Technical Forecasting of Political Conflict

Philip A. Schrodt, Ph.D.

Parus Analytics LLC
Charlottesville, Virginia, USA
http://philipschrodt.org
http://eventdata.parusanalytics.com

3rd Workshop on the EU Global Conflict Risk Index
Emergency Response Coordination Centre, Brussels
31 May 2017

PARUS
ANALYTICS

Well, this is timely...

ESSAYS

Predicting armed conflict: Time to adjust our expectations?

Lars-Erik Cederman[1,*], **Nils B. Weidmann**[2,*]

+ Author Affiliations

[*]Corresponding author. Email: lcederman@ethz.ch (L.-E.C.); nils.weidmann@uni-konstanz.de (N.B.W.)

Science 03 Feb 2017:
Vol. 355, Issue 6324, pp. 474-476
DOI: 10.1126/science.aal4483

ESSAYS

Bringing probability judgments into policy debates via forecasting tournaments

Philip E. Tetlock[1,*], **Barbara A. Mellers**[1], **J. Peter Scoblic**[2]

+ Author Affiliations

[*]Corresponding author. Email: tetlock@wharton.upenn.edu

Science 03 Feb 2017:
Vol. 355, Issue 6324, pp. 481-483
DOI: 10.1126/science.aal3147

Prediction and explanation in social systems

Jake M. Hofman[*], **Amit Sharma**[*], **Duncan J. Watts**[*]

+ Author Affiliations

[*]Corresponding author. Email: jmh@microsoft.com (J.M.H.); amshar@microsoft.com (A.S.); duncan@microsoft.com (D.J.W.)

Science 03 Feb 2017:
Vol. 355, Issue 6324, pp. 486-488
DOI: 10.1126/science.aal3856

And in the *Washington Post*

Monkey Cage | **Analysis**

Where are coups most likely to occur in 2017?

By Michael D. Ward and Andreas Beger January 31

Supporters of Tukish President Tayyip Erdogan celebrate after soldiers involved in the failed coup attempt surrendered on the Bosporus Bridge in Istanbul on July 16, 2016. (Yagiz Karahan/Reuters)

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. The typical “policy relevant forecasting interval” is 6 to 24 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 (2012) 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”

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 (PITF) 2002-present
- DARPA Integrated Conflict Early Warning System (ICEWS) 2007-present
- IARPA ACE and OSI
- Peace Research Center Oslo (PRIO) and Uppsala University UCDP models
- US Holocaust Memorial Museum Prediction Poll

Is political behavior predictable? Yes! Good Judgment Project (Tetlock, Meller et al)

- Evaluated about 2000 forecasts, typically with a 6 to 12 month window, across a wide variety political and economic domains
- Most forecasters—more than 90%—were simply "dart-throwing chimps"
- "Superforecasters", however, consistently were about 80% to 85% accurate. This held across multiple years: unlike managed mutual funds, it did not regress to the mean
- Teams of superforecasters were more effective than individuals, and behaved differently than random teams
- Superforecasters have distinct psychological profiles: "foxes rather than hedgehogs"
- Prediction markets, SMEs and ensemble models provided only marginal improvements

Political behaviors are predictable! Superforecaster accuracy is similar to that of the PITF and ICEWS models.

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

Conjecture

For the possibly first time in history, we may be entering an era when foreign policy can be based on relatively accurate projections of the future rather than random guesses and ideologly

"Possibly" since the superforecaster approach may have been independently discovered earlier, for example in Confucian and Venetian bureaucracies

Three other cases where "professional" advice was random or worse

- Medicine prior to sometime in the 20th century
- Managed mutual funds
- GRE scores

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 state capacity—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% to 85% 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."

Why have predictive models improved?

Data!


The Economist
Theresa May v Brussels
Ten years on: banking after the crisis
South Korea's unfinished revolution
Biology, but without the cells
The world's most valuable resource
Google
Data and the new rules of competition

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

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

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

Open Source Software

MySQL

mozilla Firefox

python

ubuntu

source forge

perl

Ruby PROGRAMMING Language

WordPress.com

APACHE HTTP SERVER

LibreOffice The Document Foundation

GNU run free run GNU

stackoverflow

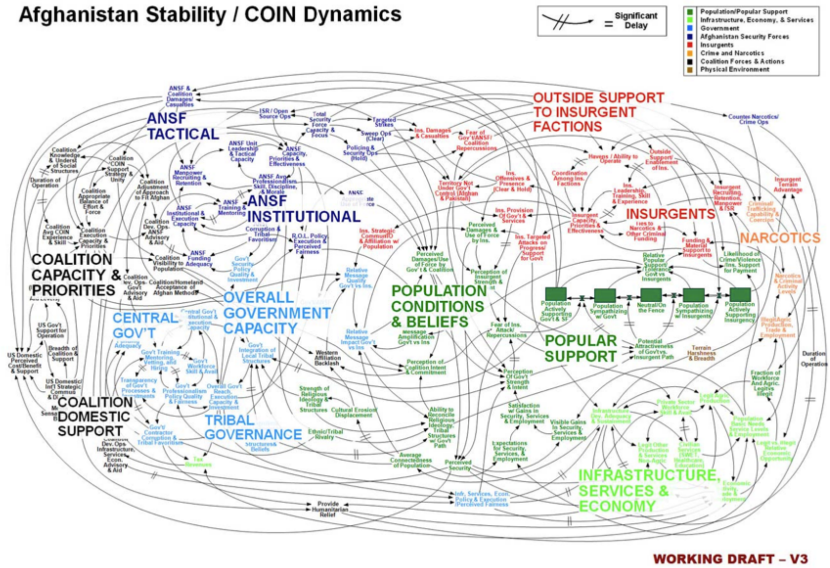
FSF FREE SOFTWARE FOUNDATION

Some challenges

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.

Challenge: distinguishing black swans from rare events

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

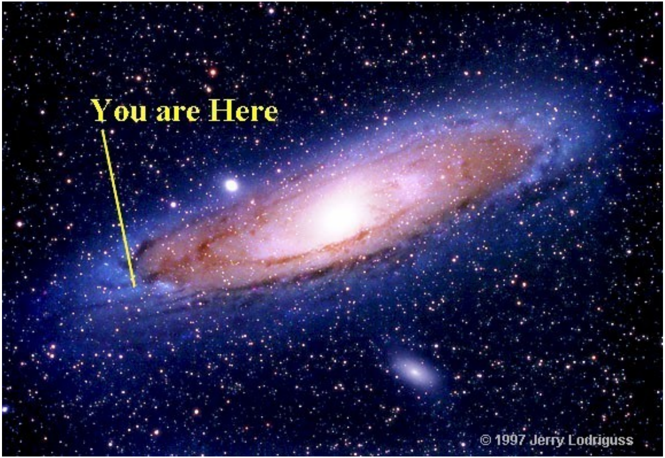
Rare event: an event that occurs infrequently, but conditional on an appropriate set of variables, does not have a low probability

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.

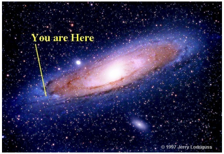
Another important category: high probability events which are ignored. The “sub-prime mortgage crisis” was the result of the failure of a large number of mortgage which models had completely accurately identified as “sub-prime” and thus likely to fail. This was not a low probability event.
Upton Sinclair: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

Black swans

Ideal forecasting targets are neither too common nor too rare



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.

Models matter

Arab Spring is an unprecedented product of the new social media

- Model used by Chinese censors of new social media: 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

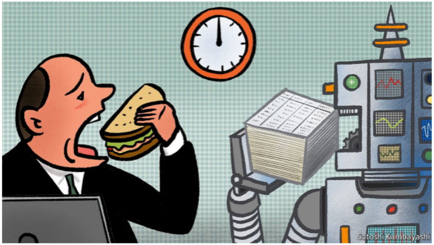
Some opportunities

Machine learning: hype or the new frontier?

Unshackled algorithms

Machine-learning promises to shake up large swathes of finance

In fields from trading to credit assessment to fraud prevention, machine-learning is advancing



Print edition | Finance and economics ›
May 25th 2017

MACHINE-LEARNING is beginning to shake up finance. A subset of artificial intelligence (AI) that excels at finding patterns and making predictions, it used to be the preserve of technology firms. The financial industry has jumped on the bandwagon. To cite just a few examples, "heads of machine-learning" can be found at PwC, a consultancy and auditing firm, at JP Morgan Chase, a large bank, and at Man GLG, a hedge-

New opportunities from machine learning

- ML methods recently have been successful in a number of "artificial intelligence" problems previously thought to be unsolvable
- Most statistical models have already been extensively explored, and in any case are not optimized for prediction (Ward, Greenhill and Bakke 2010)
- The parameter spaces of many of these models are vastly larger than those of statistical models
- ML models generally work well with heterogeneous cases
- Most ML models are relatively insensitive to missing values, or treat it as information
- Software is readily available and open source

Risks in machine learning models

- Over-fitting
- It is not clear that political early warning has a sufficient number of cases to take advantage of methods which require large amounts of data
- ML models are generally atheoretical, and the rich parameter spaces mean it is often difficult to impossible to ascertain the relative importance of independent variables
- Some models—notably "deep learning"—are quite new and may have features we don't fully understand
- In many instances, ML models show only marginal improvements over well-understood methods such as logistic regression when applied across a wide set of out-of-sample problems

The very finite set of widely used ML methods

- Support vector machines
- Clustering, typically using k-means
- Random forests, a relatively recent ensemble variation on the older method of decision trees
- Neural networks
  - A very old method which is now being used with vastly greater hardware and a few new algorithmic tricks to create “deep learning”
- Genetic algorithms
- Logistic regression, which not infrequently is ”embarrassingly effective”

Some interesting open questions

- Under what circumstances does climate change increase versus reduce conflict?
  - Contrary to the ubiquitous "Battle at the water hole" analogies, there is ample evidence to support both effects
- How can event data and structural data be combined to increase predictive accuracy?: to date, they largely just seem to be substitutable
- Are "trigger models" real or simply a cognitive illusion?
- How many theoretically distinct forms of sub-state conflict should be analyzed?
- What is the optimal level of detail in event data and geospatial data (which will depend on the question, of course)

Memo to potential funding agencies: We aren't exactly over-spending on this topic

- A \$1-million investment in research *might* avoid a \$10-million mistake in policy. Or a \$10-million investment in research might avoid a \$4-trillion mistake in policy.
- Every half hour of every business day, the amount Google spends on the study of human behavior is roughly the same as the entire political science research budget of the United States National Science Foundation (\$8-million).

Thank you

Email:
schrodt735@gmail.com

Slides:
`http://eventdata.parusanalytics.com/presentations.html`

Links to data and software: `http://philipschrodt.org`

Blog: `http://asecondmouse.org`