

---

# **Econometrics Navigator Documentation**

***Release 0.0.5***

**Evgeny Pogrebnyak**

**Nov 08, 2019**



## CONTENTS



## ANALYSIS OF VARIANCE (ANOVA)

- ANOVA can mean several things: actual decomposition of variance, comparing the group means or representation of regression results
- Boils down to a regression with dummy (categorical) variables
- Heavy traction in terminology from design of experiments (see definitions section [here](#))
- Standardised result tables with SS, DF, MSS, F, p
- Frightening multitude of R packages
- May want to look at a simple reference case [here](#)

Quote:

*ANOVA can be seen as “syntactic sugar” for a special subgroup of linear regression models. ANOVA is regularly used by researchers who are not statisticians by training. They are now “institutionalized” and its hard to convert them back to using the more general representation [suncoolsu](#)*

### 1.1 Code examples

- <https://stackoverflow.com/questions/25537399/anova-in-python-using-pandas-dataframe-with-statsmodels-or-scipy>
- <http://www.statsmodels.org/devel/anova.html>
- <https://stats.stackexchange.com/a/175265/211794>
- also possibly in Think Stats and Hadley Wickham

### 1.2 Links

Intro by Jim

Cross-Validated has several general discussions:

- [why-is-anova-taught-used-as-if-it-is-a-different-research-methodology-compared](#)
- [how-to-visualize-what-anova-does](#)
- [how-to-interpret-f-and-p-value-in-anova](#)
- [good-resource-to-understand-anova-and-ancova](#)

... followed by ANOVA vs regression:

- [difference-between-regression-analysis-and-analysis-of-variance](#)

- [why-is-anova-equivalent-to-linear-regression](#)

NIST Handbook deals with ANOVA assumptions and interpretations, as well as provides reference datasets:

- [The one-way ANOVA model and assumptions](#)
- [Interpretation of the ANOVA table](#)
- [Reference datasets and regression results](#)

Very simple and illustrative case NIST reference case [here](#).

Comoact Julia package [ANOVA.jl](#) at about 150 lines of code, but not as much documentation yet.

ANOVA is again a case where [Russian wikipedia](#) is more concise and clear on the subject.

‘Types’ of sum of squares and associated confusion:

- <https://mcfromnz.wordpress.com/2011/03/02/anova-type-iiiiii-ss-explained/>
- [https://rcompanion.org/rcompanion/d\\_04.html](https://rcompanion.org/rcompanion/d_04.html)
- <https://stats.stackexchange.com/questions/20452/how-to-interpret-type-i-type-ii-and-type-iii-anova-and-manova>

## 1.3 References

Gelman, A. (2005). Analysis of variance: why it is more important than ever (with discussion). *Annals of Statistics* 33, 1–53. doi:10.1214/009053604000001048

## **BIAS-VARIANCE TRADEOFF**

The bias variance trade-off is maybe not an ideal name, it should maybe have better been called interpolation/extrapolation trade-off. Anyway, the motivation for the name is that that when adding more parameters / complexity, you have

- Less systematic error (bias) in your model (supposedly, because it is more flexible, I would argue it depends on what you call error / bias)
- More variance in the estimation of the model parameters (because it is more flexible)

Florian Hartig at Cross Validated. [What is the difference between bias and residuals?](#)





## BOOTSTRAP

Bootstrapping is a method to construct empiric distributions of various statistics (mean, confidence intervals, deviation, etc) by using repeated sampling from an observed dataset.

A little magic is *why exactly* taking random samples like  $[1, 1, 2]$ ,  $[3, 2, 2]$ ,  $[2, 1, 3]$ , etc is a good idea to approximate properties of a dataset  $[1, 2, 3]$ .

Bootstrap originally proposed by Bradley Efron in 1979. For formal introduction see Horowitz chapter in *Handbook of Econometrics* (2001) and a usage overview by MacKinnon 2006.

### 3.1 Toy example

Bootstrap confidence intervals by Jeremy Orloff and Jonathan Bloom, pp. 4-6 provides the following basic code example for bootstrap. Their full code for this exercise is [here](#).

```
# Bootstrap
# Adapted from https://math.mit.edu/~dav/05.dir/class24-empiricalbootstrap.r
cat("Example. Empirical bootstrap confidence interval for the mean.",'\n')
x = c(30,37,36,43,42,43,43,46,41,42)
n = length(x)
set.seed(1) # for repeatability

# sample mean
xbar = mean(x)
cat("data mean = ",xbar,'\n')
nboot = 20
# Generate 20 bootstrap samples, i.e. an n x 20 array of
# random resamples from x.
tmpdata = sample(x,n*nboot, replace=TRUE)
bootstrapsample = matrix(tmpdata, nrow=n, ncol=nboot)

# Compute the means xbar*
xbarstar = colMeans(bootstrapsample)

# Compute delta* for each bootstrap sample
deltastar = xbarstar - xbar

# Find the 0.1 and 0.9 quantiles for deltastar
d = quantile(deltastar,c(0.1,0.9))

# Calculate the 80% confidence interval for the mean.
ci = xbar - c(d[2],d[1])
cat('Bootstrap confidence interval: [',ci,']','\n')
```

### 3.1.1 Bootstrap do's and don'ts by Anna Mikusheva

- If you have a pivotal statistic, bootstrap can give a refinement. So, if you have choice of statistics, bootstrap a pivotal one.
- Bootstrap may fix a finite-sample bias, but cannot help if you have inconsistent estimator.
- In general, if something does not work with traditional asymptotics, the bootstrap cannot fix your problem. For example, if we have an inconsistent estimate, the bootstrap bias correction does not fix anything.
- Bootstrap could not fix the following problems: weak instruments, parameter on a boundary, unit root, persistent regressors.
- Bootstrap requires re-centering (the null hypothesis to be true).

Source: [MIT lecture notes](#)

## 3.2 More links (preliminary)

- <https://core.ac.uk/download/pdf/6494253.pdf>
- <https://github.com/wmutschl/GMMIndirectInferenceBootstrap>
- <https://www.schmidheiny.name/teaching/bootstrap2up.pdf>
- <http://rosetta.ahmedmoustafa.io/bootstrap/>
- <http://www.cs.cornell.edu/courses/cs1380/2018sp/textbook/chapters/11/2/bootstrap.html>
- <http://economics.fundamentalfinance.com/bootstrap.php>

## 3.3 Editor notes

- Russian article in Wikipedia on bootstrap is much more concise and understandable than [English one](#).

## CAUSATION, CAUSALITY

---

**Important:** Correlation is not causation.

---

- causality (not ‘casuality’)

### 4.1 Book of Why by Judea Pearl

### 4.2 History

- Pearl, J. (2014). TRYGVE HAAVELMO AND THE EMERGENCE OF CAUSAL CALCULUS. *Econometric Theory*, 31(1), 152–179. <https://doi.org/10.1017/s0266466614000231>



**CENTRAL LIMIT THEOREM, CLT**



## MAXIMUM LIKELIHOOD

The probability density function  $p = f(x, \theta)$  tells you a probability of occurrence of a random value near  $x$ . Likelihood is essentially a reverse operation of estimating unknown parameter  $\theta$  from the same equation using  $p$  and  $x$ .

### 6.1 Lead by example

- Observations
- Probability of observations
- Observed sample is considered the most likely one
- Maximisation of probability allows to compute distribution parameters

### 6.2 Generalisation

We usually denote a set of parameters like  $\theta$  and  $\phi$  as  $\theta$ , a vector of parameters. Our task is to estimate parameter  $\theta$  given:

- a sample of observations of random variable  $X = (x_1, x_2, \dots, x_n)$ , and
- a pre-defined probability density function  $f(x, \theta)$ .

**Solution steps:**

1. collect observations  $X = (x_1, x_2, \dots, x_n)$
2. make a decision which probability density function  $f(x, \theta)$  is appropriate for this data
3. compose joint probability of observations as a function of  $\theta$ :  $L(\theta) = f(x_1, \theta) \cdot f(x_2, \theta) \cdot \dots \cdot f(x_n, \theta)$ .
4. Come to terms with a principle “if we observed this event, we consider it was the most probable outcome of all possible events in this distribution”
5. Find which  $\theta$  maximises joint probability of observations

### 6.3 Code

Python code below relies on `scipy.optimize.minimize` solver to find parameters of a normal distribution based on two measurements of mice weights. It can be easily applied to more observations.

```

"""
Maximum likelihood with two mice.
"""

import numpy as np
from scipy.optimize import minimize

def dnorm(x, mu=0, sigma=1):
    """Normal distribution probability density fucntion."""
    const = 1 / (sigma * np.sqrt(2 * np.pi))
    power = - (x - mu)**2 / (2 * sigma**2)
    return const * np.exp(power)

def log_likelihood(observed_x):
    """Sum of logs of probability densities at *observed_x*.
    Return:
        function of mu and sigma
    """
    def foo(mu, sigma):
        logs = [np.log(dnorm(x, mu, sigma)) for x in observed_x]
        return sum(logs)
    return foo

def maximise(f, start_mu, start_sigma):
    """Return mu and lambda, which maximise *f*."""
    f = lambda p: -1 * l_func(mu=p[0], sigma=p[1])
    res = minimize(f, x0=[start_mu, start_sigma])
    return res.x[0], res.x[1]

# two mice weigths are given, similar to https://www.ncbi.nlm.nih.gov/pmc/articles/
# ↪ PMC6143748/
events = [30, 50]
# construct likelihood as a function of unknown mu and sigma
l_func = log_likelihood(events)
# run maximisation procedure
# attention: need a sensible pick for start variables, eg (0, 1) will fail
estimated_mu, estimated_sd = maximise(l_func, start_mu=30, start_sigma=3)

# test outcomes
#estimated_mu is 39.99999527669165
assert np.isclose(estimated_mu, 40)
#estimated_sd is 9.999976480910071
assert np.isclose(estimated_sd, 10)

# Result: observed values [30, 50] were most likely coming from
#         normal distribution with parameters =40 and =10.

```

Other code examples:

- Annotated R code by Andrew Collier (2013)
- Doing Maximum Likelihood Estimation by Hand in R by John Myles White (2010)
- Julia vs R vs Python Simple Optimization by Zhuo Jiadai (2018)



## 6.4 Links

- [Nice video with weight of mice](#)
- [Maximum likelihood estimation in layman terms](#)
- [Why is maximum likelihood estimation considered to be a frequentist technique](#)
- [Very accessible math treatment \(in Russian\)](#)
- [Tourist sees a fountain \(also in Russian\)](#)



---

CHAPTER  
**SEVEN**

---

**MODE**



## ORDINARY LEAST SQUARES, OLS

OLS is at the core of econometrics curriculum, it is easily derived and highly practical to familiarise a learner with regression possibilities and limitations.

The usual way to teach OLS is to present assumptions and show how to deal with their violations as indicated below in a review chart from Kennedy's textbook.

Math:

$Y = \beta X + \epsilon$ ,  $\epsilon$  is iid, normal with finite variance.

Common steps:

1. specify model: select explanatory variables, transform them if needed
2. estimate coefficients
3. elaborate on model quality (the hardest part)
4. go to 1 if needed
5. know what model *does not* show (next harder part)

What may go wrong:

- residuals are not random
- variables are cointegrated
- multicollinearity in regressors
- residuals depend on  $x$  (heteroscedasticity)
- inference is not causality
- wrong signs, insignificant coefficients
- variable normalisation was not described

Discussion:

- why sum of squares as a loss function?
- connections to bayesian estimation
- is  $R^2$  useful or dangerous?

Replication examples:

- <https://www.kaggle.com/nicapotato/in-depth-simple-linear-regression>

Implementations:

- [lm function in R](<https://github.com/wch/r-source/blob/0f07757ad10ca31251b28a2c332812e63c0acf38/src/library/stats/R/lm.R>)
- [OLS class in python statsmodels]([https://github.com/statsmodels/statsmodels/blob/master/statsmodels/regression/linear\\_model.py](https://github.com/statsmodels/statsmodels/blob/master/statsmodels/regression/linear_model.py))
- [python scipy least squares](<https://github.com/scipy/scipy/blob/v1.1.0/scipy/linalg/basic.py#L1048-L1265>)
- julia [Alistair](<https://github.com/giob1994/Alistair.jl>), GLM.jl, Regression.jl

Links (preliminary):

- <https://stats.stackexchange.com/questions/296514/is-ols-estimator-the-only-blue-estimator>
- <https://cran.r-project.org/doc/contrib/Faraway-PRA.pdf>
- <https://github.com/marcoalt/Regression-parameter-estimates>
- <https://blog.minitab.com/blog/how-to-choose-the-best-regression-model>
- <https://towardsdatascience.com/simple-linear-regression-2421076a5892>
- [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1736184](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1736184)
- [https://cfss.uchicago.edu/persp012\\_regression\\_diagnostics.html](https://cfss.uchicago.edu/persp012_regression_diagnostics.html)
- <https://stats.stackexchange.com/questions/218156/what-are-some-of-the-most-common-misconceptions-about-linear-regression>
- <https://pdfs.semanticscholar.org/a410/ec58bd5ff80a4e7978955ab11d096af6d138.pdf>
- COMMON MISTEAKS MISTAKES IN USING STATISTICS: Spotting and Avoiding Them
- <http://www2.sas.com/proceedings/sugi22/STATS/PAPER267.PDF>
- <https://towardsdatascience.com/simple-linear-regression-2421076a5892>
- <http://jeromyanglim.blogspot.com/2013/12/using-r-to-replicate-common-spss.html>
- <https://stat.hevra.haifa.ac.il/~gweiss/courses/Forecasting/CarPricesAnalysis.pdf>
- [http://www.nrcresearchpress.com/doi/10.1139/f73-072#.W-SvWuK\\_Pcs](http://www.nrcresearchpress.com/doi/10.1139/f73-072#.W-SvWuK_Pcs)
- <https://www.nejm.org/doi/full/10.1056/NEJM198512263132604>
- <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2992018/> - medicine, nice!
- [http://replication.uni-goettingen.de/wiki/index.php/Ordinary\\_least\\_squares\\_\(OLS\)](http://replication.uni-goettingen.de/wiki/index.php/Ordinary_least_squares_(OLS)) - list of replications
- <https://stats.stackexchange.com/questions/211707/replicating-a-linear-regression-example-from-hastie-tibshirani-and-friedman> (replication from Hastie)

Intuitive learning for OLS

- ‘Stackoverflow hot

topics <<https://stats.stackexchange.com/questions/tagged/linear-model?sort=votes&pageSize=15>>‘\_\_ - In Depth: Linear Regression - Stachurski math intro:

- TOC
- slides
- <https://lectures.quantecon.org/py/ols.html>
- <http://www.statsmodels.org/stable/examples/index.html#regression>
- <https://docs.scipy.org/doc/scipy/reference/generated/scipy.linalg.lstsq.html>
- <https://stats.stackexchange.com/questions/1829/what-algorithm-is-used-in-linear-regression>

## PRINCIPAL COMPONENTS ANALYSIS, PCA

Math:

Assumptions:

Usual steps:

What may go wrong:

Discussion:

Replication examples:

Links:

- <https://stats.stackexchange.com/questions/2691/making-sense-of-principal-component-analysis-eigenvectors-eigenvalues/2700#2700>
- [https://scikit-learn.org/stable/auto\\_examples/decomposition/plot\\_pca\\_3d.html#sphx-glr-auto-examples-decomposition-plot-pca-3d-py](https://scikit-learn.org/stable/auto_examples/decomposition/plot_pca_3d.html#sphx-glr-auto-examples-decomposition-plot-pca-3d-py)
- <https://stats.stackexchange.com/questions/48214/replicating-shalizio-new-york-times-pca-example?rq=1>

Projection, rejection, PCA:

- <https://stackoverflow.com/questions/52288029/function-that-computes-projection-and-reconstruction-error-using-numpy-python/52290082#52290082>
- <http://www.cs.cmu.edu/~guyton/Class/15781/slides/pca-mdps-annotated.pdf>
- [https://ocw.mit.edu/courses/mathematics/18-06sc-linear-algebra-fall-2011/least-squares-determinants-and-eigenvalues/projections-onto-subspaces/MIT18\\_06SCF11\\_Ses2.2sum.pdf](https://ocw.mit.edu/courses/mathematics/18-06sc-linear-algebra-fall-2011/least-squares-determinants-and-eigenvalues/projections-onto-subspaces/MIT18_06SCF11_Ses2.2sum.pdf)





## **SIMULATION**

A very useful tool is simulation – with that we can examine the properties of our tools in situations very like those it appears our data may have arisen from, and so either use them in the comforting knowledge that they have good properties in those cases (or, sometimes, see that they don't work as well as we might hope).

From an [answer to Why do we care so much about normally distributed error terms](#) by Glen\_b.

---

**To cover next: Concepts**

- identification
  - inference
  - overfitting
  - [spurious regression](#)
- 

---

**To cover next: Theorems**

- [Bayes theorem](#)
  - Gauss-Markov theorem
- 

---

**To cover next: Research design**

- replication, replicability
  - [complexity bias](#)
  - [publishing and promotion](#)
-



## 2. TEXTBOOKS AND COURSES

### 11.1 Mathematic preliminaries

Typical prerequisites for statistics and econometrics are:

- linear algebra
- calculus
- probability

They usually take 2-4 semester in college. Linear algebra is fully covered by [VMLS](#), and probability is exposed in [PSC](#), even though latter is a compact reference, not formally a textbook. Scipy lectures is a great one-stop resource for numerical computing basics. I do not have a one single source to recommend for calculus yet.

#### 11.1.1 Linear Algebra

- [Vectors, Matrices, and Least Squares \(VLMS\)](#)

#### 11.1.2 Probability and statistics

- [Probability and Statistics Cookbook \(PSC\)](#)

#### 11.1.3 Numerical computing

- [Scipy lectures: one document to learn numerics, science, and data with Python](#)

### 11.2 Statistical Inference

This page is a draft.

Some discussions about a good book:

- [Math Overflow: Statistics for Mathematicians](#)
- [Stack Exchange: What to learn after Casella/Berger](#)

### 11.2.1 Topics

Review of probability:

- random variables, outcomes, probability
- distributions, pdf/cdf

Math.stat itself:

- idea: learn about data generating process (DGP) from a sample
- sample/realisation as a random vector
- statistic as a function of sample
- parameter inference:
  - point estimation and methods (estimators)
  - estimator quality (bias, consistency, efficiency)
- confidence intervals (CI)
- hypothesis testing (HT), types of errors
- analysis of variance and regression

See an example of [stated learning objectives](#)

### 11.2.2 Key textbook

Casella, Berger. Statistical Inference.

- [link](#)

### 11.2.3 Other choices

Larry Wasserman. All of Statistics: A Concise Course in Statistical Inference

- [text 1] (<https://www.ic.unicamp.br/~wainer/cursos/ls2013/ml/livro.pdf>)
- [text 2] ([http://static.stevereads.com/papers\\_to\\_read/all\\_of\\_statistics.pdf](http://static.stevereads.com/papers_to_read/all_of_statistics.pdf))
- [Amazon] (<https://www.amazon.com/All-Statistics-Statistical-Inference-Springer/dp/0387402721>)

Efron, Hastie. Computer Age Statistical Inference: Algorithms, Evidence and Data Science.

- [https://web.stanford.edu/~hastie/CASI\\_files/PDF/casi.pdf](https://web.stanford.edu/~hastie/CASI_files/PDF/casi.pdf)

### 11.2.4 Supplements

James Gentle. [A theory of statistics.](#)

De Groot and Shervish%20(1).pdf)

Herman J. Bierens. [Introduction to the Mathematical and Statistical Foundations of Econometrics](#)

Aris Spanos. [Probability Theory and Statistical Inference: Econometric Modeling with Observational Data.](#)

- [link] (<https://wilfridomtz.files.wordpress.com/2014/08/cambridge-university-press-probability-theory-and-statistical-inference-842pg.pdf>)

Probability and Statistics for Engineers and Scientists (PSES)

## 11.2.5 Other reading

Statistical Inference as Severe Testing: How to Get Beyond the Statistics Wars

- [Amazon] (<https://www.amazon.com/Statistical-Inference-Severe-Testing-Statistics/dp/1107664640>)  
 - very philosophic reading

Peter M. Aronow, Benjamin T. Miller. Foundations of Agnostic Statistics.

- [Google Book (fragments)] ([https://books.google.ru/books?id=u1N-DwAAQBAJ&printsec=frontcover&hl=ru&source=gbs\\_ge\\_summary\\_r&cad=0#v=onepage&q&f=false](https://books.google.ru/books?id=u1N-DwAAQBAJ&printsec=frontcover&hl=ru&source=gbs_ge_summary_r&cad=0#v=onepage&q&f=false))  
 - [book proposal] (<http://aronow.research.yale.edu/aronowmillerproposal.pdf>)  
 - simplistic

## 11.2.6 Papers

Statistical Inference: The Big Picture. Robert E. Kass (2006)

A short history of probability theory and its applications. International Journal of Mathematical Education. January 2015. Kanadpriya Basu.

## 11.2.7 In Russian

Cramer, 1946, Russian translation, timeless! also has dataset if wheat yeilds.

Also concise version of lectures [here](#) (in Russian)

# 11.3 Econometrics

## 11.3.1 General textbooks

- Thread about [picking a textbook](#)
- Favorite introductory book: [Peter Kennedy. A Guide To Econometrics](#)
- Very condensed synopsis of [undergrad econometrics](#)

Some well-known texts:

- Greene
- Dougherty
- Hayashi
- Stock and Watson
- Verbeek

- Hansen. Econometircs
- Sargent-Hansen

### 11.3.2 Cross Section and Panel Data

- Wooldridge. Econometric Analysis of Cross Section and Panel Data
- , with criticisms by:
  - Diebold: “sub-sub-sub-area of applied econometrics”
  - J. Pearl

Note that MHE is one of few books with excercises reproduced on github.

### 11.3.3 Time Series

- Hamilton
- Diebold
- Granger/Watson time series chapter in Handbook of Econometrics
- Enders
- Jeffrey Parker. Fundamental Concepts of Time-Series Econometrics in *Theory and Practice of Econometrics*. Reed College course

Macroeconomic time series:

- John H. Cochrane. Time Series for Macroeconomics and Finance.
- Eric Sims. Graduate Macro Theory II: Notes on Time Series
- Lars Peter Hansen. Time Series Econometrics in Macroeconomics and Finance
- Hilde C. Bjørnland and Leif Anders Thorsrud. Applied time series for macroeconomics. (not open source, but good table of contents and has MATLAB code)
- Karl Whelan. Time Series and Macroeconomics. (also see RATS code and other lectures on the web site)

## 11.4 Machine learning and deep learning

### 11.4.1 Books

#### ML

- An Introduction to Statistical Learning. Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani

#### DL

- Deep Learning. Ian Goodfellow and Yoshua Bengio and Aaron Courville

## 11.4.2 Courses

### ML

- [ods](#)
- [Andrew Ng course](#)

### DL

- [fast.ai](#)
- [deeplearning.ai](#)





## 3. THE SCIENCE OF TEACHING

### 12.1 Technical pedagogy

#### 12.1.1 Ideas

- “Overarching education” (plenty of math and fundamentals, aimed at systemic thinking, practical applications derived later) is expensive method of teaching
- Student motivation is scarce, information is abundant. Teaching is a guidance (esp masters level).
- Can teach programming first, and followup with more solid math second, if ever (programming allows experimenting).
- Need open source textbooks, able to update and share parts of text as well as interactive exercises. Static site generators are not fully there yet.

#### 12.1.2 Links

1. [10 rules of teaching](#) by Greg Wilson (@gvwilson) starts with rule #1 “Be kind: all else is details.” More detail is at [teachtogether.tech](#).
2. Nick Huntington-Klein [proposes a course structure](#) based on statistical programming, and causal inference/research design, with regressions postponed.
3. Allen Downey has a [presentation about teaching physical modelling](#) and sequencing of math and programming.
4. Very introductory courses are good for building student confidence and making simple things simple. They prevent gate-keeping (maybe a reason why they are attacked). See a thread by [Rochelle Terman on teaching computational social science](#).

### 12.2 Econometrics body of knowledge

What exactly is a body of knowledge of econometrics? Surely it is a set of teaching curricula at universities and accompanying textbooks and reading lists, but what if you needed construct a review of the field rather quickly? The following approaches might be useful.

#### 12.2.1 1. Textbook structure

The undergraduate textbook structure often repeats itself: OLS, estimator properties, some of OLS deviations/extensions, logit/probit + maximum likelihood, time series and maybe a bit of panels and simultaneous equations.

More textbook analysis may be found in 2017 Angrist and Pischke article [Undergraduate Econometrics Instruction: Through Our Classes, Darkly](#), below is a summary table.

### Topics Coverage in Econometrics Texts, Classic and Contemporary

	<i>1970s</i>	<i>1970s Excl. Grad</i>	<i>Contemporary</i>
Topic	(1)	(2)	(3)
Bivariate Regression	2.5	3.6	2.8
Regression Properties	10.9	11.9	9.9
Regression Inference	13.2	13.3	14.6
Multivariate Regression	3.7	3.7	6.4
Omitted Variables Bias	0.6	0.5	1.8
Assumption Failures and Fix-ups	18.4	22.2	16.0
Functional Form	10.2	9.3	15.0
Instrumental Variables	7.4	5.1	6.2
Simultaneous Equations Models	17.5	13.9	3.6
Panel Data	2.7	0.7	4.4
Time Series	12.3	15.2	15.6
Causal Effects	0.7	0.7	3.0
Differences-in-differences	--	--	0.5
Regression Discontinuity Methods	--	--	0.1
Empirical Examples	14.0	15.0	24.4

Notes: This table reports percentages of page counts by topic. Column (2) excludes Kmenta, Johnston, and Intriligator. Dashes indicate no coverage.

#### 12.2.2 2. Econometric software manuals

Gretl and EViews have quite comprehensive manuals covering principal applications of the software. They both qualify as textbooks in econometrics:

- [gretl](#)
- [course based on gretl](#)
- [Eviews I](#)
- [Eviews II](#)

Additionally one can look into [R package system for econometrics] (<https://cran.r-project.org/web/views/Econometrics.html>), MATLAB [manual](#) and [course](#).

Manulas of some less popular packages:

- [RATS](#)
- [PcGive](#)

- Shazam
- econtools (STATA flavour)

### 12.2.3 3. Handbook of Econometrics

Elsivier [Handbook of Econometrics](#) is a renowned publication series runniing since 1983. It now features 77 chapters in 6 volumes. Many earlier articles are foundational, but quite a few are about some are really narrow subjects areas.

## 12.3 Econometrics mindmap

It would be great to show a modern roadmaps in econometrics starting from mathematic foundations (linear algebra, calculus, probability) to econometrics to computationally intensive data processing tasks. I've seen this being approached as [clusters of courses](#), but I think there is more that can be done.

### 12.3.1 An (over)simplified view of econometrics curriculum

- linear algebra, calculus, probability and statistical inference
- OLS (assumptions, violations, fixes + estimatore quality)
- limited depenedent variables + maximum likelihood
- intrumental variables
- time series, state space representation
- panel data
- classifications
- systems of equations

### 12.3.2 Key areas

- data structures (crosssection, time series, panel)
- inference methods
  - model specification
  - estimation procedure
  - model evaluation
- use cases

### 12.3.3 Additional topics 1

- simulation (Monte Carlo, bootstrap)
- transformations (PCA)
- darker side of macro (DSGE)
- bayesian estimation

### 12.3.4 Additional topics 2

OLS Extensions:

- GMM,
- 2,3 stage OLS
- quantile regressions
- lasso, ridge

Estimation:

- maximum likelihood
- bayesian estimation
- mcmc (see reddit post)

Time series:

- time series, stationarity, unit root
- state space representation, Kalman filter
- fractional integration
- seasonal adjustment
- (vector) error correction model, VECM
- structural breaks

### 12.3.5 User profiles

1. “Nurate biologists” - solve a domain problem in biology, psychology, social sciences
2. “Want to hit a ‘Run’ button” - quick results without thinking, typical of students
3. “I’m doing XYZ now!” - excited adopters, writing a piece on Medium full of acronyms
4. “Sane econometrics” - appropriate methods with clear, accessible explanation, rare trait
5. “Asymptotics” - publish evermore sophisticated articles to secure academic career

### 12.3.6 Discussion

Review: [Undergraduate Econometrics Instruction: Through Our Classes, Darkly](#). [NBER/IZA](#) and a criticism of G1/G2 goals

## 4. HISTORY OF ECONOMETRICS

### 13.1 Landmark events

- [Galton and Pearson](#)
- [Tinbergen, Haavelmo](#)
- [Cowles Commission](#)
- [Lucas and Sims critique](#)

### 13.2 Overviews

- [Econometrics: A Bird's Eye View](#)
- [Econometrics: An Historical Guide for the Uninitiated](#)
- [The First Fifty Years of Modern Econometrics](#)

### 13.3 By topic

- [Pearl, J. \(2014\). Tygve Haavelmo and the emergence of causal calculus. \*Econometric Theory\*, 31\(1\), 152–179.](#)
- [A Short History of Markov Chain Monte Carlo \(arxiv\)](#)
- [Working on 1960s macroeconometrics \(blog\)](#)
- [Criticizing the Lucas Critique: Macroeconometricians' Response to Robert Lucas](#)
- [W N Venables. \*Exegeses on Linear Models\* \(1998\)](#)



## 14.1 Large collections

- Python libraries [statsmodels](#), [sci-learn](#) and [seaborn](#) and [R](#) itself have build-in datasets.
- [gretl](#) has extensive selection of datasets, including data from key textbooks.
- Well known dataset repository is [UCI](#).
- [Kaggle](#) obviously has plenty of datasets.
- Sheffield University made a good listing of [datasets for teaching](#).
- Australian National Centre for Econometric Research website has a useful [Data and code](#) section.
- [Google Dataset search](#) is not yet as good as main Google search, but will surely improve.
- [FRED](#), [Quandl](#) and [dbnomics](#) are standard sources for macroeconomic data. My own effort for clean Russian macro time series - [mini-kep](#).

## 14.2 Individual datasets

- [Tidy Tuesday](#) publishes weekly datasets and accompanying articles.
- Nick Huntington-Klein provides a variety of extra examples of [data importable into R](#).
- I extracted a small, but illustrative dataset about [lightbulb survival rate](#).
- [Sunspot data](#) is listed at variety of sources.





## SOFTWARE

Below is a small chart that outlines common options for statistics / econometrics software. gretl, EViews and MATLAB tend to have better-organised documentation. My personal choices:

- if I had to choose just one: python;
- if I had to choose other two: R (libraries) and gretl (documentation);
- before open source era: Eviews;
- if I had more time: Julia.

Open source	Proprietary
<ul style="list-style-type: none"><li>• R (RStudio), derived from S</li><li>• Python (Anaconda)</li><li>• Julia</li></ul>	
<ul style="list-style-type: none"><li>• gretl</li></ul>	<ul style="list-style-type: none"><li>• EViews (derived from TSP)</li></ul>
<ul style="list-style-type: none"><li>• Octave</li></ul>	<ul style="list-style-type: none"><li>• MATLAB</li></ul>
	<ul style="list-style-type: none"><li>• SAS</li><li>• SPSS</li><li>• Stata</li></ul>
	RATS, Ox, PcGive
Stan, PyMC3, Turing.jl	
JASP	

### 15.1 SAS and terminology

SAS, it seems, has become the gold standard, the output of SAS programs the ultimate point of reference for correct and appropriate statistical calculations and the SAS terminology is rapidly taking over as standard terminology. This is very Microsoft-like indeed and very worrying for anyone who cares about the profession.

WN Venables. *Exegeses on Linear Models* - on early adoption of S Plus

## 15.2 Example: software used in black hole discovery

Reading *First M87 Event Horizon Telescope Results. III. Data Processing and Calibration*:

## BLOGS

Blogs may seem as an outdated fashion of communication, but they often present a wider story than a twitter post.

Below there are some personal blogs that enlight and inspire about statistics and econometrics. They are regularly updated.

- [Matt Bogard](#)
- [Francis Diebold](#)
- [Dave Giles](#)
- [Rob Hyndman](#)

[Simply Statistics](#) by biostatistics professors Rafa Irizarry, Roger Peng, and Jeff Leek cover data management, analysis and teaching stats.

[Medium](#) has a variety of posts on statistics, but rarely tags econometrics. PCA is a widely popular topic.



## GOOD CLUES FROM TWITTER

A secret subtitle for this publication is “*Can you learn econometrics from Twitter and Stack Overflow alone without distracting yourself to data science tutorials*”.

Links collected in no particular order, some will show in other sections of the Navigator.

### 17.1 No links between leading macrotextbooks

### 17.2 OLS interactively exposed

### 17.3 R language guide is an econometrics guide

### 17.4 OLS, MML, Bayes and MCMC(!) for linear regression

<https://peterroelants.github.io/posts/linear-regression-four-ways/>



**17.5 Very true on R2**

**17.6 Instrumental variables**

**17.7 Interpreting coefficients 1**

**17.8 Interpreting coefficients 2**

**17.9 Doing PCA approach**

**17.10 OLS explained for social scientists**

**17.11 Suggested stat exercises**

**17.12 Coding a tree**

**17.13 Consumer demand modelling**

**17.14 Scott Cameron learning method**

**17.15 Model evaluation compendium (on classifier)**

**17.16 Causality by Judea Pearl**

**17.17 A thesis turned tutorial on probabilistic programming and MC inference by Tom Rainforth**

**17.18 Value of logit**

**17.19 Traditional statistics vs ML**

**17.20 Program evaluation by John Holbein**

**17.21 218 pages on probabilistic programming**

**17.22 A4 econometric art**

**17.23 A 1910 must-read**

**17.24 Amazing statistics animation**

**17.5. Very true on R2**

**17.25 Undergrad econometrics condensed to 3 pages**





## ACRONYMS

Econometrics is full of fancy abbreviations that one can juggle with.

**ANOVA** analysis of variance

**ARIMA** autoregression, integration, moving average

**DID** difference-in-differences

**FE** feature engineering

**GARCH** generalised (a)uto(r)egressive conditional heteroscedasticity

**GLS** generalised least squares

**GMM** generalised method of moments

**iid** independent identically distributed

**IV** instrumental variable

**OLS** ordinary least squares

**PCA** principal components analysis

**RDD** regression discontinuity design

**VECM** vector error correction model

**WLS** weighted least squares



## ROADMAP

Updated May 13, 2019.

In next versions, I would like to do the following:

### 19.1 Labels

- label 'heavy' vs 'easy' content
- adjust structure accordingly

#### 19.1.1 General

- draw a mindmap for econometrics (as described in text)
- put key textbooks on a roadmap
- pay more attention to bayesian / causality

#### 19.1.2 Interactivity

- repl.it
- colab (no Julia)
- notebooks/binder

#### 19.1.3 Data science textbooks

- Add [deep learning links](#)
- [Data Science fo Economists](#)

#### 19.1.4 Articles and code

- write more articles and code for the main section
- translate some RATS/MATLAB code to open source (esp time series)

Specific tasks:

- clean ols

- [run pca example](#)

### 19.1.5 Reproducibility

Add resources on reproducible research and why it has such a poor traction in economics

- [this - 1](#)
- [this - 2](#)

DAG tools:

- [waf.io](#)
- [invoke](#)

### 19.1.6 How to publish a textbook

Need an opensource textbook with interactive code examples, translation into Russian desired. Hard to see a combination of a static site generator, good theme and PDF export working smoothly.

- [Allen Downey on Medium](#)
- [jupyter-book](#)
- [bookdown](#)

### 19.1.7 Our publishing process

- things mentioned in [todo.txt](#)

### 19.1.8 Other goals

- review ‘deprecated’ folder, and ‘history’ page
- ‘vednorize’ [LessOLS.jl](#)

## WELCOME TO ECONOMETRICS NAVIGATOR!

### 20.1 Goals

The Econometrics Navigator (EN) is aims to provide you with:

- open access textbooks
- community knowledge (reddit, StackOverflow/Cross Validated, Twitter threads)
- minimal code examples in Python, R, gretl or Julia

for better understanding of statistics and econometrics topics.

### 20.2 Changelog

**v.0.0.5 (November 2019):**

**v.0.0.4 (May 2019):**

- Science of teaching: quoting @gvwilson, @nickchk, @AllenDowney, @RochelleTerman at <https://tinyurl.com/em-nav-teach>
- Data: added data from @stlouisfed/@quandl/@DBnomics along with several R data sources by @nickchk at <https://tinyurl.com/em-nav-da>
- Books:
  - WM Venables. *Exegeses on Linear Models*
  - Walpole, Myers, Myers, Ye. *Probability and Statistics for Engineers and Scientists*

**v.0.0.3 (April 2019)** scraps several unfinished articles, including a section on applications (hard to fill it quickly). Three main parts in content established (own articles, textbook annotations and how to teach resources).

**v.0.0.2 (November 2018)** original version of EN nobody understood what it is good for, had sample articles on max likelihood, bootstrap, ANOVA.

### 20.3 Content

#### 20.3.1 1. Collection of own articles

The main body of Econometrics Navigator articles is [Concepts and techniques section](#), organised alphabetically. Finished examples are:

- Maximum likelihood
- Bootstrap
- ANOVA

### 20.3.2 2. Textbooks guide

[Textbooks review](#) attempts to sort out and annotate textbooks and reference texts by several categories.

For example, I praise Kennedy's textbook and collect (constructive) criticisms of Mostly Harmless Econometrics there. The categories in econometrics are 'general' textbooks, cross-section/panel and time series texts.

### 20.3.3 3. Instructor resources

The backstage workings of the Navigator are [History of econometrics](#), [Ways to review econometrics](#) and [econometrics mindmap](#) (draft).

These documents aim to organise thinking about better teaching of econometrics in terms of sequence of topics, better analogies for the learner and a faster bridge to coding from formulas.

## 20.4 Twitter

So far twitter has been an enormously valuable source of demos, links and opinion for me. I keep a [separate page with twitter posts](#), some of my favorites are:

- [Undergrad Econometrics Cheatsheet](#) by Tyler Ransom
- [Casual graphs and XY plane animations](#) by Nick Huntington-Klein
- [Investigative time series example](#) by @cubic\_logic
- [Common statistical tests are linear models \(or: how to teach stats\)](#) by Jonas Kristoffer Lindeløv

## 20.5 Contacts

Feel free to contact me [@PogrebnyakE](#). I need help in shaping this guide. Some future steps outlined in [roadmap](#).

## 20.6 Source

The source of this publication is available at <https://github.com/epogrebnyak/econometrics-navigator> or <https://tinyurl.com/emnavig>