

# 1 Time Reversed Delay Differential Equation Based Modeling Of 2 Journal Influence In An Emerging Area

3 **Abstract** A recent independent study resulted in a ranking system which ranked Astronomy and Computing  
4 (ASCOM) much higher than most of the older journals highlighting its niche prominence. We investigate  
5 the notable ascendancy in reputation of ASCOM by proposing a novel differential equation based modeling.  
6 The modeling is a consequence of knowledge discovery from big data-centric methods, namely L1-SVD.  
7 The inadequacy of the ranking method in explaining the reason behind the growth in reputation of ASCOM is  
8 reasonable to understand given that the study is post-facto. Thus, we propose a growth model by accounting  
9 for the behavior of parameters that contributes to the growth of a field. It is worthwhile to spend some  
10 time in analyzing the cause and control variables behind rapid rise in reputation of a journal in a niche area.  
11 We intend to identify and probe the parameters responsible for its growing influence. Delay differential  
12 equations are used to model the change of influence on a journals status by exploiting the effects of historical  
13 data. The manuscripts justifies the use of implicit control variables and models those accordingly that  
14 demonstrates certain behavior in the journal influence.

15 **Keywords:**  $l_1$ -norm, Sparsity Norm, Singular Value Decomposition, Journal Ranking, Astronomy and  
16 Computing, Big Data, Delay Differential Equations (DDE)

## 17 1. Introduction

18 It is well-known that ranking of journals, whether in science, technology, engineering or in social sci-  
19 ences, such as in economics, is a contentious issue. For many subjects, there is no correct ranking, but a  
20 universe of rankings, each a result of subjective criteria included by its creators. In this regard, the follow-  
21 ing studies are instructive: (Engemann and Wall (2009), Jangid, Saha, Gupta, and Rao (2014)). With the  
22 creators' choices and rules laid out explicitly, the users of such ranking still need to use own judgments  
23 and institutional requirements to choose ranks appropriately. The subjective element in journal rankings not  
24 only complicates matters about what is correct, if any, but also about outcomes that depend crucially on  
25 adoption and analysis of rankings. For science and related subjects, SCOPUS and SCIMAGO hold some  
26 of the best journal ranking systems to this day, using their Cite Score and SJR indicators respectively, to  
27 rank journals.(Kianifar, Sadeghi, and Zarifmahmoudi (2014)) However, owing to the manner in which both  
28 these indicators are considered, it is often the case that the received ranking might not always display the  
29 true quality and outreach of a specific scientific journal. Obviously, this could be true for a large number  
30 of subjects across length and breadth of contemporary research therefore recourse to a scientifically more  
31 acceptable method should always be of interest and often beneficial for a large set of users. To demonstrate

32 this, we therefore considered the case of the Journal entitled, *Astronomy and Computing*, within the context  
 33 of SCOPUS Journals in the relevant domain of AstroInformatics (Bora et al. (2016)) , in particular and  
 34 Astronomy and Astrophysics, in general.

35 The primary focus of this case study is to determine the standing of Journal *Astronomy and Computing*  
 36 with respect to other journals which were established prior to it. **More importantly, the reasons for such**  
 37 **standing need to be investigated which is a more complex and qualitative study.** The algorithm also  
 38 tests the validity of the ranking and suggests an alternative rank that used a more holistic approach towards  
 39 the features. While this paper focuses on a specific journal, it is easy to see that the purpose of this construct  
 40 is broad-based and deep-seated at the same time, such that the applications of the algorithms can be adopted  
 by numerous other subjects grappling with the same problem.

**Table 1. Case Study: Astronomy and Computing, SJR (González-Pereira et al. (2009)) and L1-SVD ranks (Aedula et al. (2018))**

Journal Name	L1 Scheme Rank	SJR based Rank	Year
Astronomy and Computing	39	31	2013
Astronomy and Astrophysics Review	40	5	1999
Radiophysics and Quantum Electronics	41	51	1969
Solar System Research	42	48	1999
Living Reviews in Solar Physics	43	3	2005
Astrophysical Bulletin	44	45	2010
Journal of Astrophysics and Astronomy	45	55	1999
Revista Mexicana de Astronomia y Astrofisica	46	23	1999
Acta Astronomica	47	20	1999
Journal of the Korean Astronomical Society	48	32	2009
Cosmic Research	49	58	1968
Geophysical and Astrophysical Fluid Dynamics	50	46	1999
New Astronomy Reviews	51	12	1999
Kinematics and Physics of Celestial Bodies	52	65	2009
Astronomy and Geophysics	53	67	1996
Chinese Astronomy and Astrophysics	54	72	1981

41

## 42 2. Motivation: The ranking scheme

43 We implemented  $l_1$ -norm SVD scheme using the publicly available SCOPUS dataset to rank all its  
 44 corresponding journals, and simultaneously determine the potency of the algorithm. The outcome of the  
 45 ranking scheme posed interesting and compelling questions which led us to model the growing influence of  
 46 the particular journal. We discuss the detailed method in Appendix A, for the simple reason that the focus of  
 47 the manuscript is not on the ranking methods, rather on the model formulation and interpretation explaining  
 48 such rank. SCOPUS contains approximately 46,000 Journals listed in different domains. Discarding few  
 49 redundancies, SCOPUS effectively covers a large range of metrics and provides adequate resources for  
 50 verification. For this demonstration, we have considered 7 different metrics from SCOPUS to be used as  
 51 features in our algorithm. These features include *Citation Count*, *Scholarly Output*, *SNIP*, *SJR*, *Cite Score*,  
 52 *Percentile and Percent Cited*.

53 Indeed, to cross-verify the results of the algorithm these were compared to SJR based ranking of  
 54 SCIMAGO for suitable articulation of the discrepancies. It seems that the  $l_1$ -norm SVD scheme works quite  
 55 successfully (Aedula et al. (2018)) in rating the journals and approaches the data in a more comprehensive  
 56 way. The result is a ranking system which ranks *Astronomy and Computing* much higher than most of the

57 older journals and at the same time highlights the niche prominence of the particular journal. Similarly,  
58 this method also highlights the rise of other journals which were underrepresented due to the usage of the  
59 SCOPUS and SCIMAGO indicators only. This method, therefore, has been largely successful in rectifying  
60 the rank of such journals. Importantly, the  $l_1$ -norm SVD scheme can be extrapolated to other data as well. It  
61 can be used to study the impact of individual articles, for example. Utilizing similar features such as Total  
62 Citation, Self Citation, and NLIQ (Ginde et al. (2016)), the algorithm can be used to rank articles within  
63 a journal with great accuracy along with a holistic coverage. To re-appraise the scope of this research, it is  
64 important to remember that the common practice (Engemann and Wall (2009)) has been to control for the  
65 size of the journal (measures like pages, number of articles, even characters), age of article, age of citation,  
66 reference intensity, exclusion of self-citations, etc.

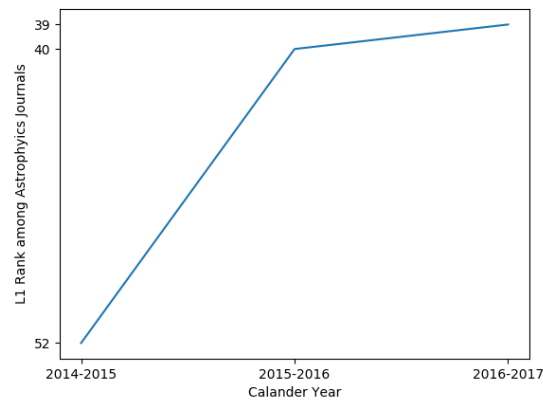
67 In order to be precise, the ranking scheme raises some important questions which can be reasonably  
68 challenging. Standard scientometric features used to study influence/reputation of journals are not adequate  
69 for explaining the ascendancy of ASCOM in influence. The importance of investigating intrinsic dynamics  
70 is rarely stressed upon in scientometric literature (Fei, Chong, and Bell (2015)). Usually, the analysis is  
71 static, based on citations and other factors. The authors intend to bring out the missing dynamics via the  
72 DDE based model. The following set of questions are addressed in this study. What are the non-quantitative  
73 factors (could be qualitative and difficult to quantify) explaining the rapid growth of this journal? What is  
74 the direction of causation, and how do we frame it? Does the big data landscape help? Can we formulate  
75 a model that reasonably accounts for such surge in influence? Are there features/factors, not statistically  
76 significant but play crucial roles as implicit control variables toward the phenomena? The proposed model  
77 (Section 4 onward) addresses these questions.

## 78 2.1. Knowledge Discovery and the Evolution of ASCOM: Key Motivation for the model

79 Albeit, Astronomy and Computing (ASCOM) has been in publication for five years only, its reputation  
80 has grown quickly as can be observed from the ranking system proposed here. This is despite the fact that  
81 ASCOM is severely handicapped in size. There is no journal focused on the interface of astronomy and  
82 computing in the same way as ASCOM. It can be observed from Table 1 that, ASCOM, in comparison with  
83 the other journals listed, is significantly younger! Unless the number of volumes and issues published are  
84 significant, a journal is unlikely to create the equivalent impact of an established journal. This is a notable  
85 handicap for any new journal, ASCOM being no exception. We define this as "size handicap".

86 Despite the "size handicap" explained above, ASCOM is ranked 39 according to our method, slightly  
87 lower than its 31 rank in SCOPUS. This is due to the fact that we have not used "citations from more presti-  
88 gious journals" as a feature (this data are not readily available). Nonetheless, it is ranked higher than many  
89 of its peers which have been in publication for over 20 years. This is also due to the fact that ASCOM is "one  
90 of its kind" and uniquely positioned in the scientific space steered by appropriate editorial support. However  
91 SUBJECTIVE the statement may sound, it seems that interdisciplinary, diversity in background of the Edi-  
92 tors and authors and novelty in theme have been instrumental in placing journals uniquely (NAKAWATASE  
93 (2017). Rodríguez (2016) Jacobs and Rebecca. (2012) Erfanmanesh (2017)). Such qualitative feature, re-  
94 grettably is not visible from the big data landscape alone. This is another significant driving factor behind  
95 framing and interpreting a novel model that explains trends arising from investigating the big data landscape.  
96

97 There is another interesting observation to take note of. By ignoring the "size does matter" paradigm,  
98 the ranks of some journals (many years in publication with several volumes and issues) suffered according  
99 to our method. A few examples include Living Reviews in Solar Physics, ranked 43 according to our scheme  
100 while it is ranked 3 in SCOPUS; and Astronomy and Astrophysics Review, ranked 40 in our scheme while  
101 it is ranked 5 according to SCOPUS. This reversal of positions should be considered as important findings,  
102 because existing methods do not offer appropriate weights to journals that are new, despite catering to a



**Fig. 1.  $l_1$  Rank Progression of ASCOM based on SCOPUS data computed by the proposed method. Since, the steady ascendancy in the journal's rank is unmistakable, it will be interesting to investigate the behavior of the journal rank in the long run once enough data is gathered. Please see Appendix A for details.**

103 niche and important area of research. In other words, the results indicate that years in publication may  
 104 sometimes dominate over other indicators of quality and may not capture the growth of journals in "short  
 105 time windows". Our study also reveals that ASCOM is indeed a quality journal as far as early promise is  
 106 concerned.

107 Scientometrics deals with analyzing and quantifying works in science, technology, and innovation. It is a  
 108 study that focuses on quality rather than quantity. The journals are evaluated against several metrics such  
 109 as the impact of the journals, scientific citation, SJR, SNIP indicators as well as the indicators used in  
 110 policy and management contexts. The practice of using journal metrics for evaluation involves handling a  
 111 large volume of data to derive useful patterns and conclusions (Manyika et al. (n.d.)). These metrics play  
 112 an important role in the measurement and evaluation of research performance. Due to the fact that most  
 113 metrics are easily susceptible to manipulation and misuse, it becomes essential to judge and evaluate a  
 114 journal by using a single metric or a reduced set of significant metrics. We proposed  $l_1$ -norm Singular Value  
 115 Decomposition( $l_1$ -SVD) (Aedula et al. (2018)) to efficiently solve this problem. The code of the proposed  
 116 method is available at Aedula (n.d.).

## 117 2.2. The Big Data Landscape

118 The appeal of modern-day computing is its flexibility to handle volumes of data through an aspect of  
 119 coordination and integration. Advancements in Big Data frameworks,(?) and technologies has allowed  
 120 us to break the barriers of memory constraints for computing and implement a more scalable approach  
 121 to employ methods and algorithms. The aforementioned journal ranking scheme is one such algorithm  
 122 which thrives under the improvements made to scalability in Big Data. With optimized additions such as  
 123 Apache Spark to the distributed computing family, the enactment of  $l_1$  Regularization and Singular Value  
 124 Decomposition has reached an all new height. Implementing the SVD algorithm with the help of Spark can  
 125 not only improve spatial efficiency but temporal as well. The  $l_1$ -norm SVD scheme utilizes the SVD and  
 126 regularization implementation of *ARPACK* and *LAPACK* libraries along with a cluster setup to enhance the  
 127 speed of execution by a magnitude of at least three times depending on the configuration. Collecting data  
 128 is also a very important aspect of Big Data topography. The necessity of a cluster based system is rendered  
 129 useless without the requisite data to substantiate it. Scientometric data usually deals with properties of  
 130 the journals such as Total Citation, Self-Citation etc. This data could be collected using Web Scraping

131 methodologies but also can be found by most journal ranking organizations, available for open source use;  
 132 SCOPUS and SCIMAGO. For the  $l_1$ -norm SVD scheme, we used SCOPUS as it had an eclectic set of  
 133 features which were deemed appropriate to showcase the effectiveness of the algorithm. The inclusion  
 134 of the two important factors such as Cite Score and SJR indicators gave a better enhancement over just  
 135 considering one over the other. For more information about the data and code used to develop this algorithm  
 136 (please refer to [Aedula \(2018\)](#), [Github](#) repository of the project).

137 The landscape and the framework, although attractive are insufficient to explain the rapid rate of growth  
 138 of ASCOM. The remainder of the paper is organized as follows. We begin by presenting the motivation for  
 139 Delay Differential Equation (DDE) based modeling by outlining key strengths of such modeling concept.  
 140 Next, we consider time reversed DDE to model the growth by including historical effects, a fundamental  
 141 contribution in section 4. Section 5 contains solutions, analytically and computationally investigated and  
 142 interpreted in light of the big data landscape. Section 6 considers further modifications in the model by  
 143 adding Editorial reputation and Publisher Goodwill value. We discuss the implications of these additional  
 144 factors and the fundamental assumptions in Discussion & Conclusion Sections, 7 and 8.

### 145 3. Scope of our study and Motivation for modeling via DDE

146 The manuscript strives to achieve two fundamental objectives:

- 147 ○ We establish and quantify current journal influence as a function of its past influence. If the past influ-  
 148 ence is positive (good inheritance), the present journal influence benefits immensely from it (Please  
 149 see sections 4, 5 and fig.s 2, 3 and 4).
- 150 ○ The manuscript proposes a doctrine of "self-serving incentivization" by exploiting implicit control  
 151 variables (publisher goodwill value and editorial reputation-the celebrity effect). The so-called "  
 152 incentivized model" is proposed to propagate a positive "start-up boost" to the journal influence.  
 153 Thereby, these control variables and the modifications form the second and more advanced, complex  
 154 layer in modeling journal influence (Please see sections 6, 7 and fig.s 5-10) and help quantify the  
 155 theory of "celebrity effect".

156 The factors mentioned above and the resulting model explained in the subsequent sections also account  
 157 for the remarkable growth in influence and ASCOM discussed in sections 1 and 2. We achieve this by the  
 158 DDE based model presented below.

159

160 DDE is a well known concept for over two centuries, which has found application in various problems  
 161 in the fields of dynamical modeling of biomedical systems, biochemical reactions as well as in the newer  
 162 models of interpersonal/romantic relationships!! DDEs also find useful applications like dynamic population  
 163 growth, economic growth and spread of diseases like HIV, cancer, etc. Delay Differential Equations belong  
 164 to the class of Partial Differential Equations. These are used by the scientific community for modeling  
 165 dynamic systems for many of the obvious advantages. These equations describe the rate of change of a  
 166 function, at time 't' as a function of earlier times. A DDE in its general form can be given by:

$$p'(t) = f(p(t), p(t - \tau)); p(0) = p_0 \quad (1)$$

167 considering a constant delay of  $\tau$ . Some of the advantages of DDEs are:

- 168 ○ DDE take care of the "hereditary effects" during modeling a system. This implies if the influence of  
 169 a journal is positive in the past and/or intrinsic factors have been responsible for surge in reputation,  
 170 such features are naturally modeled in DDEs.

- 171     ○ In system modeling, it is desirable that the model is closer to the real process (in our case, influence  
172     diffusion and percolation) and it has been observed that DDEs offer a better model than others.
- 173     ○ DDEs are seen to provide a better control over the system since historical data is directly modeled in  
174     to the system (using time-reversed structure). This is particularly desirable.
- 175     ○ In case of a DDE, the initial point  $p_0$  defined over the interval  $[-\tau, 0]$ , is a function and not just a  
176     point. The solution  $p(t)$  is also a function in the same interval. Hence, the solution becomes infinite  
177     dimensional, unlike an ODE. Moreover, in a dynamical system, DDE takes care of rate of growth,  
178     which is a robust form of looking at the real world problem than just reading from hereditary events  
179     and inferring from them.

#### 180 4. TIME REVERSED DDE: Our Contribution

181     Let  $p'(t)$  rate of change of influence over time,  $p(t)$  influence @ time  $t$ , and  $p(-t)$  influence @ time  
182      $t = -t$  ( $t = 1, p'(1) = ap(1) + bp(-1)$  or  $p'(2) = ap(1) + bp(-2)$  and so on). The Time Reversed equation  
183     can now be written as

$$p'(t) = ap(t) + bp(-t) \quad (2)$$

which implies the rate of change of influence is represented as a combination of present and past influence.  
Let us consider a simple growth model given as

$$\begin{aligned} ap'(t) &= b + cp(t) \\ p(0) &= c \end{aligned}$$

where  $p(0)$  is not the Initial condition but is the value at the instant of time under the interval of considera-  
tion. We represent this linear growth in the form of time reversed structures as follows:

$$\begin{aligned} \implies p'(t) &= \frac{b}{a} + \frac{c}{a}p(t) \\ &= \frac{b}{a} + \frac{c}{2a}p(t) + \frac{c}{2a}p(t) \\ &= \frac{b}{a} + \frac{c}{2a}p(t) + \frac{c}{2a}p(-t) \end{aligned}$$

We assume symmetric influence function; there are two possibilities, symmetric and non-symmetric influ-  
ence. Differentiating w.r.t  $t$ ,

$$\begin{aligned} p''(t) &= \frac{c}{2a}p'(t) - \frac{c}{2a}p'(-t) \\ &= \frac{c}{2a} \left( \frac{b}{a} + \frac{c}{2a}p(t) + \frac{c}{2a}p(-t) \right) - \frac{c}{2a} \left( \frac{b}{a} + \frac{c}{2a}p(-t) + \frac{c}{2a}p(t) \right) \\ &= 0 \end{aligned}$$

184     Note:  $p(t)$  may exhibit linear growth under the assumption that there is a certain repeatability in the journal  
185     influence.

186 **4.1. The model under non-symmetric influence:**

187 Let us not consider the symmetric influence function since it is too strong an assumption to begin with  
 188 (fluctuations are absent, unidirectional slope, elements of uncertainty almost absent). Let us consider the  
 189 same model given as

$$ap'(t) = b + cp(t); p(0) = c \quad (3)$$

190 without the assumption of symmetric influence ( $p(t)=p(-t)$ ). Here also,  $p(0)$  is not the Initial condition but  
 191 is the value at the instant of time under the interval of consideration. Reorganizing equation 3,

$$p'(t) + \left(-\frac{c}{a}\right)p(t) = \frac{b}{a} \quad (4)$$

192 Assuming  $\left(-\frac{c}{a}\right)$  and  $\left(\frac{b}{a}\right)$  are continuous functions, we fix  $\left(-\frac{c}{a}\right) = r(t)$  and  $\left(\frac{b}{a}\right) = s(t)$ . Putting this in the  
 193 equation, we obtain

$$p'(t) + r(t)p(t) = s(t) \quad (5)$$

Let  $\mu(t)$  be an integrating factor(Saha (2011)). Multiply both sides of equation with  $\mu(t)$  and integrating,  
 we arrive at the following form:

$$\mu(t) = Ke^{\int r(t)dt}$$

194 and eventually the expression for journal influence is written as

$$p(t) = \frac{\int e^{\int r(t)dt} s(t)dt + \frac{c}{K}}{e^{\int r(t)dt}} \quad (6)$$

195 Under the assumption of non-symmetric influence (more realistic), the influence seems exponential growth  
 196 or decay depending on the coefficients but not a combination of both in a single expression. We shall see  
 197 a different picture in the next section when we encounter non-linear growth in influence for a slightly more  
 198 complicated, time reversed model.

199 **Remark:** Please note the above model does not contain "history" functions. Hence the solution does not  
 200 display a convex combination of exponential functions, which can be easily interpreted in light of historical  
 201 data. This is in contrast to the simple case (we assume a symmetric influence) where we can safely conclude  
 202 that if either the historic influence or the current influence of the journal is high then the journal is most  
 203 likely going to experience further rise in influence in the near future.

204 **4.2. Modeling Non-linear growth using symmetric influence effects**

Let us consider eq.(1) with the condition  $p(0) = c$  by mapping these to the following DDE:

$$\begin{aligned} y'(t) &= a_1(t)y(t) + a_2(t)y(t-d), t \geq 0 \\ y(t) &= p(t), t \in [-d, 0] \end{aligned}$$

Consider  $d = -2t; a = a_2(t), b = a_1(t); y(t) \equiv p(t) \forall t \in [-d, d]$ . Our proposed model is a special case of  
 DDE and it will be shown later that eq.(1) has at least one solution, which may not be necessarily unique.

**Solution Methodology:** Let us consider the time reversed model eq.(2):

$$\begin{aligned} p'(t) &= ap(-t) + bp(t) \\ p(0) &= k \\ p'(0) &= (a+b)k \end{aligned}$$

205 By Symmetry, we have

$$p'(-t) = -ap(t) - bp(-t) \quad (7)$$

Differentiating Eq.(2) wrt t, we get,

$$\begin{aligned} p''(t) &= ap'(-t) + bp'(t) \\ &= a(-ap(t) - bp(-t)) + b(ap(-t) + bp(t)) \\ &= -a^2p(t) - abp(-t) + bap(-t) + b^2p(t) \\ &= (b^2 - a^2)p(t) \end{aligned}$$

206 where  $r = \sqrt{b^2 - a^2}$

207 Again, by symmetry,

$$p''(-t) = r^2 p(-t) \quad (8)$$

208 Solution is of the form,

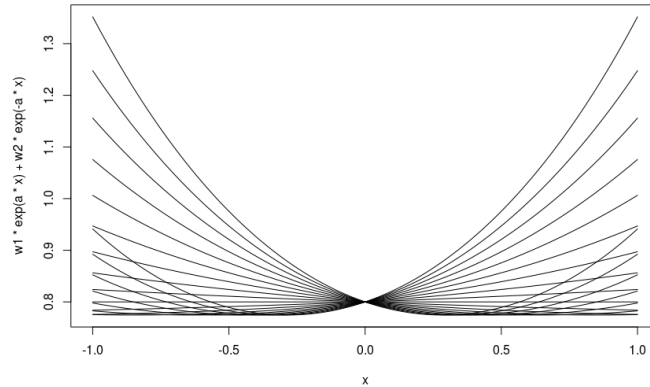
$$p(t) = A \exp(rt) + B \exp(-rt) \quad (9)$$

209 It is evident that  $p(t)$  is an exponential function. Using initial conditions, solving for A & B in terms of a &  
210 b we get,

$$p(t) = \frac{c}{2r} (r + a + b) \exp(rt) + \frac{c}{2r} (r - a - b) \exp(-rt) \quad (10)$$

211 Depending on the coefficient values, either positive or negative exponents will dominate. The two possible  
212 solutions depend on the value of  $r$ .

- When  $r > 0$  (i.e.,  $b > a$ ), we can expect an exponential real solution



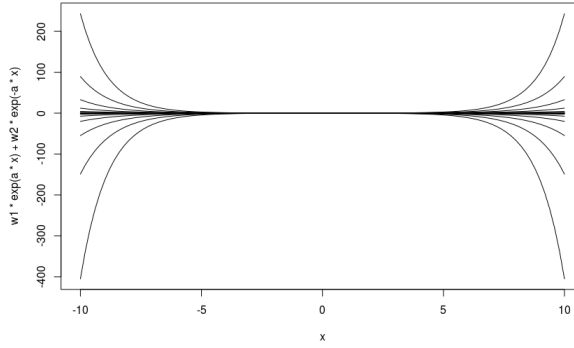
**Fig. 2. Plot of eq. (9) where  $p(t)$  is represented by the Y-axis and  $t$  is represented by the X-axis. The present influence,  $p(t)$  is controlled by past influence,  $p(-t)$**

213

- When  $r < 0$  (i.e.,  $b < a$ ), there will be oscillatory solutions, due to  $r$  being imaginary. Again, these solutions are deemed infeasible due to lack of fixed periodicity.

216  $t = 0$  is considered to be in the middle of a short time frame, at which, we are measuring the influence.  
217 Hence, this is not considered as initial value problem and hence we are not guaranteed of a unique solution.  
218  $p(-t)$  is the mirror image of  $p(t)$  and it will result in a sharp spike in influence provided its value is high.  
219 This is typically observed in a short time window and averages out in the longer time span. We see that  
220 depending on the values of the parameters and  $b$  either the historical or the current data dominates. The  
221 curve shows that in first few years the influence is largely dominated by the past reputation of the editors,  
222 represented by the historical part of the DDE. After a certain point (we have assumed this point to be at the  
223 center of time series data), other parameters such as the current journal citations and the current reputation  
224 of the editors begin to reflect on the influence.





**Fig. 3. Plot of eq. (9) where  $p(t)$  is represented by the Y-axis and  $t$  by the X-axis. Imaginary solutions are obtained when  $a$  and  $b$  are varied such that  $w1 < 0$ . This is infeasible as the model explains real solutions for obvious reasons.**

225 **5. Model Fitting:**

Let us recall Eq.(1):

$$\begin{aligned} p'(t) &= ap(t) + bp(-t) \\ p(0) &= c \\ p'(0) &= (a + b)c \end{aligned}$$

We also know, by approximation that,

$$\begin{aligned} p'(t) &\approx \frac{p(t+h) - p(t)}{h} \\ &\approx \frac{p(t) - p(t-h)}{h} \end{aligned}$$

where  $h$  is the step size. Let us consider the spread at discrete time intervals corresponding to one to five years (obtained from the dataset), indicated as  $p'(1), p'(2), p'(3), p'(4)$  and  $p'(5)$  respectively. Here, we can write  $p'(1) = ap(-1) + bp(1)$ . Also,

$$\begin{aligned} \implies \frac{p(1) - p(0)}{1} &= ap(-1) + bp(1) \\ &= a[A \exp(-r) + B \exp(r)] + b[A \exp(r) + B \exp(-r)] \\ &= (aA + bB) \exp(-r) + (aB + bA) \exp(r) \end{aligned}$$

226 The value on LHS is obtained from the dataset. Similarly, we can compute  $p'(\frac{3}{4}), p'(\frac{1}{2}), p'(\frac{1}{4})$ , etc., can  
 227 be obtained from the dataset, where the fractions represent the quarters in a year. We are now required to  
 228 estimate the coefficients  $a, b, A$  &  $B$ . This is an overestimation problem with number of equations exceeding  
 229 number of unknowns. We can solve this by method of Least Squares and hence use the solution to predict  
 230 future influence in rate of journal influence spread.

231 **5.1. Least Square Method to fit the data:**

From eq. (1), we obtain

$$p'(t) = ap(t) + bp(-t)$$

Let  $p'(t) = z, p(t) = x, p(-t) = y$ . Therefore, eq. (1) becomes

$$z = ax + by$$

Let

$$S = \sum (z - (ax + by))^2$$

Differentiating w.r.t a,

$$S = 2 \sum (z - (ax + by))(-x) = 0$$

$$S = \sum (-zx + ax^2 + bxy) = 0$$

232

$$\sum zx = a \sum x^2 + b \sum xy \quad (11)$$

Differentiating w.r.t b,

$$S = 2 \sum (z - (ax + by))(-y) = 0$$

$$S = \sum (-zy + axy + by^2) = 0$$

233

$$\sum zy = a \sum xy + b \sum y^2 \quad (12)$$

On solving eq. (11) and eq. (12), we obtain the values of  $a$  and  $b$ .

**ESTIMATING 'A' and 'B':**

We have found that

$$p(t) = (aA + bB) \exp(-rt) + (aB + bA) \exp(rt)$$

Let,

$$p(t) = y$$

$$aA + bB = w1$$

$$aB + bA = w2$$

$$\exp(rt) = x$$

Taking log

$$\log(x) = rt$$

$$\log(x^{-1}) = -rt$$

$$\exp(-rt) = \frac{1}{x}$$

Therefore,

$$y = w1 * x + \frac{w2}{x}$$

$$xy = w1 * x^2 + w2$$

Let,

$$Y = xy$$

$$X = x^2$$

Now,

$$Y = w1 * X + w2$$

234

$$\sum Y = w1 \sum X + w2 * n \quad (13)$$

235

$$\sum X \sum Y = w1 \sum X^2 + w2 \sum X \quad (14)$$

236

On solving eq. (13) and eq. (14) we can obtain values of w1 and w2. Hence, we can also find the values of A and B. We present the algorithm below.

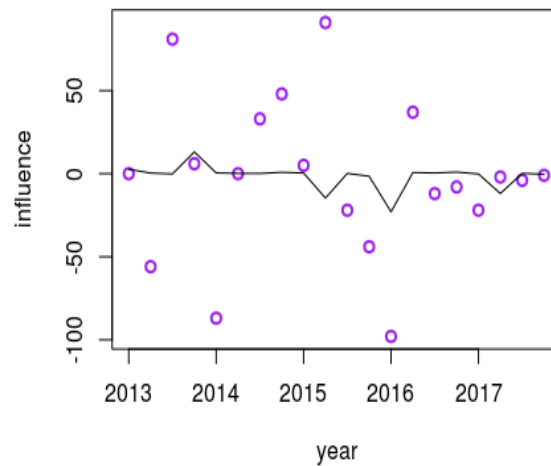
---

**Algorithm 1** Model Fit using Least Square Method

---

- 1:  $p(t) \leftarrow$  Input journal influence data
  - 2:  $EQU \leftarrow$  Model  $EQ(p(t))$
  - 3:  $NEW\_EQU \leftarrow$  LSM( $EQU$ )
  - 4: **procedure** MODEL\_EQ(P(T))
  - 5:  $p'(t) \leftarrow ap(t) + bp(-t)$
  - 6: Discretize the derivative using present and past data
  - 7:  $p'(t) \leftarrow p(t+h)-p(t)/h$
  - 8:  $(p(1)-p(0))/1 \leftarrow ap(-1) + bp(1)$
  - 9:  $= a[A \exp(-r) + B \exp(r)] + b[A \exp(r) + B \exp(-r)]$
  - 10: Return  $(aA + bB) \exp(-r) + (aB + bA) \exp(r)$
  - 11: **end procedure**
  - 12: **procedure** LSM(EQU)
  - 13: Derive the values of 'a' and 'b' of EQU using Equations
  - 14:  $\sum zx = a \sum x^2 + b \sum xy$  and
  - 15:  $\sum zy = a \sum xy + b \sum y^2$
  - 16: Derive the values of 'A' and 'B' of EQU using Equations
  - 17:  $\sum Y = w1 \sum X + w2 * n$
  - 18:  $\sum X \sum Y = w1 \sum X^2 + w2 \sum X$
  - 19: return EQU with derived values
  - 20: **end procedure**
- 

237



**Fig. 4.  $p'(t)$  VS  $t$  (Non-linear least square curve): The rate of change in influence over the span of 5 years shows small fluctuations but maintains an overall steady value. This means that the influence in 5 years will not suffer drastically.**

## 238 6. Model Modification to accommodate implicit control variables

239 Additionally, we consider implicit control variables which play important roles in the growth of any  
 240 journal. These variables pose challenges to the modeling set up and without these, the scope is limited to  
 241 empirical verification at a minor scale only. Next step in modeling data is to carry out modifications to  
 242 this structure in order to accommodate implicit control parameters such as publisher goodwill value and  
 243 start-up initiative by editors (editorial reputation). We define this initiative as the reputation of editors who  
 244 steered the journal and offered a strong attraction for quality submissions from scholars across the globe.  
 245 It is realistic to hypothesize that reputed scholars acting as editors add value and credibility to an emerging  
 246 journal. This value however is extremely hard to quantify and therefore modeling such phenomenon is novel  
 247 and imperative to understand the journals growth pattern. We propose to present the model and the analytical  
 248 solution, repeat the exercise of sections 3 and 4 and discuss the implication of the proposed modification.

249 The Time Reversed equation with the additive influence term (Publisher goodwill value) can now be  
 250 re-written as

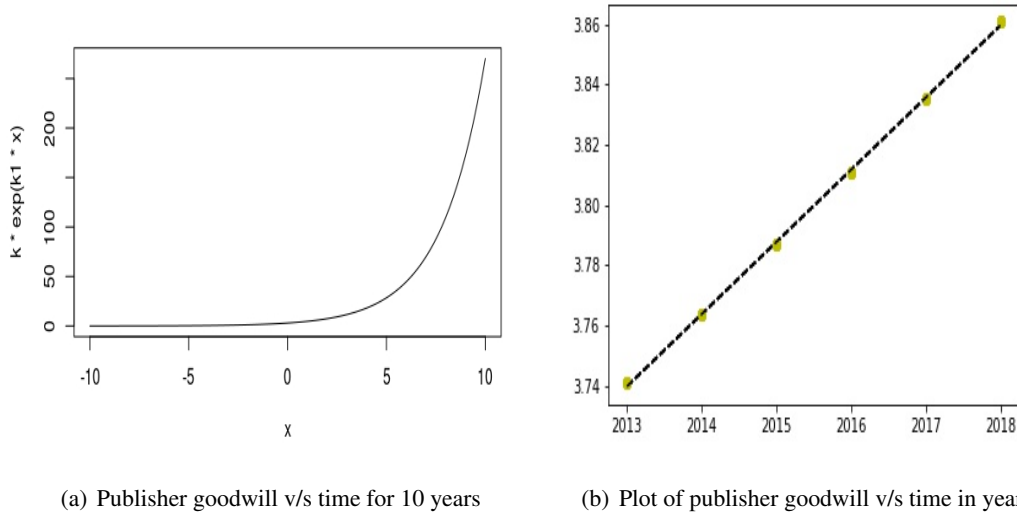
$$p'(t) = ap(t) + bp(-t) + \eta + \theta \quad (15)$$

251 where  $\eta$  is an additive term implying goodwill of the publishing house, Elsevier, in our case!  $\theta$ , OTOH  
 252 represents Editors' reputation.

### 253 6.1. Additional Considerations

254 ○ Let us assume  $\eta$  to be either linear or exponential. Such considerations are justified since any reputed  
 255 publisher, in order to remain competitive, would strive to enhance goodwill. Thus,  $\eta$  can't possibly  
 256 be a constant.

257 ○ We pose the next question pertinent to quantification of goodwill. It is modeled as a function of  
 258 the percentage of accepted papers over time, a trend that accommodates a fixed number of accepted  
 259 articles and the selection criteria of additional papers becomes increasingly stringent. It is modeled as



**Fig. 5. Plot of publishers goodwill VS time. We observe that the publishers good will shows a linear rise in the span of the 5 years between 2013 and 2017. Extrapolated to 10 years, the linear trend becomes non-linear and eventually impact the overall influence of the journal by a margin.**

260

$$\eta(.) = \exp(-art) + \alpha(a - b) \tag{16}$$

261

where *art* is the percentage of articles accepted after the initial threshold of  $\alpha$  articles.  $\alpha(a - b)$  is the initial threshold, conveniently set to ensure that the influence doesn't hover to the negative.

262

263

o Thus,  $\eta(.)$  is a control variable in the formulation and explanation of publisher goodwill. This implies, increasingly the percentage of accepted articles will diminish. Such stringent measures in peer-review bolster publisher goodwill.

264

265

266

o The formulation being in place, we now integrate  $\eta(.)$  with the modified model.

267

o Editorial reputation may be any of the three: a constant function, linear or exponential growth. The first one is more likely since the Editors of ASCOM are well established in their fields. Therefore, it is less likely that their phase of influence is still growing at quadratic rate or higher. In fact, we have observed that the influence pattern (citations) is steady. Nonetheless, we have considered all three possibilities and discuss the implications after integrating Editorial reputation,  $\theta$  (which is a function of time) in to the model.

268

269

270

271

272

## 273 6.2. Temporal evolution of publisher goodwill value

274

Figures 5(a) and 5(b) throw some useful insights. We hypothesize that the linear graph (fig. 5(b)) is a subset of the non-linear one (fig. 5(a)). Fig 5(b), which is a time-series plot of publisher goodwill value is linear upon fitting the ASCOM data. Fig. 5(a) is an extended time window plot of the same journal which is accomplished by simulating the data available from 5 years, extended to 10 years. The 5-year trend, if we take the time-slice off from fig. 5(a), produces fig. 5(b). This is done to establish the hypothesis that, available data to understand and predict longer time average behavior is insufficient.

278

279

280

281

This synthetic experiment implies that, if the good work in the past continues (good inheritance in terms of positive influence of the implicit control variable, the publisher goodwill, it shall continue to grow in

282 non-linear fashion). The observation is in agreement with the publisher in question, Elsevier, who pursues  
 283 aggressive and stringent quality practices toward the larger goal of monopoly in the business of publishing.  
 284 At this point, we may note that, the nonlinear, time dependent trend shall influence the overall journal  
 285 growth in influence to a greater proportion in comparison with the model we assumed in eq. (16) (which  
 286 is time-independent). We draw such inferences from the goodwill value as a time series plot by resolving  
 287 the equation with fitted goodwill value model from time-series data. We show that in the ensuing discussion  
 288 and the figures below (fig. 6, 7, 8). Let us now consider eq. (16). On adding the publishers goodwill as a  
 289 function of time we obtain the equation,

$$290 \quad p''(t) - (b^2 - a^2)p(t) = (a + b)\theta(t) + k * \exp^{k_1 t}$$

291 On solving the above equation on similar lines outlined in Appendix C, we obtain expressions of journal  
 292 influence as solutions for the three different cases of  $\theta(t)$  being constant, linear and exponential and  $\eta(t)$   
 293 being the time dependent function instead of a function of accepted articles as discussed earlier.

294 (1). CASE 1 (Fig. 6): Let us assume that  $\theta(t) = \theta = \text{constant}$

$$295 \quad \implies p(t) = c_1 \exp^{t\sqrt{b^2 - a^2}} + c_2 \exp^{-t\sqrt{b^2 - a^2}} + \frac{\theta}{(a - b)} + \frac{k * \exp^{(k_1) * t}}{(k_1)^2 - (b^2 - a^2)}$$

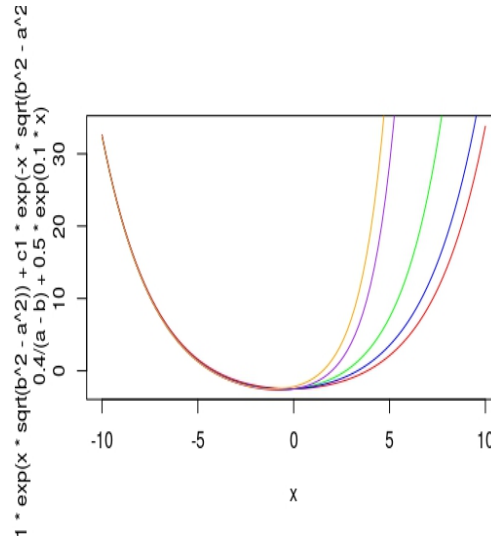


Fig. 6. Plot of  $p(t)$  VS.  $t$  when  $p(t)$ , the influence is the solution to the above equation. We observe that the slope of the graph becomes steeper as  $k_1$  increases.

296 (2). CASE 2 (Fig. 7): Let us assume that  $\theta(t)$  is linear:  $\theta(t) = At + B$

$$297 \quad \implies p(t) = c_1 \exp^{t\sqrt{b^2 - a^2}} + c_2 \exp^{-t\sqrt{b^2 - a^2}} + \frac{At + B}{a - b} + \frac{k * \exp^{(k_1) * t}}{(k_1)^2 - (b^2 - a^2)}$$

298 (3). CASE 3 (Fig. 8): Let us assume that  $\theta(t)$  is exponential:

$$299 \quad \theta(t) = \exp^{At} \implies p(t) = c_1 \exp^{t\sqrt{b^2 - a^2}} + c_2 \exp^{-t\sqrt{b^2 - a^2}} + \frac{(a + b) \exp^{At}}{A^2 - (b^2 - a^2)} + \frac{k * \exp^{(k_1) * t}}{(k_1)^2 - (b^2 - a^2)}$$

300 These plots (shown in Figure 5, 6, 7 and 8) demonstrate clearly, as publisher goodwill value is modeled as a  
 301 time dependent evolution, the influence of the journal grows at a faster pace in the longer run. Therefore, it  
 302 complements our observation that, publisher goodwill value has a small role to play in the growth of journal  
 303 influence in short time span but evolves gradually as time elapses.

### 304 6.3. Temporal evolution of Editorial reputation

305 We observe the celebrity effect here (Fei et al. (2015)). Editors are well established scholars and by the  
 306 time they assumed editorial responsibility, they are in the "cool off state" implying the surge in reputation

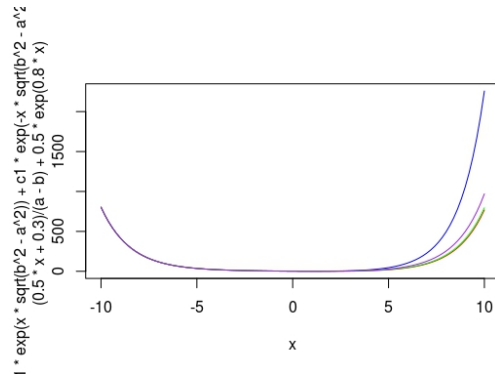


Fig. 7. Plot of  $p(t)$  VS.  $t$  when  $p(t)$ , the influence is the solution to the above equation. We observe that the slope of the graph becomes steeper as  $k_1$  increases.

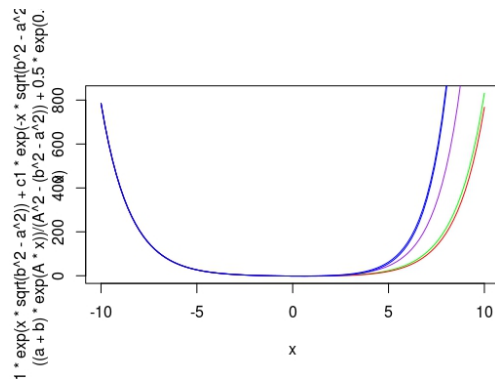


Fig. 8. Plot of  $p(t)$  VS.  $t$  when  $p(t)$ , the influence is the solution to the above equation. We observe that the slope of the graph becomes steeper as  $k_1$  increases.

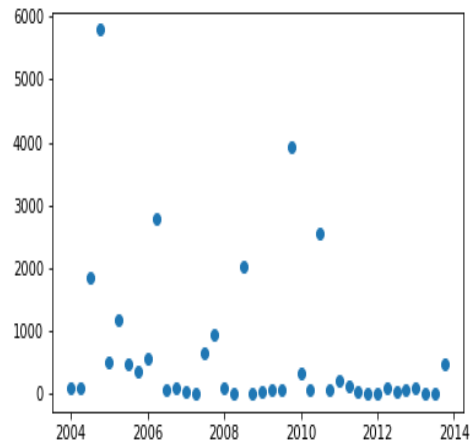
307 they experienced when they were rising stars stabilized. Therefore, steep gradient shall no longer be ex-  
 308 pected. This is what we observe in Fig 9 where the editorial influence between 2004 and 2014 is plotted.  
 309 Please note ASCOM was founded in 2013. The influence trend of all the editors during that time (2010-14)  
 310 is approximately constant.

311 Next section will deliberate on the contributions of these variables, in particular and model modification,  
 312 in general on the rate of change in influence observed in ASCOM. The role of control variables are evident  
 313 in the visualization we present below.

## 314 7. DISCUSSION

315 We develop a model to study its effect on astronomy and computer science domains and analyze pa-  
 316 rameters that have contributed in building the reputation of ASCOM. In this specific case study of journal  
 317 influence, the spread is clearly dependent on present as well as history dependent functions. This strength-  
 318 ens the motivation of using DDE model for the study. The model explains the growth pattern of the journal  
 319 well by capturing the intrinsic attributes and historical data. The time reversed model works as a mirror and  
 320 helps carry over the good deeds of the past (quality of articles in niche areas and open problems solved by  
 321 interdisciplinary efforts reflected in citation history). Our model takes care of the hereditary effects and since  
 322 the phenomenon of observing a journal in an emerging and interdisciplinary area is modeled as a function  
 323 of spatial variables renders the system infinite degrees of freedom. Thus, the proposed model is robust and  
 324 provides better control over the system.

325 However, the data is limited since the journal is in publication for just over five years. Therefore the



**Fig. 9. The above plot represents editors influence VS time. We see that the editorial influence is almost constant with time. This is possible because the editors are already well established. Hence, the influence is steady with little fluctuations.**

326 influence of historical data does not translate to overwhelming quantitative evidence in the way we liked  
 327 it to. Nonetheless, if we extrapolate the interval by extending the time window of consideration (since  
 328 the historical data is assumed to influence the present one), we observe profound effects (Please see the  
 329 discussion on temporal evolution of publisher goodwill value where the observed linear growth in goodwill  
 330 is really a 5-year snapshot subset of the longer window; (please see figs. 5(a) and (b) and the discussion  
 331 in section 6.2). Additionally, we considered implicit control variables such as Editorial reputation and  
 332 Publisher goodwill which play important roles in the growth of any journal. These variables pose challenges  
 333 to the modeling set up and without those, the scope is limited to empirical verification at minor scale. We  
 334 observe,

335 ○ The graphs for the equations for  $p(t)$  vs time where  $Y$  axis represents  $p(t)$  and  $X$  axis is time has a  
 336 parabolic shape. This shape is due to the presence of the symmetric history function  $p(-t)$ .

337 ○ From the results of this study, it can be established that the journal citations vary in a non-linear  
 338 fashion. Initially, the citation score is usually less as the journal will have less popularity. This can be  
 339 seen when we do a comparison of the citations of the editor v/s time and the influence of the journal  
 340 v/s time.

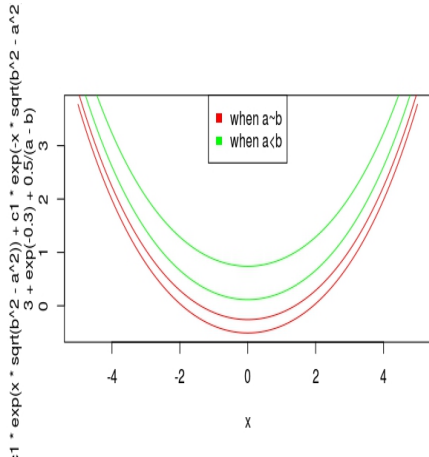
341 Fig 9. shows that the rate of change in influence is more or less constant over time. There is an  
 342 initial irregularity as the initial change in influence is directly related to the current influence. But  
 343 with time, the graph smooths out because, the other parameters such as citations and readership of  
 344 the journal also begin to affect the rate of change of influence. This is a testimony of consistent and  
 345 largely positive rate of change in influence. This hypothesis is verified by the graphs discussed below.  
 346 First, we plot the the editorial influence with time without considering the publishers goodwill (as  
 347 time series data) and editorial influence. We see that the curve is a simple parabola. This confirms our  
 348 assumption that the influence has a global minima and it stays upward elsewhere.

349 ○ After adding publishers Goodwill which we assumed to be an exponential distribution, the plot shows  
 350 a small change in shape. Since publisher's goodwill is a constant followed by a -ve exponential  
 351 function, it can be conveniently added to  $ap(t) + bp(-t)$ , as those are also exponential. This shows  
 352 that the publishers goodwill plays a small but non-negligible role in governing the journals growth in

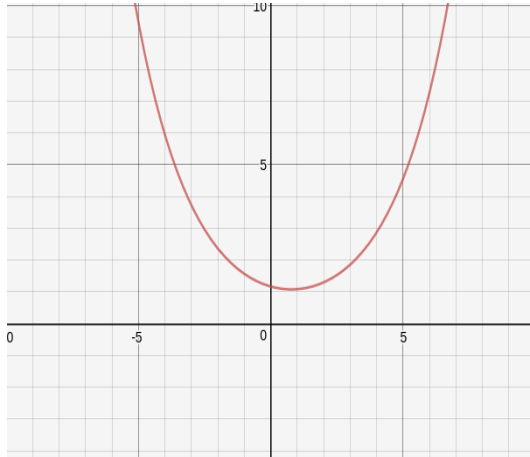


353  
354  
355  
356  
357

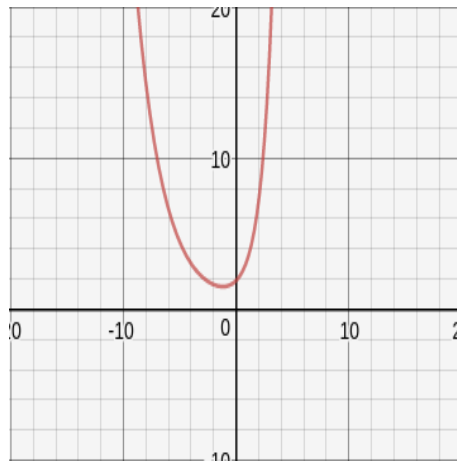
influence. However, in line with our argument presented in section 6, the small change will translate to larger increment if the time window is expanded i.e. the goodwill will have a larger contribution with the elapse of time. On plotting the curve for influence VS time after considering different Editorial influence functions (i.e.  $\theta(t)$ ) along with the publishers goodwill, we obtain the following graphs (fig. 10.(a), (b) and (c)).



(a)  $p(t)$  (i.e journal influence) VS  $t$  is positive, including the initial influence. This is because of the "start-up boost" provided by the control variables.



(b)  $p(t)$  VS  $t$  when  $\theta(t)$  (i.e editorial influence) is linear. Initial influence is shifted further up.



(c)  $p(t)$  VS  $t$  when  $\theta(t)$  (i.e editorial influence) is exponential

**Fig. 10. Journal Influence  $p(t)$  v/s  $t$  for different variations in  $\theta$**

358  
359  
360  
361  
362  
363  
364

- Note,  $\eta/(a - b) \geq \alpha$  when  $c_1 = c_2$ . If this condition is not satisfied the values of the influence becomes negative for initial values of time which is not possible. The model ensures that. When the editorial influence is exponential it gets added to the positive part of the delay differential equation.
- We simulated/computed goodwill value,  $\eta$  for every year. The value changes, as expected and therefore,  $\eta$  can now be interpreted as time series data, fitted as a function of time and used in the modified equation as a time-dependent function. With this modification, we resolve the DDE equation and noted changes in the trend. The gradient (rate of change in influence) is greater as compared to the

- 365 earlier model where  $\eta$  is considered as a function of percentage of accepted articles!
- 366 ○ The magnitudes of the constants  $c_1$  and  $c_2$  determine if the recent or the historic part of the equation  
367 dominates the curve.
- 368 ○ **Effect of implicit control variables:** Editorial reputation and publisher goodwill value are indeed  
369 positive factors. When the figures are compared, (figs. 2, 6, 7 and 10), we observe a non-zero start-up  
370 boost in figs. 6, 7, 8 and 10. On the contrary, when we investigated the original DDE (fig. 2) without  
371 these implicit variables, influence evolution begins from 0. This non-negative boost (graphs of figs. 8,  
372 9 and 10 start from some positive value above 0) is due to influence of the control variables, which are  
373 otherwise hard to quantify and model. Therefore, our arguments in favor of including these variables  
374 in the influence model stand vindicated.

## 375 8. CONCLUSION

376 model suitably explains the growth pattern of the journal by capturing the intrinsic attributes and his-  
377 torical data. We observe effects of celebrity authorship in the role of editors contribute to the growth in  
378 influence of ASCOM s well as the goodwill established publishing house brings. These effects, dynamic in  
379 nature, haven't been studied before. The contribution of the manuscript is therefore two-fold. Firstly, a novel  
380 model of DDE is exploited to study the influence of a journal in an emerging area. Secondly, qualitative and  
381 dynamic control variables (Editorial reputation and publisher goodwill value), hitherto unexploited, for the  
382 simple reason of complexity have been quantified and integrated in to the model. The time reversed model  
383 works as a mirror and helps to carry over the achievements of the past (quality of articles in niche areas and  
384 open problems solved by interdisciplinary efforts). As a final note, it might be useful to remind that our  
385 model takes care of the hereditary effects by exploiting the function,  $p(-t)$ . The phenomenon of observing  
386 a journal in an emerging and interdisciplinary area is modeled as a function of spatial variables renders the  
387 system infinite degrees of freedom. Therefore, the proposed model is robust and it provides better control  
388 over the system. Note however, that the data is limited given the low age of the journal. Therefore the in-  
389 fluence of historical data does not translate into overwhelming quantitative evidence. The model also holds  
390 promise because of its control structure and ability to accommodate implicit control variables such as Edi-  
391 torial reputation and publisher goodwill value have been found to generate significant implications, overall.  
392 We establish a fact, no less remarkable, that the implicit control variables act as incentives to the new journal  
393 in an emerging area. This is a much needed boost that the journal enjoys, in the absence of which, it may  
394 have to struggle much harder to attract readership and scholarship! We conclude by stating that unlike most  
395 of the scholarly work in scientometric literature, which are post-facto statistical studies, our work is focused  
396 on investigating the background responsible for certain trends observed in a journal in niche area. This is  
397 where the manuscript is markedly different!

## 398 9. Acknowledgements

399 We would like to thank the Science and Engineering Research Board (SERB)-Department of Science  
400 and Technology (DST), Government of India for supporting our research. The project reference number  
401 is: SERB-EMR/2016/005687.

## 402 References

403 Aedula, R. (n.d.). Retrieved 2018-04-24, from <https://github.com/rahul-aedula95/L1>  
404 [\\_Norm](#)

- 405 Aedula, R. (2018). *Github repository*.
- 406 Aedula, R., Yashasvi Madhukumar, Snehanshu Saha, Mathur, A., Kakoli Bora, & Surbhi Agrawal.  
407 (2018). L1 norm svd based ranking scheme: A novel method in big data mining. , -. doi:  
408 [10.13140/rg.2.2.23721.29280](https://doi.org/10.13140/rg.2.2.23721.29280)
- 409 Bach, F., Jenatton, R., Mairal, J., & Obozinski, G. (2012, nov). Structured sparsity through convex op-  
410 timization. *Statistical Science*, 27(4), 450–468. Retrieved from [https://doi.org/10.1214/](https://doi.org/10.1214/12-sts394)  
411 [12-sts394](https://doi.org/10.1214/12-sts394) doi: [10.1214/12-sts394](https://doi.org/10.1214/12-sts394)
- 412 Bora, K., Saha, S., Agrawal, S., Safonova, M., Routh, S., & Narasimhamurthy, A. (2016, October). CD-  
413 HPF: New habitability score via data analytic modeling. *Astronomy and Computing*, 17, 129-143.  
414 doi: [10.1016/j.ascom.2016.08.001](https://doi.org/10.1016/j.ascom.2016.08.001)
- 415 Engemann, K. M., & Wall, H. J. (2009). A journal ranking for the ambitious economist. *Review*(May),  
416 127-140. Retrieved from [https://ideas.repec.org/a/fip/fedlrv/y2009imayp127](https://ideas.repec.org/a/fip/fedlrv/y2009imayp127-140nv.91no.3.html)  
417 [-140nv.91no.3.html](https://ideas.repec.org/a/fip/fedlrv/y2009imayp127-140nv.91no.3.html)
- 418 Erfanmanesh, A. (2017, June). Composition of journals editorial board members as an indicator of the inter-  
419 disciplinary: The case of iranian journals in social sciences and humanities. *Library and Information*  
420 *Science*, 19(1), 81-106.
- 421 Fei, Q., Chong, H. G., & Bell, R. (2015, 03). The diminishing influence of celebrity authors in a diversified  
422 world of accounting journals. , 15, 37-57.
- 423 Ginde, G., Aedula, R., Saha, S., Mathur, A., Dey, S. R., Sampatrao, S., & Sagar, B. S. (2017). Big data  
424 acquisition, preparation and analysis using apache software foundation projects. In A. K. S. Ganesh  
425 Chandra Deka (Ed.), *Big data analytics tools and technology for effective planning* (chap. 9). CRC  
426 Press.
- 427 Ginde, G., Saha, S., Mathur, A., Venkatagiri, S., Vadakkepat, S., Narasimhamurthy, A., & Daya Sagar, B. S.  
428 (2016, Sep 01). Scientobase: a framework and model for computing scholastic indicators of non-  
429 local influence of journals via native data acquisition algorithms. *Scientometrics*, 108(3), 1479–1529.  
430 Retrieved from [https://doi.org/10.1007/s11192-](https://doi.org/10.1007/s11192-016-2006-2)  
431 [016-2006-2](https://doi.org/10.1007/s11192-016-2006-2) doi: [10.1007/s11192-](https://doi.org/10.1007/s11192-016-2006-2)
- 432 González-Pereira, B., Bote, V. P. G., & de Moya Anegón, F. (2009). The SJR indicator: A new indicator of  
433 journals' scientific prestige. *CoRR*, *abs/0912.4141*. Retrieved from [http://arxiv.org/abs/](http://arxiv.org/abs/0912.4141)  
434 [0912.4141](http://arxiv.org/abs/0912.4141)
- 435 Jacobs, A., & Rebecca, H. (2012, August). Interdisciplinarity in recently founded academic journals.  
436 *Scholarly Common Repository*.
- 437 Jangid, N., Saha, S., Gupta, S., & Rao, J. M. (2014). Ranking of journals in science and  
438 technology domain: A novel and computationally lightweight approach. *IERI Procedia*, 10,  
439 57 - 62. Retrieved from [http://www.sciencedirect.com/science/article/pii/](http://www.sciencedirect.com/science/article/pii/S2212667814001397)  
440 [S2212667814001397](http://www.sciencedirect.com/science/article/pii/S2212667814001397) (International Conference on Future Information Engineering (FIE 2014))  
441 doi: <https://doi.org/10.1016/j.ieri.2014.09.091>
- 442 Kalman, D. (1996). A singularly valuable decomposition: The svd of a matrix. *College Math Journal*, 27,  
443 2–23.
- 444 Kianifar, H., Sadeghi, R., & Zarifmahmoudi, L. (2014). Comparison between impact factor, eigenfactor  
445 metrics, and SCimago journal rank indicator of pediatric neurology journals. *Acta Informatica Med-*  
446 *ica*, 22(2), 103. Retrieved from <https://doi.org/10.5455/aim.2014.22.103-106> doi:  
447 [10.5455/aim.2014.22.103-106](https://doi.org/10.5455/aim.2014.22.103-106)
- 448 Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (n.d.). *Recursive*  
449 *object model (ROM)-modelling of linguistic information in engineering design*. Retrieved 2018-04-  
450 24, from [https://www.mckinsey.com/business-functions/digital-mckinsey/](https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation)  
451 [our-insights/big-data-the-next-frontier-for-innovation](https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation)
- 452 NAKAWATASE, H. (2017). The present state analysis of emerging interdisciplinary research ar-

- 453 eas: An analysis based on a classified table of grant-in-aid for scientific research. *Joho Chishiki*  
 454 *Gakkaishi*, 27(1), 1–5. Retrieved from [https://doi.org/10.2964/jsik.2017\\_007](https://doi.org/10.2964/jsik.2017_007) doi:  
 455 [10.2964/jsik.2017\\_007](https://doi.org/10.2964/jsik.2017_007)
- 456 Rodríguez, J. M. (2016, nov). Disciplinarity and interdisciplinarity in citation and reference dimensions:  
 457 knowledge importation and exportation taxonomy of journals. *Scientometrics*, 110(2), 617–642.  
 458 Retrieved from <https://doi.org/10.1007/s11192-016-2190-0> doi: [10.1007/s11192-](https://doi.org/10.1007/s11192-016-2190-0)  
 459 [016-2190-0](https://doi.org/10.1007/s11192-016-2190-0)
- 460 Saha, S. (2011). *Ordinary differential equations: A structured approach*. Cognella. Retrieved from  
 461 [https://www.amazon.com/Ordinary-Differential-Equations-Structured](https://www.amazon.com/Ordinary-Differential-Equations-Structured-Approach/dp/160927704X?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=160927704X)  
 462 [-Approach/dp/160927704X?SubscriptionId=0JYN1NVW651KCA56C102&tag=](https://www.amazon.com/Ordinary-Differential-Equations-Structured-Approach/dp/160927704X?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=160927704X)  
 463 [techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=](https://www.amazon.com/Ordinary-Differential-Equations-Structured-Approach/dp/160927704X?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=160927704X)  
 464 [160927704X](https://www.amazon.com/Ordinary-Differential-Equations-Structured-Approach/dp/160927704X?SubscriptionId=0JYN1NVW651KCA56C102&tag=techkie-20&linkCode=xm2&camp=2025&creative=165953&creativeASIN=160927704X)

## 465 Appendix

### 466 A. First Appendix

#### 467 A.1. Singular Value Decomposition

468 Singular Value Decomposition is the factorization of a real or complex matrix. Large scale of Sciento-  
 469 metric data is mined using suitable web scraping techniques and is modeled as a matrix in which the rows  
 470 represent the articles in a journal published over the years, and the columns represent various Scientometrics  
 471 or indicators proposed by experts of evaluation agencies (Kalman (1996)). The original data matrix, say  $\mathbf{A}$   
 472 of dimension  $m \times n$  and rank  $k$  is factorized into three unique matrices  $\mathbf{U}$ ,  $\mathbf{V}$  and  $\mathbf{W}^H$ .

- 473 ○  $\mathbf{U}$  - Matrix of Left Singular Vectors of dimension  $m \times r$
- 474 ○  $\mathbf{V}$  - Diagonal matrix of dimension  $r \times r$  containing singular values in decreasing order along the diago-  
 475 nal
- 476 ○  $\mathbf{W}^H$  - Matrix of Right Singular Vectors of dimension  $n \times r$ . The Hermitian, or the conjugate transpose of  
 477  $\mathbf{W}$  is taken, changing its dimension to  $r \times n$  and hence the original dimension of the matrix is maintained  
 478 after the matrix multiplication. In this case of Scientometrics, since the data is represented as a real  
 479 matrix, Hermitian transpose is simply the transpose of  $\mathbf{W}$ .

480  $r$  is a very small number numerically representing the approximate rank of the matrix or the number of  
 481 "concepts" in the data matrix  $\mathbf{A}$ . *Concepts* refer to latent dimensions or latent factors showing the association  
 482 between the singular values and individual components (Kalman (1996)). The choice of  $r$  plays a vital role  
 483 in deciding the accuracy and computation time of the decomposition. If  $r$  is equal to  $k$ , then the SVD is said  
 484 to be a Full Rank Decomposition of  $\mathbf{A}$ . Truncated SVD or Reduced Rank Approximation of  $\mathbf{A}$  is obtained  
 485 by setting all but the first  $r$  largest singular values equal to zero and using the first  $r$  columns of  $\mathbf{U}$  and  $\mathbf{W}$   
 486 (Ginde et al. (2017)).

487 Therefore, choosing a higher value of  $r$  closer to  $k$  would give a more accurate approximation whereas  
 488 a lower value would save a lot of computation time and increase efficiency.

#### 489 A.2. Regularization Norms

490 In the case of Big Data, parsimony is central to variable and feature selection, which makes the data  
 491 model more intelligible and less expensive in terms of processing.

492  $l_p$ -norm of a matrix or vector  $\mathbf{x}$ , represented as  $\|\mathbf{x}_p\|$  is defined as,  $\|\mathbf{x}_p\| = \sqrt[p]{\sum_i |x_i|^p}$  i.e the  $p^{\text{th}}$  root of  
 493 summation of all the elements raised to the power  $p$ . Hence, by definition,  $l_1$  norm =  $\|\mathbf{x}\|_1 = \sum_i |x_i|$

494 Sparse approximation, inducing structural sparsity as well as regularization is achieved by a number  
 495 of norms, the most common ones being  $l_1$  norm and the mixed group  $l_1$ - $l_q$  norm. The relative structure  
 496 and position of the variable in the input vector, and hence the inter-relationship between the variables is  
 497 inconsequential as a variable is chosen individually in  $l_1$  regularization. Prior knowledge aids in improving  
 498 the efficacy of estimation through these techniques.

499 The  $l_1$  norm concurs to only the cardinality constraint and is unaware to any other information available  
 500 about the patterns of non-zero coefficients.(Bach, Jenatton, Mairal, and Obozinski (2012))

### 501 A.3. Sparsity via the $l_1$ norm

502 Most variable or feature selection problems are presented as combinatorial optimization problems. Such  
 503 problems focus on selecting the optimal solution through a discrete, finite set of feasible solutions. Addi-  
 504 tionally,  $l_1$  norm turns these problems to convex problems after dropping certain constraints from the overall  
 505 optimization problem. This is known as convex relaxation. Convex problems classify as the class of prob-  
 506 lems in which the constraints are convex functions and the objective function is convex if minimizing, or  
 507 concave if maximizing.

508  $l_1$  regularization for sparsity through supervised learning involves predicting a vector  $\mathbf{y}$  from a set of  
 509 usually reduced values/observations consisting a vector in the original data matrix  $\mathbf{x}$ . This mapping function  
 510 is often known as the hypothesis  $\mathbf{h} : \mathbf{x} \rightarrow \mathbf{y}$ . To achieve this, we assume there exists a joint probability  
 511 distribution  $P(\mathbf{x}, \mathbf{y})$  over  $\mathbf{x}$  and  $\mathbf{y}$  which helps us model anomalies like noise in the predictions.

In addition to this, another function known as a loss function  $L(\mathbf{y}', \mathbf{y})$  is required to measure the difference  
 in the prediction  $\mathbf{y}' = \mathbf{h}(\mathbf{x})$  from the true result  $\mathbf{y}$ . Consider the resulting vectors consisting of the predicted  
 value and the true value to be  $\mathbf{y}'$  and  $\mathbf{y}$  respectively. A characteristic called *Risk*,  $R(\mathbf{h})$  associated with loss  
 function, and hence in turn with the hypothesis- $\mathbf{h}(\mathbf{x})$  is defined as the expectation of the loss function.

$$R(\mathbf{h}) = \mathbf{E}[L(\mathbf{y}', \mathbf{y})] = \int L(\mathbf{y}', \mathbf{y}) dP(\mathbf{x}, \mathbf{y})$$

Thus, the hypothesis chosen for mapping should be such that the risk,  $R(\mathbf{h})$  is minimum. This refers to as  
 risk minimization. However, in usual cases, the joint probability distribution of the problem in hand,  $P(\mathbf{x}, \mathbf{y})$   
 is not known. So, an approximation called *empirical risk* is computed by taking the average of the loss  
 function of all the observations. Empirical Risk is given by :

$$R_{emp}(\mathbf{h}) = \frac{1}{n} \sum_{i=1}^n L(\mathbf{y}'_i, \mathbf{y}_i)$$

The empirical risk minimization principle states that the hypothesis( $\mathbf{h}'$ ) selected must be such it that reduces  
 the empirical risk  $R_{emp}(\mathbf{h})$ :

$$\mathbf{h}' = \min_{\mathbf{h}} R_{emp}(\mathbf{h})$$

While mapping observations  $\mathbf{x}$  in  $n$  dimensional vector  $\mathbf{x}$  to outputs  $\mathbf{y}$  in vector  $\mathbf{y}$ , we consider  $p$  pairs  
 of data points -  $(\mathbf{x}_i, \mathbf{y}_i) \in \mathbb{R}^n \times \mathbf{y}$  where  $i = 1, 2, \dots, p$ . Thus the optimization problem for the data matrix in  
 Scientometrics takes the form:

$$\min_{\mathbf{w} \in \mathbb{R}^n} \frac{1}{p} \sum_{i=1}^p L(\mathbf{y}'_i, \mathbf{w}^T \mathbf{x}_i) + \lambda \Omega(\mathbf{w})$$

$L$  is a loss function which can either be square loss for least squares regression,  $L(\mathbf{y}', \mathbf{y}) = \frac{1}{2}(\mathbf{y}' - \mathbf{y})^2$ , or a  
 logistic loss function. Now, the problem thus takes the form:

$$\min_{\mathbf{w} \in \mathbb{R}^n} \|\mathbf{y}' - \mathbf{A}\mathbf{w}\|^2$$

512 Since the variables in the vector space/groups can overlap, it is ideal to choose  $\Omega(\mathbf{w})$  to be a group norm for  
 513 better predictive performance and structure. The  $m$  rows of data matrix  $\mathbf{A}$  are treated as vectors or groups( $g$ )  
 514 of these variables, forming a partition equal to the vector dimension,  $[1:n]$ . If  $\mathbf{G}$  is the set of all these groups  
 515 and  $d_g$  is a scalar weight indexed by each group  $g$ , the norm is said to be a  $l_1-l-q$  norm where  $q \in [2, \infty)$  (Bach  
 516 et al. (2012)),  $\Omega(\mathbf{w}) = \sum_{g \in \mathbf{G}} d_g \|\mathbf{w}_g\|_q$   
 517 The choice of the indexed weight  $d_g$  is critical because it is responsible for the discrepancies of sizes between  
 518 the groups. It must also compensate for the possible penalization of parameters which can increase due to  
 519 high-dimensional scaling. The factors that affect the selection are the choice of  $q$  in the group norm and the  
 520 consistency that is expected of the result. In addition to this, accuracy and efficiency can be enhanced by  
 521 weighing each coefficient in a group rather than weighing the entire group as a whole. The initial sparse  
 522 data matrix is first manipulated using the  $l_1$ -norm (Bach et al. (2012)).

#### 523 A.4. Methodology

524 An estimate of a journal's scholastic indices is necessary to judge its effective impact. The nuances of  
 525 scientometric factors such as Total Citation Count and Self-citation Count come into play when deciding the  
 526 impact of a journal. However, these factors unless considered in ideal circumstances don't by themselves  
 527 become a good indicator to represent the importance of a journal. Many anomalies arise when considering  
 528 these indices directly which may misrepresent or falsify a journal's true influence. The necessity to use  
 529 these indices in context with a ranking algorithm is imperative to better utilize these indices. The resulting  
 530 transformation of  $l_1$ -norms gives rise to a row matrix which is of the length equal to the number of features  
 531 of the pristine Scientometric data. This row matrix effectively represents the entire dataset at any given  
 532 iteration. The application of the Singular Value Decomposition operation on this row matrix is key in  
 533 determining the necessary norm values to remove through a recursive approach.

534 The *singval* array contains the Normalized Singular Values of all the individual  $l_1$ -norm transformed  
 535 columns. These values act as scores while addressing the impact of any given journal. In the context of  
 536 Singular Values the one with the lowest *singval* score is the most influential journal.

**Algorithm 2** Recursive  $l_1$ -norm SVD

---

```

1:  $A \leftarrow$  Input Transposed Feature Matrix  $A$ 
2: procedure LASSO
3:    $row\_matrix \leftarrow$  Coefficients of Lasso Regression
4:    $return$   $row\_matrix$ 
5: end procedure
6: procedure SVD
7:    $U, \Sigma, V \leftarrow$  Matrices of SVD
8:    $return$   $\Sigma$ 
9: end procedure
10: procedure NORMALIZE
11:    $Norm\_Data \leftarrow$  Normalized using  $l_1$ -norms
12:    $return$   $Norm\_Data$ 
13: end procedure
14: procedure RECURSIVE
15:    $L1\_row \leftarrow$  LASSO( $A$ )
16:    $singval$  [ $] \leftarrow$  SVD( $L1\_row$ )
17:    $Row\_Norm \leftarrow$  Normalize( $L1\_row$ )
18:    $Col\_Norm \leftarrow$  Normalize(All columns of  $A$ )
19:    $Col\_i \leftarrow$  Closest  $Col\_Norm$  Value to  $Row\_Norm$ 
20:   Delete  $Col\_i$  from  $A$ 
21:    $goto$  RECURSIVE
22: end procedure

```

---

537 **B. Lemma :** The equation  $p'(t) = ap(-t) + bp(t)$ , where  $p(0) = K$  and  $p'(0) = (a + b)K$  has  
538 a unique solution in  $[-d, d]$ .

Proof: Let us consider two solutions to this equation,  $y_1(t), y_2(t)$ . Then,

$$\begin{aligned}
g(t) &= y_1(t) - y_2(t) \\
y_1'(t) &= ay_1(-t) + by_1(t) \\
y_2'(t) &= ay_2(-t) + by_2(t) \\
\implies (y_1 - y_2)'(t) &= a(y_1 - y_2)(-t) + b(y_1 - y_2)(t) \\
\implies g'(t) &= ag'(-t) + bg(t) \\
g'(0) &= ag(0) + bg(0) \neq 0 \\
g(0) &= y_1(0) - y_2(0) \\
&= k - k = 0 \\
\implies g'(t) &= ag(-t) + bg(t) \\
g(0) &= 0 \\
\implies g(t) &= 0 \forall t \in [-d, d] \\
\implies y_1(t) &= y_2(t) \forall t \in [-d, d]
\end{aligned}$$

539 Therefore, the DDE has a unique solution, which implies that if there exist a oscillatory solution, there can  
540 not be an exponential or linear family of solutions depending on the parameters.





541 C. DDE solution

Let us consider a non-linear homogeneous DDE:  $p''(t) - (b^2 - a^2)p(t) = (a + b)\theta(t)$  The solution for this equation depends on the value of  $\eta$ . Auxiliary equation is:

$$AE = (D^2 - (b^2 - a^2))p(t) = 0$$

$$AE = (m^2 - (b^2 - a^2))p(t) = 0$$

$$m = (b^2 - a^2)^{1/2}$$

$$c_y = c_1 \exp^{t\sqrt{b^2 - a^2}} + c_2 \exp^{-t\sqrt{b^2 - a^2}}$$

$$\text{Since } \sqrt{b^2 - a^2} = r, c_y = c_1 \exp^{rt} + c_2 \exp^{-rt}$$

The PI (particular integral) is calculated as follows

CASE1 : Let us assume that  $\theta(t) = \theta = \text{constant}$

$$y_p = \frac{(a + b)\theta}{D^2 - (b^2 - a^2)}$$

$$y_p = \frac{\exp^0(a + b)\theta}{D^2 - (b^2 - a^2)}$$

$$y_p = \frac{(a + b)\theta}{0 - (b^2 - a^2)}$$

$$y_p = \frac{(a + b)\theta}{-(b - a)(b + a)}$$

$$y_p = \frac{\theta}{(a - b)}$$

$$\implies p(t) = c_1 \exp^{t\sqrt{b^2 - a^2}} + c_2 \exp^{-t\sqrt{b^2 - a^2}} + \frac{\theta}{(a - b)}$$

CASE2 : Let us assume that  $\eta(t)$  is linear :  $\eta(t) = At + B$

$$y_p = \frac{(a + b)(At + B)}{D^2 - (b^2 - a^2)}$$

$$y_p = \frac{(a + b)(At + B)}{(b^2 - a^2)\left(\frac{D^2}{(b^2 - a^2)} - 1\right)}$$

$$y_p = \frac{(a + b)}{a^2 - b^2} * \left(1 - \frac{D^2}{(b^2 - a^2)}\right)^{-1} * (At + B)$$

$$y_p = \frac{(a + b)}{a^2 - b^2} * \left(1 + \frac{D^2}{(b^2 - a^2)} + \frac{D^4}{(b^2 - a^2)^2}\right) * (At + B)$$

$$y_p = \frac{At + B}{a - b}$$

$$\implies p(t) = c_1 \exp^{t\sqrt{b^2 - a^2}} + c_2 \exp^{-t\sqrt{b^2 - a^2}} + \frac{At + B}{a - b}$$

CASE3 : Let us assume that  $\theta(t)$  is exponential :  $\theta(t) = \exp^{At}$

$$y_p = \frac{(a + b)\exp^{At}}{D^2 - (b^2 - a^2)}$$

$$y_p = \frac{(a + b)\exp^{At}}{D^2 - (b^2 - a^2)}$$

$$y_p = \frac{(a + b)\exp^{At}}{A^2 - (b^2 - a^2)}$$

$$\implies p(t) = c_1 \exp^{t\sqrt{b^2 - a^2}} + c_2 \exp^{-t\sqrt{b^2 - a^2}} + \frac{(a + b)\exp^{At}}{A^2 - (b^2 - a^2)}$$