Interpretable Statistical Learning Methods

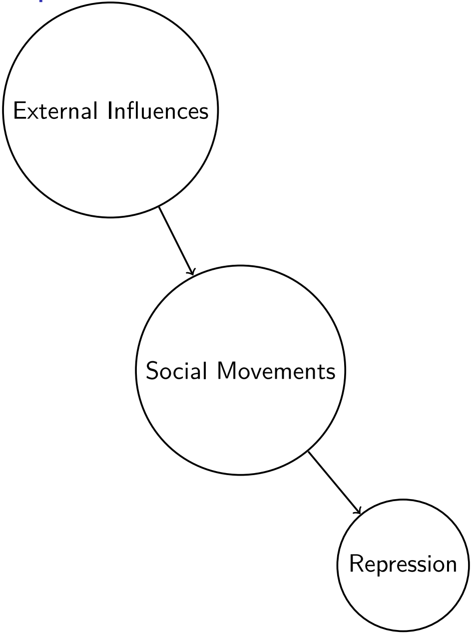
Zachary M. Jones

Penn State

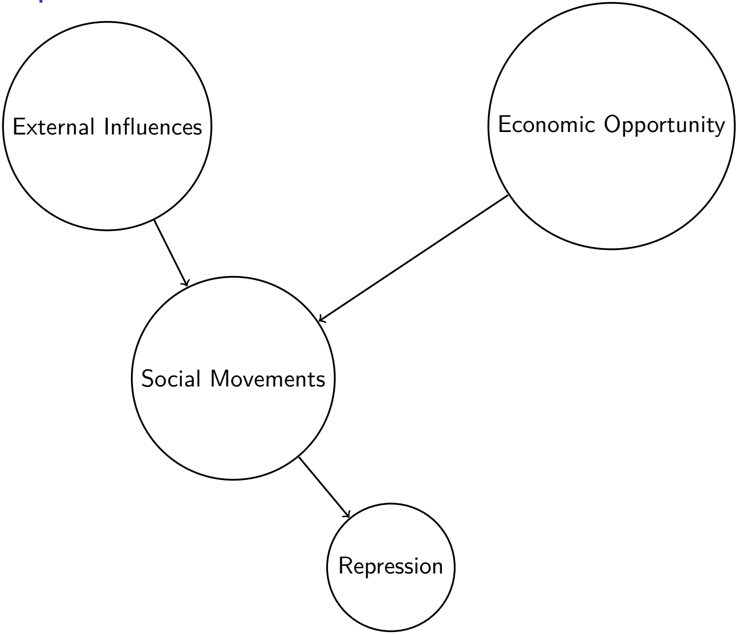
Introduction

- statistical learning $\approx$ flexible generalization of regression
  - learns interactions, nonlinearities, without prespecification
  - not generally interpretable
- my contribution: methods for making statistical learning methods usable in political science
  - implemented in easy-to-use software
  - empirical applications

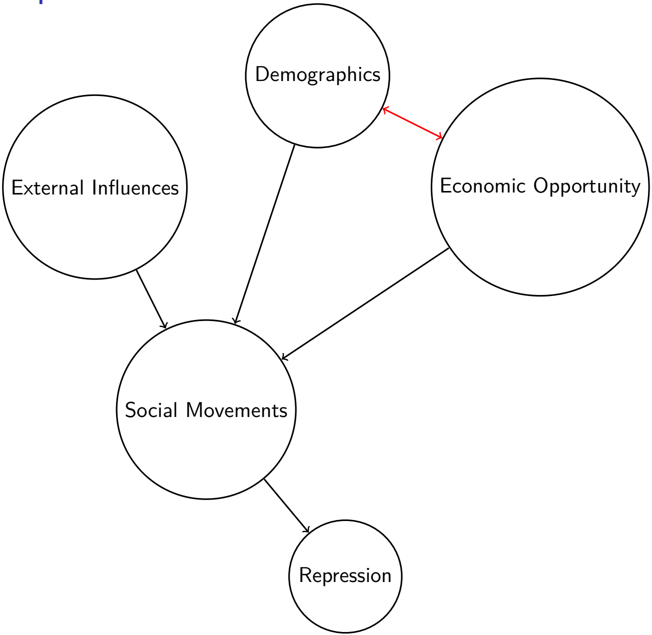
Repression



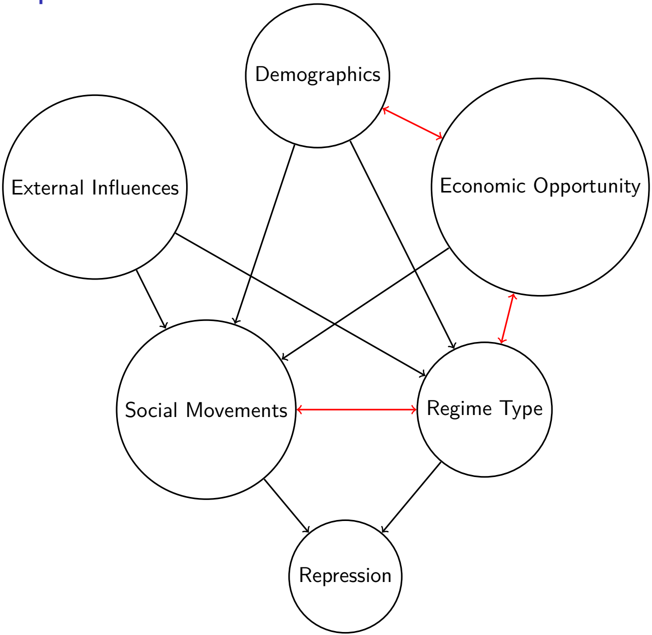
Repression


External Influences
Economic Opportunity
Social Movements
Repression

Repression


Demographics
External Influences
Economic Opportunity
Social Movements
Repression

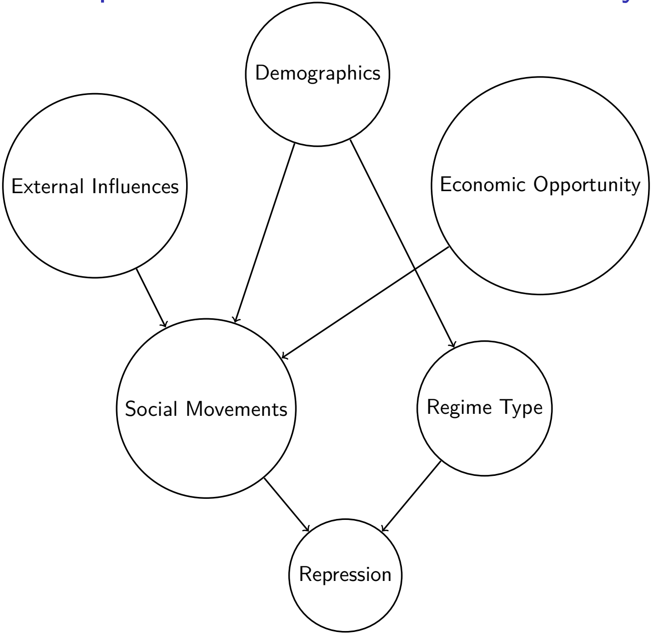
Repression


Demographics
External Influences
Economic Opportunity
Social Movements
Regime Type
Repression

Conventional Data Analysis in the Social Sciences

- linear models and generalizations thereof are most common
- substantive interpretation is the focus

Assumptions of Conventional Data Analysis



Problems with Conventional Data Analysis

- functional form assumptions often not implied by theory
- robustness check = informal specification search

Statistical Learning Methods

machine learning with statistical properties

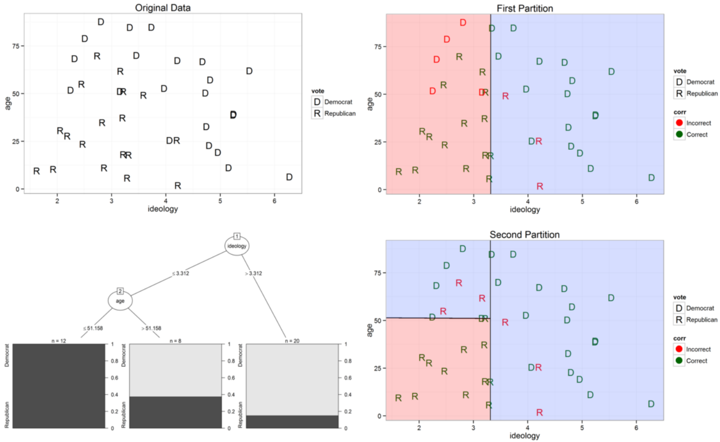
- developed with a focus on model generalizability
- few limits on functional form

Definitions

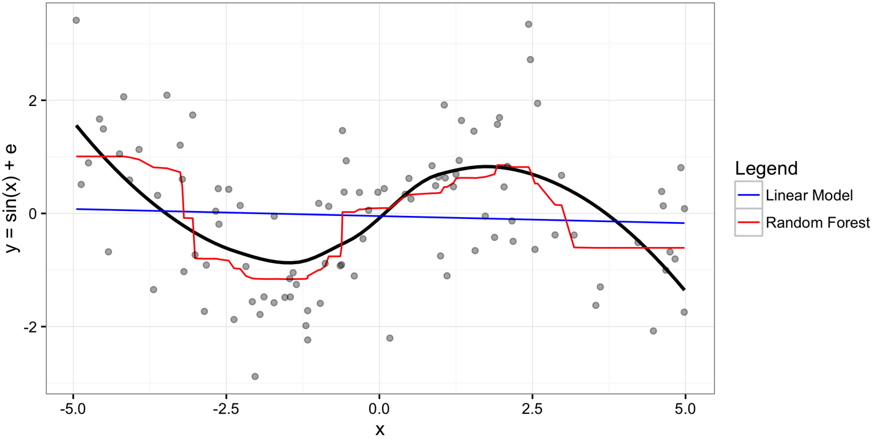
conventional methods have a finite, fixed number of parameters

statistical learning methods have a data-adaptive number of parameters

Decision Trees



Conventional versus Statistical Learning Methods



Why Don't We Use Statistical Learning Methods to Analyze Our Data?

interpretability!

solved by my software

Interpreting Models

- interpretation is how we test theory
- interpretability $\approx$ simplicity
- simplicity is inappropriate if the data are generated by a complex process

Minimize the cost of interpretability!

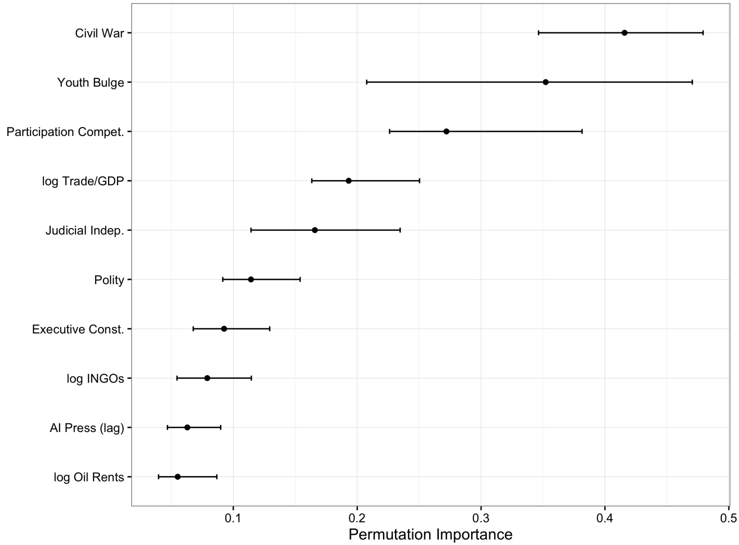
Components of Model Interpretation

component	conventional	statistical learning
size	$\lvert\beta\rvert$	permutation importance
shape	sign of β or marginal effect	partial dependence
variability	prespecified	interaction detection
reliability	standard error, CI	extrapolation, variance

Permutation Importance

- identifies covariates important for variation in predictions
- randomly shuffling covariates makes them useless to the model (if they are important)

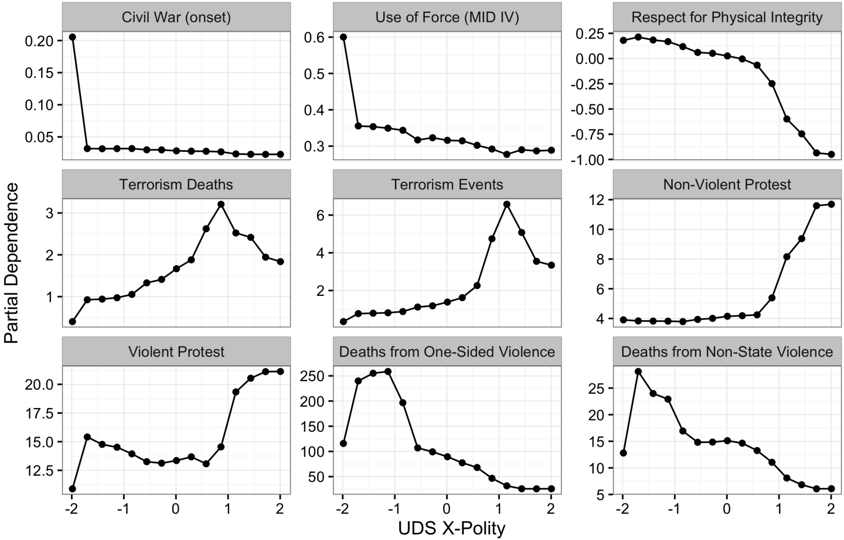
Permutation Importance (Hill and Jones 2014)



Partial Dependence (Friedman 2001)

- Monte-Carlo method for estimating how a model depends on a subset of the covariates
- similar to marginal effects

Partial Dependence (Jones and Lupu 2016)



Software

- using interpretable statistical learning methods requires substantial time, programming skill
  - using statistical learning methods well
  - implementing interpretation methods
- two easy-to-use packages in `R`
  - **M**achine **L**earning with **R** (`mlr`)
  - **E**xploratory **D**ata **A**nalysis using **R**andom **F**orests (`edarf`)

Future Research

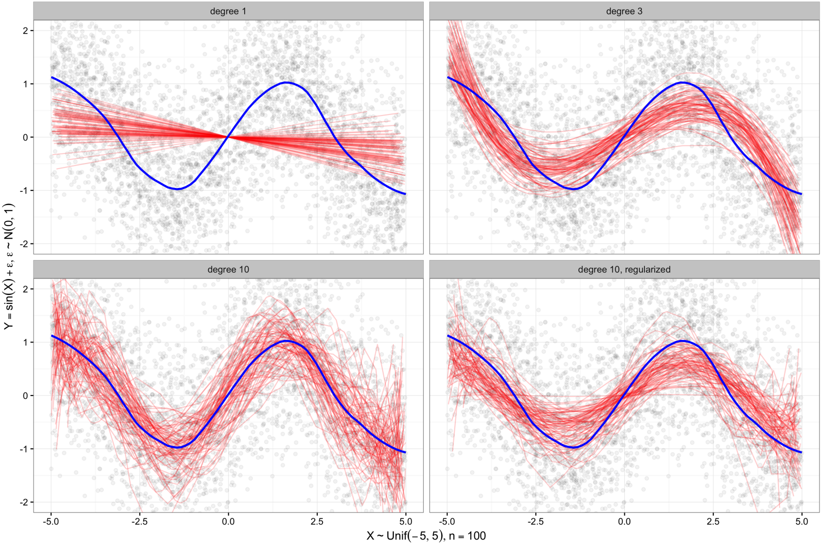
- further methodological development
  - enhancements
  - statistical properties
  - faster (bigger data)
- applications! collaboration!

Conclusion

- statistical learning methods can replace or be used as a robustness check for conventional methods
- try out my software!
- more information at `zmjones.com/hire_me`

Thanks!

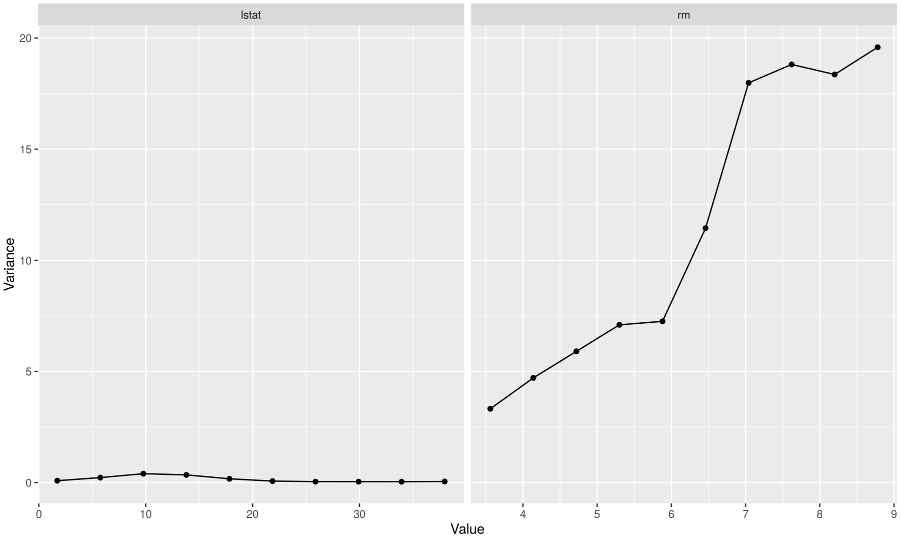
The Bias/Variance Tradeoff



Partial Dependence

$$\hat{f}_u(x_u) = \frac{1}{n}\sum_{i=1}^{n}\hat{f}(x_u, x_{-u,i})$$

Variations on Partial Dependence (Goldstein et. al., 2015)



Functional ANOVA

$$f(x) = f_0 + \sum_{i=1}^{p} f_i(x_i) + \sum_{i \neq j} f_{ij}(x_{ij}) + \ldots$$

$$f_u(x) = \int_{x_{-u}} \left(f(x) - \sum_{v \subset u} f_v(x) \right) dx_{-u}$$

Extrapolation

- Behavior of models at points far from the sample data is variable.
- Different models behave differently far from the data.

Methods for Detecting Extrapolation

- joint density estimation
- data depth
- anomoly/outlier detection

Weighted Functional ANOVA (Hooker 2012)

$$f_u(x) = \int_{x_{-u}} \left(f(x) - \sum_{v \subset u} f_v(x) \right) w(x) dx_{-u}$$

Weighted Functional ANOVA