



Predicting Conflict

Jamie Kalil

Historical & Operational Data Analysis (HODA)

PCS Department

Dstl

The Presentation

- Background and objectives
- The dependent variable – i.e. internal conflict
- The independent variables
- The method
- The output
- The way forward

Predicting Conflict

- The objective is to improve HMG internal conflict prediction capabilities
 - Internal and not international (i.e. civil war)
 - Prediction of conflict onset and not forecasting of ongoing conflict
- Three years of research funded by DG(S&A)
- Increasing emphasis across Government upon Conflict Prevention as opposed to Conflict Resolution

The Objective

- To create a quantitative model relying upon historical data through which it is possible to predict the future onset of internal conflict with a high degree of accuracy.
- A provisional model was created in FY 2007-2008 and the 'performance' was improved during spring 2008.
- It is believed that the output from this model will be of value to various OGDs, such as the Cabinet Office, DfID, the FCO, etc.

The Field – What Currently Exists?

- 1. Models that predict broader phenomena
 - PITF – predicts *political instabilities* (including *adverse regime change, genocide, and internal conflict*)
 - ACTOR – predicts internal and international *crises and conflicts*
- 2. Studies that identify correlations between the onset of internal conflict and a range of input variables but that do not predict
 - Collier & Hoeffler
 - Fearon & Laitin

Thus.....

To our knowledge this is the only existing
quantitative model that exclusively seeks
to predict internal conflict

Issue 1 – The Dependent Variable

- Issues with measuring conflict intensity:
 - Should fatality figures be used?
 - Should fatalities of non-combatants be included?
 - Should indirect fatalities be included?
 - Should fatalities resulting from genocide be included?
 - Should fatalities figures be cumulative / year-by-year?
- Issues with locating conflict onset on the intensity scale:
 - Thresholds (e.g. 1,000 fatalities) are arbitrary and impose a binary distinction (onset / non-onset) upon a continuous variable
 - No distinction between, say, 1,000 fatalities in China and Cyprus

Existing Measures of Conflict Onset

- Fearon & Laitin (*Civil War / Violent Civil Conflicts*)
 - At least 1,000 cumulative deaths ('including civilians')
 - At least 100 deaths per year
 - At least 100 deaths caused by each of at least two sides
- ACTOR divide conflict into subjective categories:
 - Latent conflict – “totally non-violent”
 - Crisis – “predominantly non-violent”
 - Severe crisis – “sporadic, irregular use of force”
 - War – “systematic and collective use of violence”

Issue 2 – The Independent Variables

- **Structural Variables** – most existing studies exclusively focus upon variables such as GDP per capita, population size, ethnic fractionalisation, ‘youth bulge’, etc
- **Actor-Centric Variables** – relate to characteristics of the key actors involved (insurgents / counterinsurgents), including ideology, strategy / tactics, capabilities, attacks
- **Trigger Variables** – include assassinations, military coups, humanitarian disasters
- The current version of the model is reliant upon 3 structural variables and 1 trigger variable

Issue 3 – Methodology

- The basic unit of analysis was the country-year (e.g. Iran 1978)
- The original sample included all country-years between 1970 and 1995
- The original sample excluded country-years experiencing ongoing conflict
- Predictions relate to conflict onset within the subsequent 5 years for each of the country-years in question
- The dependent variable was Fearon & Laitin *Civil War* (binary)
- Variations of 16 independent variables were tested (data was collected for each of the country-years)

Tested Independent Variables

The 'optimal' combination of variables

GDP per capita	Population Size
Mountainous territory	Non-Contiguous State
Geological Disasters	Oil Exporter
Regime Type	New State
Democracy	Democratisation
Religious Fractionalisation	Ethnic Fractionalisation
Instability	Bordering Conflict
Prior Conflict Termination	Peace Duration

Logistic Regression

- Three samples (S1, S2, S3) were created that were composed of twenty of the twenty-six years (1970-1995) selected at random
- The out-of-sample (OOS) sets were composed of all remaining country-years for each of the three samples
- Stepwise addition / deletion was undertaken on these three samples using the 16 potential independent variables (model metrics shall be discussed shortly)

Optimal Combination of Variables

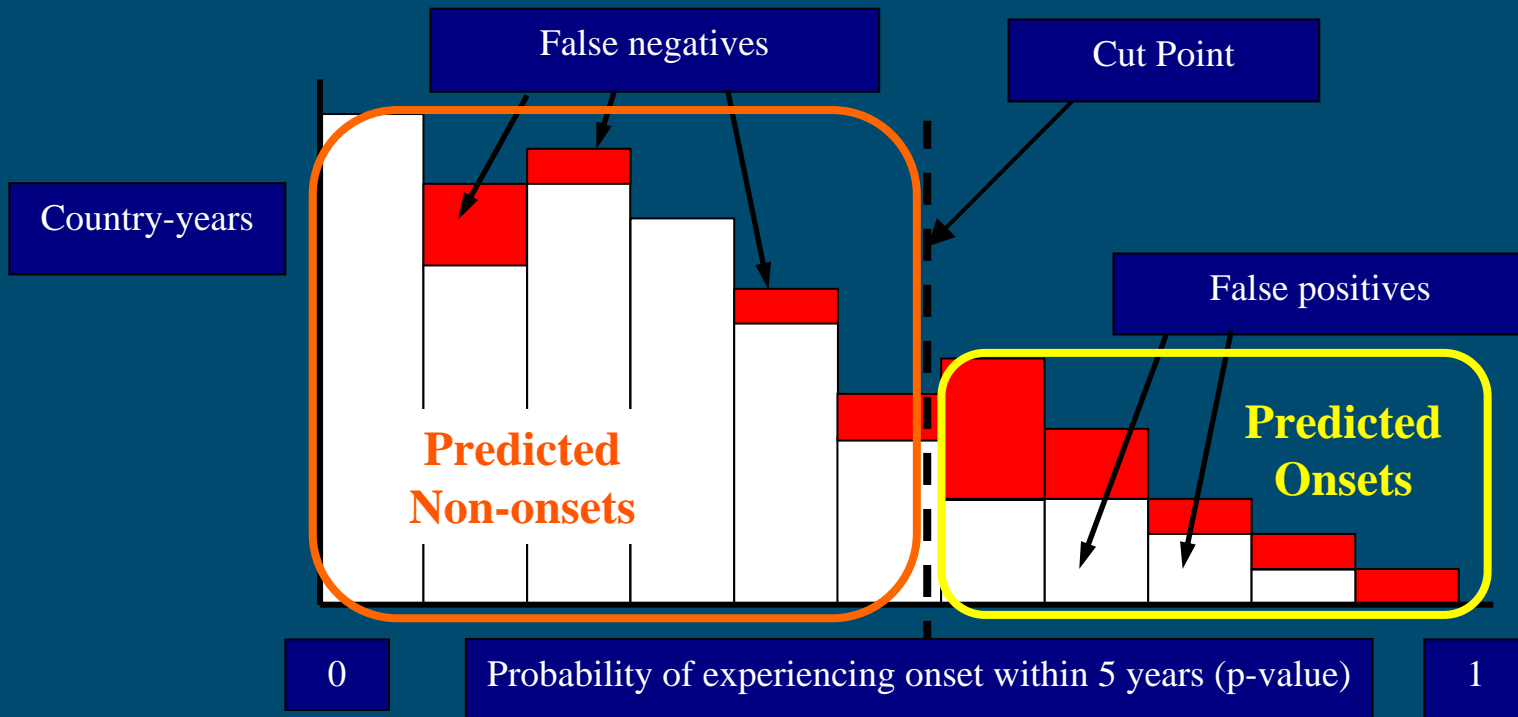
- The 'optimal' combination of independent variables (resulting from the process of stepwise addition / deletion) included four variables for each of the three samples

Variable	Coefficient (S2)	SE (S2)
GDP per capita (log)	-0.35	0.05
Mountainous (log)	0.26	0.06
Geological disasters (log)	3.0	0.66
Regime type	-1.4	0.33
Constant	-1.1	0.32

The P-Value

- The coefficients were used to generate a p-value for each of the OOS country-years
 - The p-value is ‘the probability that a country-year will experience conflict onset within 5 years’

Assessing Performance



= Actual non-onsets (within 5 years)

= Actual onsets (within 5 years)

Model Metrics

	Actual Onsets	Actual Non-onsets
Predicted Onsets	Correct Predictions	False Positives
Predicted Non-onsets	False Negatives	Correct Predictions

- **Accuracy** = total correct predictions / total predictions
- **Recall** = correctly predicted onsets / actual onsets
- **Precision** = correctly predicted onsets / predicted onsets

The Cut Point

- There is no 'correct' location for the cut point
- Through fixing the cut point it is possible to provide standardised metrics in order to enable comparisons:
 - at the p-value at which the number of predicted onsets is equal to the number of actual onsets (M1) – i.e. at the p-value at which recall equals precision

Results for the S2 Model

Cut Point	Accuracy	Recall	Precision
0.3			42.9%
0.16 (M1)			
0.1	80.0%		

At lower cut points we “over-predict” (i.e. there are lots of false positives) but points we miss onsets (i.e. there are lots of false negatives)

The higher the cut point the lower the number of predicted onsets
 The lower the cut point the higher the number of predicted onsets

Output – An ‘At Risk’ Shortlist

- This provides the following ‘typical’ output (for data from 1987):

	Actual Onsets	Actual Non-onsets
Predicted Onsets	<u>Correct Predictions</u> (3) Burundi, China, Papua NG	<u>False Positives</u> (3) Ecuador, Nepal, Pakistan
Predicted Non-onsets	<u>False Negatives</u> (8) Algeria, Haiti, Liberia, Mali, Rwanda, Senegal, Sierra Leone, Yugoslavia	<u>Correct Predictions</u> 95 Countries

The Way Forward 1

- Future work will involve:
 - Collecting and analysing of post-1995 data
 - Assessing 'performance' using alternative dependent variables
 - Assessing 'performance' using alternative onset timeframes
 - Assessing the feasibility of developing a regression model that predicts scales of internal conflict (e.g. 'low', 'medium', 'high')

The Way Forward 2

- Most importantly future work will also involve an analysis of a range of untested variables:

Political assassination	Military coups
Political rights	Civil liberties
Infant mortality	State-led discrimination
State repression	'Youth bulge'
Natural resources	Change in GDP per capita
Terrorist incidents	Active terrorists
Historical internal conflict	

QUESTIONS?

Jamie Kalil

Analyst, Historical Analysis Team

Historical and Operational Data Analysis Team

Dstl Policy & Capability Studies Department

Rm 1014, Bldg A3, Dstl,

Ively Rd, Farnborough, HANTS GU14 0LX

Tel: +44(0) 1252 455595

Email: jhkalil@dstl.gov.uk

[dstl] is part of the UK Ministry of Defence



08 September 2008
© Dstl 2008

UK UNCLASSIFIED



Dstl is part of the
Ministry of Defence