Syllabus

Data Analysis 2: Finding Patterns with Regressions – Business Analytics track

- **Instructor:** Agoston Reguly

    Email: reguly_agoston@phd.ceu.edu
    Office hours: Monday 10:30-12:00 or by appointment

- **Credits:** 2 (4 ECTS)

- **Term:** Fall 2020-2021

- **Course level:** [MA/MS]

- **Prerequisites:** Mathematics and Informatics Pre-session for Business Analytics;
    Data Analysis 1: Exploration – Business Analytics track

- **Course drop:** Course can be dropped free of charge 24 hours after the first session. After this date drop is possible until the course is halfway over (late drop fee applies). No changes are allowed past that date.

1. COURSE DESCRIPTION

Uncovering patterns in the data can be an important goal in itself, and it is the prerequisite to establishing cause and effect and carrying out predictions. The course starts with simple regression analysis, the method that compares expected y for different values of x to learn the patterns of association between the two variables. It discusses nonparametric regressions and focuses on the linear regression. It builds on simple linear regression and goes on to enriching it with nonlinear functional forms, generalizing from a particular dataset to other data it represents, adding more explanatory variables, etc. We also cover regression analysis for time series data, binary dependent variables, as well as nonlinear models such as logit and probit.

The seminars will focus on selected case studies introduced in lectures. In classes, interpretation and coding solutions will both be discussed using R.

2. LEARNING OUTCOMES

Key outcomes:

By successfully completing the course the students will be able to:

- Successfully formulate research questions that are answerable by empirical analysis;
- Produce meaningful descriptive statistics and informative graphs;
- Carry out simple regression analysis;
- Discuss and interpret results, understand validity and constraints;
- Present empirical analysis and write short reports with data;

Other outcomes. The course will also help develop skills in the following areas:

Learning Area	Learning Outcome
Critical Thinking	Evaluating statistical models and reports. Be able to find possible weaknesses and strength of different quantitative analysis.
Quantitative Reasoning	Arguments based on data and statistical models. Interpreting the results of these models.
Technology Skills	Use available packages in R and employing them for data analysis.
Interpersonal Communication Skills	Share ideas and talk about statistical and modelling problems or challenges with other statisticians.
Management Knowledge and Skills	Understand the needs of managers and be able to create such reports that is easily readable.
Cultural Sensitivity and Diversity	While working together on specific tasks students are encouraged to learn and respect cultural sensitivity and diversity.
Ethics and Social Responsibility	Ethics of statistical modelling and reporting. Consciously articulate the truth which can be or can not be found in the data.

3. READING LIST

Required:
- Békés – Kézdi [BK]: Data Analysis for Business, Economics and Policy (forthcoming) - available as handouts; Chapters 7-12
- Data, codes will be provided.

Optional Reading:
- James, Witten, Hastie & Tibshirani (2013): Introduction to Statistical Learning. Chapter 1-4; Online version and supplementary codes are available here.
- Wooldridge (2018): Introductory Econometrics, Part 1-2
- Angrist – Pischke (2009): Mostly Harmless Econometrics, Chapter 2-3

Databases. The CEU Library boasts a range of databases covering financial and company data, market and industry reports, global news and more. For a full list of databases visit the CEU Library.
- Refinitiv (Thomson Reuters) Eikon for Students + Datastream/Thomson ONE
  - Eikon: Platform used by finance practitioners including market traders to monitor and analyze financial information. Information, analytics and news on all major financial markets including real-time pricing data, financial research, global financial news and commentary, financial estimates, fundamentals analysis, visual analysis through charting. Import/export from Excel.
  - Datastream: Range of economic, securities and company financial data. Excel add-in.
  - Thomson ONE: Global overviews on 55,000 public companies, one million private companies. Reuters News, ownership, deals, private equity, key ratios, company filings, officers and directors. Investext analyst reports, active and historical research from 1,600 independent research firms, brokerages, investment banks.
- Standard & Poor's Capital IQ
  - Web and Excel-based platform combining deep global company information, credit ratings and research, and market research with powerful tools for risk assessments. Real-time and historical information on markets, industries, companies, transactions and people. Tearsheet data.

- Lexis Nexis Academic
  – Global database of news, business, legal and other sources. Full text of 350 newspapers, 300 magazines and journals, 600 newsletters. Wire services including Associated Press, Business Wire and PR Newswire. Company financial information, market research, industry reports.

4. TEACHING METHODS AND LEARNING ACTIVITIES

The course will involve a mix of
- Lectures to present the ideas of statistical modelling.
- Seminar-type classroom to understand how to carry out data analysis with statistical models. Through this we demonstrate the use of the learned theories and concepts.
- Optional homework to practice through extra problems and deepen the student's knowledge.

5. ASSESSMENT

- Start-of-the-class Quizzes (10%)
- Assignments (40%)
- Closed book exam (50%)

Grading policy

- Students may not miss more than 2 sessions. Failing to do so will yield an automatic Fail grade.
- To pass, students will need to get at least 50% of the overall grade AND at least 50% of the exam. Failure to do so, will yield a Fail grade

6. TECHNICAL/LAPTOP REQUIREMENT

Having laptops on the seminar sessions is highly advised.

7. TOPIC OUTLINE AND SCHEDULE

Session	Topics	Readings
October 21	Simple regression analysis 1. (non-parametric regression, linear regression, OLS)	BK Ch 07
October 28	Simple regression analysis 2. (predicted values and residuals, regression and causality)	BK Ch 07
November 4	Complicated patterns and messy data 1. (taking log and other transformations of variables, piecewise linear splines and polynomials)	BK Ch 08
November 9	Complicated patterns and messy data 2. (measurement error in variables, influential observations, using weights)	BK Ch 08

November 11	Generalizing results of regression analysis 1. (standard error, confidence interval, prediction interval, testing, external validity)	BK Ch 09
November 18	Multiple linear regression (linear regression mechanics, binary and other qualitative right-hand-side variables)	BK Ch 10
November 18	Seminar 1: Replicating case studies from BK Ch 07-09 with R	-
November 23	Multiple linear regression (interactions, ceteris paribus vs. conditioning in multiple regression)	BK Ch 10
December 2	Probability models (linear probability, logit and probit, marginal differences, goodness of fit)	BK Ch 11
December 2	Seminar 1: Replicating case studies from BK Ch 10-11 with R	-
December 7	Time series regressions (trends, seasonality, lags, serial correlation)	BK Ch 12
December 14	Putting together and interpreting reports, based on data analysis. Overview.	-

8. SHORT BIO OF THE INSTRUCTOR

Ágoston Reguly is a fifth-year Economics PhD student at the Department of Economics and Business. He has worked more than three years at the Hungarian Government Debt Management Agency as an expert analyst. His research topics are in econometrics, policy evaluation with machine learning techniques and optimal survey designs.