Fast Estimation of Ideal Points with Massive Data

Kosuke Imai† James Lo† Jonathan Olmsted‡

†Princeton University ‡NPD Group

Asian Political Methodology Conference

January 9, 2015

Motivation

- Since NOMINATE, widespread interest in spatial voting models
- Extensions compare actors across time and institutions
  - Presidents, legislators, and justices (Bailey 2007)
  - State legislators and representatives (Shor and McCarty 2011)
  - Voters and representatives (Bafumi and Herron 2010)
  - Agencies, presidents, and representatives (Clinton et al 2012)
- But computational challenges are order of magnitude larger
  - MCMC estimation extremely slow (Martin and Quinn 2002)
  - Shortcuts using subsets of data (Shor and McCarty 2011)
  - Compromise in the model (Bailey 2007)
  - Difficulty in convergence (Bafumi and Herron 2010)
  - Supercomputer center usage (Carroll et al 2009)
- Models attractive but practically unusable with large data sets

Ideal Points from Large Data Sets

	# of subjects	# of items	Data types
DW-NOMINATE scores (1789 – 2012)	37,511	46,379	roll calls
Common Space scores (1789 - 2012)	11,833	90,609	roll calls
Martin and Quinn scores (1937 – 2013)	697	5,164	votes
Gerber and Lewis (2004)	2.8 million	12	votes
Bailey (2007)	27,795	2,750	roll calls & votes
Bafumi and Herron (2010)	8,848	4,391	survey & roll calls
Shor and McCarty (2011)	6,201	5,747	survey & roll calls
Tausanovitch and Warshaw (2013)	275,000	311	survey
Peress (2013)	700	16,000	co-sponsorship & roll calls
Bonica (2014)	4.2 million	78,363	contribution

Courtesy of Will Lowe

Our Solution

- EM algorithms for exact or approximate posterior inference
  - deterministic algorithm
  - variational EM algorithm for approximate inference
- Derive EM algorithms for popular bayesian ideal point models
  1. Standard binary choice (Clinton, Jackman, Rivers 2004)
  2. Ordinal choice (Jackman and Treier 2008)
  3. Dynamic random walk (Martin and Quinn 2002)
  4. Hierarchical model (Bafumi *et al.* 2005)
- Parametric bootstrap for uncertainty (Lewis and Poole 2004)
- EM algorithms yield nearly identical results to standard estimates
- Fast and scalable algorithms
  - 5.5 day processes run in under 10 seconds
  - Simulated data $>$ 500 times in size run in under 25 minutes

Standard Two-Parameter Ideal Point Model

- Implemented as `ideal()` in R, similar to `wnominate()` and `oc()`
- Legislators $i = 1 \ldots N$ and roll call votes $j = 1 \ldots J$
  - Observed votes: $y_{ij} \in \{0, 1\}$
  - Bill parameters: $\tilde{\boldsymbol{\beta}}_j^\top = (\alpha_j, \boldsymbol{\beta}_j^\top)$
  - Ideal point: $\tilde{\mathbf{x}}_i^\top = (1, \mathbf{x}_i^\top)$
- Latent propensity to vote yea:

$$y_{ij}^* = \tilde{\mathbf{x}}_i^\top \tilde{\boldsymbol{\beta}}_j + \epsilon_{ij} \quad \text{with } y_{ij} = \mathbf{1}\{y_{ij}^* > 0\}$$

- Posterior distribution (with normal priors on $\tilde{\mathbf{x}}_i^\top$ and $\tilde{\boldsymbol{\beta}}_j^\top$):

$$p(\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\boldsymbol{\beta}}_j\}_{j=1}^J \mid \mathbf{Y})$$

$$\propto \prod_{i=1}^N \prod_{j=1}^J \left(\mathbf{1}\{y_{ij}^* > 0\}\mathbf{1}\{y_{ij} = 1\} + \mathbf{1}\{y_{ij}^* \leq 0\}\mathbf{1}\{y_{ij} = 0\}\right) \phi_1\left(y_{ij}^*; \tilde{\mathbf{x}}_i^\top \tilde{\boldsymbol{\beta}}_j, 1\right)$$

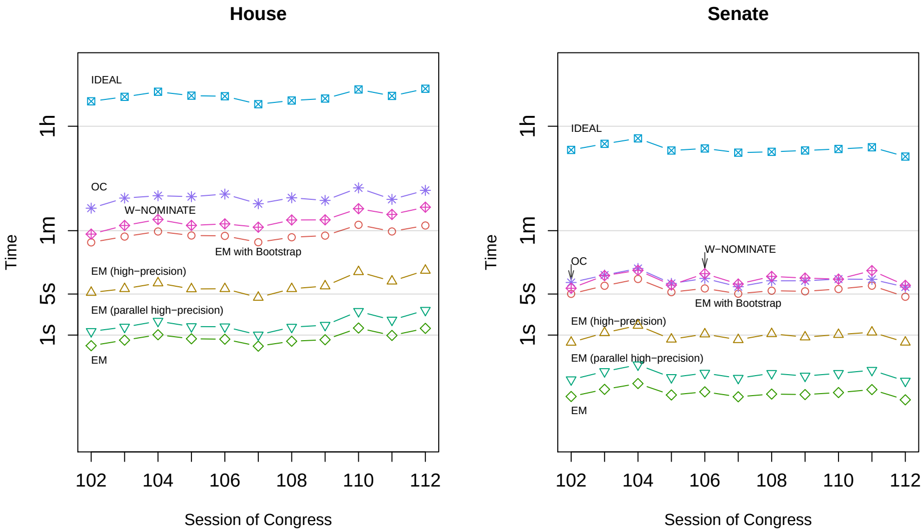
$$\times \prod_{i=1}^N \phi_K\left(\mathbf{x}_i; \boldsymbol{\mu}_{\mathbf{x}}, \boldsymbol{\Sigma}_{\mathbf{x}}\right) \prod_{j=1}^J \phi_{K+1}\left(\tilde{\boldsymbol{\beta}}_j; \boldsymbol{\mu}_{\tilde{\boldsymbol{\beta}}}, \boldsymbol{\Sigma}_{\tilde{\boldsymbol{\beta}}}\right)$$

EM Algorithm for Exact Posterior Inference

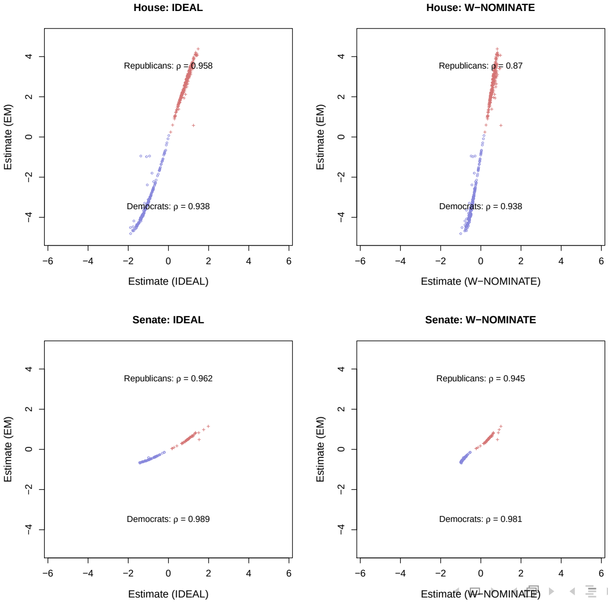
- Treat y_{ij}^* as missing data and $\tilde{\beta}$ and $\mathbf{x}_i$ as parameters
- Iterative algorithm with starting values for $\{\tilde{\beta}_j\}_{j=1}^J$ and $\{\mathbf{x}_i\}_{i=1}^N$
- E-step: compute the "Q-function": $Q(\{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\beta}_j\}_{j=1}^J)$

  $$= \mathbb{E}\left[\log p(\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\beta}_j\}_{j=1}^J \mid \mathbf{Y}) \mid \mathbf{Y}, \{\mathbf{x}_i^{(t-1)}\}_{i=1}^N, \{\tilde{\beta}_j^{(t-1)}\}_{j=1}^J\right]$$
- compute $y_{ij}^{*(t)} = \mathbb{E}(y_{ij}^* \mid \mathbf{x}_i^{(t-1)}, \tilde{\beta}_j^{(t-1)}, y_{ij})$ using a truncated normal
- M-step: maximize Q-function
  1. Bayesian regression of $y_{ij}^{*(t)} - \alpha_j^{(t-1)}$ on $\beta_j^{(t-1)}$
  2. Bayesian regression of $y_{ij}^{*(t)}$ on $(1, \mathbf{x}_i^{(t)\top})$
- Repeat until convergence: correlation of every parameter between two successive iterations is greater than $1 - 10^{-6}$

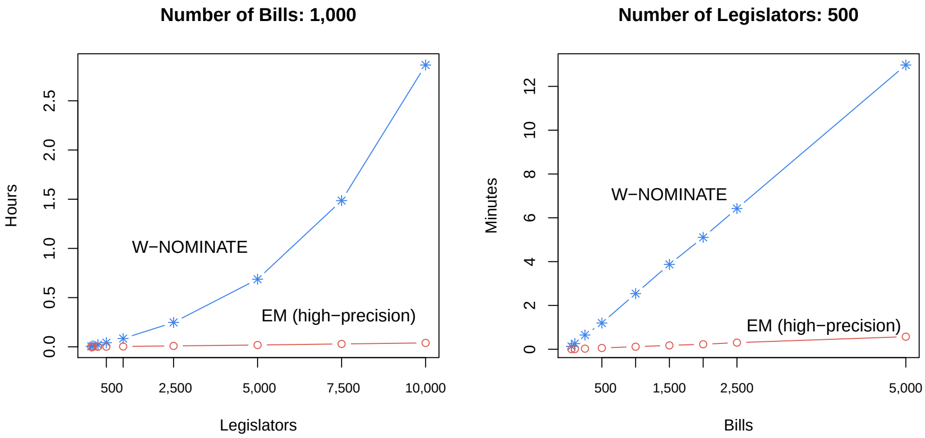
Computational Performance, $102^{nd} - 112^{th}$ Congress



Estimated Ideal Points, 112^{th} Congress



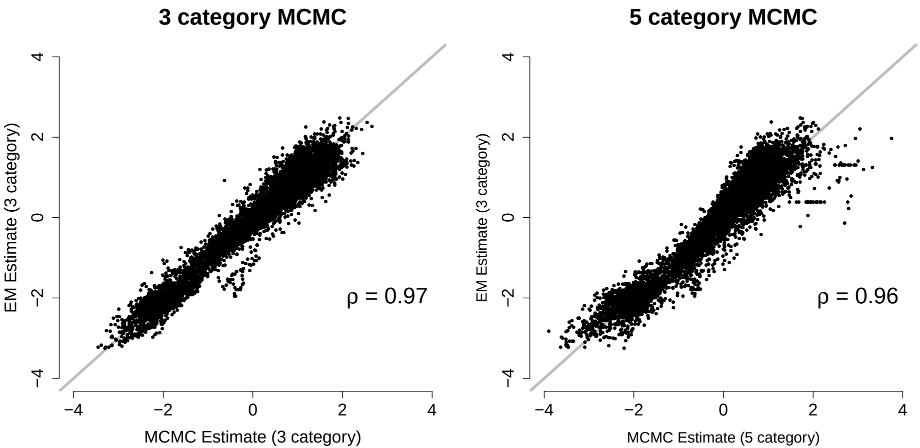
Computational Scalability



Ordinal Ideal Point Model

- A new EM algorithm for the 3-category ordinal choice
  - Particularly applicable to Likert scales from survey data
  - Collapse response categories if there are more than 3
- Devise a parameter transformation to obtain a closed-form E-step
- **Application:** Asahi-Todai Elite Survey
  - Spans 8 Japanese Upper & Lower House elections, 2003-2013
  - In 6 waves, survey administered to N=1,000-2,000 voters
  - Total $N = 19,443$ respondents (7,734 politicians + 11,709 voters)
  - $J = 98$ items, 5-category items collapsed to 3
- Standard MCMC runtime: 4 hours
- Ordinal EM runtime: Under 2 minutes

Politician Ideal Points, Asahi-Todai Survey



Dynamic Ideal Point model

- Dynamic Ideal Point model (Martin and Quinn 2002)
- Flexible dynamic modeling with random walk prior:

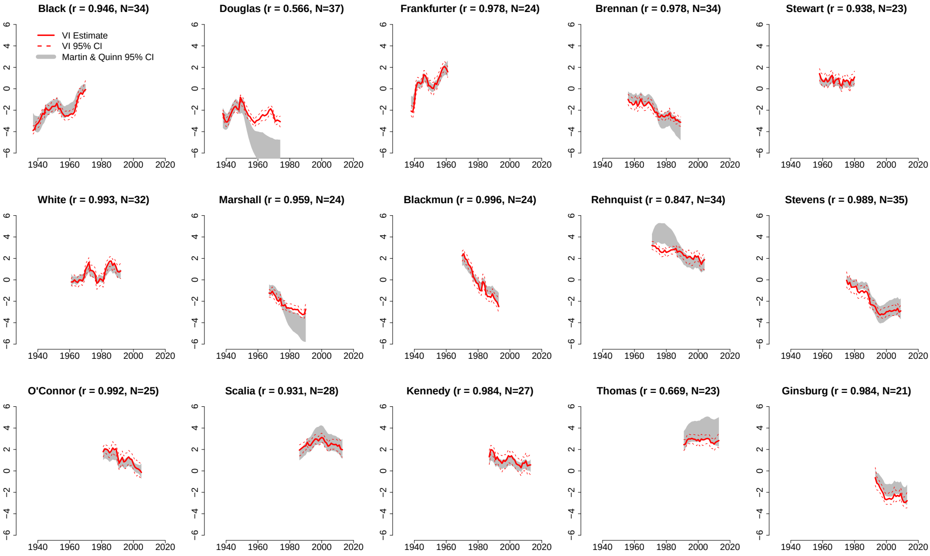
$$x_{it} \sim N(x_{i,t-1}, \omega_x^2)$$

- Derive a variational EM algorithm for this model
- Approximate inference under the factorization assumption

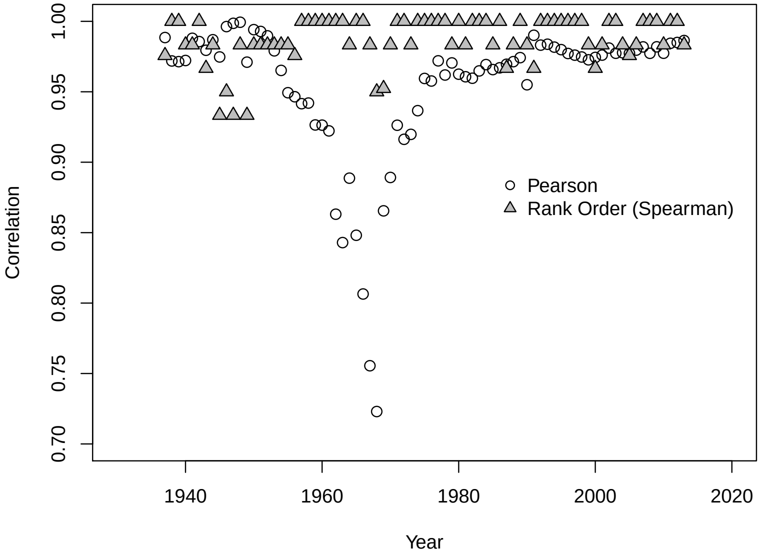
$$q(\mathbf{Y}^*, \{\mathbf{x}_i\}_{i=1}^N, \{\tilde{\beta}_j\}_{t=1}^T) = \prod_{i=1}^N \prod_{t=\underline{T}_i}^{\overline{T}_i} q(y_{it}^*) \prod_{i=1}^N q(\mathbf{x}_i) \prod_{t=1}^T \prod_{j=1}^{J_t} q(\tilde{\beta}_{jt})$$

- **Application:** U.S. Supreme Court MQ Scores
  - N=45 justices, J=5,164 votes, T=77 periods
  - 9 active justices per term
- `MCMCdynamicIRT1d()` runtime: 5.5 days
- EM runtime: Under 10 seconds

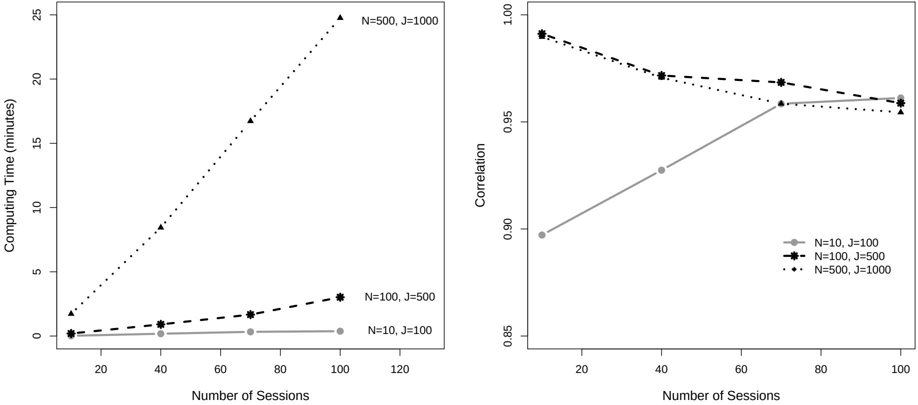
Correlation by Justice, US Supreme Court



Correlation by Term, US Supreme Court



Scalability and Accuracy of Dynamic VI Model



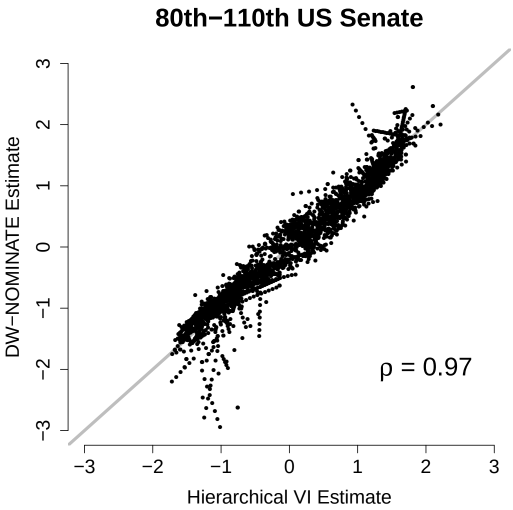
Hierarchical Ideal Point Model

- Enables the use of covariates to predict ideal points
- Given:
  - index of observed vote: ℓ
  - associated legislators $i[\ell]$ and bills $j[\ell]$
  - legislator membership in groups: $g[i[\ell]]$
  - observed covariate(s) associated with specific legislator: $z[i[\ell]]$
- Then we get:

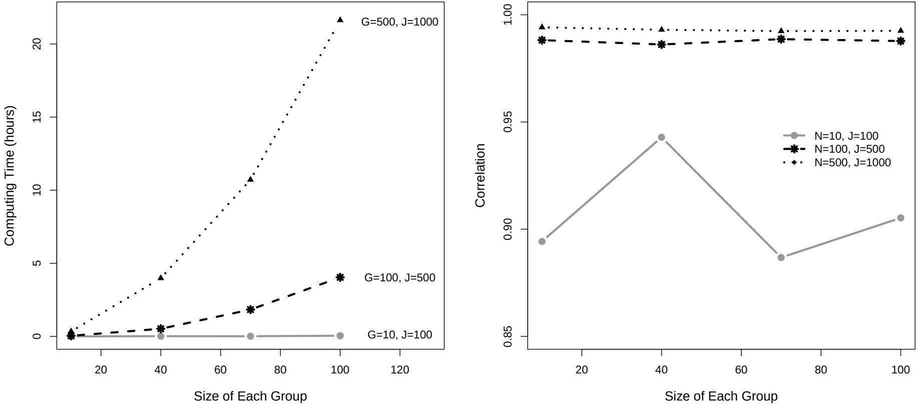
$$\begin{aligned} y^*_\ell &= \alpha_{j[\ell]} + \beta_{j[\ell]} x_{i[\ell]} + \epsilon_\ell \quad \text{where} \quad \epsilon_\ell \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0,1) \\ x_{i[\ell]} &= \gamma^\top_{g[i[\ell]]} \mathbf{z}_{i[\ell]} + \eta_{i[\ell]} \quad \text{where} \quad \eta_{i[\ell]} \stackrel{\text{indep.}}{\sim} \mathcal{N}(0, \sigma^2_{g[i[\ell]]}) \end{aligned}$$

- Parametric time trend model as a special case
- Derive a variational EM algorithm for approximate posterior inference
- DW-NOMINATE runtime: Over 14 hours for 80th-110th Senate
- EM runtime: 8.3 minutes (using 8 threads)

80th-110th U.S. Senate Ideal Points



Scalability and Accuracy



Concluding Remarks

- Ideal point models as an essential tool in political science
- Recent trend: scaling across time and institutions
- Need for speed: fast estimation with massive data
- We propose various EM algorithms that:
  1. produce nearly identical results to standard procedures
  2. are much faster: reducing runtime from 6 days to 10 seconds
  3. scale well to even larger data sets
- Open-source R package `fastideal` will be made available