradford analysis and its application to the periodical literature in systems librarianship. *Libr. Hi Tech* **2011**, *29*, 751–763. [[CrossRef](#)]
51. Cline, G.S. Application of Bradford's Law to citation data. *Coll. Res. Libr.* **1981**, *42*, 53–61. [[CrossRef](#)]
52. Bornmann, L. How are excellent (highly cited) papers defined in bibliometrics? A quantitative analysis of the literature. *Res. Eval.* **2014**, *23*, 166–173. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.