

# An Empirical Evaluation of Explanations for State Repression

Daniel W. Hill Jr. <sup>1</sup>     Zachary M. Jones <sup>2</sup>

<sup>1</sup>University of Georgia

<sup>2</sup>Penn State University

# The Empirical Literature on State Repression is Prolific

- ▶ Goal is to discover political, economic, social conditions related to human rights abuse
- ▶ The most seminal study is Poe and Tate (1994)
  - ▶ Democracy
  - ▶ Civil/international war
  - ▶ Economic development
  - ▶ Population size
- ▶ International economic factors
- ▶ Civil society/NGO activity
- ▶ Domestic legal institutions
- ▶ International law

# Theoretical Claims Could be Better Evaluated

- ▶ Importance of theoretically relevant indicators assessed by null hypothesis significance tests
- ▶ Statistical significance  $\neq$  predictive power
- ▶ This does not prevent overfitting

# What We Contribute

- ▶ Cross-validation
  - ▶ Does inclusion of the indicator improve out-of-sample fit?
  - ▶ Assessing out-of-sample fit guards against overfitting
- ▶ Random forests
  - ▶ No need to choose “baseline” model
  - ▶ Allows for interactions/non-linear relationships
  - ▶ Shows relative importance of variables

# Cross-Validation: What Is It?

- ▶ A way to assess how good a model is at predicting the outcome of interest
- ▶ Split the data into  $k$  folds, estimate a model on  $k - 1$  of the folds, and then predict the held out fold (saving the discrepancy statistic)
- ▶ Do this a lot of times to make sure that the discrepancy statistic that you got for a particular iteration isn't dependent on a particular split
- ▶ We use RMSE for OLS models and Somer's  $D$  (a rank correlation coefficient) for the ordinal logit models

# Cross-Validation: How Do We Do It?

- ▶ Dependent variables: aggregated CIRI, PTS, CIRI sub-components
- ▶ Covariates: base model (two types) + variable of interest
- ▶ 7 dependent variables (sort-of), 62 model specifications, 10 folds, 1000 resampling iterations, on 5 imputed data sets = 21,700,000 models estimated (ack!)

# CV Ordinal Logit: Political Imprisonment

`./figures/cv-polpris.png`

# CV Ordinal Logit: Torture

`./figures/cv-tort.png`



## CV OLS: Aggregated CIRI Index

`./figures/cv-cwar-physint.png`

## CV OLS: Political Terror Scale

`./figures/cv-cwar-pts-ols.png`

# Random Forests: Theory

- ▶ Random forests are an ensemble of decision trees (we use a variation on the normal variety)
  - ▶ Select a set of observations
  - ▶ Select a set of variables
  - ▶ Find which variable is the most strongly related to the outcome variable
  - ▶ Find the split in the selected variable which optimally classifies observations
  - ▶ Repeat until stopping criteria is met
  - ▶ Repeat independently for each tree in the forest
- ▶ Predicted class based on consensus vote from all trees in the forest

# Random Forests: Variable Importance

- ▶ If a variable is an important predictor, then randomly permuting its value should decrease classification/regression performance
- ▶ For each variable, test the null hypothesis that  $P(Y, X_j, Z) = P(Y, Z)P(X_j)$  where  $Y$  is the dependent variable,  $X_j$  is the variable being permuted, and  $Z$  are all  $X_i, i \neq j$

# Random Forests: Political Imprisonment

`./figures/polpris-imp-sig.png`

# Random Forests: Torture

`./figures/tort-imp-sig.png`

# What Do the Results Tell Us?

- ▶ Civil war matters a lot
- ▶ Democracy matters a lot, but matters a lot more for some things
  - ▶ Some Polity components measure repression
  - ▶ This requires attention in the future
- ▶ Concepts that receive far less attention matter a lot
  - ▶ Domestic legal institutions
    - ▶ Judicial independence
    - ▶ Common law
    - ▶ Fair trial provisions
  - ▶ Oil rents
  - ▶ Youth bulges

# Thanks!

- ▶ Data & code online: <http://github.com/zmjones/eeesr>
- ▶ A draft of the paper (at [zmjones.com](http://zmjones.com) or [myweb.fsu.edu/dwh06c](http://myweb.fsu.edu/dwh06c)) and these slides are also online (on the PSS website or our personal sites)
- ▶ Email us comments if you have any: Danny ([dwhill@uga.edu](mailto:dwhill@uga.edu)), Zach ([zmj@zmjones.com](mailto:zmj@zmjones.com))



# Random Forest Implementations

- ▶ Decision trees are high variance estimators (hence the use of ensembles, aka random forests)
- ▶ Many implementations of decision trees/random forests are biased towards variables with more possible splits (`rpart`, `randomForest` in R, for example)
- ▶ Many implementations of random forests overfit (hence pruning after fitting)
- ▶ We use random forests with conditional inference trees (via `party`, in R)
  - ▶ Separates variable selection and variable splitting
  - ▶ Uses linear statistics for selection and stopping rules
  - ▶ This prevents biased variable selection and overfitting

# Covariate Missingness

- ▶ Multiple Imputation Using Chained Equations (MICE)
  - ▶ Classification and Regression Trees for categorical data
  - ▶ Random Indicator method for non-ignorably missing numeric data
- ▶ Run cross-validation/random forests on each imputed data-set
- ▶ Pool results, and compute summary statistics

# CV Ordinal Logit (with Civil War): Political Imprisonment

`./figures/cv-cwar-polpris.png`

# CV Ordinal Logit: Disappearances

`./figures/cv-disap.png`

# CV Ordinal Logit (with Civil War): Disappearances

`./figures/cv-cwar-disap.png`

## CV Ordinal Logit: Killings

`./figures/cv-kill.png`

# CV Ordinal Logit (with Civil War): Killings

`./figures/cv-cwar-kill.png`

# CV Ordinal Logit: Political Terror Scale

`./figures/cv-pts-lrm.png`



# CV Ordinal Logit (with Civil War): Political Terror Scale

`./figures/cv-cwar-pts-lrm.png`

# CV OLS: Political Terror Scale

`./figures/cv-pts-ols.png`

## CV OLS (with Civil War): Political Terror Scale

`./figures/cv-cwar-pts-ols.png`

# CV OLS: Aggregated CIRI Index

`./figures/cv-physint.png`

# CV OLS: Aggregated CIRI Index

`./figures/cv-physint.png`

# Random Forests: Disappearances

`./figures/disap-imp-sig.png`

# Random Forests: Killings

`./figures/kill-imp-sig.png`

# Random Forests: Political Terror Scale

`./figures/pts-imp-sig.png`



# Random Forests: Aggregated CIRI Index

`./figures/physint-imp-sig.png`