New York University
Wilf Family Department of Politics

Quantitative Methods III (Fall 2014)

Mon & Wed 10:00-11:50, Room 217, 19W4
Instructor: Arturas Rozenas
ar199@nyu.edu
Office Hours: Tue 10:00-12:00, Room 215, 19W4
Lab session: Thu 10:00-11:00, 3d floor computer lab. Instructor: Elad Zippory

OVERVIEW & PREREQUISITES

The course is intended to provide students with training in statistical inference and modeling, with focus on generalized linear models, panel data, parametric and non-parametric frequentist and Bayesian methods. The goal of the class is to enable students to read state of the art political science and some statistics/econometrics literature, employing more advanced data modeling methods and innovate new models. The course requires that students have working knowledge of probability theory, matrix algebra, calculus, and statistical inference at the level of Quant I. The course is restricted to PhD students in Politics Department, but students from other department may audit the class if space permits.

TEXTBOOKS & READINGS

There is no one specific textbook used for this class. However, a lot of the material in the class is covered in the following textbooks, which I recommend you purchase, if possible. The first is one among many treatments of GLM; you can look for alternatives, that differ in price and the depth of exposition. This one is reasonably priced as written at about the level we need for this class. The second book is perhaps the most widely used textbook by applied researchers working with multilevel data.

- Annette J Dobson. *An introduction to generalized linear models*. 3rd ed. CRC press, 2008.
- Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press, 2007.

In addition, it might be useful to have access to these textbooks as well:

- George Casella and Roger S. Berger. *Statistical Inference*. Duxbury: Thompson Learning, 2002. A standard masters-level textbook on statistical inference, covers mostly prerequisites for this class plus some of the MLE and robust inference material.
- J Scott Long. "Regression models for categorical and limited dependent variables". In: *Advanced quantitative techniques in the social sciences* 7 (1997). A gentle and highly applied introduction to GLM's. Focuses a lot on parameter interpretation.
- William H Greene. *Econometric Analysis*. 7th ed. Prentice Hall, 2010. Widely used PhD-level econometrics textbook, covers many topic in this class, but with greater emphasis on panel data and causal inference.
- Gary King. *Unifying Political Methodology. The Likelihood Theory of Statistical Inference*. Ann Arbor: University of Michigan Press, 1989. Covers many GLM's and some likelihood theory, with many examples from political science.
- Peter Mccullagh and John A. Nelder. *Generalized Linear Models*. London: Chapman and Hall, 1989. A standard reference text for GLM's. Quite outdated, but still widely referenced.
- Dipak K Dey, Sujit K Ghosh, and Bani K Mallick. *Generalized linear models: A Bayesian perspective*. CRC Press, 2000. An authoritative Bayesian treatment of GLM's.

Each week you will have to read 1-3 papers, designated below as required readings. With each topic, I also provide a short list additional readings. Some of these are theoretical papers and others are applications of the methods at hand. Course materials will be posted on the Dropbox folder.

SOFTWARE

The class will use R software. If you do not know this software, start familiarizing yourself with it by googling letter 'R.'

GRADING

30% of your grade will be based on six **homework assignments**. Depending on the number of people in the class, I might divide you into groups to facilitate collective work on the assignments. 60% of grade will be based on the in-class **midterm test** (30%) and a **final test** (30%), consisting of in-class and take-home parts,. The midterm test is scheduled to take place during the class of October 15. The remaining 10% is going to be for an in-class paper presentations. Each student will be expected to make an in-class presentation of a paper assigned to that week's class. The papers students will choose for presentation are listed in "Further readings" list, marked with an asterisk.

SCHEDULE & TOPICS

1. (SEPT 3) INTRODUCTION

Limitations of linear regression, introduction to GLMs, distribution theory, estimation and inference, sufficiency, properties of some frequently encountered distributions, exponential family of distributions, a brief intro to Monte Carlo.

Required readings:

- Gary King, Michael Tomz, and Jason Wittenberg. "Making the most of statistical analyses: Improving interpretation and presentation". In: *American journal of political science* (2000), pp. 347–361

Further readings:

- Dobson, Ch. 3

2. (SEP 8/10) MAXIMUM LIKELIHOOD ESTIMATION

Likelihood function, analytical and numerical optimization, score function, Fisher's information, Cramer-Rao bound, statistical properties of MLE's, confidence interval estimation (asymptotic, simulation-based, bootstrapping).

Required readings:

- M: Dobson, Ch. 4.
- W: Dobson, Ch. 5.

Further readings:

- William H Greene. *Econometric Analysis*. 7th ed. Prentice Hall, 2010, Ch. 14.
- Gary King. *Unifying Political Methodology. The Likelihood Theory of Statistical Inference*. Ann Arbor: University of Michigan Press, 1989, Ch. 3, 4.
- Bradley Efron. "Better bootstrap confidence intervals". In: *Journal of the American statistical Association* 82.397 (1987), pp. 171–185.

3 (SEP 15/17) BINARY AND ORDINAL RESPONSE DATA I

The nature of categorical data, parametric models, model estimation and interpretation, separability problem, hypothesis testing and goodness-of-fit, in- and out-of-sample prediction, ROC and AUC.

Required readings:

- M: Gelman & Hill, Ch. 5.
- W: Dobson, Ch. 7.

Further readings:

- *Justin Esarey and Andrew Pierce. "Assessing Fit Quality and Testing for Misspecification in Binary-Dependent Variable Models". In: *Political Analysis* 20.4 (2012), pp. 480–500.
- *Gary King and Langche Zeng. "Logistic regression in rare events data". In: *Political analysis* 9.2 (2001), pp. 137–163.
- Christopher Zorn. "A solution to separation in binary response models". In: *Political Analysis* 13.2 (2005), pp. 157–170.

4. (SEP 22/24) BINARY AND ORDINAL RESPONSE DATA II

Ordinal models, parallel slopes and generalized ordinal models, nested binary/ordinal models, interactions in GLM's, multivariate-response binary data.

Required readings:

- M: Dobson, Ch. 8.

- W: William Greene. "Testing hypotheses about interaction terms in nonlinear models". In: *Economics Letters* 107.2 (2010), pp. 291–296.

Further readings:

- *William D Berry, Jacqueline HR DeMeritt, and Justin Esarey. "Testing for interaction in binary logit and probit models: is a product term essential?" In: *American Journal of Political Science* 54.1 (2010), pp. 248–266.

- C. Signorino and K. Yilmaz. "Strategic Misspecification in Regression Models". In: *American Journal of Political Science* 47.3 (2003), p. 551.

- Michael D Ward, Brian D Greenhill, and Kristin M Bakke. "The perils of policy by p-value: Predicting civil conflicts". In: *Journal of Peace Research* 47.4 (2010), pp. 363–375.

5. (SEP 29/OCT 1) MULTINOMIAL AND COUNT DATA

Multinomial data models, independence of irrelevant alternatives, Poisson distribution, over-dispersion, negative binomial model, estimation using quasi-likelihood functions, asymptotic theory, zero-inflated data.

Readings:

- M: Dobson, Ch. 8.
- W: Dobson, Ch. 9.

Further readings:

- *Benjamin E Lauderdale. "Compound Poisson–Gamma Regression Models for Dollar Outcomes That Are Sometimes Zero". In: *Political Analysis* (2012), mps018.

- Jeffrey T Grogger and Richard T Carson. "Models for truncated counts". In: *Journal of applied econometrics* 6.3 (1991), pp. 225–238.

- Gary King. "Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential Poisson regression model". In: *American Journal of Political Science* (1988), pp. 838–863.

- Emilie M Hafner-Burton, Edward D Mansfield, and Jon CW Pevehouse. "Human Rights Institutions, Sovereignty Costs and Democratization". In: *British Journal of Political Science* (2011), pp. 1–27.

6. (OCT 6/8) BAYESIAN INFERENCE I

Principles of Bayesian inference, conjugacy, constructing uninformative priors, Jeffrey's prior, Gelman's weakly informative priors, improper priors, analytical posterior computation, interpretation of Bayesian estimates, credible intervals, highest posterior density intervals.

Required readings:
- M: Jackman, Chapter 1.
- W: Jackman, Chapter 2.

Further readings:

- *Bruce Western. "Causal heterogeneity in comparative research: A Bayesian hierarchical modelling approach". In: *American Journal of Political Science* (1998), pp. 1233–1259.

- Dipak K Dey, Sujit K Ghosh, and Bani K Mallick. *Generalized linear models: A Bayesian perspective*. CRC Press, 2000.

- Andrew Gelman et al. "Objections to Bayesian statistics". In: *Bayesian Analysis* 3.3 (2008), pp. 445–449.

7. (OCT 15) BAYESIAN INFERENCE II

Maximum a priori likelihood approach and comparison to penalized likelihood, joint vs marginal posteriors, analytical posterior for regression analysis, predictive distributions.

Required readings:
- W: Jackman, Chapter 2.

Further readings:

- *Jeff Gill and Lee D Walker. "Elicited priors for Bayesian model specifications in political science research". In: *Journal of Politics* 67.3 (2005), pp. 841–872.

8. (OCT 20/22) MIDTERM & COMPUTATIONAL BAYES

Markov chains, introduction to Monte Carlo methods for posterior computation.

Required readings:

- M: Midterm.
- W: Jackman, Chapter 5.

Further readings:

- Allan E. Gelfand and Adrian F. M. Smith. "Sampling-Based Approaches to Calculating Marginal Densities". In: *Journal of the American statistical association* 85.410 (1990), pp. 398–409. ISSN: 0162-1459.
- Simon Jackman. "Estimation and inference are missing data problems: Unifying social science statistics via Bayesian simulation". In: *Political Analysis* 8.4 (2000), pp. 307–332.

9. (OCT 27/OCT 29) TIME-SERIES DATA

Stochastic processes, stationarity, spurious regression, co-integration, univariate time-series models (ARMA, ARCH, GARCH), VAR, impulse response functions, error correction models.

Required readings:

- M: Green, Chapter 20.
- W: Suzanna De Boef and Luke Keele. "Taking time seriously". In: *American Journal of Political Science* 52.1 (2008), pp. 184–200.

Further readings:

- *David B Carter and Curtis S Signorino. "Back to the future: Modeling time dependence in binary data". In: *Political Analysis* 18.3 (2010), pp. 271–292.
- Nathaniel Beck and Jonathan N Katz. "Modeling dynamics in time-series-cross-section political economy data". In: *Annual Review of Political Science* 14 (2011), pp. 331–352.
- Patrick T Brandt and John R Freeman. "Advances in Bayesian time series modeling and the study of politics: Theory testing, forecasting, and policy analysis". In: *Political Analysis* 14.1 (2006), pp. 1–36.
- Luke Keele and Nathan J Kelly. "Dynamic models for dynamic theories: The ins and outs of lagged dependent variables". In: *Political Analysis* 14.2 (2006), pp. 186–205.

10. (NOV 3/NOV 5) HIERARCHICAL/MULTILEVEL MODELS

Hierarchical data structures, inferential advantages and challenges, hierarchical linear regression models, random vs fixed effect, marginal vs conditional models, multi-level data with time-invariant covariates, inference and prediction in multilevel models, hierarchical GLM models, dynamic models models for panel data, random-effects distributions, generalized estimating equations.

Required readings:

- M: Gelman & Hill Ch. 11-13.
- W: Gelman & Hill Ch. 14-15.

Further readings:

- Green, Chapter 11
- Tom S Clark and Drew A Linzer. "Should I Use Fixed or Random Effects?" In: *Mimeo, Emory University* (2012).
- Gelman & Hill Ch. 16-17.
- *Christopher JW Zorn. "Generalized estimating equation models for correlated data: A review with applications". In: *American Journal of Political Science* (2001), pp. 470–490.
- *Boris Shor et al. "A Bayesian multilevel modeling approach to time-series cross-sectional data". In: *Political Analysis* 15.2 (2007), pp. 165–181.

11. (NOV 10/NOV 12) LATENT VARIABLES

Ideal point estimation, EM algorithm, Bayesian ideal point estimation, expert opinion aggregation, Bayesian latent variable models for dynamic binary/ordinal data, Markov switching models, state-dependence

Required readings:

- M: John S Ahlquist and Christian Breunig. "Model-based clustering and typologies in the social sciences". In: *Political Analysis* 20.1 (2012), pp. 92–112.

- W: Shawn Treier and Simon Jackman. "Democracy as a Latent Variable". In: *American Journal of Political Science* 52.1 (2008), pp. 201–17.

Further readings:

- *Thomas König, Moritz Marbach, and Moritz Osnabrügge. "Estimating Party Positions across Countries and TimeA Dynamic Latent Variable Model for Manifesto Data". In: *Political Analysis* 21.4 (2013), pp. 468–491.

- *Royce Carroll et al. "The Structure of Utility in Spatial Models of Voting". In: *American Journal of Political Science* 57.4 (2013), pp. 1008–1028.

- *Mark R Beissinger. "The Semblance of Democratic Revolution: Coalitions in Ukraine's Orange Revolution". In: *American Political Science Review* 107.03 (2013), pp. 574–592

- Andrew D. Martin and Kevin M. Quinn. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court, 1953–1999". In: *Political Analysis* 10.2 (2002), pp. 134–153.

- Joseph Bafumi et al. "Practical Issues in Implementing and Understanding Bayesian Ideal Point Estimation". In: *Political Analysis* 13.2 (2005), pp. 171–187.

12. (NOV 17/NOV 19) MIXTURE MODELS

Density estimation using mixtures, E-M algorithm, finite gaussian mixtures for regression analysis, mixture models for random effects, parametric flexibility and robustness, split-population models, profile regression, component selection and cross-validation.

Required readings:

- M: Kosuke Imai and Dustin Tingley. "A statistical method for empirical testing of competing theories". In: *American Journal of Political Science* 56.1 (2012), pp. 218–236.

Additional readings:

- *Benjamin E Bagozzi and Bumba Mukherjee. "A mixture model for middle category inflation in ordered survey responses". In: *Political Analysis* 20.3 (2012), pp. 369–386

- Jong Hee Park. "Changepoint analysis of binary and ordinal probit models: An application to bank rate policy under the interwar gold standard". In: *Political Analysis* 19.2 (2011), pp. 188–204.

- Isobel Claire Gormley and Thomas Brendan Murphy. "A mixture of experts model for rank data with applications in election studies". In: *The Annals of Applied Statistics* (2008), pp. 1452–1477.

- Jennifer L Hill and Hanspeter Kriesi. "Classification by opinion-changing behavior: A mixture model approach". In: *Political Analysis* 9.4 (2001), pp. 301–324.

- Nils B Weidmann. "Violence "from above" or "from below"? The Role of Ethnicity in Bosnia's Civil War". In: *The Journal of Politics* 73.04 (2011), pp. 1178–1190.

13. (NOV 24/26) SEMI-PARAMETRIC AND NON-PARAMETRIC METHODS

Generalized additive models, spline basis functions, penalized spline regression, cross-validation, dimensionality problem, kernel regularization methods, Bayesian Gaussian process regression, applications for causal inference.

Required readings:

- M: Nathaniel Beck and Simon Jackman. "Beyond linearity by default: Generalized additive models". In: *American Journal of Political Science* (1998), pp. 596–627.

Further readings:

- *Jens Hainmueller and Chad Hazlett. "Kernel Regularized Least Squares: Reducing Misspecification Bias with a Flexible and Interpretable Machine Learning Approach". In: *Political Analysis* 22.2 (2014), pp. 143–168.

- Simon Wood. *Generalized additive models: an introduction with R*. CRC press, 2006.

- Anjishnu Banerjee, David B Dunson, and Surya T Tokdar. "Efficient Gaussian process regression for large datasets". In: *Biometrika* (2012), ass068.

- Abel Rodríguez, David B Dunson, and Alan E Gelfand. "Bayesian nonparametric functional data analysis through density estimation". In: *Biometrika* 96.1 (2009), pp. 149–162.

14. (DEC 1/3) MODEL SELECTION, AVERAGING, AND PREDICTION

Specification uncertainty, Bayesian model selection and averaging, choice of priors for model selection, modeling averaging approach to non-parametric regression modeling, Bayesian ensemble learning for improved prediction.

Required readings:

- M: Larry M Bartels. "Specification uncertainty and model averaging". In: *American Journal of Political Science* (1997), pp. 641–674.

Further readings:

- *Jacob M Montgomery, Florian M Hollenbach, and Michael D Ward. "Improving predictions using ensemble Bayesian model averaging". In: *Political Analysis* 20.3 (2012), pp. 271–291.

- Kevin A Clarke. "A simple distribution-free test for nonnested model selection". In: *Political Analysis* 15.3 (2007), pp. 347–363.

- Adrian E Raftery, David Madigan, and Jennifer A Hoeting. "Bayesian model averaging for linear regression models". In: *Journal of the American Statistical Association* 92.437 (1997), pp. 179–191.

15. (DEC 8) SPATIAL STATISTICS

Spatial stochastic process, covariance functions, kriging for continuous measurement locations, spatial smoothing of discretely measured data, regression models for spatially structured data.

Final test (Dec 10), in class.