Technical Forecasting of Political Conflict

Philip A. Schrodt

Parus Analytics
schrodt735@gmail.com

Workshop on "Big Data and Death"
University of Wisconsin
7 November 2014

PARUS ANALYTICS

The Debate

FP
Foreign Policy

THURSDAY, NOVEMBER 29, 2012

▸ ARGUMENT

PRINT | TEXT SIZE | EMAIL | SINGLE PAGE

Why the World Can't Have a Nate Silver

The quants are riding high after Team Data crushed Team Gut in the U.S. election forecasts. But predicting the Electoral College vote is child's play next to some of these hard targets.

BY JAY ULFELDER | NOVEMBER 8, 2012

Vs.

FP
Foreign Policy

THURSDAY, NOVEMBER 29, 2012

▸ ARGUMENT

PRINT | TEXT SIZE | EMAIL | SINGLE PAGE

Predicting the Future Is Easier Than It Looks

Nate Silver was just the beginning. Some of the same statistical techniques used by America's forecaster-in-chief are about to revolutionize world politics.

BY MICHAEL D. WARD , NILS METTERNICH | NOVEMBER 16, 2012

Two approaches that did not work well in the past

Two approaches that did not work well in the past

Qualitative

Two approaches that did not work well in the past

Qualitative

Quantitative

Problems with qualitative approaches

Tetlock: Experts typically do about as well as a “dart-throwing chimp”

Problems with qualitative approaches

Tetlock: Experts typically do about as well as a “dart-throwing chimp”

Except for television pundits, who do even worse. Ask President Romney. Or pretty much every Democratic Party pundit prior to the 2014 election.

The media want things to be dramatic. “We’re all going to die! Details follow *American Idol*”

Problems with qualitative approaches

Tetlock: Experts typically do about as well as a "dart-throwing chimp"

Except for television pundits, who do even worse. Ask President Romney. Or pretty much every Democratic Party pundit prior to the 2014 election.

The media want things to be dramatic. "We're all going to die! Details follow *American Idol*"

Qualitative *theory* isn't much better:
Remember the hegemonic US seizure of undefended Canadian and Mexican oil fields in response to the 1973 OPEC oil embargo?

Problems with qualitative approaches

Tetlock: Experts typically do about as well as a "dart-throwing chimp"

Except for television pundits, who do even worse. Ask President Romney. Or pretty much every Democratic Party pundit prior to the 2014 election.

The media want things to be dramatic. "We're all going to die! Details follow *American Idol*"

Qualitative *theory* isn't much better:
Remember the hegemonic US seizure of undefended Canadian and Mexican oil fields in response to the 1973 OPEC oil embargo?

Neither do I.

SMEs and the “narrative fallacy”

SME = “subject matter expert”

Hegel: the owl of Minerva flies only at dusk

Taleb (*Black Swan*): seeking out narratives is an almost unavoidable cognitive function and it generates a dopamine hit

Tetlock (Good Judgement Project): prior knowledge as a SME contributes only 2% to improved forecasting accuracy

This is your brain on narratives

Problems with quantitative approaches

Ward, Greenhill and Bakke (2010): Models based on significance tests don't predict well because that is not what a significance test is supposed to do.

Gill, Jeff. 1999. The Insignificance of Null Hypothesis Significance Testing. *Political Research Quarterly* 52:3, 647-674.

Problems with quantitative approaches

Ward, Greenhill and Bakke (2010): Models based on significance tests don't predict well because that is not what a significance test is supposed to do.

Gill, Jeff. 1999. The Insignificance of Null Hypothesis Significance Testing. *Political Research Quarterly* 52:3, 647-674.

Frequentism is logically inconsistent and has been characterized in Meehl (1978) as "a terrible mistake, basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology"

- Hey, dude, tell us what you *really* think…
- But that is another lecture…

Problems with quantitative approaches

Ward, Greenhill and Bakke (2010): Models based on significance tests don't predict well because that is not what a significance test is supposed to do.

Gill, Jeff. 1999. The Insignificance of Null Hypothesis Significance Testing. *Political Research Quarterly* 52:3, 647-674.

Frequentism is logically inconsistent and has been characterized in Meehl (1978) as "a terrible mistake, basically unsound, poor scientific strategy, and one of the worst things that ever happened in the history of psychology"

- Hey, dude, tell us what you *really* think. . .
- But that is another lecture. . .

The norm in political science has been to do full-sample evaluation, whereas the norm in machine-learning has been split-sample, which is usually more robust and is certainly more credible

Kahneman et al: people are really bad at statistical reasoning

- Everyone, including statisticians unless they focus very hard
- Example: managed mutual funds, which both theory and evidence indicate cannot work
- Example: opposition to “evidence based medicine” in the US, with a preference for clinical intuition even when this has been demonstrated to be less effective
- Probabilitistic weather forecasts seem to be the one major exception: rain likelihood, hurricane tracks

The Necessity of Prediction in Policy

Feedforward: policy choices must be made in the present for outcomes which may not occur for many years

Planning Times: even responses to current conditions may require lead times of weeks or months

Factors encouraging technical political forecasting

- Conspicuous failures of existing methods: end of Cold War, post-invasion Iraq, Arab spring
- Success of forecasting models in other behavioral domains
  - Macroeconomic forecasting [maybe...]
  - Elections: Nate Silver effect
  - Demographic and epidemiological forecasting
  - Famine forecasting: USAID FEWS model
  - Example: statistical models for mortgage repayment were quite accurate
- Technological imperatives
  - Increased processing capacity
  - Information available on the web
- Decision-makers now expect visual displays of analytical information, which in turn requires systematic measurement
  - "They won't read things any more"

This must be important: it's in *The Economist*!

The science of civil war

What makes heroic strife

Computer models that can predict the outbreak and spread of civil conflict are being developed

Apr 21st 2012 | from the print edition

Like 95 Tweet 40

And *The Chronicle*!

THE CHRONICLE OF HIGHER EDUCATION
November 5, 2014

The Chronicle Review

Comment (1)

October 13, 2014

Can We Predict the Next War?

Mark Shaver for The Chronicle Review

Enlarge Image

By Beth McMurtrie

Every minute of every day, computers overseen by Virginia Tech process billions of bits of data in an attempt to predict the future.

Tweets from politicians, satellite images of hospital parking lots, news stories about rising bus fares: All are mined, categorized, and fed into algorithms designed to anticipate the next flu outbreak or which candidate will win a city election. As many as 50 computer-generated alerts flash daily on computer monitors and are evaluated for accuracy at the end of each month.

Most Popular

Most Viewed | Most Comme

1. Seeking a Climate Change
2. What the Midterm Election
3. Outcome of Governors' Rac
Policy in Several States
4. Confessions of a Young, Pr

Large Scale Conflict Forecasting Projects

- State Failures Project 1994-2001
- Joint Warfare Analysis Center 1997
- FEWER [Davies and Gurr 1998]
- Center for Army Analysis 2002-2005
- Swiss Peace Foundation FAST 2000-2008
- Political Instability Task Force 2002-present
- DARPA ICEWS 2007-present
- IARPA ACE and OSI
- Peace Research Center Oslo (PRIO) and Uppsala University UCDP models
- US Holocaust Memorial Museum Prediction Poll

Political Instability Task Force

- US government, multi-agency: 1995-present
- Statistical modeling of various forms of state-level instability
- Forecasting models actively used since about 2005
  - Two year probability forecasts with roughly 80% accuracy (AUC)
  - Predominantly logistic models with a simple “standard PITF”set of variables; shifting to Bayesian approaches
  - (PITF has accumulated a set of 2700 variables but only a small number end up being important predictors)

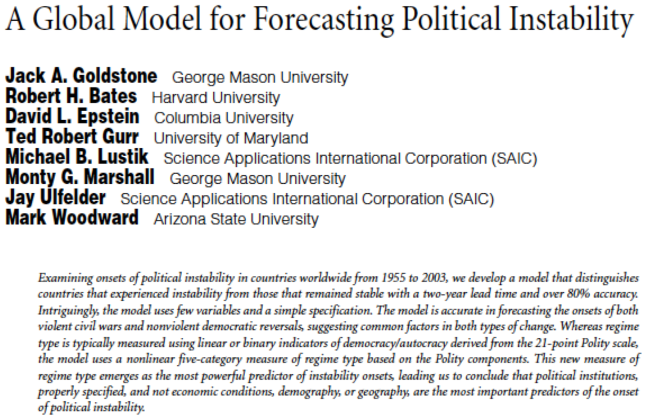
PITF Variables

Variables Tested

CONCEPT	SELECTED EXAMPLES OF MEASURES TESTED
state capacity	infant mortality, population, GDP, military personnel, polity durability
violent conflict	civil war, armed attacks, regional conflicts, reported fatalities in political violence, government mass killing
non-violent challenges to state authority	protests, strikes, government crises
government institutions	democracy, autocracy, factionalism, other polity measures
ethnic relations	ethnic diversity, elite ethnicity, state-led discrimination
demographics	youth-bulge
international ties	GATT/WTO membership, trade-openness

Two-year time horizon tends to favor structural variables Source: Ben Valentino and Chad Hazlett, "Forecasting Non-state Mass Killings", October 2012

PITF Results, ca. 2005


A Global Model for Forecasting Political Instability

Jack A. Goldstone George Mason University
Robert H. Bates Harvard University
David L. Epstein Columbia University
Ted Robert Gurr University of Maryland
Michael B. Lustik Science Applications International Corporation (SAIC)
Monty G. Marshall George Mason University
Jay Ulfelder Science Applications International Corporation (SAIC)
Mark Woodward Arizona State University

Examining onsets of political instability in countries worldwide from 1955 to 2003, we develop a model that distinguishes countries that experienced instability from those that remained stable with a two-year lead time and over 80% accuracy. Intriguingly, the model uses few variables and a simple specification. The model is accurate in forecasting the onsets of both violent civil wars and nonviolent democratic reversals, suggesting common factors in both types of change. Whereas regime type is typically measured using linear or binary indicators of democracy/autocracy derived from the 21-point Polity scale, the model uses a nonlinear five-category measure of regime type based on the Polity components. This new measure of regime type emerges as the most powerful predictor of instability onsets, leading us to conclude that political institutions, properly specified, and not economic conditions, demography, or geography, are the most important predictors of the onset of political instability.

Source: Amer J of Pol Sci Vol 54, no. 1, Jan 2010 pg. 190

Political Instability Task Force (*AJPS* 2010)

Table 2 **Out-of-Sample Prediction Exercise for Observed Onsets of Instability, 1995–2004**

A. Countries That Had Instability Onsets, 1995–2004. Quintile/decile in model score rankings based on 2-yr. prior data

Year	Top Decile	Second Decile	Second Quintile	Third Quintile
1995	Armenia, Comoros	Belarus		
1996	Albania, Niger, Zambia		Nepal	
1997	Cambodia, Congo-Brazz.			
1998	Guinea-Bissau, Lesotho			Serbia/Montenegro
1999	Ethiopia, Haiti			
2000		Solomon Ils., Guinea*		
2002	Cote d'Ivoire			
2003	Central African Republic			
2004	Iran*	Yemen*		Thailand*

B. Tabulation of All Country-years, 1995–2004. Model estimates based on censored data, using only sample data from prior to year of forecast (countries w/population over 500,000, no ongoing conflict, at least two years old)

	Countries with Instability in $t + 2$	Countries Remaining Stable
Predicted for Instability (Top Quintile)	18	233
Predicted for Stability (Not Top Quintile)	3	992
N = 1,246 Percent Classed Correctly	85.7%	81.0%

Number of instability onsets, 1995–2004: 21. Number of instability onsets in top quintile of model scores: 18 (86%).
*Cases added to the problem set in 2005 update.

This is ca. 2010

PITF Results, ca. 2005

Table 1 Results of Global Analysis of Onsets of Instability

Independent Variables	Full Problem Set		Civil War Onsets		Adverse Regime Change Onsets	
	Coefficient (S.E.)	Odds Ratio (95% CI)	Coefficient (S.E.)	Odds Ratio (95% CI)	Coefficient (S.E.)	Odds Ratio (95% CI)
Regime Type (Full Autocracy as Reference)						
Partial Autocracy	1.85***	6.37	1.94***	6.98	2.85***	17.32
	(0.47)	(2.53, 16.02)	(0.62)	(2.05, 23.8)	(0.86)	(3.19, 94.0)
Partial Democracy with	3.61***	36.91	3.35***	28.5	5.06***	157.0
Factionalism	(0.51)	(13.5, 101)	(0.73)	(6.86, 118)	(1.02)	(21.1, 1164)
Partial Democracy without	1.83***	6.22	.981	2.67	2.58***	13.23
Factionalism	(0.54)	(2.17, 17.8)	(0.79)	(0.57, 12.4)	(0.91)	(2.20, 79.5)
Full Democracy	0.981	2.67	.545	1.73	1.26	3.51
	(0.68)	(0.70, 10.2)	(0.92)	(0.29, 10.4)	(1.09)	(0.42, 29.5)
Infant Mortality†	1.59***	6.59	1.64***	4.19	1.38*	4.56
	(0.35)	(2.91, 14.9)	(0.48)	(1.82, 9.60)	(0.58)	(1.30, 16.0)
Armed Conflict in 4+	3.09***	22.0	2.81***	16.7	.091	1.10
Bordering States	(0.95)	(3.42, 142)	(0.82)	(3.36, 83.0)	(1.49)	(0.06, 20.4)
State-Led Discrimination	0.657*	1.93	1.17***	3.23	−.502	0.61
	(0.30)	(1.08, 3.45)	(0.36)	(1.59, 6.55)	(0.62)	(0.18, 2.04)
N = Total (Problems, Controls)	468 (117, 351)		260 (65, 195)		196 (49, 147)	
Onsets Correctly Classified	80.3%		80.0%		87.8%	
Controls Correctly Classified	81.8%		81.0%		87.8%	

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$. †Odds ratios for continuous variables compare cases at the 75th and 25th percentiles.

Source: Amer J of Pol Sci Vol 54, no. 1, Jan 2010 pg. 190

“Schrodt should do everything in ‘sevens”’

http://asecondmouse.org

Opportunities

- Totalitarian law of the universe: whatever is not forbidden is mandatory. Prediction *is* scientific [now]
- Data: small, big, fast
- You can't solve everything with more machine cycles, but it ~~never~~ rarely hurts
- Successful large-scale projects: PITF, ICEWS, ACE, ENCoRe
- (Mostly) Convergent models
- Location, location, location
- Open source, open access, open collaboration

Challenges

- Determining credible metrics
- Black swans
- Heterogeneous environments
- Absence of theories indicating what is and is not predictable
- Pournelle's Law: no task is so virtuous that it will not attract idiots
- Ethical concerns

OPPORTUNITIES

The Forecaster's Quartet

- Nassem Nicholas Taleb. *The Black Swan*
  (most ~~entertaining~~ obnoxious)
- Daniel Kahneman. *Thinking Fast and Slow*
  (30 years of research which won Nobel Prize)
- Philip Tetlock. *Expert Political Judgment*
  (most directly relevant)
- Nate Silver. *The Signal and the Noise*
  (high level of credibility after perfect 2012 electoral vote predictions)

Prediction is cool.

Data!


Science
11 February 2011 | $10
scientific
information
climate
science
research
analysis
visualization
researchers
access
many
new
human
challenges
sharing
model
brain
data
example
knowledge
AAAS

PRIO

ACLED Armed Conflict Location & Event Dataset

Minorities at Risk

CORRELATES OF WAR

UPPSALA UNIVERSITET

PENNSTATE

1855

UCDP GEOREFERENCED EVENT DATASET

CIRI Human Rights Data Project

www.humanrightsdata.org

The Dataverse Network™ Project

An Open-Source Application for Publishing, Citing and Discovering Research Data

Polity IV

GTD

Global Terrorism Database

Universität Konstanz

Though this may be going a little far...

WIRED MAGAZINE: 16.07

SCIENCE : DISCOVERIES

The End of Theory: The Data Deluge Makes the Scientific Method Obsolete

By Chris Anderson 06.23.08

Illustration: Marian Bantjes

Computing power

Computing Power

Control Data Corporation 3600
(ca.1965)
32 K (48-bit) RAM memory
1 processor
~1-million operations per second
Output: line printer

Computing Power

Control Data Corporation 3600
(ca.1965)
32 K (48-bit) RAM memory
1 processor
~1-million operations per second
Output: line printer

Penn State High Performance Computing Facility
15 cluster computers
100 to 2000 2.66 Ghz processors in each cluster
~50 Gb RAM accessible to each processor
130 Tb disk space
4 interactive visualization rooms

Computing Power

Control Data Corporation 3600
(ca.1965)
32 K (48-bit) RAM memory
1 processor
~1-million operations per second
Output: line printer

Penn State High Performance Computing Facility
15 cluster computers
100 to 2000 2.66 Ghz processors in each cluster
~50 Gb RAM accessible to each processor
130 Tb disk space
4 interactive visualization rooms

Motorola Razr
16 Gb RAM memory
Dual processor
~500-million operations per sec
540 x 860 color display

Open Source Software

Computationally-intensive methods

- Bayesian estimation using Markov chain Monte Carlo methods
- Bayesian model averaging ("*AJPS*-as-algorithm")
- random forest models
- large-scale textual databases
- machine translation
- geospatial visualization
- real-time automated coding
- remote sensing data such as nightlight density

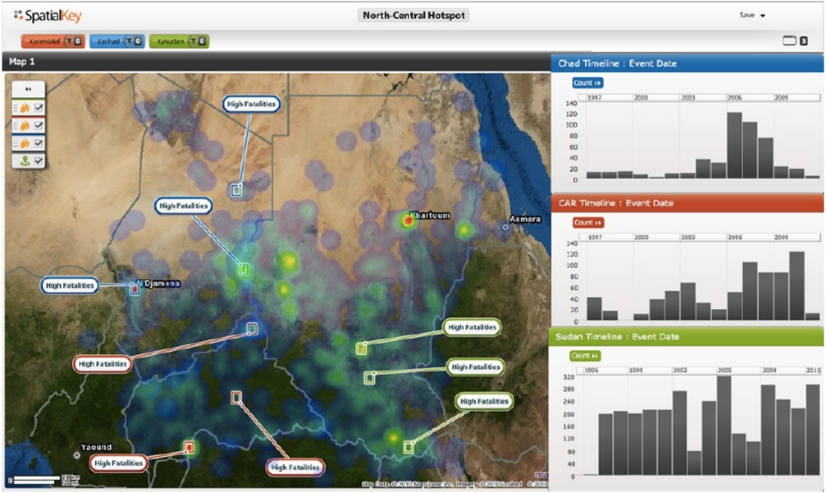
Convergent Results from Forecasting Projects-1

- Most models require only a [very] small number of variables
- Indirect indicators—famously, infant mortality rate as an indicator of development—are very useful
- Temporal autoregressive effects are huge: the challenge is predicting onsets and cessations, not continuations
- Spatial autoregressive effects—“bad neighborhoods”—are also huge
- Multiple modeling approaches generally converge to similar accuracy

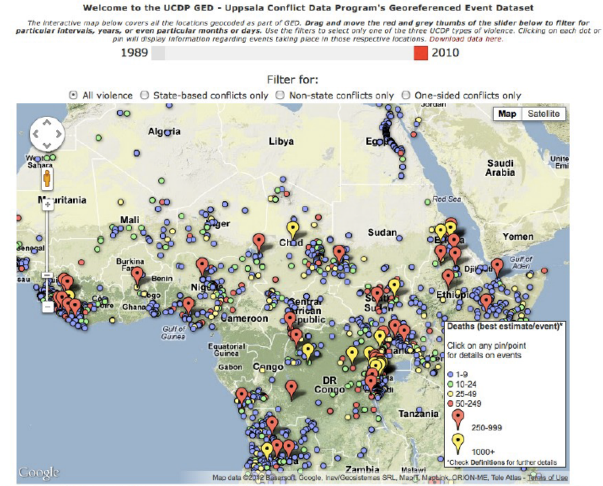
Convergent Results from Forecasting Projects-2

- 80% accuracy—in the sense of AUC around 0.8— in the 6 to 24 month forecasting window occurs with remarkable consistency: few if any replicable models exceed this, and models below that level can usually be improved
- Measurement error on many of the dependent variables—for example casualties, coup attempts—is still very large
- Forecast accuracy does not decline very rapidly with increased forecast windows, suggesting long term structural factors rather than short-term "triggers" are dominant. Trigger models more generally do poorly except as *post hoc* "explanations."

Location: ACLED Geospatial



Location: UCDP Geospatial



The GED is the product of two and a half years of work at the **Department of Peace and Conflict Research, Uppsala University**. The UCDP GED contains conflict data disaggregated spatially and temporally down to the level of the individual incidents of violence. For more details please see the About UCDP link above.

Open source, open access, open collaboration

- There is a strong if incomplete norm towards open sharing of data and methods
  - Unintended consequence: PITF “forecasting tournament” cannot be published in a major journal because it used proprietary data—the baseline data has 2,700 variables—that cannot be archived in replication sets. The results are, however, still available on SSRN.
  - The inability to share source texts is clearly a concern in news-report-based datasets such as ICEWS and MID, though URLs can be shared.
- By all available evidence, US government forecasting projects are using similar methodologies to those available in open sources; in fact they are probably lagging somewhat behind this
- We now have significant NGO and academic work, and an international “epistemic community” has developed around the topic.

Welcome to the new normal

- Rifkin [NYT March 2014]: The most disruptive technologies in the current environment combine network effects with zero marginal cost
- Key: zero *marginal* costs: open source software is "free-as-in-puppy"
- Examples
  - Operating systems: Linux
  - Statistical software: *R*
  - Encyclopedia: Wikipedia
  - Commercial photography: Shutterstock (55K photographers; 30K new images per day) vs Getty Images in $5B/year market

CHALLENGES

Metrics

What is being predicted

- Probability of binary outcomes by a fixed date
- Quintile rankings of risk / probability-based “watch lists”
- Survival and hazard models
- Switching and phase models
- Networking—both social and geospatial—models

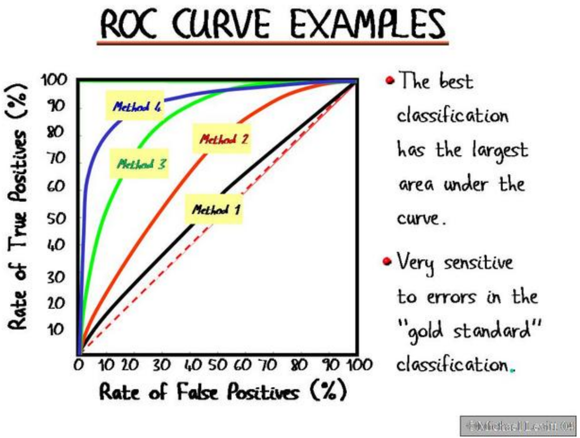
All of these can be used as input to ensemble methods

Classification Matrix

Relationships among terms

		Condition (as determined by "Gold standard")		
		Condition Positive	Condition Negative	
Test Outcome	Test Outcome Positive	**True Positive**	**False Positive** (Type I error)	Positive predictive value = $\frac{\Sigma \text{ True Positive}}{\Sigma \text{ Test Outcome Positive}}$
	Test Outcome Negative	**False Negative** (Type II error)	**True Negative**	Negative predictive value = $\frac{\Sigma \text{ True Negative}}{\Sigma \text{ Test Outcome Negative}}$
		Sensitivity = $\frac{\Sigma \text{ True Positive}}{\Sigma \text{ Condition Positive}}$	**Specificity** = $\frac{\Sigma \text{ True Negative}}{\Sigma \text{ Condition Negative}}$	

ROC Curve



Source:
http://csb.stanford.edu/class/public/lectures/lec4/Lecture6/Data_Visualization/images/Roc_Curve_Exa

Separation plots

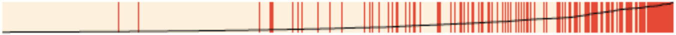
TABLE 4 Rearrangement (and Coloring) of the Data Presented in Table 1 for Use in the Separation Plot

Country	Fitted Value ($\hat{p}$)	Actual Outcome (y)
B	0.364	0
F	0.422	1
D	0.728	0
A	0.774	0
E	0.961	1
C	0.997	1

FIGURE 2 Separation Plot Representing the Data Presented in Table 1

FIGURE 3 Separation Plot for a Larger Data Set

FIGURE 4 Adding a Graph of $\hat{p}$ to the Separation Plot



The Separation Plot: A New Visual Method for Evaluating the Fit of Binary Models

Brian Greenhill[1], Michael D. Ward[2], Audrey Sacks[3]

Article first published online: 20 JUN 2011
DOI: 10.1111/j.1540-5907.2011.00525.x

Issue

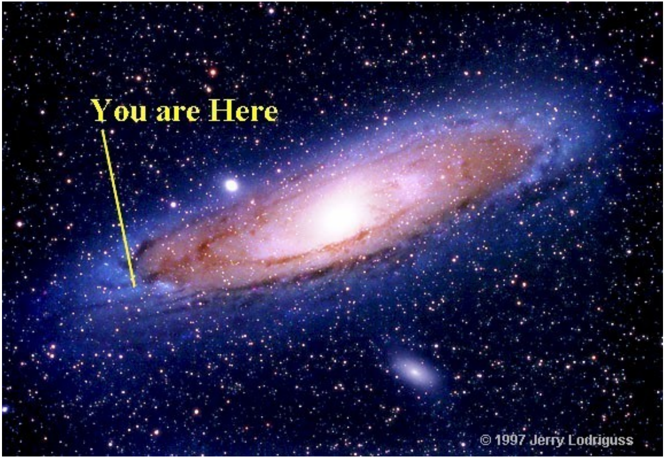
American Journal of Political Science
Volume 55, Issue 4, pages 991–1002, October 2011

And wait, there's still more!

- Recall / True Positive Rate/Sensitivity
- Precision / Positive predictive value (PPV)
- Specificity / True Negative Rate
- F1 score: harmonic mean of precision and recall
- Beier scores
- Posterior probabilities
- Proportional reduction of error or entropy

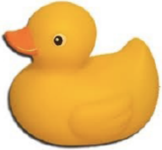
Black swans

Ideal forecasting targets are neither too common nor too rare

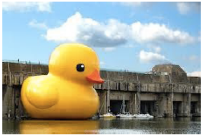


Black swan: Irene Country Lodge, 19 May 2014

The Forecasting Zoo

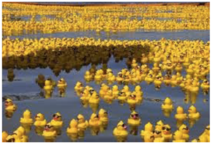


Ducks can be interesting...



Size

Variety

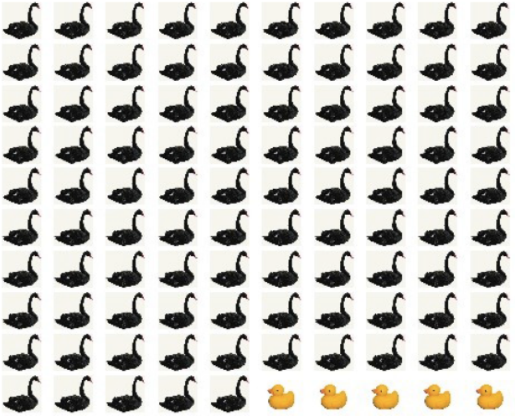


Quantity

Suspicious behaviors

And this is going too far. . .

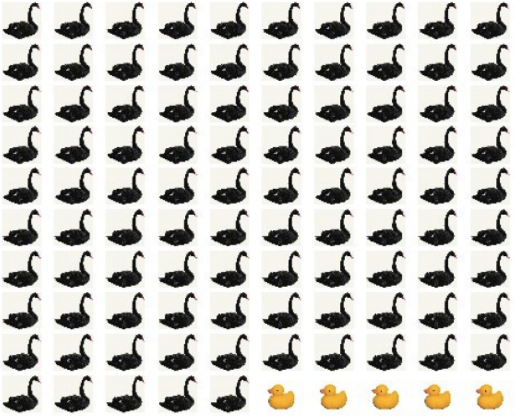
DARPA-World!



By definition, most black swans *will not occur*! So there is little point in investing a large amount of effort trying to predict them.

And this is going too far...

DARPA-World!



By definition, most black swans *will not occur*! So there is little point in investing a large amount of effort trying to predict them.
"Can your model predict a chemical attack by self-recruited Mexican jihadis working as rodeo clowns in Evanston, Wyoming? Why not?!"

Challenge: distinguishing black swans from rare events

Black swan: an event that has a low probability even conditional on other variables

Rare event: an event that occurs infrequently, but conditional on an appropriate set of variables, does not have a low probability

Medical analogy: certain rare forms of cancer appear to be highly correlated with specific rare genetic mutations. Conditioned on those mutations, they are not black swans.

Another important category: high probability events which are ignored. The "sub-prime mortgage crisis" was the result of the failure of a large number of mortgage which models had completely accurately identified as "sub-prime" and thus likely to fail. This was not a low probability event.
Upton Sinclair: It is hard to persuade someone to believe something when he can make a great deal of money not believing it.

Heterogeneous environments

- Per Pinker, Goldstein, Mueller, etc, is the system changing significantly while we are trying to model it? How far back are data still relevant?
- How different are various types of militarized non-state actors? For example, how much do al-Qaeda and international narcotics networks have in common?
- We are also using a more heterogenous set of forecasting methods, and probably do not understand their weak points as well as we understand those of regression-based models.
- Threats tend to occur in small number of "hot-spots"
  - Europe 1910-1945
  - Middle East 1965-present
  - Balkans in 1990s
  - Internal conflict in India

Note that all of these are complicated by rare events—some of which may be black swans—since it limits the number of observations we have on the dependent variable.

Changing nature of conflict-1

Threats in 1910

- "Gunboat diplomacy" was an accepted norm, as were elements of bellicism and social Darwinism
- Some competition occurred between approximate equals
- Mediation was *ad hoc* with no established international institutions
- Territorial change was credible
- Military actors are almost exclusively states

Changing nature of conflict-2

Threats in 2015

- Highly asymmetric distribution of military power
- Threats get almost immediate attention from potential mediators, including the UN
- Non-military sanctions are credible (Iraq, Iran)
- Territorial changes are rare and highly problematic
- Non-state actors can exercise substantial military force

Theory: what can and cannot be predicted?

Is astronomy scientific?

Astronomy generally has a very good record of prediction, and from the earliest days of astronomy, successful prediction has been a key legitimating factor

- relation between star positions and the Nile flood
- eclipses
- orbits
- Halley's comet
- precision steering of space-craft

Is astronomy scientific?

Astronomy generally has a very good record of prediction, and from the earliest days of astronomy, successful prediction has been a key legitimating factor

- relation between star positions and the Nile flood
- eclipses
- orbits
- Halley's comet
- precision steering of space-craft

Nonetheless, astronomy cannot predict, nor does it attempt to predict:

- solar flares, despite their potentially huge economic consequences
- previously unseen comets
- next nearby supernova: the end of the 410-year supernova peace

Determinism: The Pioneer spacecraft anomaly

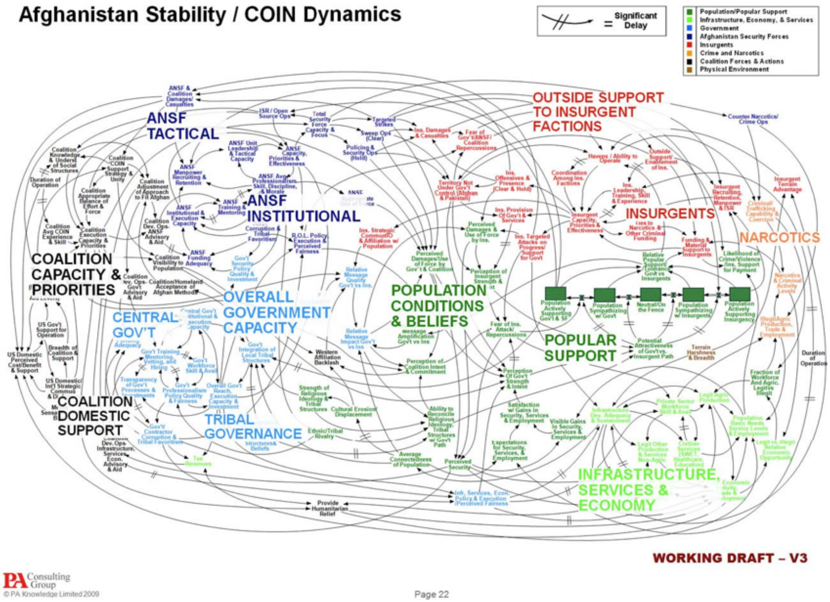
"[Following 30 years of observations] When all known forces acting on the spacecraft are taken into consideration, a very small but unexplained force remains. It appears to cause a constant sunward acceleration of $(8.74 \pm 1.33) \times 10^{-10} m/s^2$ for both spacecraft."

Source: Wikipedia

Irreducible sources of error-1

- Specification error: no model of a complex, open system can contain all of the relevant variables;
- Measurement error: with very few exceptions, variables will contain some measurement error
  - presupposing there is even agreement on what the "correct" measurement is in an ideal setting;
  - Predictive accuracy is limited by the square root of measurement error: in a bivariate model if your reliability is 80%, your accuracy can't be more than 90%
  - This biases the coefficient estimates as well as the predictions
- Quasi-random structural error: Complex and chaotic deterministic systems behave as if they were random under at least some parameter combinations .
  Chaotic behavior can occur in equations as simple as
  $x_{t+1} = a{x_t}^2 + bx_t$

Open, complex systems



Irreducible sources of error-2

- Rational randomness such as that predicted by mixed strategies in zero-sum games
- Arational randomness attributable to free-will
  - Rule-of-thumb from our rat-running colleagues: "A genetically standardized experimental animal, subjected to carefully controlled stimuli in a laboratory setting, will do whatever it wants."
- Effective policy response: in at least some instances organizations will have taken steps to head off a crisis that would have otherwise occurred.
- The effects of natural phenomenon
  - the 2004 Indian Ocean tsunami dramatically reduced violence in the long-running conflict in Aceh

(Tetlock (2013) independently has an almost identical list of the irreducible sources of error.)

Balancing factors which make behavior predictable

- Individual preferences and expectations, which tend to change very slowly
- Organizational and bureaucratic rules and norms
- Constraints of mass mobilization strategies
- Structural constraints:
  the Maldives will not respond to climate-induced sea level rise by building a naval fleet to conquer Singapore.
- Choices and strategies at Nash equilibrium points
- Autoregression (more a result than a cause)
- Network and contagion effects (same)

"History doesn't repeat itself but it rhymes"
Mark Twain (also occasionally attributed to Friedrich Nietzsche)

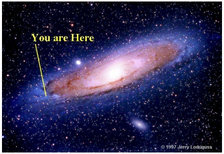
Paradox of political prediction

Political behaviors are generally highly incremental and vary little from day to day, or even century to century (Putnam).

Nonetheless, we *perceive* politics as very unpredictable because we focus on the unexpected (Kahneman).

Consequently the only “interesting” forecasts are those which are least characteristic of the system as a whole. However, only some of those changes are actually predictable.

Finding a non-trivial forecast



- Too frequent: prediction is obvious without technical assistance
- Too rare: prediction may be correct, but the event is so infrequent that
  - The prediction is irrelevant to policy
  - Calibration can be very tricky
  - Accuracy of the model is difficult to assess
- "Just right": these are situations where typical human accuracy is likely to be flawed, and consequently these could have a high payoff, but there are not very many of them.

Differing attitudes towards error

Geography:

- Progress consists of ever more accurate data

Political science:

- Trust nothing—everything has error, just control for the systematic biases

Machine learning::

- it is what it is: goal is improving prediction

Statistics::

- signal to noise: Perfect is the enemy of "good enough"
- mathematically approximate the characteristics of the error
- Taleb, Mandelbrot: don't be a Gaussian in a power-law world

Models matter

Arab Spring is an unprecedented product of the new social media

- Model used by Chinese censors of NSM: King, Peng, Roberts 2012
- Next likely candidates: Africa

Arab Spring is an example of an instability contagion/diffusion process

- Eastern Europe 1989-1991, OECD 1968, CSA 1859-1861, Europe 1848, Latin America 1820-1828
- Next likely candidates: Central Asia

Arab Spring is a black swan

- There is no point in modeling black swans, you instead build systems robust against them

Statistical and modeling challenges

Rare events

- Incorporate much longer historical time lines?—Schelling used Caesar's *Gallic Wars* to analyze nuclear deterrence
- New approaches made possible by computational advances

Analysis of event sequences, which are not a standard data type

- There are, however, a large number of available methods, and it is just possible that these will work with very large data sets
- This possibility will be discussed in detail in Lecture 5

Causality

- Oxford *Handbook of Causation* is 800 pages long

Integration of qualitative and qualitative/subject-matter-expert (SME) information

- Bayesian approaches using prior probabilities are promising but to date they have not really been used

Pournelle's Law: No task is so virtuous that it will not attract idiots

- Need to establish with the media and policy-makers that not every forecast, ~~even~~ especially those made using "Big Data" methods, is scientifically valid
  - It took the survey research community about thirty to forty years to establish professional credibility, though they have largely succeeded
- Conveying limitations of the methods against the hyper-confidence of pundits and individuals with secret models
  - Limitations of the data sources
  - Limitations of data coding, particularly when automated
  - Limitations of the model estimation
  - Limitations of probabilistic forecasts, particularly for rare events, even when the models are correct

Critical case: studies of climate change and conflict: people want to hear simple scary answers.

Ethical concerns

- Thus far, we've generally had the luxury of no one paying attention to any of our predictions : what if governments do start paying attention?
  - "Policy relevant forecast interval" is around 6 to 24 months
  - USAID/FAO famine forecasting model
  - It is *possible* that our models could become less accurate because crises are being averted, but I don't see that happening any time soon.
- Difficulties in getting *anyone*, including experts (see Kahneman, Tetlock), to correctly interpret probabilistic forecasts
- Possible impact on sources
  - Local collaborators
  - Journalists (cf. Mexico)
  - NGOs to the extent we are using their information

Thank you

Email: schrodt735@gmail.com

Slides: http://eventdata.parusanalytics.com/presentations.html
[for about seven hours of lectures on this topic, see the videos from the University of Konstanz, October-13]

Forecasting papers:
http://eventdata.parusanalytics.com/papers.html

PARUS
ANALYTICS