Applied Econometrics

(Macro and Finance)

by Jozef Baruník and Luboš Hanus

Lviv Data Science Summer School
July 2019

Український Католицький Університет
APPLIED SCIENCES FACULTY

Course Outline

Machine learning, Data Science, and Econometrics

Basis of the time series analysis (AR, MA, processes, concept of stationarity, etc.)

Multivariate Time Series (VAR, estimation, identification, IRFs, FEVD, causality

Application in forecasting macroeconomics Application in measuring connectedness + distributional forecasting

VARs as Networks

Connection to Machine Learning and Big Data

Big Data the Phenomenon

“How big are your data?”

Depends on WHEN you ask the question!

50 years ago: 90 observations on each of 10 variables

2010+: we do not count, report size instead: 200 GB

2030: do we even want to know? Explosive growth in available data.

Big Data the Phenomenon

Vast data available today

- hundreds of billions of observations and millions of features
- A 100,000,000,000x1,500,000 dimensional tables/matrices common

Allows analyses unthinkable 15 years ago

Presents challenges to econometricians, computer scientists and statisticians

- How to get that matrix into a computer

Can you get that matrix into a computer

- If so how do you interpret 1.5 million independent variables

Can you interpret 1.5 million independent variables

Computing Power

Cambridge, early 1950s: computation still largely done on Marchant electric calculators

takes days

Computing Power

Your smartphone is millions of times more powerful than all of NASA's combined computing in 1969

What is Data Science about?

Big data

and

Large scale computations?

What is Data Science about?

Estimation

Programming

Big Data

Processing

Machine learning

Statistics

Classification

Analysis

Data collection

Predictions

Mathematics

Algorithms

Model building

Data Mining

MODERN DATA SCIENTIST

Data Scientist, the sexiest job of 21th century requires a mixture of multidisciplinary skills ranging from an intersection of mathematics, statistics, computer science, communication and business. Finding a data scientist is hard. Finding people who understand who a data scientist is, is equally hard. So here is a little cheat sheet on who the modern data scientist really is.

MATH & STATISTICS

- Machine learning
- Statistical modeling
- Experiment design
- Bayesian inference
- Supervised learning: decision trees, random forests, logistic regression
- Unsupervised learning: clustering, dimensionality reduction
- Optimization: gradient descent and variants

PROGRAMMING & DATABASE

- Computer science fundamentals
- Scripting language e.g. Python
- Statistical computing package e.g. R
- Databases SQL and NoSQL
- Relational algebra
- Parallel databases and parallel query processing
- MapReduce concepts
- Hadoop and Hive/Pig
- Custom reducers
- Experience with xaaS like AWS

DOMAIN KNOWLEDGE & SOFT SKILLS

- Passionate about the business
- Curious about data
- Influence without authority
- Hacker mindset
- Problem solver
- Strategic, proactive, creative, innovative and collaborative

COMMUNICATION & VISUALIZATION

- Able to engage with senior management
- Story telling skills
- Translate data-driven insights into decisions and actions
- Visual art design
- R packages like ggplot or lattice
- Knowledge of any of visualization tools e.g. Flare, D3.js, Tableau

What is Data Science?

Hal Varian, chief economist at Google

“I keep saying the sexy job in the next ten years will be statisticians…

… the ability to take data, to be able to understand it, process it, extract value from it, visualize it, communicate it”

These are the key skills for the next decades

Statistician vs. Data Scientist

Depends on age of person being asked

What does it take to become a good statistician in economics?

(or data scientist)

Data Science in Economics and Finance

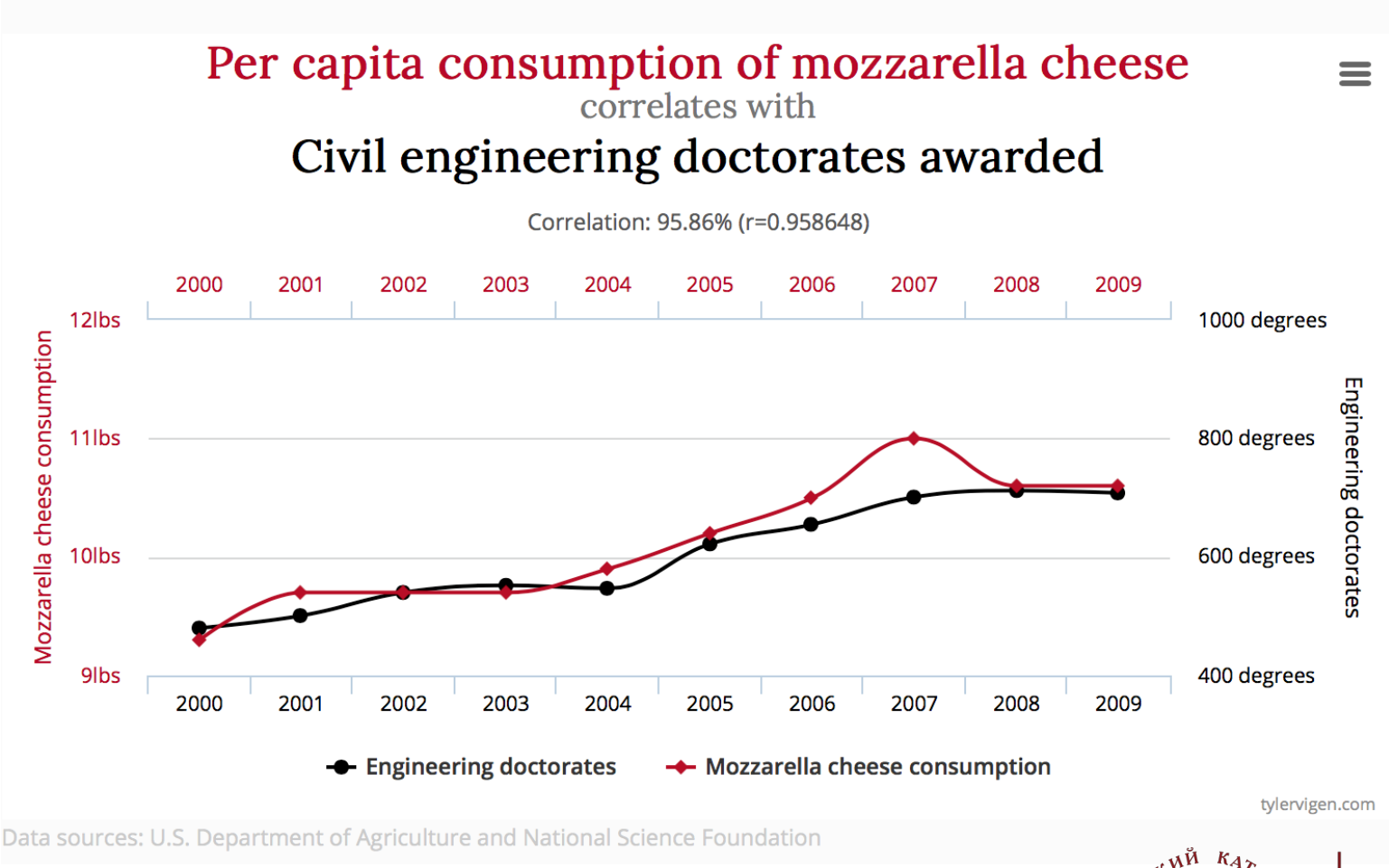
3 main ingredients:

Economic statistics
Economic Theory
Mathematics

Everything is:

Model — economic intuition
Information — data
Risk — statistics help to minimize risk

Proper techniques are crucial



Approaches

– **Econometricians** (and other social scientists)

What is the probability model for the data
Interested primarily in gaining insight into behavioral processes, sometimes prediction

– **Statisticians**

How to make inference for data from general probability models

– **Computer scientists**

How to compute

Culture Wars

Brieman(2000)

- **ML**
  - How does the model predict out-of-sample
  - Proofs of properties more-or-less useless
  - Let data reveal structure
- **Statistics**
  - What is the underlying probability model
  - How to infer structure
  - How to predict using structure
- **Economics**
  - Much like Stat
  - Though economist and nearly as famous Economist Milton Friedman was clearly amenable to the ML camp
  - I would add Sims as well.

Overview

Introduction to the intersection for economists.

Computer Science will tell you that they invented “big data” methods

Actually, a lot comes from *economics*

Much of the new stuff is quite old.

Impossible to do in its day

Made Possible by Advances in *CS* and *Statistics*

Econometrics in Machine Learning

Causality and identification

- Economics 1950
  - Leonid Hurwicz (Economics 2007, Nobel Prize winner)
- Computer Science 1998
  - Judea Pearl (UCLA)

Tree methods

- Economics 1963
  - James N. Morgan (U of Michigan)
- Statistics 1984
  - Leo Brieman (Berkeley)

Map Reduce

- Economics 1980
  - Gregory M Duncan
- Computer Science 2004
  - Jeffrey Dean and Sanjay Ghemawat (Google)

Econometrics in Machine Learning

Bagging/Ensemble Methods

- Economics 1969
  - Bates and Granger (UCSD)
- ML 1996
  - Brieman (almost of course)

Economists early contributors to neural net/deep learning literature (late 1980's early 1990's)

- Hal White, Jeff Wooldridge, Ron Gallant, Max Stinchcombe
- Quantile neural nets, learning derivatives, guarantees for convergence, distributions etc.
  - Needed for counterfactuals when out of sample validation make no conceptual sense

Econometrics in Machine Learning

Much of machine learning

– rediscovers old statistical and econometric methods
– but making them blindingly fast and practicable

Many of the tools used today have been enabled purely by computers (eg. simulation based estimators)

Outcome of machine learning is generally a data reduction

– Regression coefficients
– Summary statistics
– A prediction

Machine learning allows more complex models than economists are used to

Time Series Econometrics vs. Machine Learning

Especially in TS econometrics we need to be careful

ML methods mostly assumes independent variables, but TS have strong dynamics

As a result, most of ML efforts result in overfitting, and bad forecasts

One needs to understand the TS first, then ML applications will be much easier

So let’s get our hands on code and data..