STATISTICAL AND PREDICTIVE ANALYSIS OF R&D IN INDIA

Jalem Raj Rohit

CoE of Systems Science

Indian Institute of technology, Jodhpur

Jodhpur, India.

rajrohit@iitj.ac.in

Desidi Siva Prakash

CoE of Systems Science

Indian Institute of technology, Jodhpur

Jodhpur, India.

desidisiva@iitj.ac.in

*Abstract*— This research involves statistical analysis of the R&D data of India which includes the number of technicians, researchers, grants, number of research articles published and number of patent applications and identifying a trend and classification of the same depending upon the regression analysis. Machine Learning algorithms are implemented on the data for better and intelligent analysis. This is an attempt to learn and gain various insights of the R&D department of India from the data available and be able to tell a very interesting story.

*Index Terms*—Statistics, Regression, Machine Learning, Education, R&D.

# INTRODUCTION

The R&D department of a country is very important for the growth and scientific development of the country. The number of researchers, the number of technicians, the published and the patent applications are the factors which drive the R&D department of any country. The research and Development in a country play a very significant role in the technological advancement of a country. So, proper analysis of the R&D is very necessary. Understanding the trends of various data measures in the R&D data is important for getting a proper and clear idea about the advancements and improvements in the industry.

Many statistical techniques exist for analyzing the data including regression analysis, Deep Learning approach, etc. So, proper selection of the analysis technique is very important. Machine Learning has some great concepts with which the regression analysis can be done effectively. Proper visualization is very important for comparative and statistical analysis.

# DETAILS ABOUT DATA

The Data obtained is from an open-source Data repository. The dataset contains the previous 32 years of data about the percentage expenditure with respect to the GDP of a country, the number of researchers, the number of technicians, the number of research/scholarly articles published and also the number of patent applications. The data is in the excel and the csv formats. The datasets of the R&D departments of about 100+ countries are available and used for comparative analysis.

# FEATURES OF DATA

The data primarily consists of 5 features, basing on which the analyses are done. They are:

* Research and Development Expenditure – This data says the amount of expenditure spent on Research and Development as a percentage out of the country’s GDP. Expenditures for research and development are current and capital expenditures (both public and private) on creative work undertaken systematically to increase knowledge, including knowledge of humanity, culture, and society, and the use of knowledge for new applications. R&D covers basic research, applied research, and experimental development.
* Researchers in the R&D department -- Researchers in R&D are professionals engaged in the conception or creation of new knowledge, products, processes, methods, or systems and in the management of the projects concerned. Postgraduate PhD students (ISCED97 level 6) engaged in R&D are included.
* Technicians in R&D -- Technicians in R&D and equivalent staff are people whose main tasks require technical knowledge and experience in engineering, physical and life sciences (technicians), or social sciences and humanities (equivalent staff). They participate in R&D by performing scientific and technical tasks involving the application of concepts and operational methods, normally under the supervision of researchers.
* Scientific and Technical article journals -- Scientific and technical journal articles refer to the number of scientific and engineering articles published in the following fields: physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences.

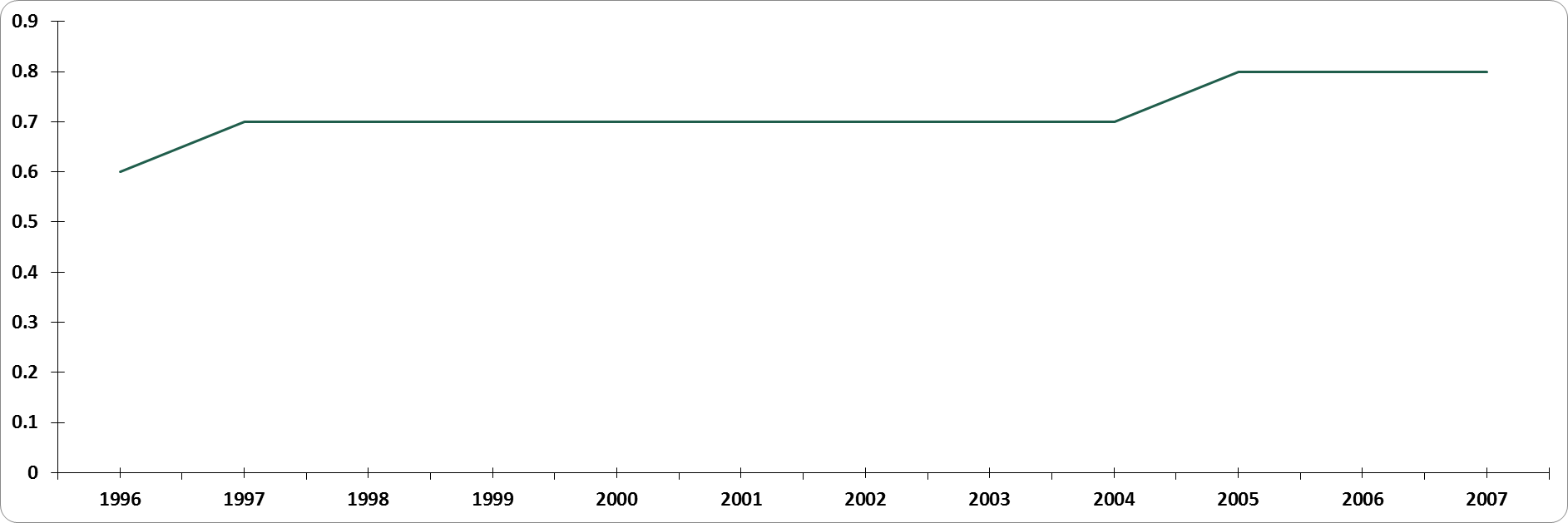
Patent Applications -- Patent applications are worldwide patent applications filed through the Patent Cooperation Treaty procedure or with a national patent office.

# INDIAN R&D DATA IS HIGHLY NON-UNIFORM

Through a detailed time-series analysis of the data of the India R&D Department, it can be concluded that the statistics of the R&D department is highly non-uniform with quite a lot of steep rises and dips. The necessary explanation of the time-series graphical analysis of all the five features is given below.

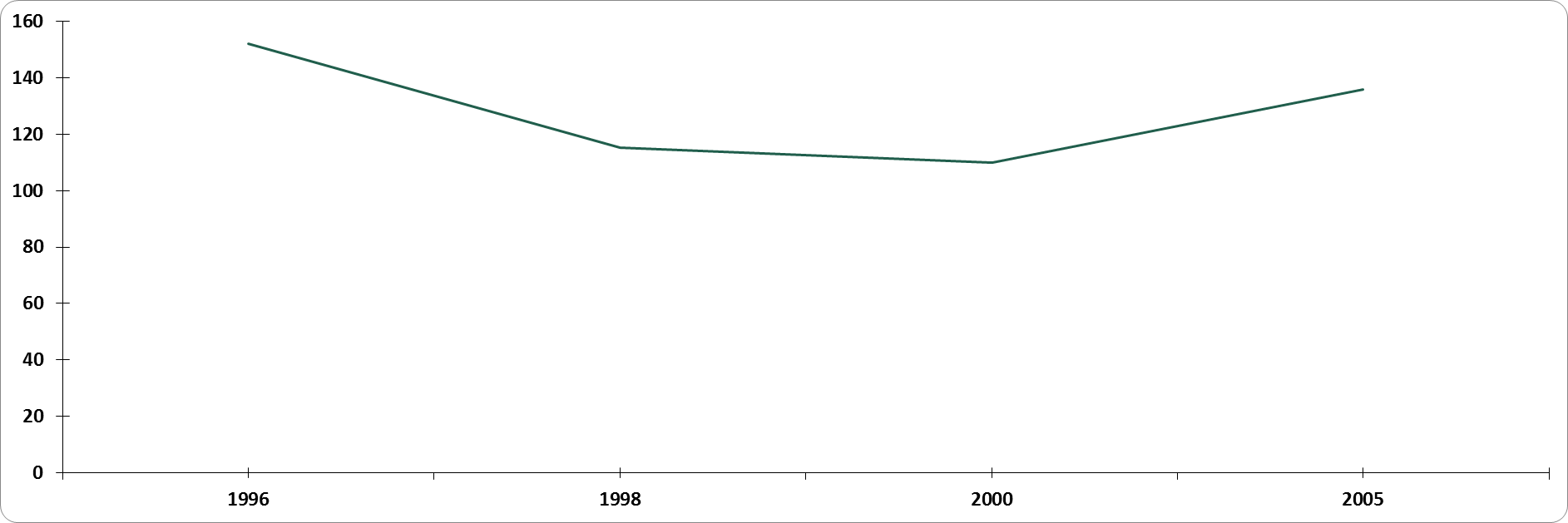
## R&D Expenditure (% of GDP)

The R&D Expenditure of India has been stably increasing from 1996 to 2000; where it experienced a sudden drop from almost 0.75% in 2000 to about 0.7% in 2003; which can be considered as a steep drop.



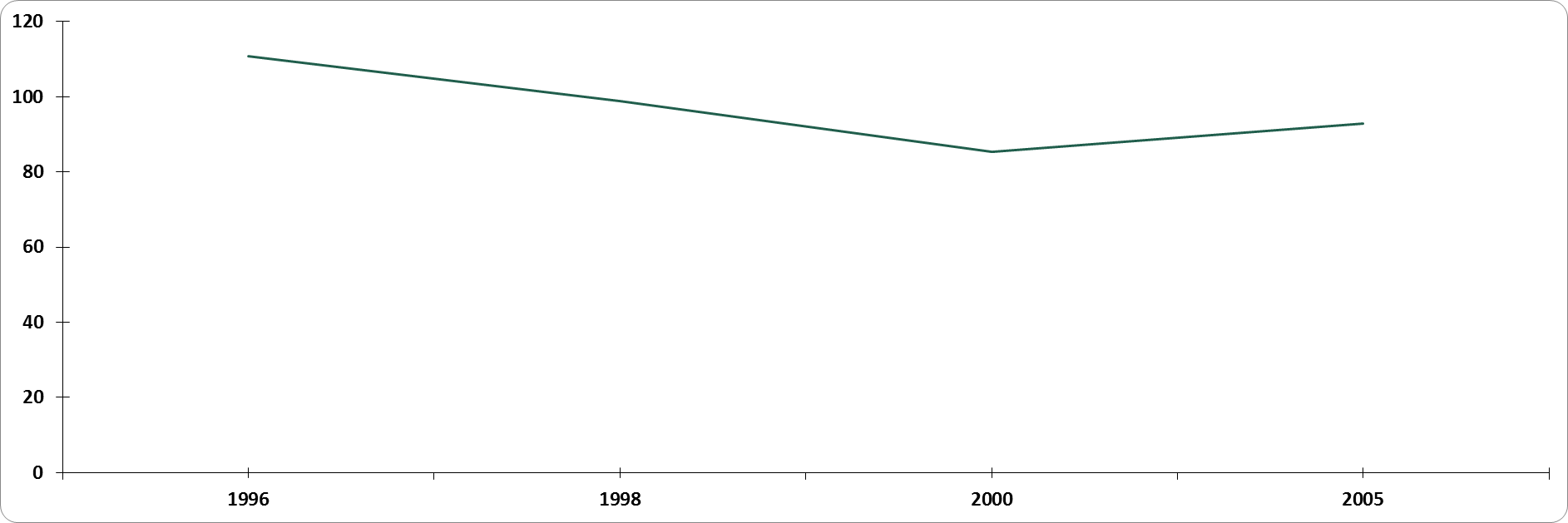
## Researchers in R&D

The number of researchers also has seen a very steep drop from about 152 per million people in 1996 to 115.4 in 1998. The value have again improved to 135.8 in 2005. Both can be considered as very steep drops and also very steep rises; keeping in mind the high fluctuations. The respective changes are -24.07% and 23.45% respectively.



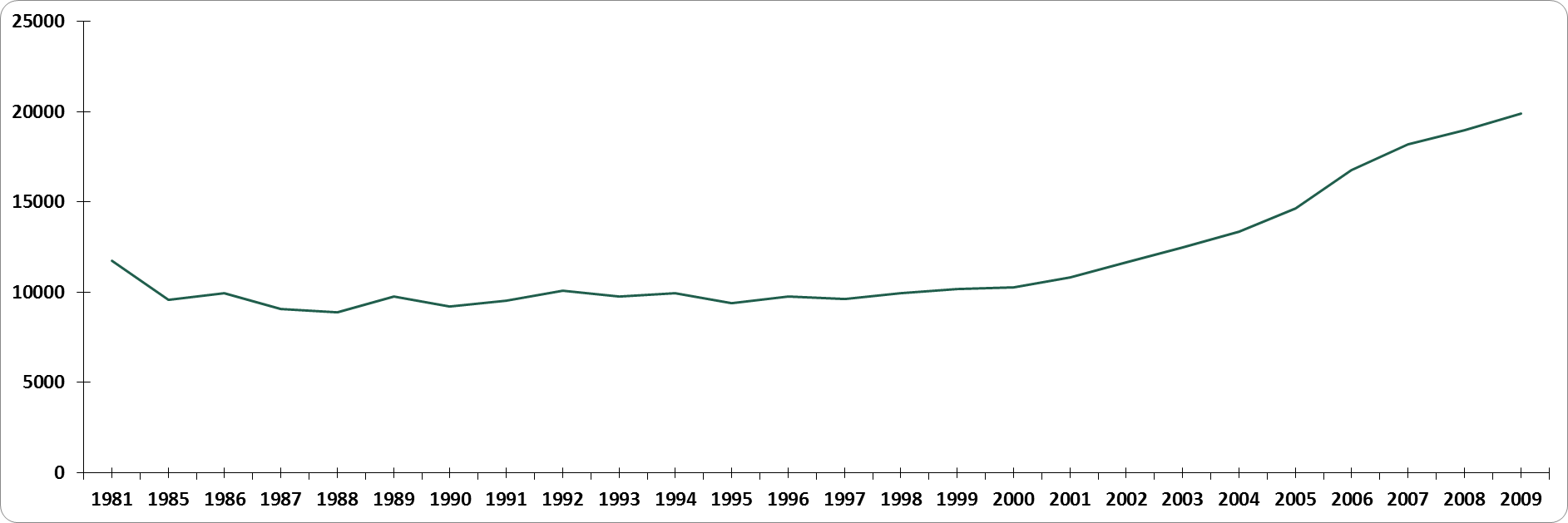
1. **Technicians in R&D**

The number of technicians per million people has taken a very steep drop from about 110.7 in 1996 to 85.4 in 2000; which can be considered a very awkward result in the data. The depreciation and the appreciation rates are close to 13.55% and 10.76% respectively; which are very steep values for the likes of a department.



# Scientific and technical journal articles

The number of technical journal articles released/published by the R&D department of India has been on a constant rise since 2000. Before the year 2000, there has been very unstable with a lot of minor fluctuations. The increase rate is approximately around 5% since 2000; which can be considered as healthy.



1. **Patent Applications**

The number of patent applications filed by the R&D department of India has also been on a constant rise since 2000. But this is not a smooth increase rate as compared to the previous data, since the increase rates are around 25%; which can be considered as very steep.

**REASONS FOR ABNORMAL BEHAVIOUR IN 2000-2001**

From the above visualizations, it can be clearly stated that all the data obtained showed an abnormal behavior in the region 2000-2001. By proper analysis of the financial sector happenings of India in the time-span of 2000-2001; it can be clearly understood that India has undergone a very steep drop in the economic indicators during the above mentioned financial year. So, the expenditure has taken a steep drop till 2000. Immediately after the financial crisis, Indian markets have shown a very appreciable growth or improvement which has resulted in the improvement of the Expenditure data and also the remaining four features as observed. From the analysis of the expenditure data, it would be very logical to assume a depreciation in the remaining features too. The steepness cannot be accurately determined; but Place figures and tables at the top and bottom of columns. Avoid placing them in the middle of columns. Large figures and tables may span across both columns. Figure captions should be below the figures; table captions should appear above the tables. Insert figures and tables after they are cited in the text. Use the abbreviation “Fig. 1” in the text, and “Figure 1” at the beginning of a sentence.

Use 8 point Times New Roman for figure labels. Use words rather than symbols or abbreviations when writing figure-axis labels to avoid confusing the reader. As an example, write the quantity “Magnetization”, or “Magnetization, M”, not just “M”.

If including units in the label, present them within parentheses. Do not label axes only with units. In the example, write “Magnetization (A/m)” or “Magnetization {A[m(1)]}”, not just “A/m”. Do not label axes with a ratio of quantities and units. For example, write “Temperature (K)”, not “Temperature/K”.

## Footnotes

Use footnotes sparingly (or not at all) and place them at the bottom of the column on the page on which they are referenced. Use Times 8-point type, single-spaced.

To help your readers, avoid using footnotes altogether and include necessary peripheral observations in the text (within parentheses, if you prefer, as in this sentence).

Number footnotes separately from reference numbers, and in superscripts. Do not put footnotes in the reference list. Use letters for table footnotes.

1. Table Type Styles

| Table Head | Table Column Head | | |
| --- | --- | --- | --- |
|  | Table column subhead | Subhead | Subhead |
| copy | More table copya |  |  |

a. Sample of a table footnote. *(table footnote)*

1. Example of a figure caption. *(figure caption)*

We suggest that you use a text box to insert a graphic (ideally 300 dpi, with all fonts embedded) because, in an MSW document, this method is somewhat more stable than directly inserting a picture.

To have non-visible rules on Example of a figure caption. *(figure caption)* your frame, use the MSWord pull-down menu, select Format > Borders and Shading > Select ”None”.

# Copyright Forms

You must submit the IEEE Electronic Copyright Form (ECF) as described in your author-kit message. THIS FORM MUST BE SUBMITTED IN ORDER TO PUBLISH YOUR PAPER.

# Acknowledgment

The preferred spelling of the word “acknowledgment” in America is without an “e” after the “g”. Avoid the stilted expression, “One of us (R. B. G.) thanks . . .” Instead, try   
“R. B. G. thanks”. Put applicable sponsor acknowledgments here; DO NOT place them on the first page of your paper or as a footnote.

# References

List and number all bibliographical references in 9-point Times, single-spaced, at the end of your paper. When referenced in the text, enclose the citation number in square brackets, for example: [1]. Where appropriate, include the name(s) of editors of referenced books. The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in “[3]”—do not use “Ref. [3]” or “reference [3]”. Do not use reference citations as nouns of a sentence (e.g., not: “as the writer explains in [1]”).

Unless there are six authors or more give all authors’ names and do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
2. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
4. K. Elissa, “Title of paper if known,” unpublished.
5. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
6. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
7. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.