

Developing Good Research Habits

Rob J Hyndman

20 March 2019

Outline

1 Citing

2 Searching

3 Scripting

4 Writing

Alternative titles

There's more to being a researcher than writing a thesis

Alternative titles

There's more to being a researcher than writing a thesis

What you should know but probably don't

Alternative titles

There's more to being a researcher than writing a thesis

What you should know but probably don't

Listen up, young padawans



Outline

1 Citing

2 Searching

3 Scripting

4 Writing

Managing references

Mendeley

- ▶ Free and on all operating systems
- ▶ Web-version and local version synced
- ▶ Browser extension for adding papers/books
- ▶ Attach notes and annotations to papers.
- ▶ Works with Word, LibreOffice or LaTeX.
- ▶ Generate bibliography automatically
- ▶ Handles all formatting for you.



To install:

- Set up account at
www.mendeley.com
- Download from
www.mendeley.com

Managing references

Zotero

- ▶ Free and on all operating systems
- ▶ Web-version and local version synced
- ▶ Browser extension for adding papers/books
- ▶ Attach notes and annotations to papers.
- ▶ Works with Word, LibreOffice or LaTeX.
- ▶ Generate bibliography automatically
- ▶ Handles all formatting for you.

The logo for Zotero, featuring the word "zotero" in a lowercase, sans-serif font. The letter "z" is red, while the rest of the letters are black.

To install:

- Set up account at
www.zotero.org
- Download from
www.zotero.org

Managing references

Paperpile

- \$3 per month and runs on Google Chrome
- Papers stored on Google Drive
- Browser extension for adding papers/books
- Works with Google Docs or LaTeX.
- Generate bibliography automatically
- Handles all formatting for you.
- Amazingly fast



Paperpile

To install:

- Set up account at paperpile.com
- Download Google chrome browser extension

What to cite?

- Cite what is important.
- Cite (only) what is relevant.
- Avoid lists of gratuitous references.
- Include proper citations for all packages and software you use.



What to cite?

- Cite what is important.
- Cite (only) what is relevant.
- Avoid lists of gratuitous references.
- Include proper citations for all packages and software you use.



```
citation("packagename")
```

Sight what you cite

- Every article cited should be sighted, & preferably read.
- At the very least, check that the article cited really does say what you think it says.
- Type the reference information yourself.
- Don't just cite what other people say about citations.
- Store accurate reference info from the start.
- Give credit where it is due.



Sight what you cite

- Every article cited should be sighted, & preferably read.
- At the very least, check that the article cited really does say what you think it says.
- Type the reference information yourself.
- Don't just cite what other people say about citations.
- Store accurate reference info from the start.
- Give credit where it is due.
 - ▶ Diebold did not invent PITs.
 - ▶ Hyndman did not invent exponential smoothing or ARIMA models.



Outline

1 Citing

2 Searching

3 Scripting

4 Writing

Google Scholar

- Searching for papers
- Use advanced search
- Link GS to your reference manager
- Track citations of key papers
- Star papers for your own library
- Check recommended articles
- Check author profiles, especially highly cited authors
- Create your own GS profile once you have (at least) one paper
- Follow key authors in your area

Outline

1 Citing

2 Searching

3 Scripting

4 Writing

Reproducibility

Not reproducible:

- Data edited in a spreadsheet
- Click and point analysis
- Copy and paste graphs and tables
- Tables typed by hand

Reproducible

- All data edits scripted
- All analysis scripted
- Graphs and tables automatically pulled in to the thesis
- Tables generated with scripts



Reproducibility

Someone should be able to reproduce your thesis without having to guess what software you had installed, what versions, which files do what, etc.

Reproducibility

Someone should be able to reproduce your thesis without having to guess what software you had installed, what versions, which files do what, etc.

- Stay organized.
- One system for doing this using R is to write your thesis in an Rmarkdown file.
- Track software versions

Reproducibility

Someone should be able to reproduce your thesis without having to guess what software you had installed, what versions, which files do what, etc.

- Stay organized.
- One system for doing this using R is to write your thesis in an Rmarkdown file.
- Track software versions

```
{r setup, include=FALSE}  
knitr::opts_chunk$set(echo=FALSE)  
writeLines(capture.output(sessionInfo()),  
          "sessionInfo.txt")
```

Version control

I HEARD YOU DON'T USE VERSION
CONTROL



I ALSO LIKE TO LIVE
DANGEROUSLY

Version control

- `thesis_v1`, `thesis_v2`, etc., is not adequate version control.
- You need to track changes over time, have a *remote* repository, and be able to roll back as required.
- Your repository should contain *everything* required to produce your thesis including computer code, references, writing.
- Your repository should have an obvious structure and be fully documented.
- **Github** solves these problems
- Read “Happy git with R”: happygitwithr.com



Version control with git

- RStudio integrates with github, so version control is built in.
- But github can be used with *any* text-based language including Stata, Python, LaTeX, R, Rmarkdown, markdown, etc.
- Git allows you to:
 - ▶ track changes
 - ▶ experiment in branches
 - ▶ undo
- Github provides:
 - ▶ backup and restore
 - ▶ synchronisation



Some rules for good graphics

- 1 Design graphics to highlight comparisons you want the reader to make.
- 2 Choose scales to facilitate comparisons.
- 3 Use a sensible aspect ratio. width:height of about 1.6, unless axes are on the same scale.
- 4 Prepare graphics in final aspect ratio. Distorted fonts look awful.
- 5 Use vector graphics such as eps or pdf.
- 6 Clear labels and detailed captions.
- 7 Avoid cluttered legends.
- 8 Specify units.
- 9 Label axes.

Outline

1 Citing

2 Searching

3 Scripting

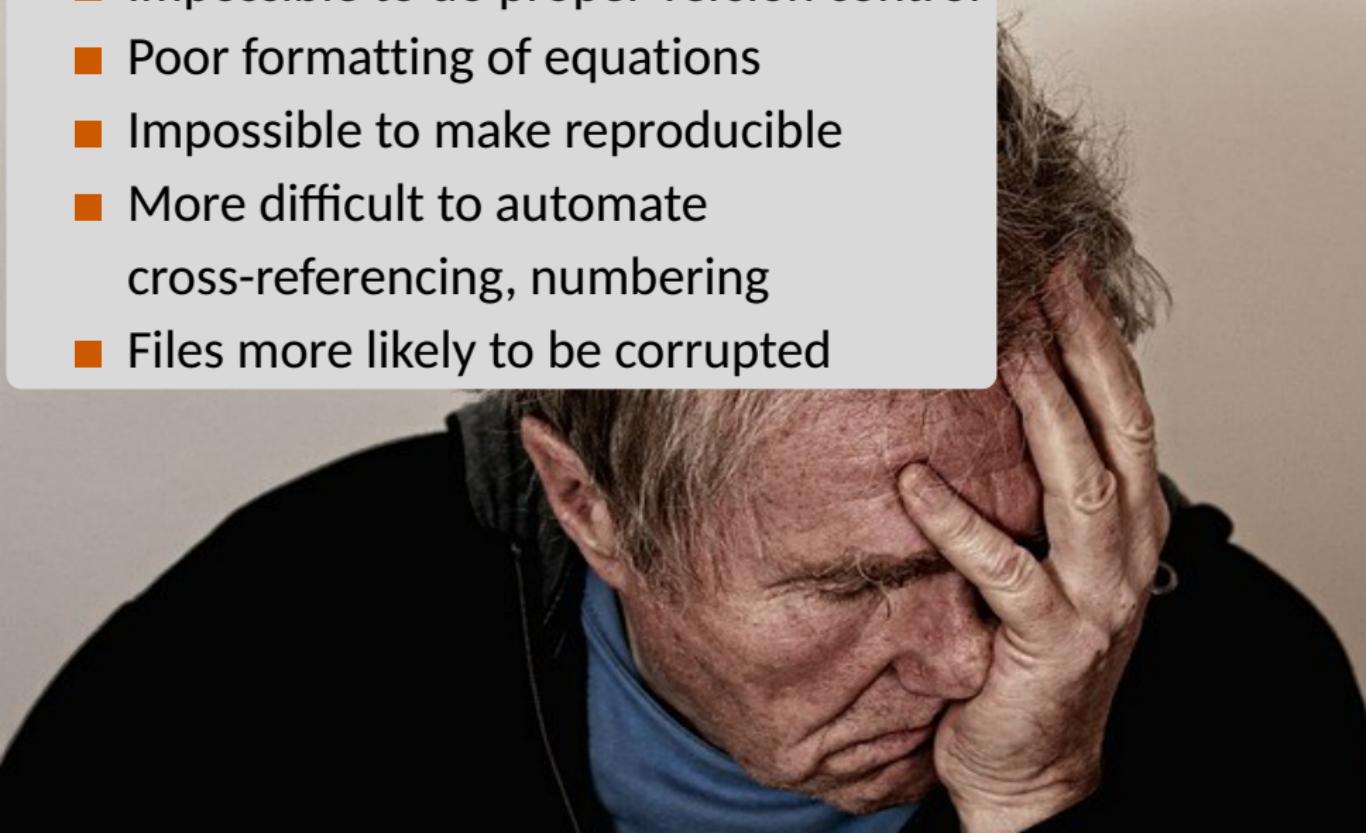
4 Writing

Microsoft Word



Microsoft Word

- Impossible to do proper version control
- Poor formatting of equations
- Impossible to make reproducible
- More difficult to automate
cross-referencing, numbering
- Files more likely to be corrupted



LATEX

To install:

- Download MikTeX,
MacTeX or TeXlive.
- Download TeXStudio
from texstudio.sourceforge.net/

Document processing

- ➔ Free and open-source
- ➔ Available on all operating systems
- ➔ Used by every mathematical publisher
- ➔ Separate content from style
- ➔ Format complex equations
- ➔ Automatic numbering
- ➔ Automatic bibliography
- ➔ Almost every language

RMarkdown

My Thesis

A thesis submitted for the degree of
Bachelor of Commerce (Honours)

by

Sarah Baggart



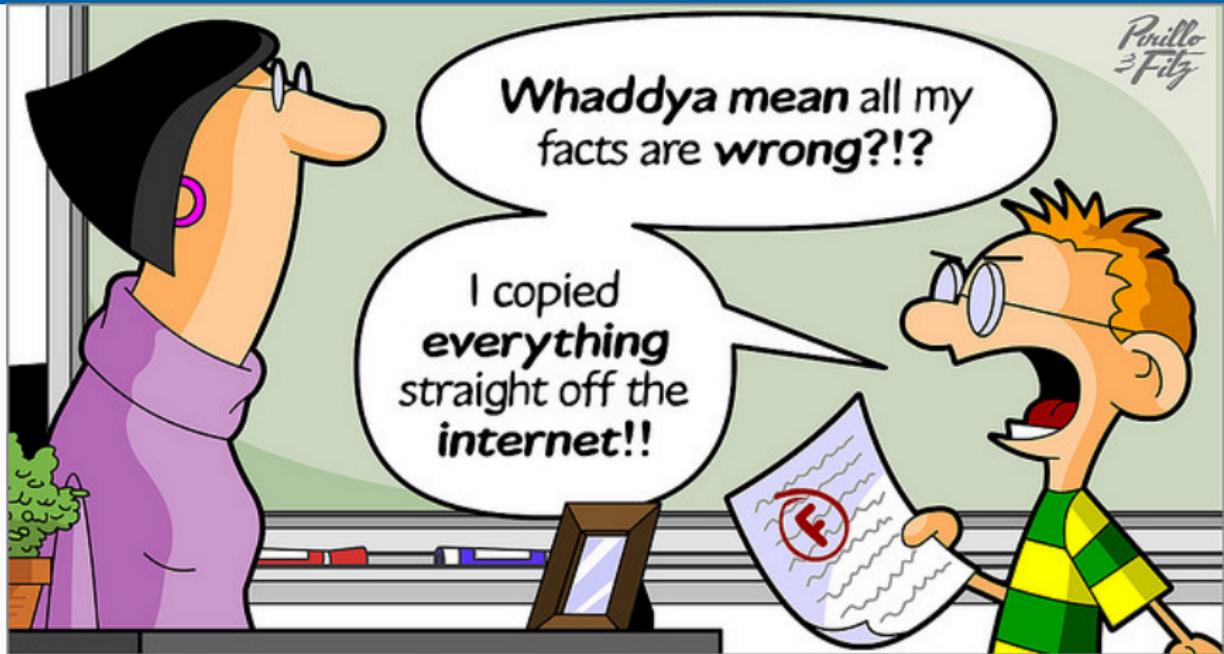
Department of Econometrics and Business Statistics
Monash University
Australia

April 2017

RMarkdown

- Combines R and LaTeX into one system
- Reproducible research
- Monash Honours Template
- Useful for assignments too

Plagiarism



- Never plagiarise from other papers — not even sentence fragments. Use your own words.

Words to avoid

According to Andrew Gelman

- Note that
- Interestingly
- It is interesting to note that
- Obviously
- It is clear that
- very
- quite
- of course
- Notice that

Writing an abstract

- 1 What did you do?
- 2 Why did you do it? What question were you trying to answer?
- 3 How did you do it? State your methods.
- 4 What did you learn? State your major results.
- 5 Why does it matter? Point out at least one significant implication.

Writing an abstract

STATISTICAL ANALYSIS AND DATA MINING

Original Article

Visualizing statistical models: Removing the blindfold

Abstract

Visualization can help in model building, diagnosis, and in developing an understanding about how a model summarizes data. This paper proposes three strategies for visualizing statistical models: (i) display the model in the data space, (ii) look at all members of a collection, and (iii) explore the process of model fitting, not just the end result. Each strategy is accompanied by examples, including MANOVA, classification algorithms, hierarchical clustering, ensembles of linear models, projection pursuit, self-organizing maps, and neural networks.

Writing an abstract



International Journal of Fore

Volume 23, Issue 2, April–June 2007, Pag

- 1 What did you do?
- 2 Why did you do it?
What question were
you trying to answer?
- 3 How did you do it?
State your methods.
- 4 What did you learn?
State your major
results.
- 5 Why does it matter?
Point out at least one
significant implication.

Bias in macroeconomic forecasts

Roy Batchelor

Show more

<https://doi.org/10.1016/j.ijforecast.2007.01.004>

Abstract

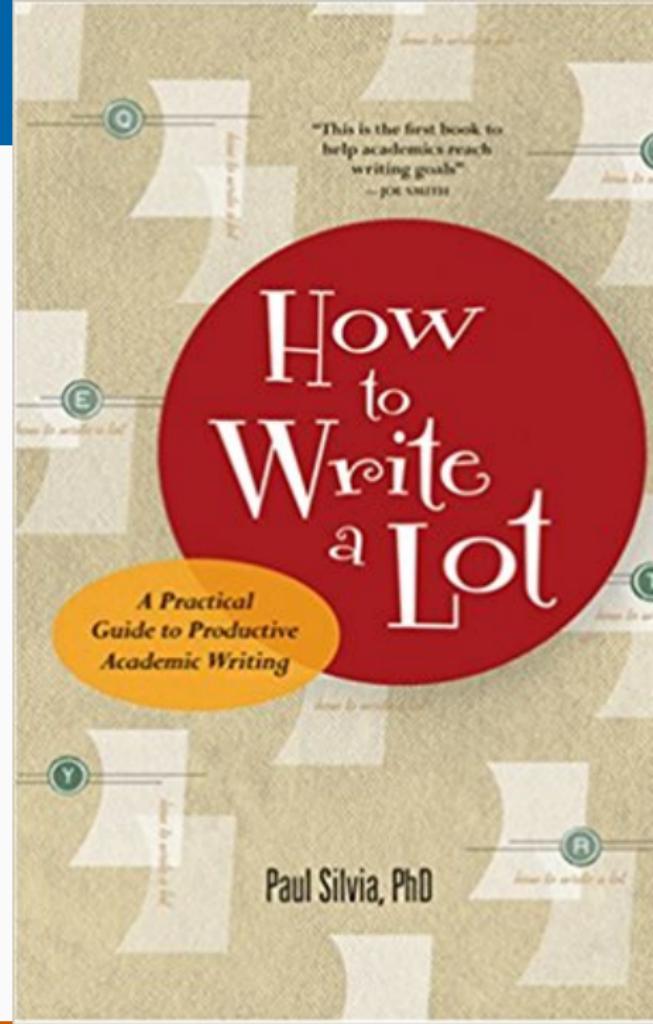
This paper documents the presence of systematic bias in growth and inflation forecasts of private sector forecasters in the years 1990–2005. The data come from the Consensus Economics forecasting service, and bias is measured by significance using parametric fixed effect panel regression and nonparametric tests on accuracy ranks. We examine individual countries and forecasters to establish whether the bias reflects inefficient use of information, or whether it reflects financial, reputational and other incentives operating in several G7 countries – Japan, Italy, Germany and the United States. We find evidence of a change in the trend growth rate. In all countries the standard tests for rationality are inappropriate, and the optimism in the consensus forecast is inevitable. We learn about the new trend. In all countries there is evidence that forecasters converge on the consensus forecast. The persistent optimism of some forecasters, and the pessimism of others, is not consistent with the predictions of models that have become popular in the finance and econometrics literature.

Writing an abstract

- Should be a stand-alone summary. It is the only thing most people will read.
- No references or citations.
- Write in the past tense.
- Be explicit, precise and concise.
- Stick to a single paragraph.
- Restrict background information to a sentence or two at most.
- Make sure that your abstract is consistent with what you reported in the thesis.
- Write the abstract last

How to write a lot

- Block out a regular writing time (e.g., 1 hour every morning).
- Write something every day.
- Set a word or paragraph goal for each session.
- Practice makes perfect better.
- Writing clarifies thinking.
- Build on a scaffold.



Resources

- **Slides:**

robjhyndman.com/seminars/research_habits

- **Github:**

happygitwithr.com

- **Rmarkdown thesis template:**

github.com/robjhyndman/MonashHonoursThesis