Statistical Methods III: Spring 2013

Jonathan Wand

Stanford University

model select + inequalities

Outline

Campaign finance and interest groups

A puzzle:

Why do private interests fund public elections?

Classical theories:

1. spot market for favors/access (service-induced)
2. ideological or policy battles (position-induced)
3. consumption

Campaign finance and interest groups

- Investor Behavior:
  Austin-Smith (1995), Baron (1989,1994) , Chappell (1982) , Langbein (1996), Mebane (1999) , McCarty & Rothenberg (1996, 2000), Morton and Cameron (1992) , Denzau and Munger (1986) Grier and Munger (1986,1993) Hinich and Munger (1989), Snyder (1990,1991,1992) Snyder and Groseclose (1996), Stratmann (1992,1995), Wayman and Hall (1990) , Wawro (2000) , Wright (1989)
- Investor and Ideological
  Evans (1988) , Grenzke (1986) , Grossman and Helpman (1999) , Jacobson and Kernell (1981), Magee (2002), Welch (1978, 1980) , Wright (1985, 1990, 1996)
- Ideological
  Poole and Romer (1985), McCarty and Poole (1998), McCarty, Poole and Rosenthal (200), Bonica (2010)

Campaign finance and interest groups

Problem: little evidence of actual buying favors

- little or mixed effect on roll call votes
- also mixed in committee actions
- main evidence of impact have not been replicated/generalized

Response

- of course, favors will be hidden
- look for indirect evidence...

Investor model of PACs

Q: How would PACs allocate contributions to candidates across House races *if* buying favors?

Theory (Welch 1980; Baron 1989; Snyder 1990)

- contributions buy promises of favors from candidate
- receiving favors contingent on candidate winning
- perfectly competitive market for $ and favors

Q: How to test this theory?

Observe

- contributions to each candidate
- which candidate won

Investor model of PACs

Let's define some notation:

For each district i

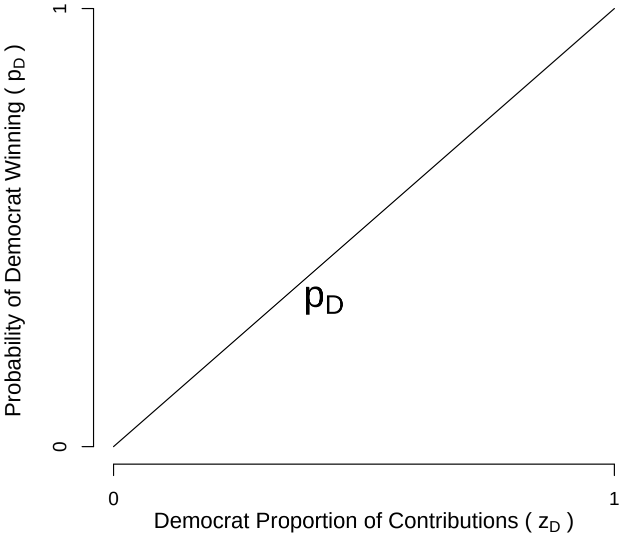
$z_{iD} \in [0,1]$: proportion \$ to Dem
$y_{iD} \in \{0,1\}$: indicator, Dem wins
$P(y_{iD} = 1)$: prob. Dem wins

Hypotheses generated by investor theories

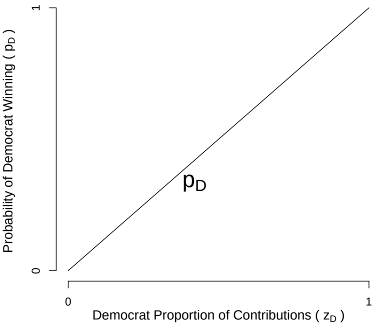
$H_0 : P_{iD} = z_{iD}$ (investor)
$H_A : P_{iD} \neq z_{iD}$ (non-investor)

(See Snyder, 1990; also Baron 1989; Welch 1980)

Investor model of PACs



Investor model of PACs



Model with
investors only
funding candidates

Hypotheses:

$H_0 : P_{iD} = z_{iD}$ (investor)

$H_A : P_{iD} \neq z_{iD}$ (non-investor)

Model I,

$$P(y_{iD} = 1) = \beta_0 + \beta_1 z_{iD}$$

$H_0 : \beta_0 = 0 \ \& \ \beta_0 = 1$ (investor)

$H_A : \beta_0 \neq 0 \text{ or } \beta_0 \neq 1$ (non-investor)

Model II,

$$P(y_{iD} = 1) = g(z_{iD})$$

$H_0 : z_{iD} = g(z_{iD})$ (investor)

$H_A : z_{iD} \neq g(z_{iD})$ (non-investor)

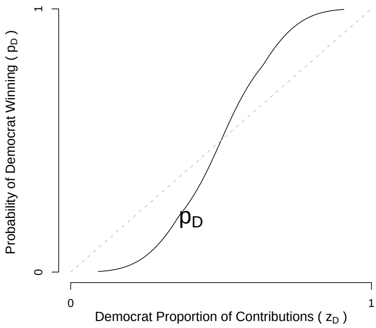
Partisan theory of PACs

Q: How would PACs allocate $ across House candidates *if* had preference over which party holds the majority?

Theory (Wand 2011; Wand 2013)

- gain benefits from preferred party if in majority (e.g., cartel theory)
- allocate $ maximize seat won by preferred party
- may also give as investors to less preferred party

Alternative: Partisan theory of PACs



Model with
investors and partisans
funding candidates

Model, with a flexible curve $f()$

$$P(y_{iD} = 1) = f(z_{iD})$$

H_0 is the same, but H_A is restricted,

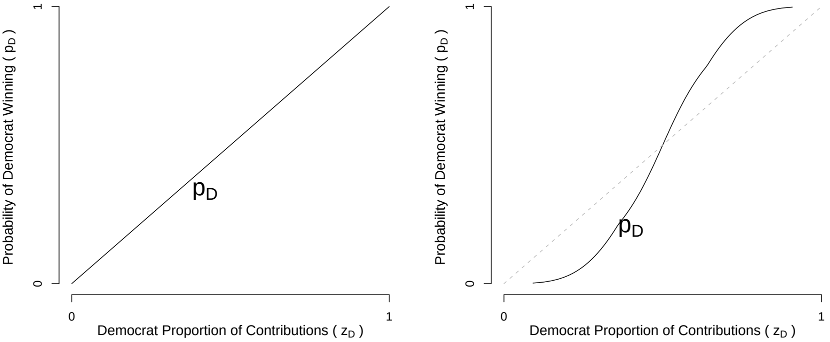
H_0 : $z_{iD} = f(z_{iD})$
(investor)

H_A : $f()$ symmetric S-shape
(partisan & investor)

What partisan theory doesn't tell us

- Steepness in middle of S-curve
- Sharpness of curve in tails

Comparing investor and partisan theories



Questions:

- How to estimate a function with shape constraints?
- How to compare such models?
- How does this change inference about PAC motives?

Motivations of PACs

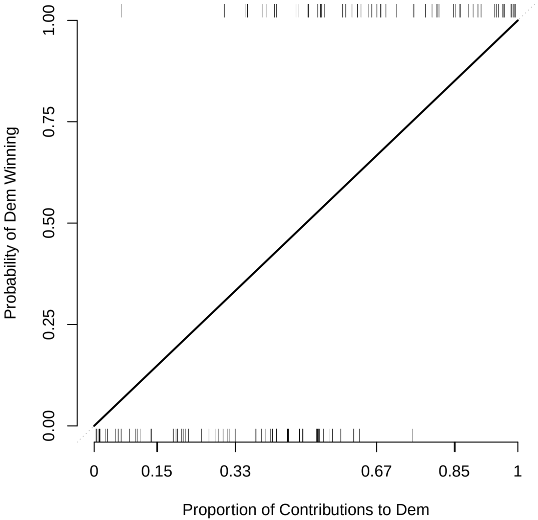
Snyder (1990, JPE)

- empirical test of investor theory of PACs
- universe of races: open seat contributions
  - avoids complications of seniority, etc
  - at cost of sample size
  - limiting case:
    if we we find investor here
    then everywhere
- universe of contributors: economic groups
  - Labor PACs
  - Corporate PACs
  - Trade/Health/Membership PACs

Open seats summary statistics

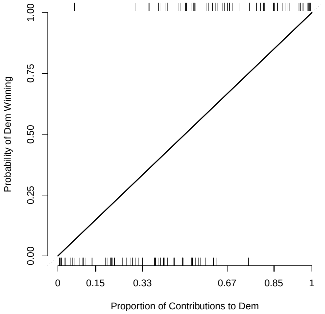
	Mean	Standard Deviation
DEM TOTAL CONTRIBUTIONS	425,877	328,428
REP TOTAL CONTRIBUTIONS	471,905	299,081
DEM INVESTOR CONTRIBS (X_{iD})	95,931	61,768
REP INVESTOR CONTRIBS (X_{iR})	102,928	75,461
DEM SHARE INVEST CONTRIBS (x_{iD})	.510	.295
DEM IDEOLOGICAL CONTRIBS	18,590	20,258
REP IDEOLOGICAL CONTRIBS	24,187	18,858
DEM INDIVIDUAL + CANDIDATE CONTRIBS	302,955	297,721
REP INDIVIDUAL + CANDIDATE CONTRIBS	319,057	232,488
DEM WIN*	.472	.501
PARTY STRENGTH†	−.013	.090

Investor theory

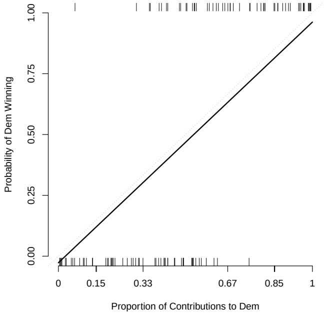


Probability winning by Proportion \$: $P(y_{iD} = 1) = z_{iD}$

Investor vs WLS



(a) Investor (0 parm)



(b) WLS (2 parm)

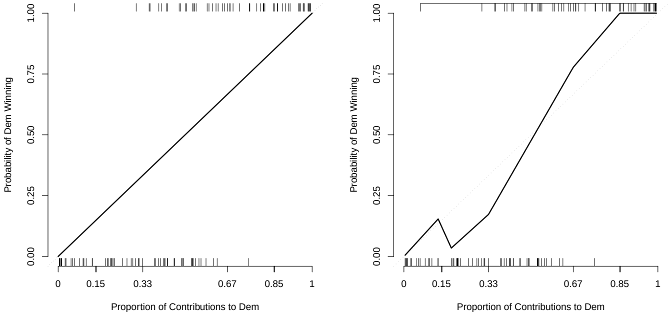
$$P(y_{iD} = 1) = z_{iD}$$

$$P(y_{iD} = 1) = -.027 + .989 z_{iD}$$

Comparing fit of investor and WLS model:

χ^2_2-statistic: 2.11; p-value (Prob $> \chi^2_2$): 0.35

Investor vs unrestricted curve



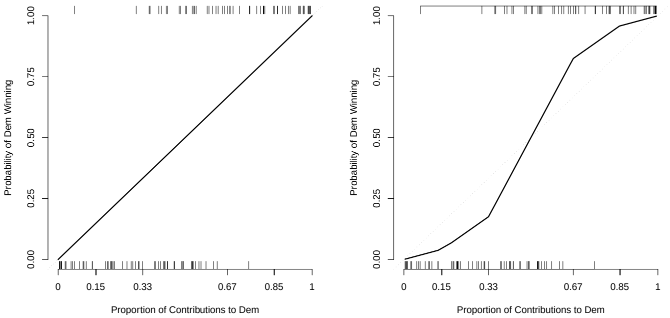
Investor (0 parm) vs unrestricted piecewise linear (8 parm)

$$P(y_i = 1) = z$$

$$P(y_i = 1) = \mathrm{f}(z)$$

χ^2_8-statistic: 9.14; p-value (Prob $> \chi^2_8$): 0.33

Investor vs partisan-mixture



Investor (0 parm) vs symmetric piecewise linear (2 parm)

$$P(y_i = 1) = x$$

$$P(y_i = 1) = \mathrm{g}(x)$$

χ^2_2-statistic: 5.18; p-value (Prob $> \tilde{\chi}^2_2$): 0.06

PAC motives: model comparisons

			$\Pr(\bar{\chi}^2 > c)$		
Model (j):	Parms m max	Log-lik. L_j	j vs Linear	j vs Dips	j vs Mono.
Linear Equality	0	-47.03			
w/ symmetric dips	1	-46.59	0.18		
Symmetric, monotonic	2	-44.44	0.06	0.03	
w/ knots at $(\frac{1}{3}, \frac{2}{3})$	3	-43.81	0.07	0.04	0.48
Unrestricted	6	-42.92	0.22	0.89	0.34
w/ knots at $(\frac{1}{3}, \frac{2}{3})$	8	-42.46	0.99	0.89	0.34

Note: $\bar{\chi}^2 = -2(L_{\text{row}} - L_{\text{column}})$

Model Confidence Set - Purpose

- Aim to find best model ...
- and all models which indistinguishable from best!
- This is the “Model Confidence Set” (MCS)
- Provides p-values for models with respect to MCS

Cf. Hansen, Lunde, and Nason (2011, Econometrica)

Model Confidence Set – Logic

- Sequentially test whether any models in a set are not among "best"
- If fail to reject, then stop declare set "best"
- Else, drop worst and repeat
- Since true "best" set is tested only once, correct size of test despite multiple comparisons

$$Pr(\mathit{TrueBestSet} \subset \mathit{EstimatedBestSet}) \geq 1 - \alpha$$

$$Pr(\mathcal{M}^* \subset \hat{\mathcal{M}}_{1-\alpha}) \geq 1 - \alpha$$

Model Confidence Set - Notation

- $\mathcal{M}$: set of all models
- $\mathcal{M}_0$: initial set of all models to test
- $\mathcal{M}^*$: true set of equally best models
- $\hat{\mathcal{M}}_{1-\alpha}$: the MCS

Model Confidence Set - Properties

As sample size grows large,

$$Pr(\mathcal{M}^* \subset \hat{\mathcal{M}}_{1-\alpha}) \geq 1 - \alpha$$

- If only one best then in limit

$$Pr(\mathcal{M}^* \subset \hat{\mathcal{M}}_{1-\alpha}) = 1$$

- If two best, α chance that at least 1 will be rejected
- Limit, all models not in $\mathcal{M}^*$ eliminated

Model Confidence Set - Components of test

1. A loss metric
   E.g., squared error from cross-validation or forecasting for model j
   $$L_{ji} = (y_i - \hat{y}_{i(j)})^2$$
   E.g., Expected KLIC...
2. Average loss for each model $\bar{L}_j$ from original sample
3. Average loss for each model $\bar{L}_{bj}$ in B boot-strapped samples
4. Centered $\eta_{bj} = \bar{L}_{bj} - \bar{L}_j$

Model comparison

Let,

$$Q(\beta_j) = -2L(\beta_j)$$

Classical,

$$Q(\hat{\beta}_j) - Q(\beta_{0j}) \sim \chi^2_k$$

Instead, estimate effective degrees of freedom,

1. *treat $\hat{\beta}_j$ as the population parameter*
2. *sample $Z_b^* = (Y_b^*, X_b^*)$*
3. *calculate $Q(Z_b^*, \hat{\beta}_j) - Q(Z_b^*, \hat{\beta}_{b,j}^*)$*

$$\hat{k}_j^* = B^{-1} \sum Q(Z_b^*, \hat{\beta}_j) - Q(Z_b^*, \hat{\beta}_{b,j}^*)$$

This is *also* critical value for LRT with inequality constraints

MCS

Test statistic for set:

$$T_{\mathcal{M}} = \max_{i,j \in \mathcal{M}} \mid [(Q(\hat{\beta}_i) + k_i^*) - (Q(\hat{\beta}_j) + k_j^*)]$$

if big, reject null that all models in set are "best".

The joint distribution for m models of

$$\{Q(\hat{\beta}_i) + k_i^* - Q(\beta_{0i}), ..., Q(\hat{\beta}_m) + k_j^* - Q(\beta_{0m})\}$$

is estimated by a bootstrap, taking the differences,

$$\{Q_b(\hat{\beta}^*_{b,i}) + k_i^* - Q_b(\hat{\beta}_i), ..., Q_b(\hat{\beta}^*_{b,m}) + k_j^* - Q_b(\hat{\beta}_m)\}$$

we have distribution of $T_{\mathcal{M}}$ under null of all models "best"

Model Confidence Set - algorithm

Pre-process / calculation

(a) Get MLE $\hat{\beta}$, gives fit $Q(\hat{\beta}_j)$

(b) Bootstrap conditional on each model being "best", gives k_j and fit $Q_b(\hat{\beta}_{b,j})$

Begin with all models as candidates in $\mathcal{M}_i$, $i = 0$

1. Given $\mathcal{M}_i$, calc $T_{\mathcal{M}_i}$ (observed differences) a and bootstrap distribution $T_{b,\mathcal{M}_i}$
2. Calculate

$$\hat{p} = B^{-1} \sum_{b}^{B} I(T_{b,\mathcal{M}_i} > T_{\mathcal{M}_i})$$

3. If $\hat{p} > \alpha$ stop
4. If $\hat{p} \leq \alpha$ eliminate model with worst fit
5. Return to step 1

Monte Carlo: Frequency of finding best model

	Frequency selecting true model			Rejecting	
True shape	$\hat{\mathcal{M}}_{1-\alpha}$	AIC^*	AIC	linearity	monotonicity
constant	97.2	82.8	81.2	4.0	4.8
Linear	96.4	77.6	78.0	4.0	0.0
Quadratic	96.0	93.2	93.2	100.0	100.0
Unimodal 0	95.2	83.6	0.0	100.0	100.0
Unimodal 1	95.2	81.2	0.0	19.6	0.0
Unimodal 2	95.2	81.2	0.0	12.8	0.0
Oscillating	100.0	100.0	100.0	99.2	0.8

Set of models:

constant, monotonic, linear, quadratic, unimodal, unrestricted.

Test size is $\alpha = 0.05$

Sample size of each MC is $N = 500$,

Number of simulations per model is $B = 250$.

An example

Q: Connection between a country's democracy score and child mortality rate?

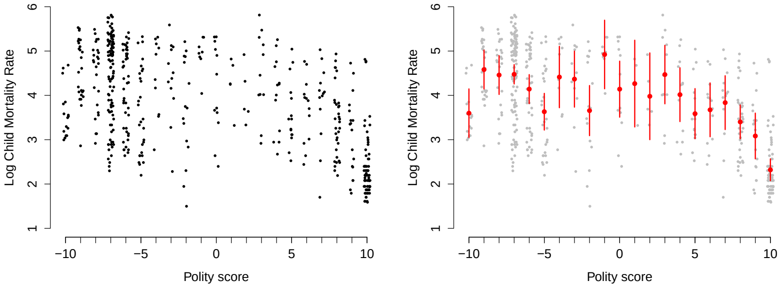
Two camps,

- Yes: Przeworski et al. (2000), BdM (2003), many more
- No: Ross "Is Democracy Good for the Poor?" (AJPS 2006)

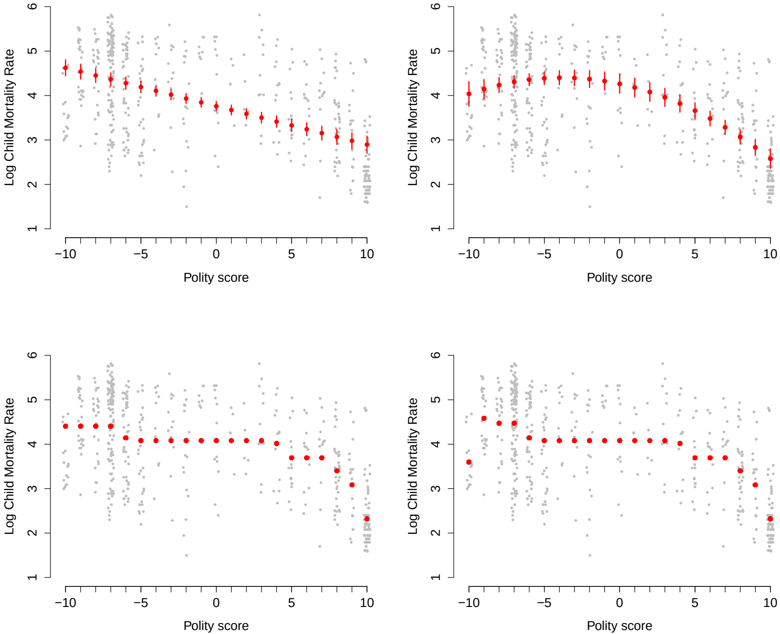
Is Democracy Good for the Poor?

	1	2	3	4
	LDV Only	LDV Only	LDV & Period Dummies	FE & Period Dummies
INCOME	−.13***	−.13***	−.15***	−.18***
	(.028)	(.025)	(.024)	(.038)
HIV	.035	.044***	.1***	.22***
	(.024)	(.016)	(.011)	(.033)
POP DENSITY	−.023***	−.022**	−.02***	−.021
	(.006)	(.0052)	(.0051)	(.016)
GROWTH	−.0046	−.0048	−.0071*	−.0033
	(.0026)	(.0023)	(.0023)	(.0035)
POLITY	–	−.0015	−.0025	−.00096
		(.0011)	(.0012)	(.003)
DEMOCRATIC YEARS	–	–	–	–
Observations	1176	1122	1122	1122
Countries	168	168	168	168

Polity scores and child mortality



Polity scores and child mortality



Challenges of isotonic, unimodal

- do you know which direction for monotonic? else must determine direction (2 choices)
- do you know where is peak? else must search.

Fit statistics for polynomial models of child mortality and polity

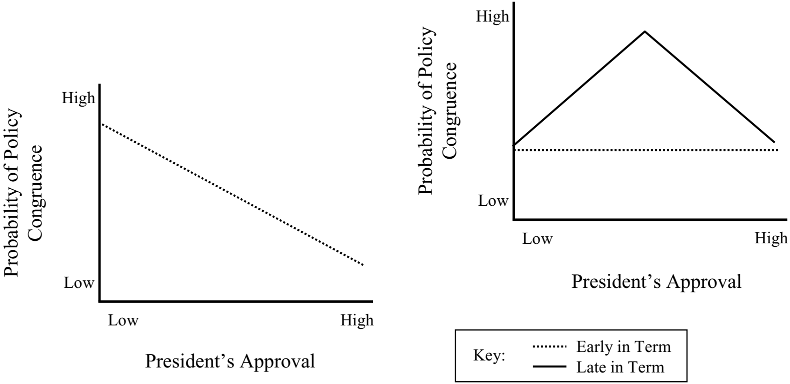
	Q	AIC	AIC*	k	k*
Constant	1575.5	1579.5	1578.5	2	1.5
Monotonic	1575.5	1582.6	1583.7	3.5	4.1
Linear	1314.5	1320.5	1320.5	3	3.0
Quadratic	1253.2	1261.2	1261.9	4	4.4
*Unimodal	1198.4	1223.9	1229.8	12.7	15.7
Unrestricted	1184.3	1228.3	1230.6	22	23.1

Sequence of MCS tests for full set of models of child mortality and polity

j	H_{0,M_k} p	MCS p	$\hat{\mathcal{M}}_j$	Model to eliminate
1	< 0.000	< 0.000	..., Unimodal, Unrestricted	Mono
2	< 0.000	< 0.000	..., Unimodal, Unrestricted	Constant
3	< 0.000	< 0.000	..., Unimodal, Unrestricted	Linear
4	< 0.000	< 0.000	..., Unimodal, Unrestricted	Quadratic
5	0.084	0.084	Unimodal, Unrestricted	(none)

Exec. policy (Canes-Wrone and Shotts 2004)

Level of Presidential policy congruence with public opinion



- model offers predictions about shape of congruence curve in late term
- but not about location of peak, or slopes

Beyond average differences ... and arbitrary flexibleness

- Have a theory, ideally more than one
- "Make your theories elaborate" (Fisher / Cochran 1965):
  - when constructing a causal hypothesis one should envisage as many different consequences of its truth as possible
  - if a hypothesis predicts that y will increase steadily as the causal variable z increases, a study with at least three levels of z gives a more comprehensive check than one with two levels
  - i.e, check shape! not just average change
- And check against omnibus alternatives
  but be clear this is for idea generation and robustness!

GLM extensions

Simple extension, back-fitting

- Bachetti (1989) Additive Isotonic Models
- Geyer, Charles J. (1991) Constrained Maximum Likelihood in Logistic

However, do you want to...

- non-linear transformation of link often unappealing, distorts shape!
- Wand (2011) uses (constrained) spline to fit binary choice

Multivariate shapes

- rather than additive (cf Stout 2011)

Testing theories based on shapes

Design: no less important here than in RCM

- case selection

  minimizing confounders

  Eg., theories of campaign finance and “open seat” races

- selection of a test / distance-metric

  identifying unique and invariant implications from theory

  E.g., agenda theories hinge on status quo locations of (potential) proposals

- sensitivity analysis: bounds from theory and data

  E.g., what (implausible) distribution of SQ could make agenda theories observationally equivalent