Week N: Forecast Modeling

POLI502

Howard Liu

2024

University of South Carolina

Outline

- Why do we make prediction/forecast?
- Examples of actual research from civil war studies
- Evaluating forecasting results

Two mindsets: causation vs. prediction

- Causal questions:
    - RQ: Why do governments kill civilians in civil war?

Two mindsets: causation vs. prediction

- Causal questions:
  - RQ: Why do governments kill civilians in civil war?
  - Theory: a causal argument ($x \rightarrow y$), information $\rightarrow$ violence

Two mindsets: causation vs. prediction

- Causal questions:
  - RQ: Why do governments kill civilians in civil war?
  - Theory: a causal argument ($x \rightarrow y$), information $\rightarrow$ violence
  - Covariates: Only one variable matters

Two mindsets: causation vs. prediction

- Causal questions:
  - RQ: Why do governments kill civilians in civil war?
  - Theory: a causal argument ($x \rightarrow y$), information $\rightarrow$ violence
  - Covariates: Only one variable matters
  - Methodology: causal identification strategy $\rightarrow$ causal inference in observational data or experimental studies

Two mindsets: causation vs. prediction

- Prediction questions (kinda like AI questions):
  - RQ: Can we **predict** civilian killing/casualties by governments in civil war? (So that we can prevent it)

Two mindsets: causation vs. prediction

- Prediction questions (kinda like AI questions):
  - RQ: Can we **predict** civilian killing/casualties by governments in civil war? (So that we can prevent it)
    - DoD cares a ton about this.

Two mindsets: causation vs. prediction

- Prediction questions (kinda like AI questions):
  - RQ: Can we **predict** civilian killing/casualties by governments in civil war? (So that we can prevent it)
    - DoD cares a ton about this.
  - RQ: **How well can we predict** civilian killing/casualties by governments in civil war?

Two mindsets: causation vs. prediction

- Prediction questions (kinda like AI questions):
  - RQ: Can we **predict** civilian killing/casualties by governments in civil war? (So that we can prevent it)
    - DoD cares a ton about this.
  - RQ: **How well can we predict** civilian killing/casualties by governments in civil war?
    - DoD cares a ton about this.

Two mindsets: causation vs. prediction

- Prediction questions (kinda like AI questions):
  - RQ: Can we **predict** civilian killing/casualties by governments in civil war? (So that we can prevent it)
    - DoD cares a ton about this.
  - RQ: **How well can we predict** civilian killing/casualties by governments in civil war?
    - DoD cares a ton about this.
  - Accuracy: we want precision in prediction
  - Covariates: all of them matter
  - Method: regression, machine learning algorithms
- Are prediction & causation not compatible?

Two mindsets: causation vs. prediction

- Prediction questions (kinda like AI questions):
  - RQ: Can we **predict** civilian killing/casualties by governments in civil war? (So that we can prevent it)
    - DoD cares a ton about this.
  - RQ: **How well can we predict** civilian killing/casualties by governments in civil war?
    - DoD cares a ton about this.
  - Accuracy: we want precision in prediction
  - Covariates: all of them matter
  - Method: regression, machine learning algorithms

- Are prediction & causation not compatible?
  - They are. While they have different goals, both can be useful.

Prediction & forecasting

Improve previous models' predictive ability

- Procedure: models give us predicted values of $\hat{Y}$ **in-sample**, and we would like our predictions to be accurate in **out-of-sample** setting
- For a usual causal story research, prediction helps evaluate a new *variable (x)* introduced by theory
- For a methodology research, people propose an alternative *model* that has a better forecasting ability than existing models
- One of the most important tasks in social (data) science

Evaluating forecast

Ways to evaluate forecasts of binary outcomes

1. Percentage correctly predicted
2. ROC curve & AUC (Area Under the ROC Curve)
3. Separation Plot
4. Precision-recall curve

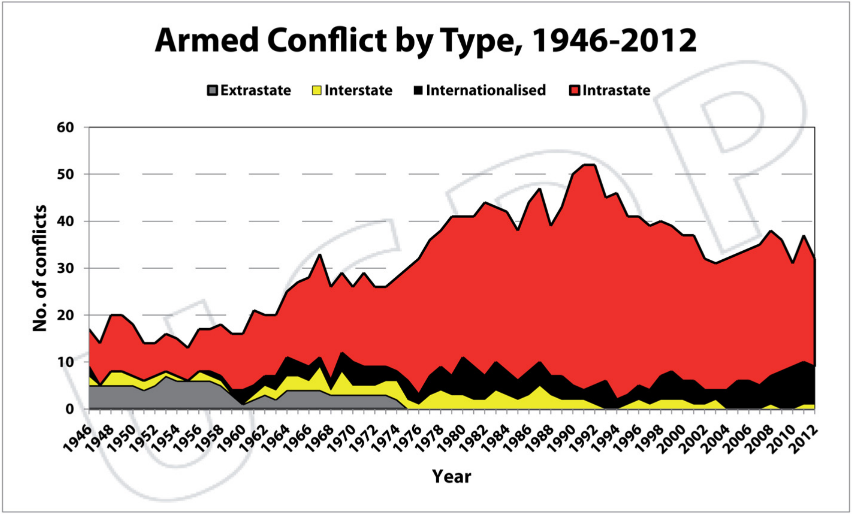
Example: civil war research

- *Inter*-state wars are on the decline since WWII

  No wars between great powers since Korean War (1950–53)

- *Intra*-state wars (civil wars) have been on the rise

- In the past 20 years, the number and quality of *intra*-war research have gone up


Armed Conflict by Type, 1946-2012
Extrastate
Interstate
Internationalised
Intrastate
No. of conflicts
0
10
20
30
40
50
60
1946 1948 1950 1952 1954 1956 1958 1960 1962 1964 1966 1968 1970 1972 1974 1976 1978 1980 1982 1984 1986 1988 1990 1992 1994 1996 1998 2000 2002 2004 2006 2008 2010 2012
Year

(c) UCDP 2013

Fearon & Laitin (2003)

James Fearon & David Laitin (2003) "Ethnicity, Insurgency, and Civil War." *APSR*

Fearon & Laitin (2003)

James Fearon & David Laitin (2003) “Ethnicity, Insurgency, and Civil War.” *APSR*

RQ: Why do civil wars occur?

Fearon & Laitin (2003)

James Fearon & David Laitin (2003) "Ethnicity, Insurgency, and Civil War." *APSR*

RQ: Why do civil wars occur?

Previous answers

- The prevalence of civil wars in the 1990s is due to the end of the Cold War

Fearon & Laitin (2003)

James Fearon & David Laitin (2003) "Ethnicity, Insurgency, and Civil War." *APSR*

RQ: Why do civil wars occur?

Previous answers

- The prevalence of civil wars in the 1990s is due to the end of the Cold War
- Ethnic, religious, or cultural diversity is the root cause of civil wars

  = Ethnic or political grievances (state oppression and discrimination) cause civil wars

Fearon & Laitin (2003)

Fearon & Laitin's (2003) argument

Fearon & Laitin (2003)

Fearon & Laitin's (2003) argument

- Civil wars are best understood as insurgency
- Insurgency = a technology of military conflict characterized by small, lightly armed bands practicing guerrilla warfare from rural base areas

Fearon & Laitin (2003)

Previous argument:

- Ethnic or political grievances (state oppression and discrimination) causes civil wars

Fearon & Laitin (2003)

Previous argument:

- Ethnic or political grievances (state oppression and discrimination) causes civil wars

If this *were* true,

- variables such as *ethnic fractionalization*, *religious fractionalization*, and *democracy* should be a strong predictor of civil wars

Fearon & Laitin (2003)

What are the observable conditions that favor insurgency?

Fearon & Laitin (2003)

What are the observable conditions that favor insurgency?

- Weak government:
  - Rebels can easily launch insurgency

Fearon & Laitin (2003)

What are the observable conditions that favor insurgency?

- Weak government:
    - Rebels can easily launch insurgency
- Rough (mountainous) terrain

Fearon & Laitin (2003)

What are the observable conditions that favor insurgency?

- Weak government:
  - Rebels can easily launch insurgency
- Rough (mountainous) terrain
- Large population

	Civil War	Ethnic War	Civil War
Prior war	-0.954^{***}	-0.935^{**}	-0.916^{***}
	(0.314)	(0.367)	(0.312)
Per capita income	-0.344^{***}	-0.344^{***}	-0.318^{***}
	(0.072)	(0.088)	(0.071)
log(population)	0.263^{***}	0.378^{***}	0.272^{***}
	(0.073)	(0.085)	(0.074)
log(% mountanious)	0.219^{***}	0.163	0.199^{**}
	(0.085)	(0.106)	(0.085)
Noncontiguous state	0.443	0.420	0.426
	(0.274)	(0.327)	(0.272)
Oil exporter	0.858^{***}	1.046^{***}	0.751^{***}
	(0.279)	(0.325)	(0.278)
New state	1.709^{***}	1.793^{***}	1.658^{***}
	(0.339)	(0.393)	(0.342)
Instability	0.618^{***}	0.462	0.513^{**}
	(0.235)	(0.296)	(0.242)
Democracy	0.021	0.022	
	(0.017)	(0.021)	
Ethnic fractionalization	0.166	0.705	0.164
	(0.373)	(0.466)	(0.368)
Religious fractionalization	0.285	1.452^{**}	0.326
	(0.509)	(0.648)	(0.506)
Anocracy			0.521^{**}
			(0.237)
Democracy			0.127
			(0.304)
Constant	-6.731^{***}	-8.864^{***}	-7.019^{***}
	(0.736)	(0.924)	(0.751)
Observations	6,327	6,327	6,327
Log Likelihood	−480.402	−338.791	−478.671
Akaike Inf. Crit.	984.803	701.582	983.342
Note:			$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

Fearon & Laitin (2003)

F & L's "favorite" variables are significant and in the expected direction

- Rough terrain (positive)
- Population (positive)
- Per capita GDP (negative)

Variables suggested by the conventional wisdom are insignificant

- Ethnic fractionalization
- Religious fractionalization
- Democracy

(**Note:** They could have done more to illustrate substantive effects by plotting the marginal effects)

Fearon & Laitin (2003)

Broader policy implications:

- "Longstanding hatred" between different ethnic groups is not really the root cause of civil wars

Fearon & Laitin (2003)

Broader policy implications:

- "Longstanding hatred" between different ethnic groups is not really the root cause of civil wars
- Democratization is not a solution

Fearon & Laitin (2003)

Criticisms

- Per capita GDP as a proxy measure of weak government?

Fearon & Laitin (2003)

Criticisms

- Per capita GDP as a proxy measure of weak government?
- Ethnic fractionalization as a measure of grievance?

Fearon & Laitin (2003)

Criticisms

- Per capita GDP as a proxy measure of weak government?
- Ethnic fractionalization as a measure of grievance?
- **Forecasting ability (Ward, Greenhill, & Bakke 2010)**: Out-of-sample forecast

Steps to generate out-of-sample prediction

- In-sample prediction: use all your data
- Out-of-sample prediction: use some data to build your model, and evaluate prediction using the remaining

Steps to generate out-of-sample prediction

- Traditional way (80-20): 80% data to train the model (in-sample stage), 20% data to test the model (out-of-sample)
- Modern way: Randomly subset data into k-fold, and use 4/5 of the data to train and 1/5 data to test, and repeat this process N times (Cross Validation)

1. Percentage correctly predicted

How well do models correctly predict civil war onset?

- Predicted values $\hat{P}$: 0.001, 0.201, 0.84, 0.335, 0.659, ...
- Actual outcomes Y: 0, 1, 0, 0, 1, 1, ...

1. Percentage correctly predicted

How well do models correctly predict civil war onset?

- Predicted values $\hat{P}$: 0.001, 0.201, 0.84, 0.335, 0.659, ...
- Actual outcomes Y: 0, 1, 0, 0, 1, 1, ...
- $\hat{P} >$ threshold (e.g., 0.5) $\rightsquigarrow \hat{Y} = 1$
- $\hat{P} \leq$ threshold $\rightsquigarrow \hat{Y} = 0$

Create a cross-tabulation of actual outcomes against predicted outcomes

1. Percentage correctly predicted

	$Y = 0$	$Y = 1$	total
$\hat{Y} = 0$	30	21	51
$\hat{Y} = 1$	5	19	24
total	35	40	75

- True positive (19), true negative (30), false positive (5), and false negative (21)
- Correctly predicted = 30 + 19
- Incorrectly predicted = 21 + 5
- Percentage correctly predicted = (30+19) / (30+19+21+5) = 0.653 = 65%

1. Percentage correctly predicted

Pro:

- Intuitive

Cons:

- Problematic with rare events (such as conflict): Easy to predict 0s, but not 1s
- Sensitive to threshold

1. Percentage correctly predicted

In R, we can easily obtain this using the `hitmiss` function (in the `pscl` package)

```
> library(pscl)
> hitmiss(mod1)
Classification Threshold = 0.5
        y=0 y=1
yhat=0 6221 106
yhat=1    0   0
Percent Correctly Predicted = 98.32%
Percent Correctly Predicted = 100%, for y = 0
Percent Correctly Predicted = 0%  for y = 1
Null Model Correctly Predicts 98.32%
[1]  98.32464 100.00000   0.00000
```

- Null model = a model that predicts "All zero"
- Fearon & Laitin's (2003) model performs no better than the null model (with 0.5 as threshold)

1. Percentage correctly predicted

```
> hitmiss(mod1, k = 0.4)
Classification Threshold = 0.4
        y=0 y=1
yhat=0 6220 105
yhat=1    1   1
Percent Correctly Predicted = 98.32%
Percent Correctly Predicted = 99.98%, for y = 0
Percent Correctly Predicted = 0.9434%  for y = 1
Null Model Correctly Predicts 98.32%
[1] 98.3246404 99.9839254  0.9433962
```

If we lower the threshold, we could get

- more true positives;
- more false positives;
- lower percentage correctly predicted.

1. Percentage correctly predicted

```
> hitmiss(mod1, k = 0.3)
Classification Threshold = 0.3
        y=0 y=1
yhat=0 6220 105
yhat=1    1   1
Percent Correctly Predicted = 98.32%
Percent Correctly Predicted = 99.98%, for y = 0
Percent Correctly Predicted = 0.9434%  for y = 1
Null Model Correctly Predicts 98.32%
[1] 98.3246404 99.9839254  0.9433962
```

If we lower the threshold, we could get

- more true positives;
- more false positives;
- lower percentage correctly predicted.

1. Percentage correctly predicted

```
> hitmiss(mod1, k = 0.2)
Classification Threshold = 0.2
        y=0 y=1
yhat=0 6216 102
yhat=1    5   4
Percent Correctly Predicted = 98.31%
Percent Correctly Predicted = 99.92%, for y = 0
Percent Correctly Predicted = 3.774%  for y = 1
Null Model Correctly Predicts 98.32%
[1] 98.308835 99.919627  3.773585
>
```

If we lower the threshold, we could get

- more true positives;
- more false positives;
- lower percentage correctly predicted.

1. Percentage correctly predicted

```
> hitmiss(mod1, k = 0.1)
Classification Threshold = 0.1
        y=0 y=1
yhat=0 6144  90
yhat=1   77  16
Percent Correctly Predicted = 97.36%
Percent Correctly Predicted = 98.76%, for y = 0
Percent Correctly Predicted = 15.09%  for y = 1
Null Model Correctly Predicts 98.32%
[1] 97.36052 98.76226 15.09434
```

If we lower the threshold, we could get

- more true positives;
- more false positives;
- lower percentage correctly predicted.

1. Percentage correctly predicted

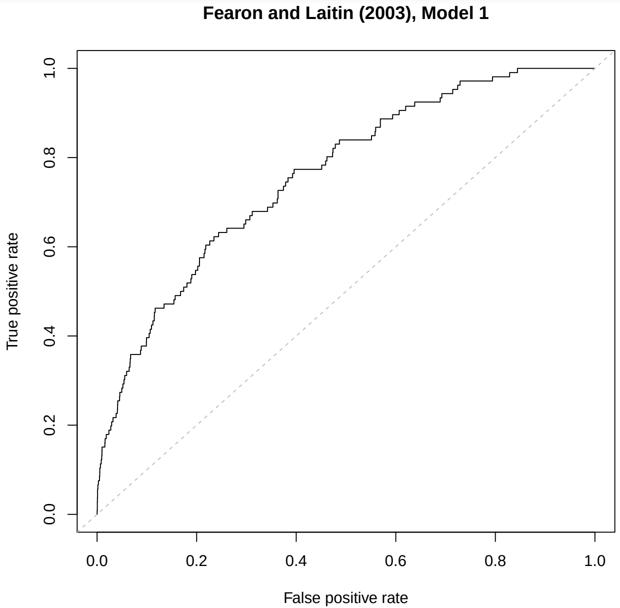
```
> hitmiss(mod1, k = 0.05)
Classification Threshold = 0.05
        y=0 y=1
yhat=0 5937  77
yhat=1  284  29
Percent Correctly Predicted = 94.29%
Percent Correctly Predicted = 95.43%, for y = 0
Percent Correctly Predicted = 27.36%  for y = 1
Null Model Correctly Predicts 98.32%
[1] 94.29429 95.43482 27.35849
```

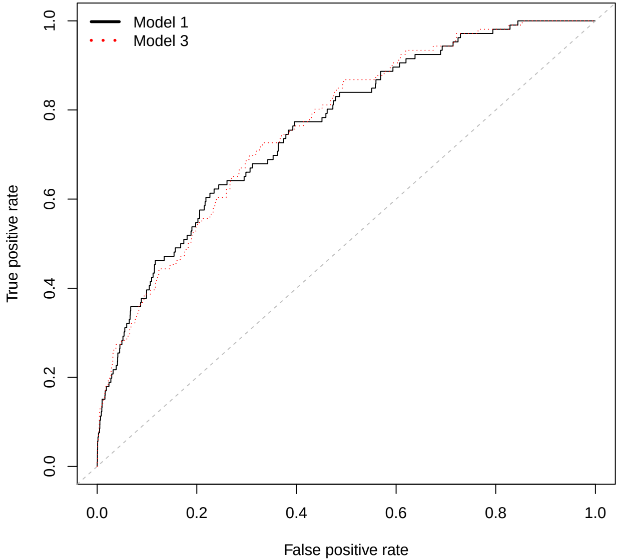
If we lower the threshold, we could get

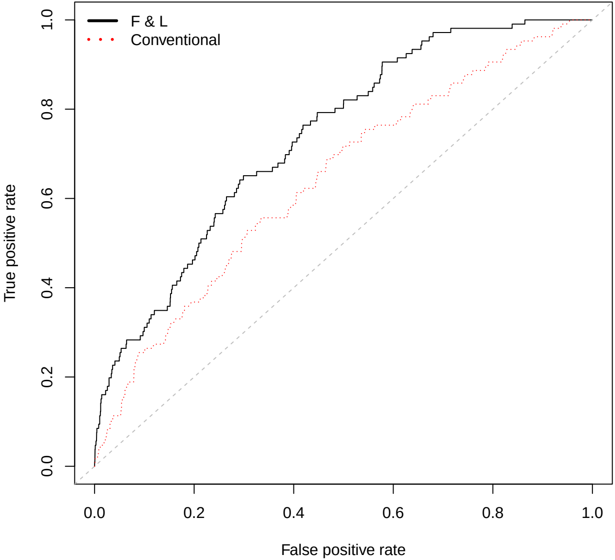
- more true positives;
- more false positives;
- lower percentage correctly predicted.

2. ROC (receiver operating characteristic) Curve

- True positive rate vs false positive rate for different thresholds
  - With a constant-only model, the two are equal
- **Threshold-independent**


Fearon and Laitin (2003), Model 1
True positive rate
0.0
0.2
0.4
0.6
0.8
1.0
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0


Model 1
Model 3
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0

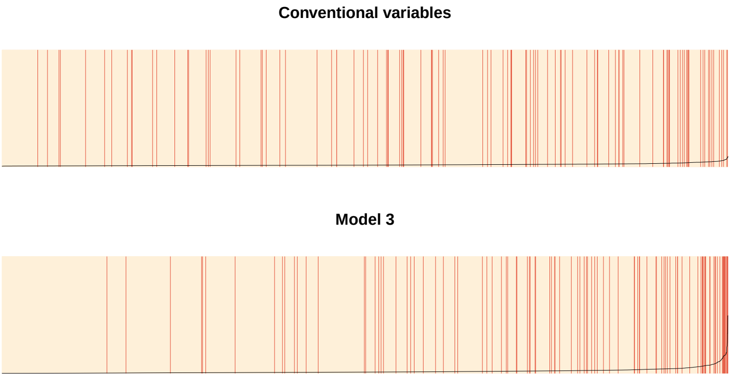

F & L
Conventional
True positive rate
False positive rate
0.0
0.2
0.4
0.6
0.8
1.0

3. AUC

Area under the ROC curve: 0 - 1

- AUC for a constant-only model is __0.5__
- AUC for a "perfect" model is __1__

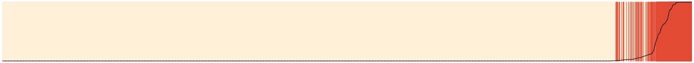
4. Separation plot



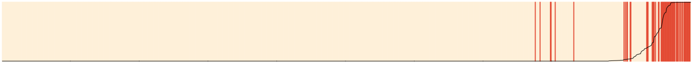
Greenhill, Ward, & Sacks (2011)

4. Separation plot

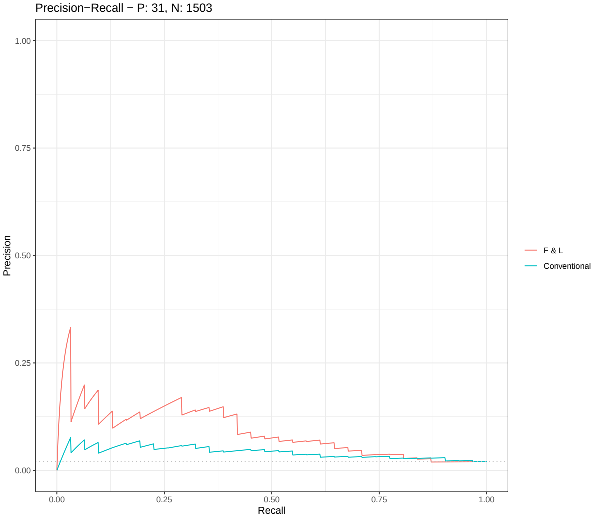
In-Sample: Ethnic Violence Model



Out-of-Sample: Ethnic Violence Model



5. Precision-Recall Curve



Summary

- Out-of sample prediction is a new way for evaluating the effect of proposed variable: in addition to the substantive effect
- A very common tool of analysis in the world of machine learning