A Statistical Method for Empirical Testing of Competing Theories

Kosuke Imai **Dustin Tingley**

Princeton University

April 14, 2010

Motivation

- Empirical testing of competing theories lies at the heart of social science research
- Need to test the validity of alternative theories explaining the same phenomena
- "theory confirmation is not possible when a theory is tested in isolation, regardless of the statistical approach" (Clarke)

- Common statistical methods used in the discipline:
  1. "Garbage-can" regressions: atheoretical (Achen)
  2. Model selection methods (e.g., AIC, BIC, Vuong test, J test): All or nothing, Independence of Irrelevant Alternatives (IIA)
- Key distinction between causal and predictive inference

The Proposed Approach

- Theoretical heterogeneity: No single theory can explain everything
- Explaining when each theory "works"
  1. Testing the entire theory including its assumptions rather than just its implications
  2. Leading to further theory development
- Finite mixture models
  1. A well-known, very general class of statistical models
  2. Can test more than two theories at the same time
  3. Under-utilized in political science except a few studies
- Quantities of interest:
  1. population proportion of observations consistent with each theory
  2. how this proportion varies as a function of observed characteristics
  3. probability that a particular observation is consistent with a theory
  4. list of observations that are consistent with each theory

An Example: Determinants of Trade Policies

- Hiscox (2002, *APSR*) analyzes US legislative voting on trade bills
- Stolper-Samuelson (SS) model: cleavages along factoral lines
  - The highly skilled favor liberalization while the low-skilled oppose it
- Ricardo-Viner (RV) model: cleavages along sectoral lines
  - Exporters favor liberalization while importers oppose it
- Key contribution: the applicability of the two models depends on the level of factor mobility in the US economy
  - If capital is highly mobile across industries, then the conditions for the SS model are satisfied
  - If capital is highly specific, then the conditions for the RV model are satisfied

Finite Mixture Models: A Review

- M competing theories, each of which implies a statistical model $f_m(y \mid x)$ for $m = 1, \dots, M$
- The data generating process:

$$Y_i \mid X_i, Z_i \;\sim\; f_{Z_i}(Y_i \mid X_i, \theta_{Z_i})$$

  where Z_i is the *latent* variable indicating the theory which generates observation i
- The observed-data likelihood function:

$$L_{obs}(\Theta, \Pi \mid \{X_i, Y_i\}_{i=1}^N) \;=\; \prod_{i=1}^N \left\{ \sum_{m=1}^M \pi_m f_m(Y_i \mid X_i, \theta_m) \right\},$$

  where $\pi_m = \Pr(Z_i = m)$ is the population proportion of observations generated by theory m
- π_m: a measure of overall performance of the theory

- Explaining theoretical heterogeneity:

$$\Pr(Z_i = m \mid W_i) \quad = \quad \pi_m(W_i, \psi_m),$$

- Predicting which theory has generated a particular observation:

$$\begin{aligned}\zeta_{i,m} &= \Pr(Z_i = m \mid \Theta, \Pi, \{X_i, Y_i\}_{i=1}^N) \\ &= \frac{\pi_m f_m(Y_i \mid X_i, \theta_m)}{\sum_{m'=1}^M \pi_{m'} f_{m'}(Y_i \mid X_i, \theta_{m'})}\end{aligned}$$

- Grouped observations:

$$\zeta_{i,m} \quad = \quad \frac{\pi_m \prod_{j=1}^{J_i} f_m(Y_{ij} \mid X_{ij}, \theta_m)}{\sum_{m'=1}^M \pi_{m'} \prod_{j=1}^{J_i} f_{m'}(Y_{ij} \mid X_{ij}, \theta_{m'})}$$

- Estimation: Expectation-Maximization or Markov chain Monte Carlo algorithm
- Implementation: `flexmix` package in R by Leisch and Gruen

Statistically Significantly Consistent with a Theory

- Identification of observations that are statistically significantly consistent with each theory
- Idea: If $\zeta_{i,m}$ is greater than a threshold λ_m, then include observation i in the list
- Problem of multiple testing: false positives
- Simple example:
  - 10 Independent 0.05 level tests
  - $1 - 0.95^{10} \approx 0.4$ chance of at least one false discovery
- Solution: choose the smallest value of λ_m such that the posterior expected value of false discovery rate on the resulting list does not exceed a prespecified threshold α_m:

$$\lambda_m^* = \inf\left\{\lambda_m : \frac{\sum_{i=1}^N (1-\hat{\zeta}_{i,m})\mathbf{1}\{\hat{\zeta}_{i,m} \geq \lambda_m\}}{\sum_{i=1}^N \mathbf{1}\{\hat{\zeta}_{i,m} \geq \lambda_m\} + \prod_{i=1}^N \mathbf{1}\{\hat{\zeta}_{i,m} < \lambda_m\}} \leq \alpha_m\right\}$$

Measuring the Overall Performance of a Theory

1. Population proportion of observations consistent with each theory: π_m or $\sum_{i=1}^{N} \hat{\zeta}_{i,m}/N$
2. Sample proportion of the observations statistically significantly consistent with the theory

Testing the Competing Theories of Trade Policy

- Data
  - Congressional voting data on 55 trade bills spanning over 150 years
  - A combined measure of factor specificity for a given year
  - State-level measures of relevant covariates for each model
- The original analysis used the J test in logistic regression with bill fixed effects
- The J test in its original form:

$$Y_i = (1-\pi) f(X_i, \beta) + \pi g(X_i, \gamma) + \epsilon_i,$$

  - The null hypothesis, $Y_i = f(X_i, \beta) + \epsilon_i$
  - The alternative hypothesis, $Y_i = g(X_i, \gamma) + \epsilon_i$
- Finite mixture models do not assume π is either 0 or 1

The Mixture Model Specification

- Assuming all votes for the same bill belong to the same model
- Stolper-Samuelson Model:

$$\text{logit}^{-1}(\beta_0 + \beta_1 \texttt{profit}_{ij} + \beta_2 \texttt{manufacture}_{ij} + \beta_3 \texttt{farm}_{ij})$$

- Ricardo-Viner Model:

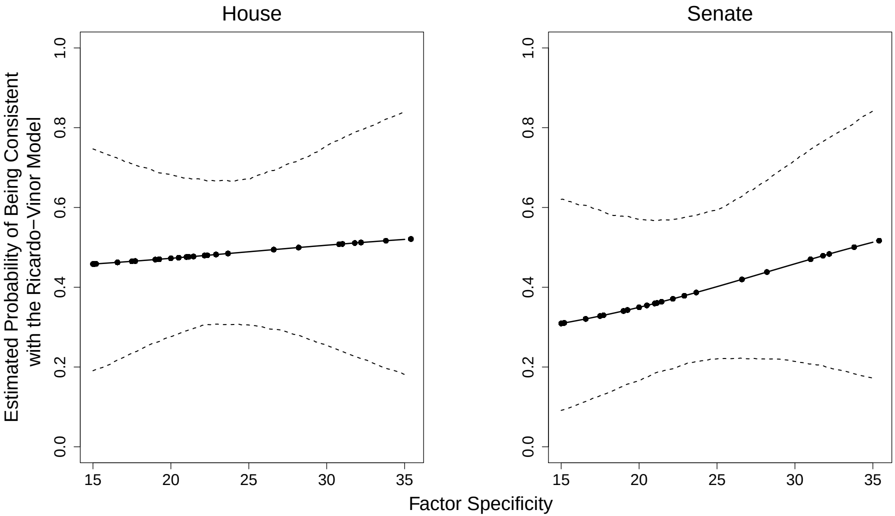
$$\text{logit}^{-1}(\gamma_0 + \gamma_1 \texttt{export}_{ij} + \beta_2 \texttt{import}_{ij})$$

- Model for mixing probability:

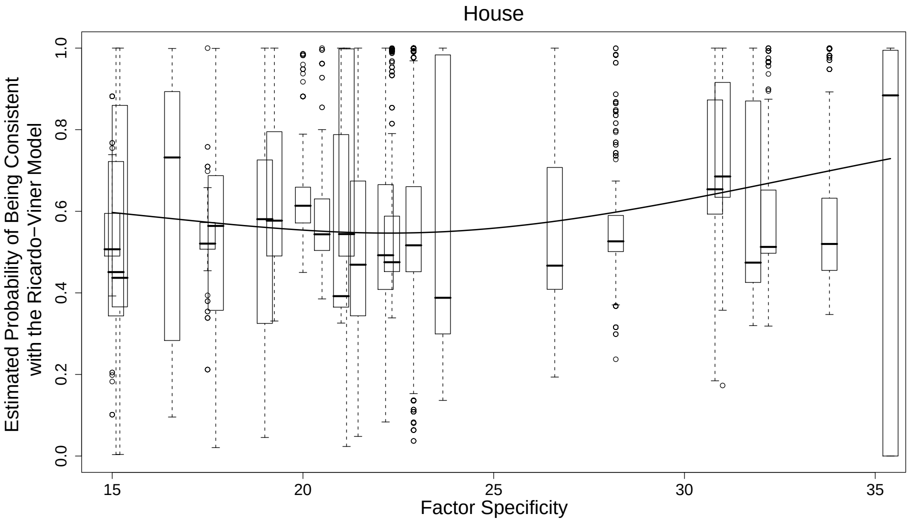
$$\text{logit}^{-1}(\delta_0 + \delta_1 \texttt{factor}_j)$$

- Implementation using `flexmix` package in R

Results with Grouped Observations



Results without Grouping and Parametric Assumption



Mixture Model vs. Garbage-can Model

		Mixture Model				"Garbage-can" Model			
		House		Senate		House		Senate	
Models	Variables	coef.	s.e.	coef.	s.e.	coef.	s.e.	coef.	s.e.
SS	profit	−1.60	0.53	−5.69	1.19	−0.42	0.33	−2.14	0.73
	manufacture	17.60	1.54	19.79	2.59	5.69	0.63	4.73	1.32
	farm	−1.33	0.29	−1.27	0.43	−0.11	0.14	−0.03	0.25
RV	import	3.09	0.33	2.53	0.80	0.63	0.21	1.21	0.43
	export	−0.85	0.16	−2.80	0.77	−0.85	0.08	−1.48	0.20
π	factor	0.01	0.06	0.05	0.07				

- All estimates have expected signs and are statistically significant for the mixture model
- Garbage-can regression has smaller and sometimes statistically insignificant coefficients
- The original analysis contains some estimates with "wrong" signs

Classification of House Trade Bills

Stolper-Samuelson Model	Ricardo-Viner Model
Adams Compromise (1832)	Tariff Act (1824)
Clay Compromise (1833)	Tariff Act (1828)
Tariff Act (1842)	Gorman Tariff (1894)
Walker Act (1846)	Underwood Tariff (1913)
Tariff Act (1857)	RTAA (1934)
Morrill Act (1861)	RTA Extension (1937)
Tariff Act (1875)	RTA Extension (1945)
Morrison Bill (1984)	RTA Extension (1955)
Mills Bill (1988)	Trade Expansion Act (1962)
McKinley Tariff (1890)	Mills Bill (1970)
Dingley Tariff (1894)	Trade Reform Act (1974)
Payne-Aldrich Tariff (1909)	Fast-Track (1991)
Fordney-McCumber Tariff (1922)	NAFTA (1993)
Smoot-Hawley Tariff (1930)	GATT (1994)
Trade Remedies Reform (1984)	

- Fitting the SS (RV) model to the SS and RV votes separately reveals an interesting pattern in terms of sign and statistical significance of estimated coefficients

Testing Agenda Control Theories in Congress

- Ongoing joint project with Josh Clinton and Dan Pemstein
- Roll call data analysis and ideal point estimation
- But, not all potential bills come to the floor
- Party cartel theory (Cox and McCubbins): there should be no proposal on the floor to which the majority of the majority party prefers the status quo

$$\text{cutpoint} \notin [x_{floor}, x_{maj}]$$

- Committee gate-keeping

$$\text{cutpoint} \notin [x_{floor}, x_{comm}]$$

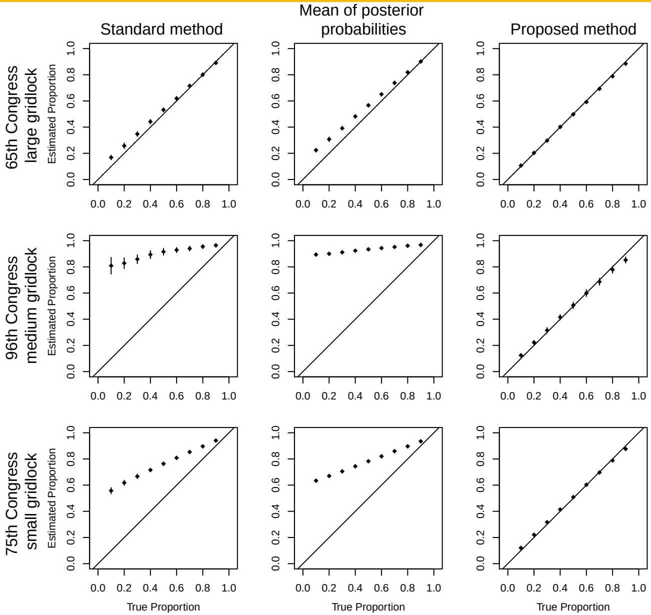
- Mixture model: some bills are consistent with majority party and/or committee gate-keeping

Measuring Party Influence in Congress

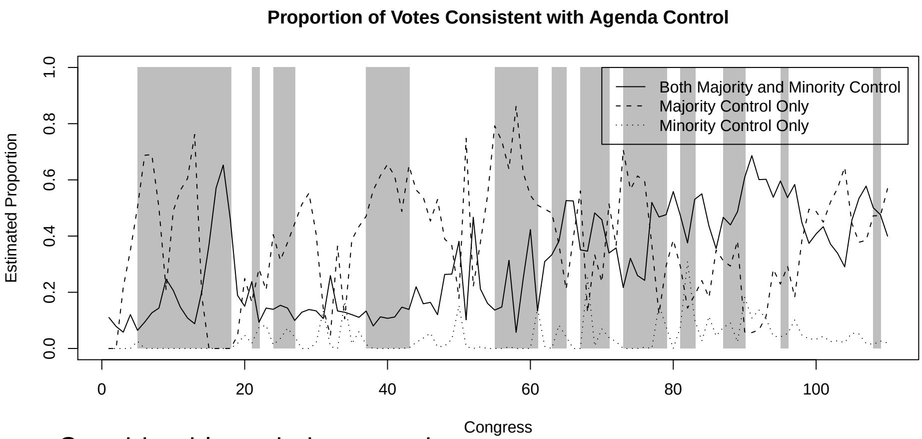
- Width of "gridlock intervals"
- Majority party roll-rates: a majority of the majority party opposes a bill but loses
- Proportion of bills whose cutpoints are in the gridlock interval

- A standard method: run an IRT model (NOMINATE or IDEAL) and count the number of bills that fall outside of the gridlock interval
- A large positive bias results when
  1. the width of the gridlock interval is narrow
  2. the number of bills is small (early Congresses)
  3. the number of legislators is small (Senate)
- Taking into account estimation uncertainty does not reduce bias
- Need to control for false discovery rate
- Challenge: develop a data-driven method to choose the value of α

Some Simulation Results based on Hirsh (2010)



Majority Party Influence in House over Time



- Considerable variation over time
- Positive correlation with united government (House party = Senate party = President's party)
- Potential importance of conference votes or presidential vetoes

Future Plans

- Development of a mixture model incorporating different agenda control theories
- Systematic analysis of factors that determine whether a particular bill is consistent with each theory
  - characteristics of bills
  - characteristics of legislators and committees
  - outside factors: proximity to elections, etc.
- Understanding where the bias of a standard method comes from
- Developing a systematic way to deal with bias

Other Potential Applications

American and Comparative Politics		International Relations
Pivotal politics vs. party cartel accounts of Congressional law making	Greed vs. grievance accounts of civil war onset	Realist vs. liberal theories of conflict
Swing vs. core voter hypotheses of distributional politics	Proximity vs. directional voting	Cultural vs. material explanations of trade and immigration public opinion
Prospective vs. retrospective economic voting	Sociotropic vs. pocket book voting	Screening vs. commitment theories of international organizations

Concluding Remarks

- Mixture models offer an effective way to test competing theories
- Particularly useful in observational studies when causal inference is difficult but predictive inference is possible

- Many advantages over the standard model selection procedures:
  1. Test any number of competing theories
  2. Include nested and/or non-nested models
  3. Conduct frequentist or Bayesian inference
  4. Quantify the overall performance of each theory
  5. Test the conditions under which each theory applies
  6. Identify observations statistically significantly consistent with theory

- Some potential pitfalls:
  1. Demands more from the data
  2. Computationally intensive
  3. Lack of statistical power