Political Science 504/597C: Advanced Topics in Political Methodology

Tuesday 1-3pm Pond 236, & Thursday 1-2pm Pond Grad Lab

Fall 2011

James Honaker	Phone: (814) 865 6901
Pond 317	tercer@psu.edu
Eitan Tzelgov	eut113@psu.edu

1 Goals of This Course

The purpose of this class is to learn a system of inference, the likelihood theory, as a framework to tackle real issues and problems found in political science data, and to make inferences to solve questions of substantive importance. At the end of this class a student should be able to read a methods article, understand which types of political science data it applies to, program an estimator, and tailor or modify it for a particular dataset or application. They should then be able to present their statistical results in a format meaningful and informative to a nontechnical audience.

In the first five weeks, we will build up a foundation in likelihood theory, and explore some common and increasingly elaborate models as our abilities grow. The next two thirds of term will be spent covering common features and problems with political science data. At first we will tackle these with estimation methods grounded in likelihood theory, and towards the end of term wander into other systems of inference–bootstrapping, Bayesian and matching methods. The topics we will focus on are common features found in political science data. They are features of data that if present and unaddressed in your analysis, will provoke audiences to murmur though your talks, and referees to summarily dismiss your work no matter the quality or originality of all other parts. These are topics that also demonstrate the growing maturity of the field, as political methodologists solve problems important to political science data instead of shoehorning methods developed by econometricians for economic data.

There are of course many issues we will not address. Most notably we will may touch briefly on panel data but will not address time series analysis. We'll also avoid hierarchical models and N-stage least squares methods. Partly this is to theoretically unify this semester around issues addressed by likelihood approaches and its near neighbours. Partly this is because these topics are commonly clustered in substantive subfields that rely on "time-series cross-sectional" datasets, and a follow on class drawing these together will be offered next term. In the meantime, students intending to use time series data in their final papers for this class should see me for additional reading and study or dissuasion.

2 Class Format and Materials

Class on Tuesday will be lecture format and focus on the fundamentals, statistical development, theory and proof. Class on Friday will be in the computer lab, and will focus on actual computational methods of estimation with real and Monte Carlo datasets. We will be using the R statistical language. No prior knowledge of R is necessary. The questions posed in the problems sets will be primarily solved by using R. The learning curve is quite steep, but by the time we begin the topics, we should be able to write programs, and by the end of the course we should consider ourselves programmers. As an aside I'll provide pointers to Stata packages that provide similar tools when they are available. Feel free to use these in your final papers if preferred.

Most of the material from the first half of class is covered in "Unifying Political Methodology" by Gary King, and also in "In All Likelihood" by Yudi Pawitan. King is very readable, intuitive and oriented towards political science. Pawitan is rigorous and newer. Pick which author you prefer, or for the ambitious, read King before lecture and Pawitan afterwards. In the second half of term we will look at a selection of models popular or valuable to political science. Specific suggested texts devoted to these models are listed at the end of the syllabus. You might track one of these down if you already know your project will focus on one of these topics.

3 Problem Sets

There will be problem sets most weeks. Problem sets should be worked on in pairs. One printed paper write-up should be submitted by each pair. Write-ups should summarize the approach used in the problem as well as any results. A good write-up should read like the third quarter of a journal article. Any presented statistics should be explained and interpreted. Graphs should have titles, captions and axis labels. Digressions are encouraged.

Please, please, give me one paper copy of your write up. Additionally with each problem set all code should be submitted electronically to `tercer@psu.edu` in one zip file.

4 Final Papers

There will be a final paper but no final exam. There will probably be a short take home midterm that resembles a problem set and be more of a diagnostic than an exam. If you do not already have a dataset you are using for a current research paper you should begin to look for a dataset immediately. A good place to start is the Dataverse `thedata.org`or ICPSR `http://www.icpsr.umich.edu/` archives. Americanists might also look at the ANES, GSS, and ROAD datasets. A very good idea is to find a paper that interests or irks you, that you believe should be done differently or for which there are obvious further additional questions to be answered in the data. Check the Dataverse and ICPSR for a replication dataset or email the author. Replications with additional improvements or developments are excellent topics for papers. This will not be looked down on. Replications are not easy. You should have your dataset by the end of the second week of class. By the end of the third week you will need to write a paragraph summarizing your dataset and research agenda. Get on this today.

In the eleventh week you will need to submit your preliminary results by printing up all graphs, tables and figures. Two copies should be printed. No write-up is necessary, but include an short introduction explaining the topic and any information necessary to understand the figures and tables. Also you need to submit a replication dataset and code to replicate your results. These will be exchanged between students.

September 8 Paragraphs due describing dataset and research agenda of paper.

November 10 Preliminary figures, tables and graphs for paper due, along with paragraph of introduction and all replication datasets and code on a disk.

December 1 Replication summaries due.

December 16 Papers due.

5 Chronology of Topics

Most of the topics in the first half of term are covered in King's UPM. A more technical treatment is also in Pawitan. Specific texts treating the latter weeks are noted below.

Week 1

Theory: *Likelihood as a theory of inference.*
The meaning of inference and uncertainty.
Inverse probability and likelihood.

Week 2

Theory: *Properties of maximum likelihood estimators.*
Numerical derivatives, Newton, Fisher, and other optimizers
Taylor Series Expansions and Newton's Method

Practice: Binary dependent variables: *Latent variables.*

Week 3

Theory: *Taylor Series Expansions and Wald and Wilk Tests*
Construction of Confidence Intervals

Week 4

Theory: *Coefficient interpretation*
Methods of "First differences."
Parametric bootstrap simulation

Practice: Event count and grouped data: *Problems of aggregation. Problems of dispersion. Binomial and Poisson Models.*

Week 5

Theory: *Sampling from likelihoods*

Practice: More event count and grouped data: *Negative Binomial Distributions, Expanded Beta Binomials, Hurdle and Zero-inflated models.*(Cameron)

Week 6

Theory: *EM algorithms and other scoring maximizers* (Givens)

Practice: Censored and truncated variables. Ordered Choice: *Selection bias. Two-stage estimation. Stochastic and unobserved thresholds. Penalty Functions.* (Maddala)

Week 7

Practice: Estimating Functional Forms: *Local regression. Splines. ROC-curves.* (Hastie, Wood)

Week 8, 9

Practice: Duration Data: *Hazard functions. Time-varying covariates. Semiparametric estimation. Markovian transition models.* (Box-Stefensmeier)

Week 10

Theory: *EM algorithms and fully unobserved variables* (Givens)

Practice: Mixture Models: *Mixed distributions. Latent category analysis.* (McLachlan)

Week 11

Practice: Missing data: *Unit and item nonresponse. Missingness mechanisms. Imputation methods.* (Schafer)

Possible Topics for Additional Weeks

- Scale Construction: *Eigenvalues, Factors, Principal Components. Projection and Information.*
- Compositional variables: *Ratio transformations. Additive logistic distributions. Robust and t-based estimation. Seemingly Unrelated Regressions.*
- Ecological inference: *Ecological fallacies. Goodman's regression. Information from constraints. Shrinkage.* (King)
- Rare Event Data: *Oversampling Bias Corrections, Small Sample Bias Corrections, Jackknife Corrections*
- Measurement Error: *Multiple proxies, SIMEX and 2SLS approaches. Multiple Overimputation.*
- Non-nested model comparison: *Out-of-sample forecasting, k-folding, leave-one-out scoring. Information Criteria. Bayesian Model Averaging.*
- Bag of Words Models: *Term frequency matricies. Text as count data. Ideal point estimation.*
- Endogeneous Relationships: *Reciprocal Causation. n-SLS. Matching and experimental design.*

6 What should I do today

You should check Angel has your correct mailing address. You should email me your name, fields of interest, background in maths, and whether you are interested in a computational or applied track. Also tell me if you are not registered but want to receive class emails. You should find your dataset.

7 Texts

King, Gary. 1989. *Unifying Political Methodology*, Michigan Press.

Pawitan, Yudi. 2001. *In All Likelihood*, Oxford University Press.

Givens Geof H. and Jennifer A. Hoeting. 2005. *Computational Statistics*, Wiley.

Greene, William H. 2007 *Econometric Analysis, 6/e*, Prentice Hall.

8 References

Box-Steffensmeier, Janet M. and Bradford S. Jones. 2004. *Event History Modeling*, Cambridge.

Cameron, A. Colin and Pravin K. Trivedi. 1998. *Regression Analysis of Count Data*, Cambridge.

Gelman, Andrew et al. 2003. *Bayesian Data Analysis, Second Edition*, Chapman and Hall.

Hastie, T.J. and T.J. Tibshirani. 1990. *Generalized Additive Models*, Chapman and Hall.

King, Gary. 1997. *A Solution to the Ecological Inference Problem.* Princeton.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*, Cambridge.

McLachlan, Geoffrey, and David Peel. 2000. *Finite Mixture Models*, Wiley.

Schafer, Joseph L. 1997. *Analysis of Incomplete Multivariate Data*, Chapman and Hall.

Tanner, Martin A. 1996. *Tools for statistical inference: observed data and data augmentation methods, 3rd edition*, New York: Springer.

Wood, Simon. 2006. *Generalized Additive Models: An Introduction with R*, Chapman and Hall.

Academic Dishonesty

The Department of Political Science, along with the College of the Liberal Arts and the University, takes violations of academic dishonesty seriously. Observing basic honesty in one's work, words, ideas, and actions is a principle to which all members of the community are required to subscribe.

All course work by students is to be done on an individual basis unless an instructor clearly states that an alternative is acceptable. Any reference materials used in the preparation of any assignment must be explicitly cited. Students uncertain about proper citation are responsible for checking with their instructor.

In an examination setting, unless the instructor gives explicit prior instructions to the contrary, whether the examination is in class or take home, violations of academic integrity shall consist but are not limited to any attempt to receive assistance from written or printed aids, or from any person or papers or electronic devices, or of any attempt to give assistance, whether the one so doing has completed his or her own work or not. Lying to the instructor or purposely misleading any Penn State administrator shall also constitute a violation of academic integrity.

In cases of any violation of academic integrity it is the policy of the Department of Political Science to follow procedures established by the College of the Liberal Arts.

Disabilities

The Pennsylvania State University encourages qualified people with disabilities to participate in its programs and activities and is committed to the policy that all people shall have equal access to programs, facilities, and admissions without regard to personal characteristics not related to ability, performance, or qualifications as determined by University policy or by state or federal authorities. If you anticipate needing any type of accommodation in this course or have questions about physical access, please tell the instructor as soon as possible. Reasonable accommodations will be made for all students with disabilities, but it is the student's responsibility to inform the instructor early in the term. Do not wait until just before an exam to decide you want to inform the instructor of a learning disability; any accommodations for disabilities must be arranged well in advance.