Developing Good Research Habits

Rob J Hyndman

22 August 2018

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 **Plotting**

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.

MENDELEY

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.

zotero

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

P aperpile

To install:
- ■ Set up account at paperpile.com
- ■ Download Google chrome browser extension

6

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

Use Google Scholar for more than searching

Use Google Scholar for more than searching

Use Google Scholar for more than searching

- Link GS to your reference manager.
- Keep everything you are ever likely to cite in your reference manager.

Use Google Scholar for more than searching

Google Scholar

feature based anomaly detection in streaming data

Articles

About 798 results (0.03 sec)

My profile   My library

Any time
Since 2018
Since 2017
Since 2014
Custom range...

Sort by relevance
Sort by date

☐ include patents
☐ include citations

✉ Create alert

Voila: Visual Anomaly Detection and Monitoring with Streaming Spatiotemporal Data
N Cao, C Lin, Q Zhu, YR Lin, X Teng... - IEEE transactions on ..., 2018 - ieeexplore.ieee.org   Paperpile
... feature descriptors along multiple facets, with the advantage to examine features associated
with ... from a traffic tensor; however, like other methods, the produced feature matrices and ... Based
on these system requirements, we developed Voila, an inte- grated system solution for ...
☆ 🗩 Cited by 2   Related articles   All 8 versions

A New Model for Evaluating Range-Based Anomaly Detection Algorithms   [PDF] arxiv.org
N Tatbul, TJ Lee, S Zdonik, J Gottschlich - arXiv preprint arXiv:1803.03639, 2018 - arxiv.org   Paperpile
... We aim to demonstrate two key features of our model: (i) its expressive power for ... Multiple
ranged-based stochastic anomalies appear with inverted negative values ... Anomaly Detection
Algorithm: We trained a TensorFlow- implemented version of LSTM-AD, a long short-term mem ...
☆ 🗩 All 2 versions

[PDF] **Anomaly Detection in Streaming Nonstationary Temporal Data**   [PDF] researchgate.net
PD Talagala, RJ Hyndman, K Smith-Miles... - 2018 - researchgate.net   Paperpile
... (the same features as were defined in step 1 of Algorithm 1) from each of these n time
series in W[t −w,t]. This produces an n×m feature matrix, Mtest. 3. Project Mtest, onto the
2D PC space of Dt0 ... Since the test statistic based on the integrated ...
☆ 🗩 ≫

[PDF] **Anomaly Detection in Finance: Editors' Introduction**   [PDF] mlr.press
A Anandakrishnan, S Kumar... - ... - 2018 - proceedings.mlr.press   Paperpile
... Feature engineering strategies for credit card fraud detection. Expert Systems With Ap- plications,
51:134–142, 2016 ... Pd-fds: Purchase density based online credit card fraud detection system ...
A framework for constructing features and models for intrusion detection systems ...
☆ 🗩 ≫

Detecting anomalous emotion through big data from social networks based on

Use Google Scholar for more than searching

Google Scholar

feature based anomaly detection in streaming data

Articles   About 798 results (0.03 sec)   My profile   My library

Any time
Since 2018
Since 2017
Since 2014
Custom range...

Sort by relevance
Sort by date

☐ include patents
☐ include citations

✉ Create alert

Voila: Visual Anomaly Detection and Monitoring with Streaming Spatiotemporal Data
N Cao, C Lin, Q Zhu, YR Lin, X Teng... - IEEE transactions on ..., 2018 - ieeexplore.ieee.org   Paperpile
... feature descriptors along multiple facets, with the advantage to examine features associated with ... from a traffic tensor; however, like other methods, the produced feature matrices and ... Based on these system requirements, we developed Voila, an inte- grated system solution for ...
☆ 🗨 Cited by 2   Related articles   All 8 versions

A New Model for Evaluating Range-Based Anomaly Detection Algorithms   [PDF] arxiv.org
N Tatbul, TJ Lee, S Zdonik, J Gottschlich - arXiv preprint arXiv:1803.03639, 2018 - arxiv.org   Paperpile
... We aim to demonstrate two key features of our model: (i) its expressive power for ... Multiple ranged-based stochastic anomalies appear with inverted negative values ... Anomaly Detection Algorithm: We trained a TensorFlow- implemented version of LSTM-AD, a long short-term mem ...
☆ 🗨 All 2 versions

[PDF] **Anomaly Detection in Streaming Nonstationary Temporal Data**   [PDF] researchgate.net
PD Talagala, RJ Hyndman, K Smith-Miles... - 2018 - researchgate.net   ✅ Paperpile 📁
... (the same ... as ... defined in step 1 of Algorithm 1) from each of these n time series in W[t −w,t]. This produces an n×m feature matrix, Mtest. 3. Project Mtest, onto the 2D PC space of Dt0 ... Since the test statistic based on the integrated ...
☆ 🗨 ≫

[PDF] **Anomaly Detection in Finance: Editors' Introduction**   [PDF] mlr.press
A Anandakrishnan, S Kumar... - ... Anomaly Detection ..., 2018 - proceedings.mlr.press   Paperpile
... Feature engineering strategies for credit card fraud detection. Expert Systems With Ap- plications, 51:134–142, 2016 ... Pd-fds: Purchase density based online credit card fraud detection system ... A framework for constructing features and models for intrusion detection systems ...
☆ 🗨 ≫

Detecting anomalous emotion through big data from social networks based on

Use Google Scholar for more than searching

Google Scholar

Rob J Hyndman ✎

Professor of Statistics, Monash University
Verified email at monash.edu - Homepage

Forecasting Time series Applied Statistics Machine learning Exploratory data analysis

✉ FOLLOW

TITLE	CITED BY	YEAR
Forecasting methods and applications S Makridakis, SC Wheelwright, RJ Hyndman John Wiley & Sons	4798 *	1998
Another look at measures of forecast accuracy RJ Hyndman, AB Koehler International journal of forecasting 22 (4), 679-688	1815	2006
Automatic time series forecasting: the forecast package for R RJ Hyndman, Y Khandakar Journal of Statistical Software	1149	2007
Forecasting: principles and practice RJ Hyndman, G Athanasopoulos OTexts	776	2014
25 years of time series forecasting JG De Gooijer, RJ Hyndman International journal of forecasting 22 (3), 443-473	682	2006
Forecasting with exponential smoothing: the state space approach RJ Hyndman, AB Koehler, JK Ord, RD Snyder Springer Verlag	675	2008
Sample quantiles in statistical packages RJ Hyndman, Y Fan The American Statistician 50 (4), 361-365	636	1996
Detecting trend and seasonal changes in satellite image time series	581	2010

Cited by   VIEW ALL

	All	Since 2013
Citations	19636	12091
h-index	54	43
i10-index	113	91

2011 2012 2013 2014 2015 2016 2017 2018

3100
2325
1550
775
0
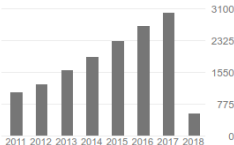
Co-authors   EDIT

Ralph Snyder
Associate Professor, Department... ›

George Athanasopoulos
Monash University ›

Han Lin Shang
Research School of Finance, Act... ›

Keith Ord
Professor of Business Statistics, ... ›

Bircan Erbas
La Trobe University ›

Use Google Scholar for more than searching

Google Scholar

Search: feature based anomaly detection in streaming data

Articles   About 798 results (0.03 sec)

Any time
Since 2018
Since 2017
Since 2014
Custom range...

Sort by relevance
Sort by date

☐ include patents
☐ include citations

✉ Create alert

Voila: Visual **Anomaly Detection** and Monitoring with **Streaming** Spatiotemporal Data
N Cao, C Lin, Q Zhu, YR Lin, X Teng... - IEEE transactions on ..., 2018 - ieeexplore.ieee.org   ⊕ Paperpile
... **feature** descriptors along multiple facets, with the advantage to examine **features** associated with ... from a traffic tensor; however, like other methods, the produced **feature** matrices and ... **Based** on these system requirements, we developed Voila, an inte- grated system solution for ...
☆ ⁹⁹   Cited by 2   Related articles   All 8 versions

A New Model for Evaluating Range-**Based Anomaly Detection** Algorithms
[PDF] arxiv.org
N Tatbul, TJ Lee, S Zdonik, J Gottschlich - arXiv preprint arXiv:1803.03639, 2018 - arxiv.org   ⊕ Paperpile
... We aim to demonstrate two key **features** of our model: (i) its expressive power for ... Multiple ranged-**based** stochastic **anomalies** appear with inverted negative values ... **Anomaly Detection** Algorithm: We trained a TensorFlow- implemented version of LSTM-AD, a long short-term mem ...
☆ ⁹⁹   All 2 versions

[PDF] **Anomaly Detection** in **Streaming** Nonstationary Temporal Data
[PDF] researchgate.net
PD Talagala, RJ Hyndman, K Smith-Miles... - 2018 - researchgate.net   ✅ Paperpile 📄
... (the same **features** as were defined in step 1 of Algorithm 1) from each of these n time series in W[t −w,t]. This produces an n×m **feature** matrix, Mtest. 3. Project Mtest, onto the 2D PC space of Dt0 ... Since the test statistic **based** on the integrated ...
☆ ⁹⁹ ᐅᐅ

[PDF] **Anomaly Detection** in Finance: Editors' Introduction
[PDF] mlr.press
A Anandkrishnan, S Kumar... - ... 2018 - proceedings.mlr.press   ⊕ Paperpile
... **Feature** engineering strategies for credit card fraud **detection**. Expert Systems With Ap- plications, 51:134–142, 2016 ... Pd-fds: Purchase density **based** online credit card fraud **detection** system ... A framework for constructing **features** and models for intrusion **detection** systems ...
☆ ⁹⁹ ᐅᐅ

Detecting anomalous emotion through big **data** from social networks **based** on

Use Google Scholar for more than searching

Articles

About 798 results (0.03 sec)

My profile   ★ My library   🔔

Voila: Visual Anomaly Detection and Monitoring with Streaming Spatiotemporal Data

N Cao, C Lin, Q Zhu, YR Lin, X Teng... - IEEE transactions on ..., 2018 - ieeexplore.ieee.org   Paperpile
... feature descriptors along multiple facets, with the advantage to examine features associated with ... from a traffic tensor; however, like other methods, the produced feature matrices and ... Based on these system requirements, we developed Voila, an inte- grated system solution for ...
☆ 🔗   Cited by 2   Related articles   All 8 versions

A New Model for Evaluating Range-Based Anomaly Detection Algorithms   [PDF] arxiv.org

N Tatbul, TJ Lee, S Zdonik, J Gottschlich - arXiv preprint arXiv:1803.03639, 2018 - arxiv.org   Paperpile
... We aim to demonstrate two key features of our model: (i) its expressive power for ... Multiple ranged-based anomalies appear with inverted negative values ... Anomaly Detection Algorithm: We trained a TensorFlow- implemented version of LSTM-AD, a long short-term mem ...
☆ 🔗   All 2 versions

[PDF] **Anomaly Detection in Streaming Nonstationary Temporal Data**   [PDF] researchgate.net

PD Talagala, RJ Hyndman, K Smith-Miles... - 2018 - researchgate.net   ✅ Paperpile 📁
... (the same features as were defined in step 1 of Algorithm 1) from each of these n time series in W[t −w,t]. This produces an n×m feature matrix, Mtest. 3. Project Mtest, onto the ... space of Dt0 ... Since the test statistic based on the integrated ...
☆ 🔗 🔗

[PDF] **Anomaly Detection in Finance: Editors' Introduction**   [PDF] mlr.press

A Anandakrishnan, S Kumar... - ... 2018 - proceedings.mlr.press   Paperpile
... Feature engineering strategies for credit card fraud detection. Expert Systems With Ap- plications, 51:134–142, 2016 ... Pd-fds: Purchase density based online credit card fraud detection system ... A framework for constructing features and models for intrusion detection systems ...
☆ 🔗 🔗

Detecting anomalous emotion through big data from social networks based on ...

Use Google Scholar for more than searching

Use Google Scholar for more than searching

Use Google Scholar for more than searching

My library   ☐ ✏️ 🗑️ ⬇️ 🏷️   🎓 My profile   ★ My library   🔔

My library
Cited by me
Trash
Manage labels...

Any time
Since 2018
Since 2017
Since 2014
Custom range...

☐ **Effective forecasting and judgmental adjustments: an empirical**   [PDF] academia.edu
evaluation and strategies for improvement in supply-chain planning
R Fildes, P Goodwin, M Lawrence... - International journal of ..., 2009 - Elsevier [Paperpile]
Abstract Demand forecasting is a crucial aspect of the planning process in supply-chain
companies. The most common approach to forecasting demand in these companies
involves the use of a computerized forecasting system to produce initial forecasts and the ...
 99   Cited by 249   Related articles   All 12 versions

☐ **Judgmental forecasting: A review of progress over the last 25 years**   [PDF] bilkent.edu.tr
M Lawrence, P Goodwin, M O'Connor... - International Journal of ..., 2006 - Elsevier [✓ Paperpile]
Abstract The past 25 years has seen phenomenal growth of interest in judgemental
approaches to forecasting and a significant change of attitude on the part of researchers in
the role of judgement. While previously judgement was thought to be the enemy of accuracy ...
 99   Cited by 342   Related articles   All 17 versions

☐ **Forecasting: principles and practice**   [PDF] robjhyndman.com
RJ Hyndman, G Athanasopoulos - 2014 - books.google.com [✓ Paperpile 📁]
Forecasting is required in many situations. Stocking an inventory may require forecasts of
demand months in advance. Telecommunication routing requires traffic forecasts a few
minutes ahead. Whatever the circumstances or time horizons involved, forecasting is an ...
 99   Cited by 776   Related articles   All 5 versions   »

☐ **Prediction by supervised principal components**   [PDF] jstor.org
E Bair, T Hastie, D Paul... - Journal of the American ..., 2006 - amstat.tandfonline.com [✓ Paperpile]
In regression problems where the number of predictors greatly exceeds the number of
observations, conventional regression techniques may produce unsatisfactory results. We
describe a technique called supervised principal components that can be applied to this ...
 99   Cited by 535   Related articles   All 24 versions

☐ Common principal components & related multivariate models

Use Google Scholar for more than searching

Use Google Scholar for more than searching

Outline

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.

IJF graphics examples
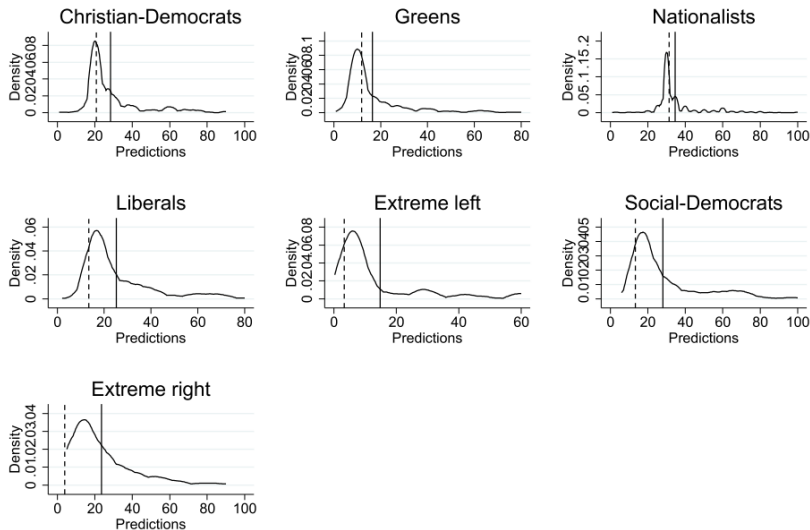
Fig. 1. Electoral prediction results.
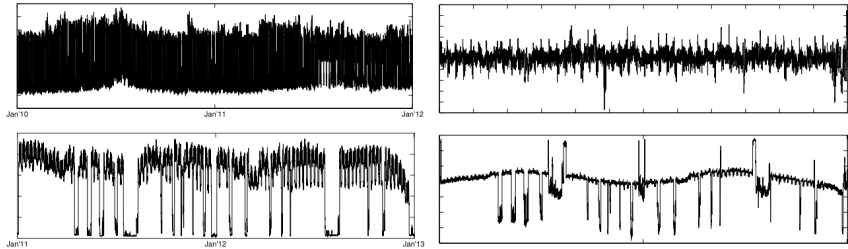
Fig. 2. Original (top left) and deseasonalized (top right) time series for a non-regime-switching load profile, versus original (bottom left) and deseasonalized (bottom right) time series of a customer with a regime-switching electricity load.

IJF graphics examples
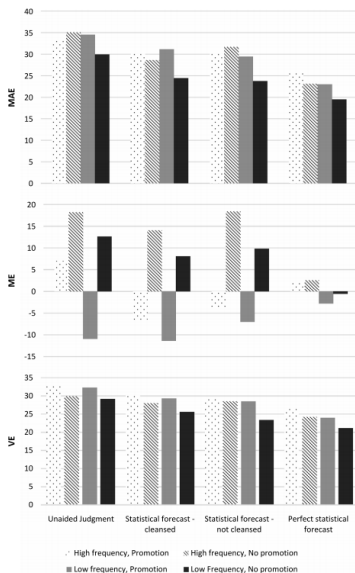
Fig. 4. Error scores associated with the different forecasting conditions studied in the two experiments: MAE (upper panel); ME (central panel); VE (lower panel).

Outline

LaTeX

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

To install:

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

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

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

- `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.
- Someone should be able to reproduce your thesis from your repository without having to guess what software you had installed, what software versions, which files do what, etc.
- **Github** solves these problems

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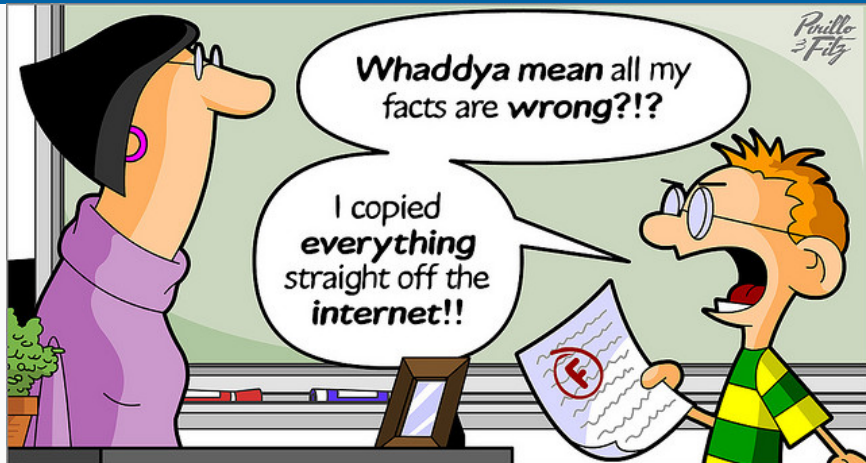
- Never plagiarise from other papers — not even sentence fragments. Use your own words.

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

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

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

Bias in macroeconomic forecasts

Roy Batchelor

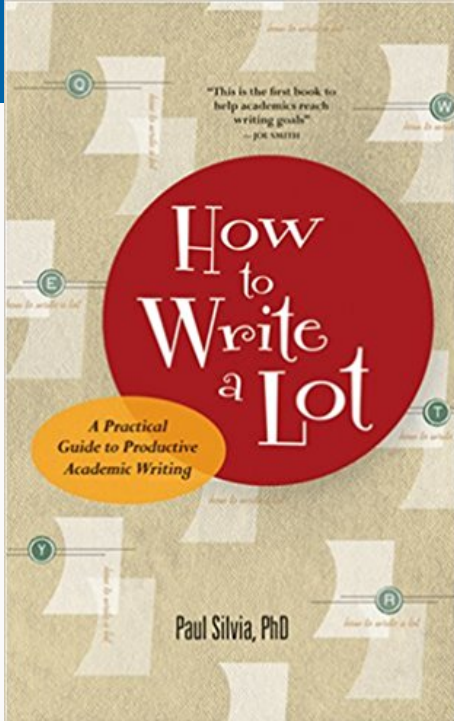⊞ Show more

Get rights and content

Abstract

This paper documents the presence of systematic bias in the real GDP and inflation forecasts of private sector forecasters in the G7 economies in the years 1990–2005. The data come from the monthly Consensus Economics forecasting service, and bias is measured and tested for significance using parametric fixed effect panel regressions and nonparametric tests on accuracy ranks. We examine patterns across countries and forecasters to establish whether the bias reflects the inefficient use of information, or whether it reflects a rational response to financial, reputational and other incentives operating for forecasters. In several G7 countries – Japan, Italy, Germany and France – there is evidence of a change in the trend growth rate. In these circumstances, standard tests for rationality are inappropriate, and a bias towards optimism in the consensus forecast is inevitable as rational forecasters learn about the new trend. In all countries there is evidence that individual forecasters converge on the consensus forecast too slowly. However, the persistent optimism of some forecasters, and the persistent pessimism of others, is not consistent with the predictions of models of "rational bias" that have become popular in the finance and economics 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.

"This is the first book to help academics reach writing goals."
— JOE SMITH

How to Write a Lot

A Practical Guide to Productive Academic Writing

Paul Silvia, PhD

Resources

- **Slides**:

  `robjhyndman.com/seminars/Honours2018`

- **Github**:

  `happygitwithr.com`

- **Rmarkdown thesis template**:

  `github.com/robjhyndman/MonashHonoursThesis`