

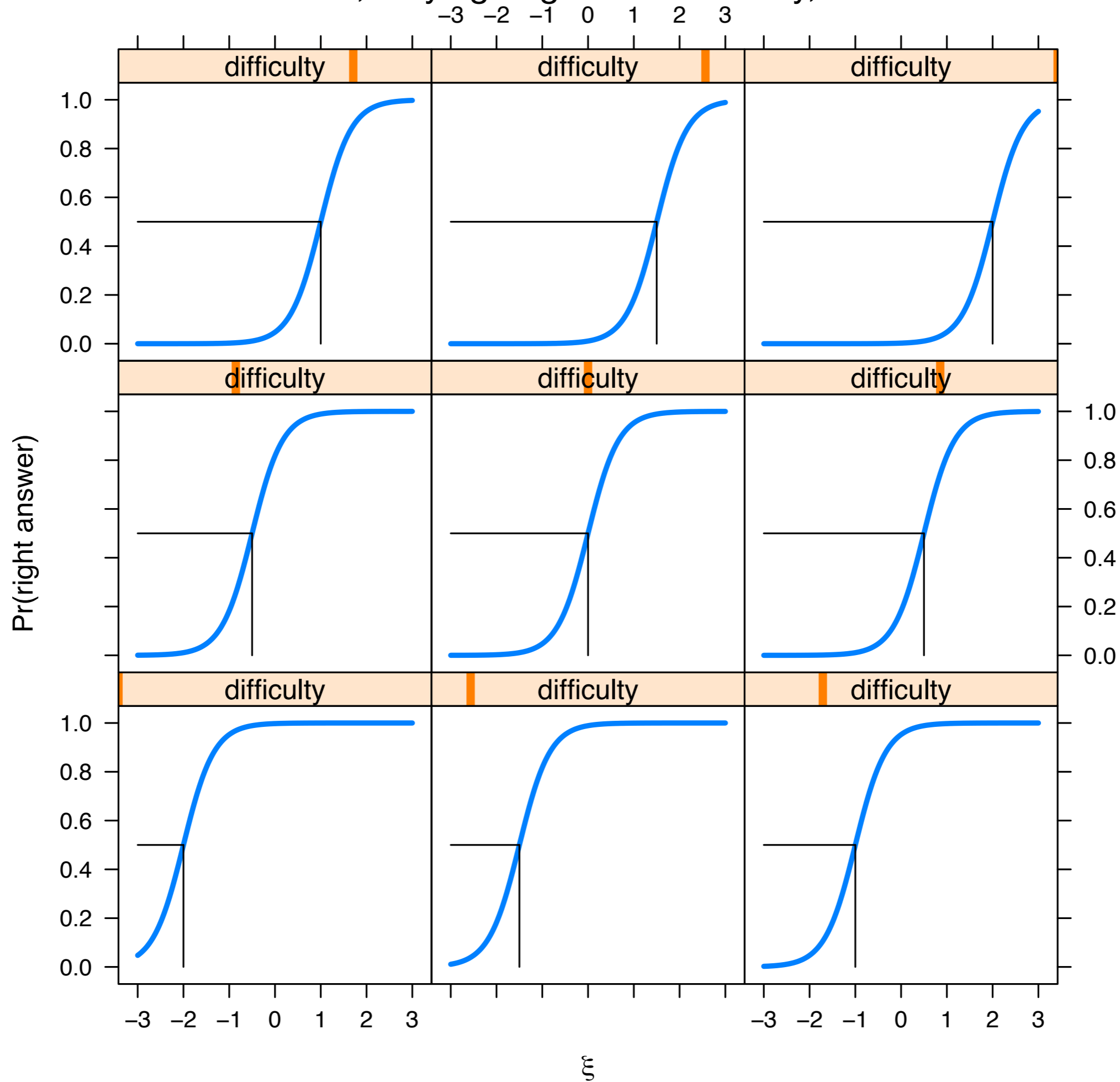
On measuring democracy and things like that

simon jackman
stanford university

validity / reliability

- my focus in memo on reliability/precision
- validity probably a bigger issue

item-characteristic curves, varying degrees of difficulty, constant discrimination



GREs

- Quant & Verbal components self-administered by computer for U.S. based test-takers
- adaptive selection of roughly 30 items
- items selected so as to provide most information regarding subject ability, given current estimate as to subject's ability

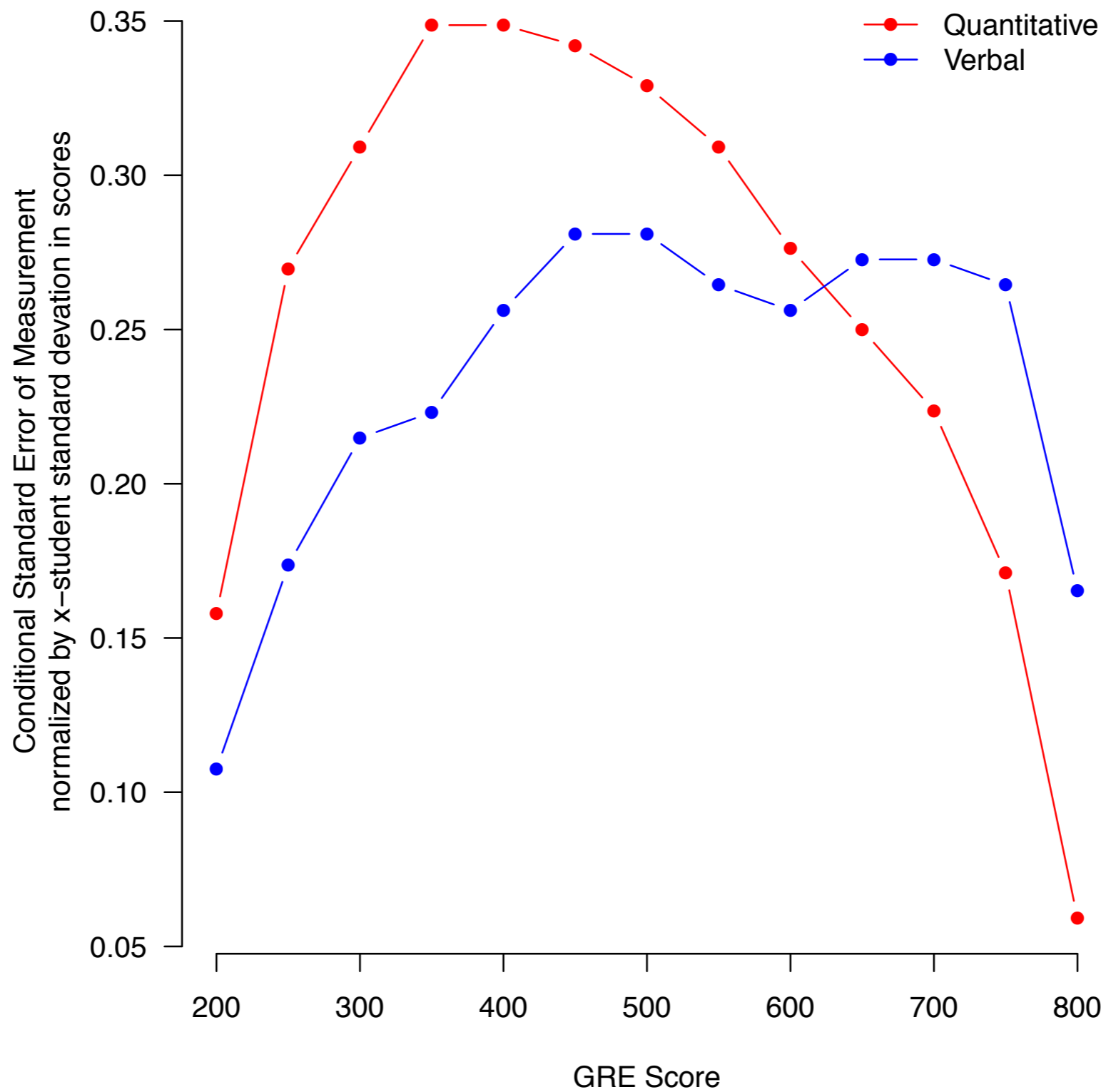
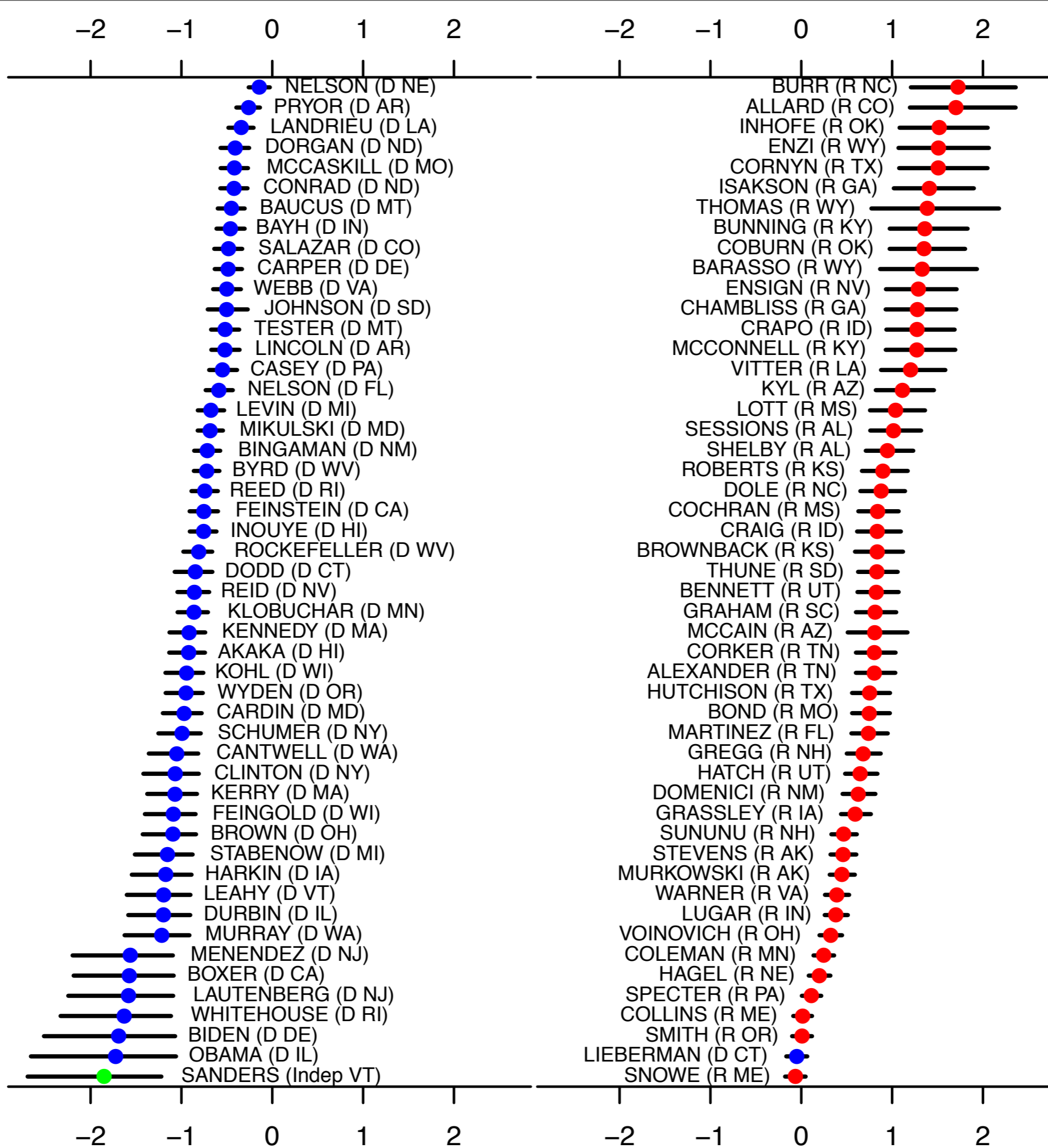


Table 6A: Conditional Standard Errors of Measurement at Selected Scores for General Test Measures*

Measure	200	250	300	350	400	450	500	550	600	650	700	750	800
Verbal	13	21	26	27	31	34	34	32	31	33	33	32	20
Quantitative	24	41	47	53	53	52	50	47	42	38	34	26	9

legislative ideal points

- 99 *National Journal* “key votes” in 2007
- Obama rated as “most liberal senator”
- Clinton & Jackman (2009; LSQ) re-fit data using 2P-IRT model
- Obama misses 1/3 of 99 key votes; Biden misses 26%; McCain misses 56%.



Measurement Error Standard Deviation

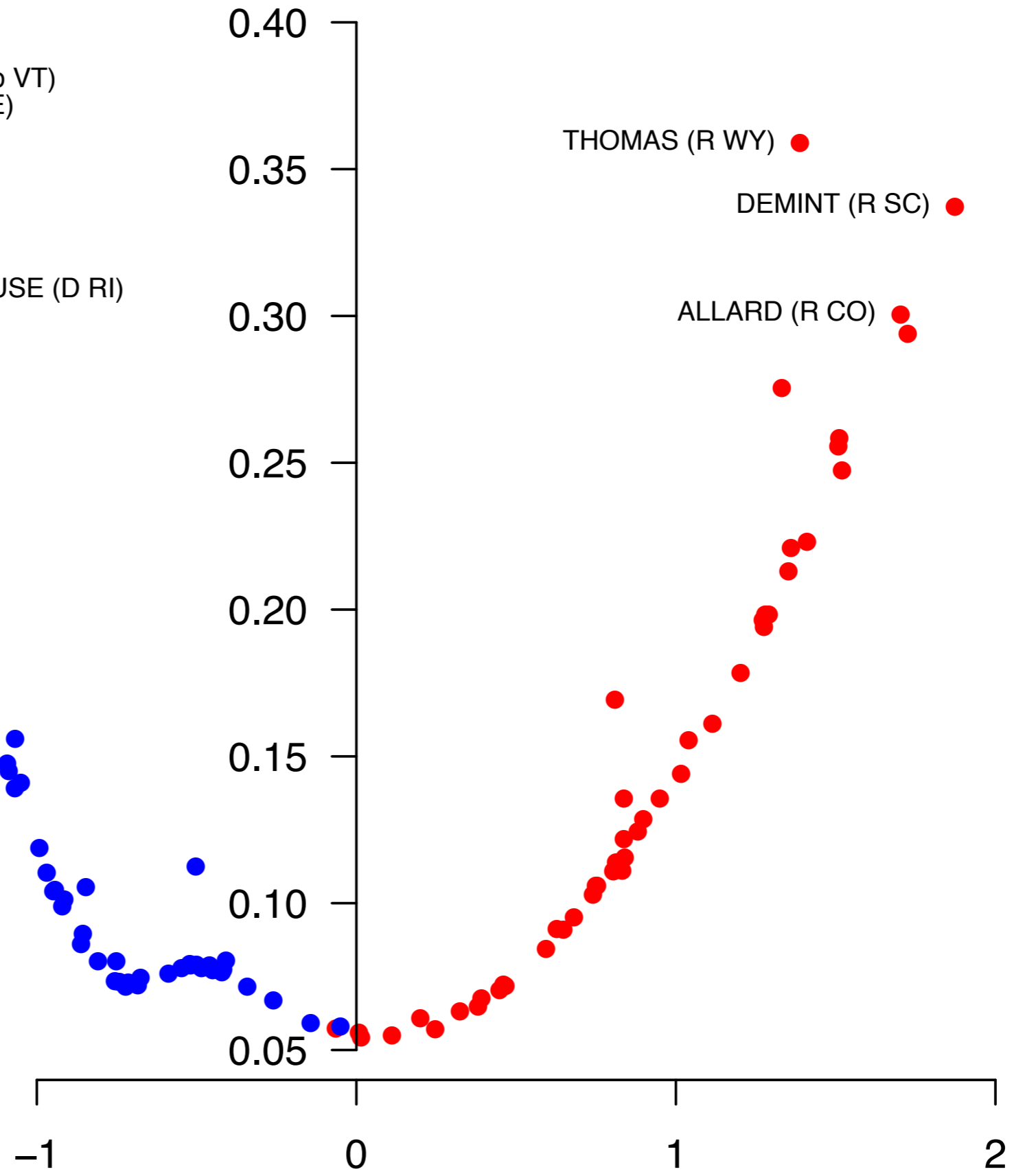
- OBAMA (D IL)
- SANDERS (Indep VT)
- BIDEN (D DE)

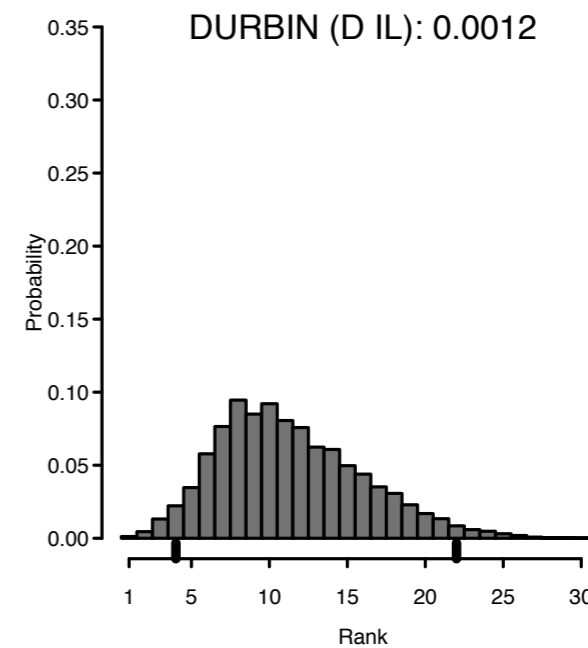
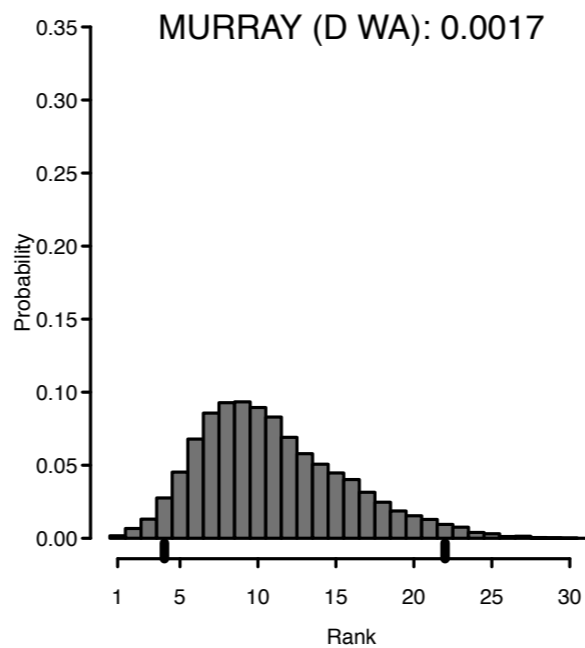
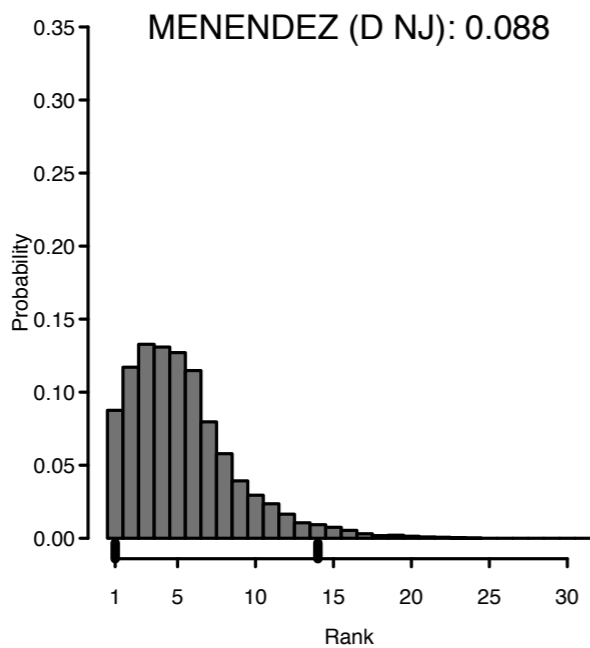
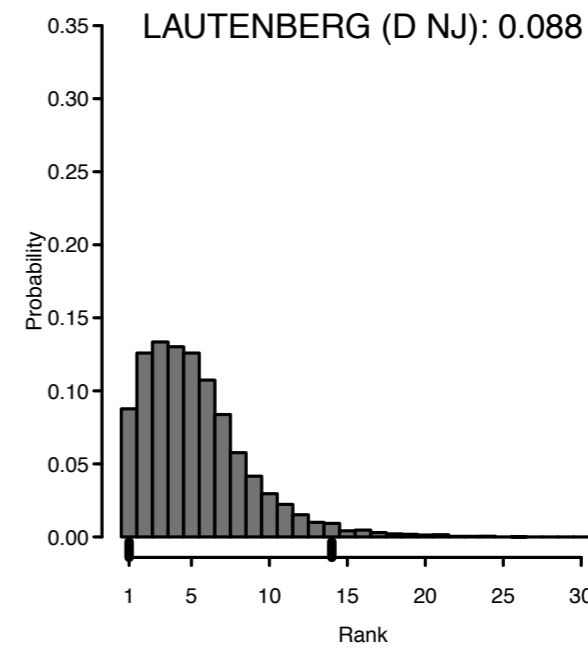
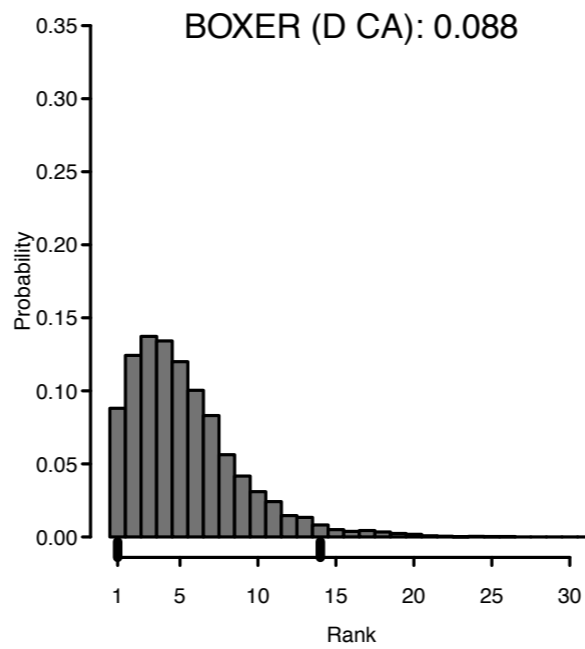
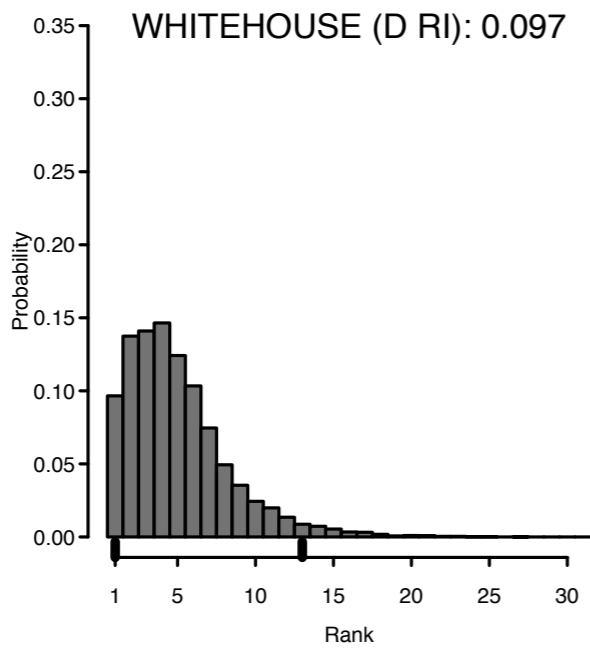
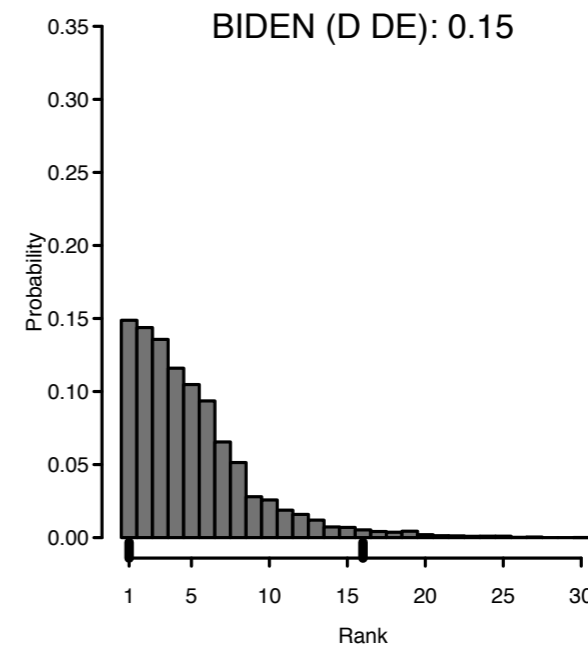
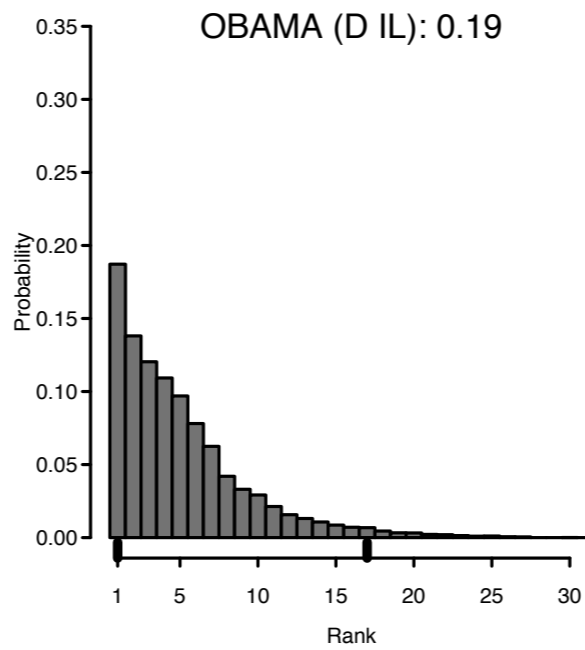
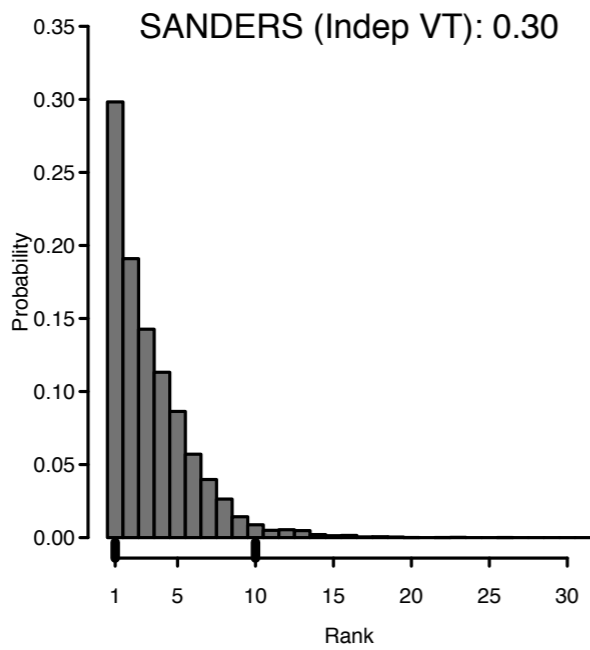
WHITEHOUSE (D RI)

THOMAS (R WY)

DEMINT (R SC)

ALLARD (R CO)





Polity IV

- 5 components, but not conditionally independent
- hard-coding across components

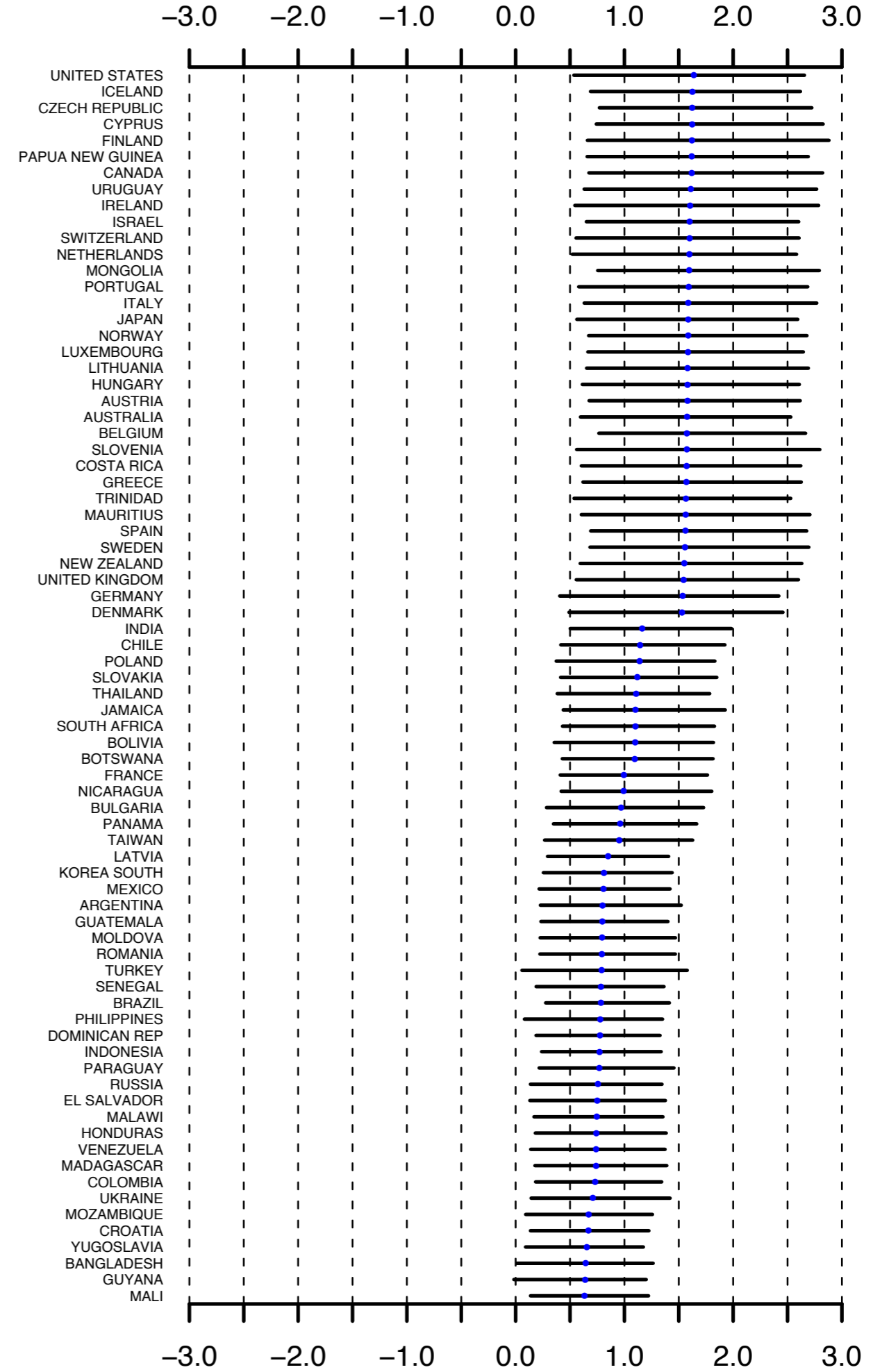
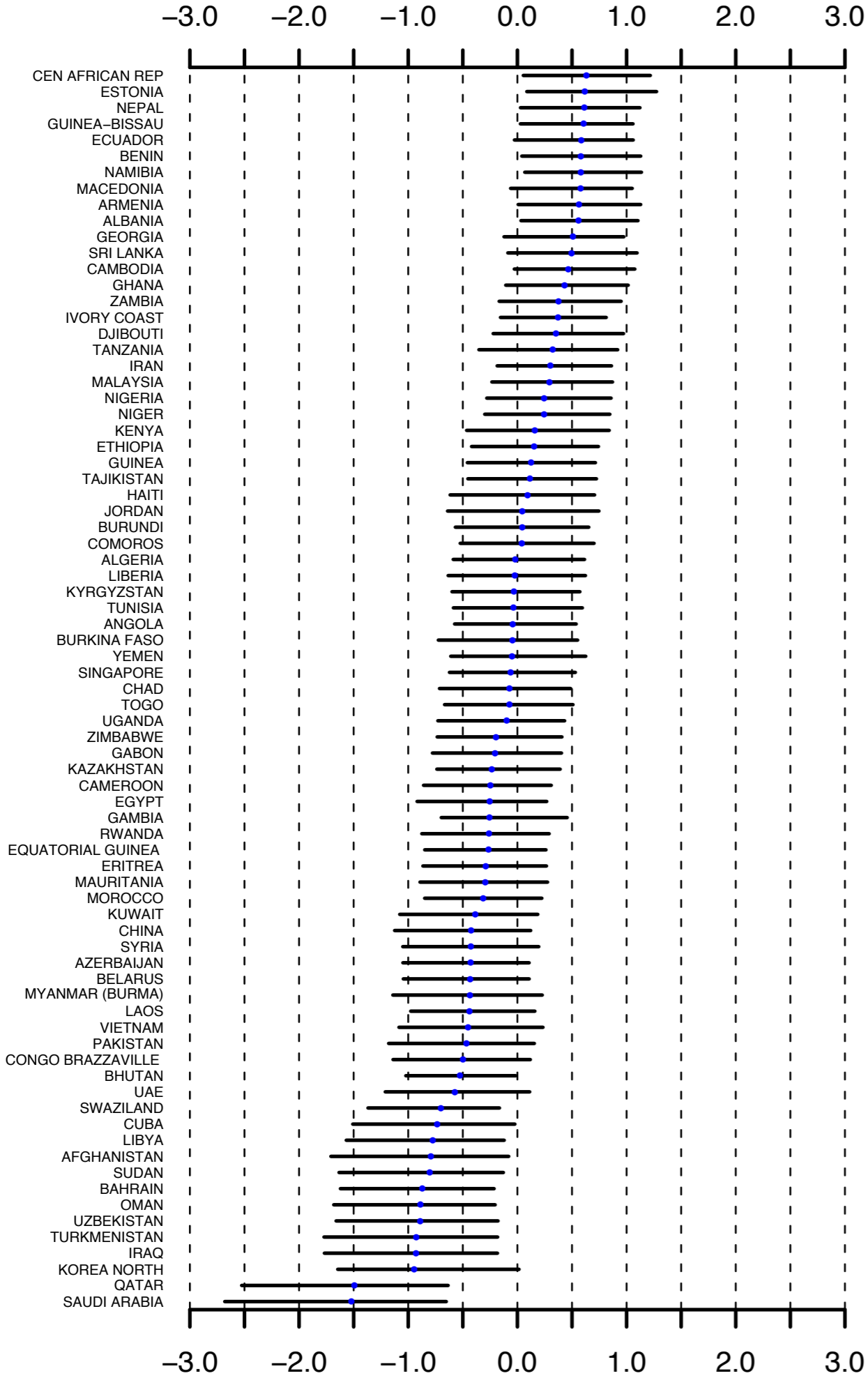
Polity IV indicators: local dependence

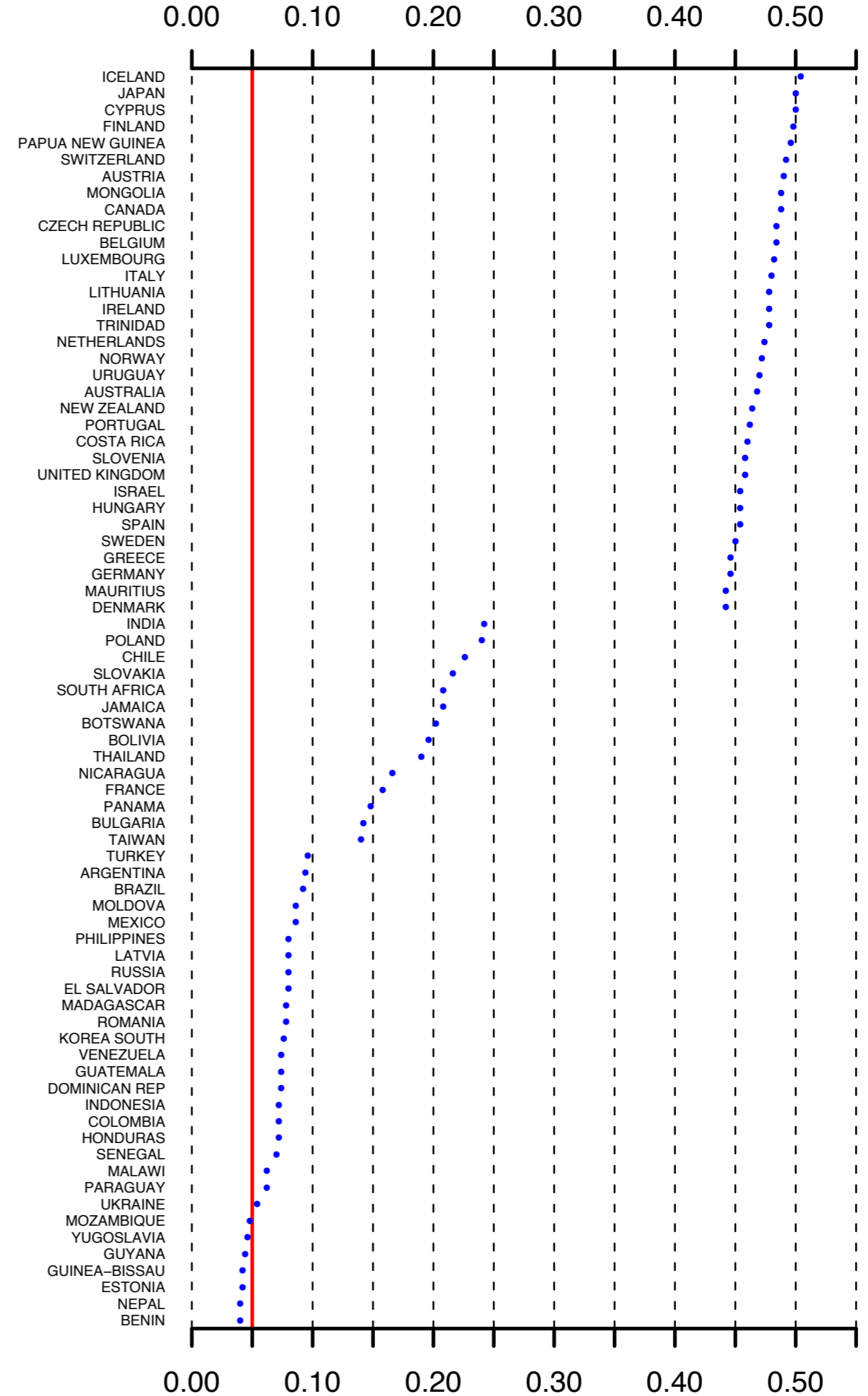
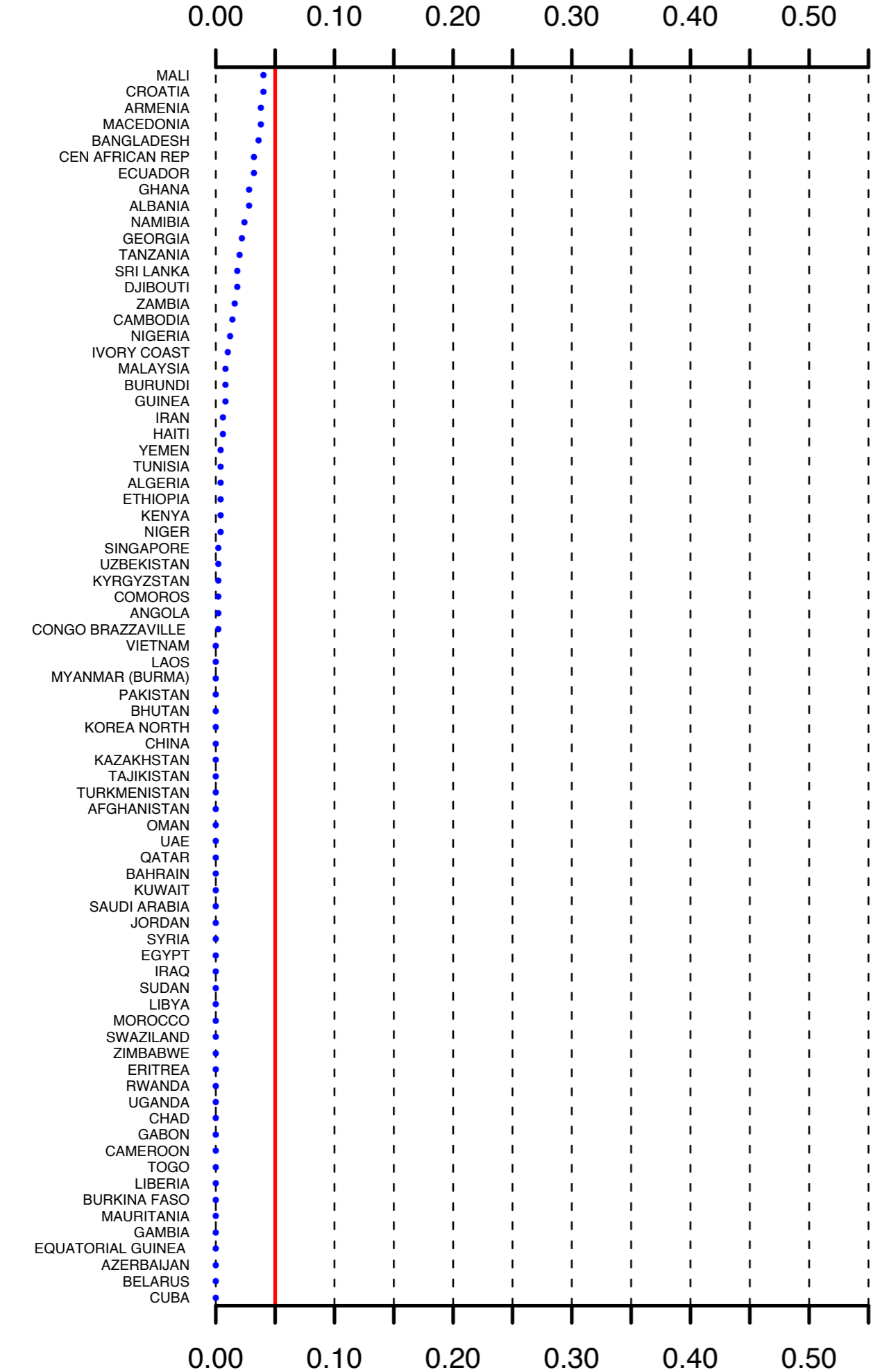
Xrcomp	Xropen					
	1	2	4 ^{NE}	0	3	4 ^E
1	2487	1175	3782	0	25	0
0	0	0	0	1600	0	0
2	0	0	0	0	130	667
3	0	0	0	0	0	4075

Parreg	Parcomp					
	1	2	0	3	4	5
4	3878	1811	0	0	0	0
3	0	299	0	3509	76	0
1	0	0	487	0	10	0
2	0	0	96	740	583	0
5	0	0	0	0	116	2336

IRT analysis of Polity IV indicators

- Treier & Jackman (2008; *AJPS*): 3 conditionally independent indicators; scale via “graded response” model
- recover estimates and 95% of scores on underlying dimension (“democracy”)





latent or observable?

- whenever possible, social & political theory should refer to observables
- is democracy necessarily latent?
- or “state capacity”, “resolve”, “threat”, “military professionalism”
- if a **valid** indicator of the phenomenon is available, then why not use it?
- avoid hand-wringing over measurement error, reliability; inferences “conditional on the indicator”

recommendations

- more data can never hurt; measurement modeling lets us bridge across multiple indicators across different sources
- “smarter” indicators: like the GREs, go get more data on the cases where we need it
- establish validity, perhaps especially **predictive validity** (“policy relevance”); cf construct validity (“political scientists think it is right”)
- “consult a professional” re scale construction

hierarchical model

data



y_{ij}

\sim

$$N(\lambda_{j1} + \lambda_{j2}\xi_i, \omega_j^2)$$

ξ_i

\sim

$$N(0, 1)$$

non-hierarchical prior

ξ_i

\sim

$$N(z'_i\gamma, \sigma^2)$$

hierarchical prior

latent variable



dynamic measurement model

$$y_{ijt} \sim N(\lambda_{j1} + \lambda_{j2}\xi_{it}, \omega_j^2)$$

$$\xi_1 \sim N(0, \sigma^2)$$

$$\xi_t \sim N(\xi_{t-1}, v^2), t = 2, \dots, T$$