**Predicting Elections: Analytical Techniques and Illustrative Case Studies**

**Instructor Clifford Young**

Election forecasting is an activity which attempts to predict the outcome of an election. Such prediction can either be very far out from the election (a year or more) or very close to election-day (a few days before). Election forecasting can also employ the most simple of analytical frameworks to highly-sophisticated multivariate statistical models.

Election prediction has many applications in practice. Indeed, reducing uncertainty about which candidate will win on election-day can be an important strategic advantage for decision-makers in the private sector as we all in politics. The examples are many but include financial services firms making bets on currency or equities, ‘brick and motor’ companies making long-term capital investments, and political parties choosing the most viable candidate. Ultimately, election- forecasting is a central tool in assessing the political risk associated with any specific decision.

Election forecasting is increasingly finding its way into the national media coverage of elections. The best example of this is the 2012 US Presidential election which pitted Obama versus Romney. During this election cycle, Nate Sliver as well as other election forecasters became important stakeholders in the national narrative about the elections and ultimately predicted the election spot on. Many are talking about the rise of the “forecaster-pundit” who will replace traditional journalist and politicos in election analysis.

So how is it done? How did Nate Silver and his cohorts do it? More generally, how is election forecasting employed in practice today? What methods are employed? And are some better under certain conditions? And what are different methods’ relative strengths and weaknesses?

The course will attempt to answer these questions through the use of both an examination of the extant literature on the subject on election forecasting as well as using illustrative case studies. The primary objective is to provide the student will a general understanding of election forecasting and a simple tool box to work from. Here it is important to stress that the course will focus on national elections (in both presidential and parliamentary systems) not on downstream races at the legislative, state, and local levels.

The course also includes four components:

First, the course will review the existing literature on forecasting as well as best-industry practices. In particular, we will explore definitions of prediction and forecasting; will survey theories behind behavioral change and behavioral prediction; will examination of the median voter model and related concepts; and will review of different types of analyst biases and how they can hinder optimal prediction.

Second, the course will review the different forecasting models used today by election forecasters, including the most naïve ‘rule of thumb’ frameworks to highly sophistical multivariate statistical models. We will explore a number of different statistical models including those which include purely economic variables to those which employ public opinion variables as well. Finally, we will also examine statistical models using both classical and Bayesian frameworks. In essence, we want to compare and contrast different methods, assessing their accuracy and determining under what conditions they perform more optimally and under which conditions they don’t. Ultimately, we will find that ‘one size does not fit all’ and that the forecaster needs to have a broad tool box of methods in order to be effective across contexts and countries.

Third, we will use the US 2012 presidential election and the Brazilian 2010 election to assess the different forecasting methods. I choose these two elections because of their unique characteristics: the US is what I consider a data-intensive election environment, while Brazil is a data-scarce one. Such relative differences in the volume of information require different forecasting solutions. When appropriate, we will also examine other elections including the 2013 Kenyan, Italian, and Venezuelan elections; the 1992 British elections; 2011 Nigerian Elections as well as the upcoming Brazilian 2014 election.

Fourth, the course will also include lab sessions which will be used to run different models and methods. Specifically, we will analyze data both from the 2012 US presidential and 2014 Brazilian elections. The lab is meant to give ‘hands on’ data experience and practical tools.

Prerequisites:

* Some familiarity with SPSS, STATA R, or SAS. We will be using a combination of SPSS, Excel, and R.
* At least one course in multivariate statistics

**Day 1**

**Morning**

Introductions and Objectives of the Course

What is Election Forecasting? When is it used and why?

Analyst Bias, Model Validation and Model Training, Law of Large Numbers

**Afternoon 1**

Polling and Voting Intention Questions

Behavioral Change Model (Attitude, Intention, Barriers)

Median Voter Model

Heuristic Based Modeling (Main Problems)

Case Study: Gallup US Presidential Election 2012 and Brazil Presidential 2010 (Serra v. Dilma)

Introduction to Lab: data sets and statistical packages

**Afternoon 2 🡺 A look at the datasets.**

**Readings:**

Graefe and Armstrong (2009) “Predicting elections from the most important issue: A test of the take-the-best heuristic” White Paper

**Nate Silver (2012) The Signal and the noise: why so many predictions fail—but some don’t; (Chapters 1-3)**

**Recommended**

Congleton (2010) “The Median Voter Model” White Paper Center for Study of Public Choice George Mason University

**Day 2: Simple and Model-Based Poll Averaging**

**Morning**

Simple Poll Averaging

Model-based Poll Averaging

Averaging versus Forecasting

House effects, Timing of Poll and Other Issues

Assessing Poll Acuracy

Strengths and Weaknesses

**Afternoon**

Lab Session using the 2012 US Presidential Election data set

**Readings:**

Jackman, S (2005) “Pooling the Polls over an Election Campaign

Wlezien and Erikson (2007) “Temporal Horizons and Presidential Elections Forecasts” American Politics Research 24, 492-505.

Young, C.A. (2014*). “The rise of the forecaster-pundit big data, data aggregation and the 2012*

*US Presidential Election”.* Ipsos Understanding Society. July, London.

**Day 3: Economic And Other Structural Electoral Forecasting Models**

**Morning**

Economic and structural election forecasting models

Approval Ratings and Economic Variables

Proxy Models

Historical Performance

Strengths and Weaknesses

**Afternoon**

Lab Session using the 2012 US Presidential Election data set

**Readings:**

Abramowitz, A (2008) Forecasting the 2008 presidential election with the time-for-change model PS: political science & politics 41: 691-5

Mueller (1970) The American Political Science Review, Vol. 64, No. 1, pp. 18-34Published

**Day 4: False Positives, Bayesian Models, and combining polls and other model inputs**

**Morning**

Campaign and Other effects: Discerning False Positives and Herding

Bayesian Estimations Techniques: Model Based Poll Averaging

Combining multiple estimates

More Data: State and National Level Polls

Some Discussion of Smoothing

**Afternoon**

Lab Session using the 2012 US Presidential Elections dataset

**Readings:**

Linzer, D (2013) “Dynamic Bayesian Forecasting of presidential elections in the states” Journal of the American Statistical Association March 2013 pp124-134

**Recommended (Not required)**

Holbrook (1996) Do Campaigns Matter? Sage Publications

Jackman (2000) Estimation and Inference Are Missing Data Problems: Unifying Social Science Statistics via Bayesian Simulation Political Analysis, Vol. 8, No. 4 (Autumn)

**Day 5: the 2014 Brazilian Presidential Election as a Case Study**

**Morning (Lab Session)**

Heuristic Models

Simple Poll Averaging

Model-Based Poll Averaging

Economic and Structural Models

**Afternoon (Lab Session)**

Bayesian Model-Based Poll Averaging

Combining of Estimates

Discussion of estimates