Issues and Challenges in Estimating Political Preferences from Text

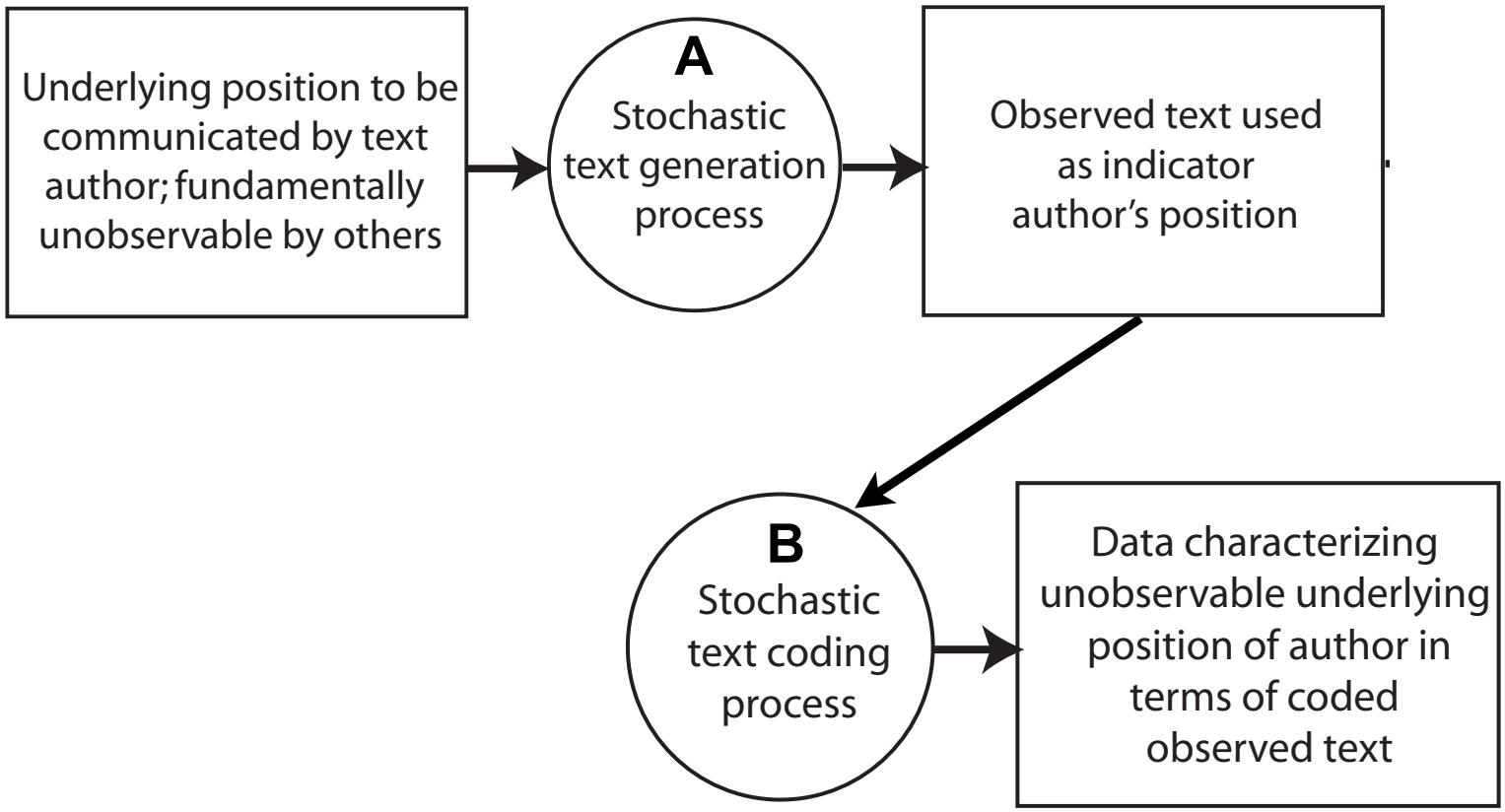
Kenneth Benoit
Trinity College, Dublin

April 2, 2009

Text as data

- Text offers huge, largely untapped possibilities to inform us about the preferences of political actors
- Text must be reduced to summary, quantitative information in order to be useful as data on estimating actors' positions
- Good quantitative information always comes with associated estimates of uncertainty
- Another key concern is reliability: Could another researcher achieve similar results using the text as data?
- Note: based on research with Slava Mikhaylov and Michael Laver

From positions to coded text: a stochastic process



The CMP: Brief overview

- A few of you may have heard of it?
- Covers 3,000+ party programmes from 1948-2000, 650+ parties, 52 countries, three books; very widely used in best comparative research
- 56 individual categories, plus uncoded; represented as percentages of total manifesto text
- Lengths vary widely, from 5 to 5,000 "quasi-sentences"
- Many quasi-sentences are VERY short example: "We are all in it together." (coded as 606: Social Harmony: Positive)
- Many categories known to be especially error-prone (e.g. 305: Political Authority)
- Most commonly used quantity is 26-category additive left-right scale known as "Rile"

The "Dublin papers"

- Introduce and demonstrate that observed political texts are generated by a stochastic process of authorship, from unobservable policy positions to observed text
- Capitalize on a basic intuition: More information increases our confidence — we are more certain about actors' positions estimated from longer texts — and apply this to the CMP
- Test the reliability of the CMP coding scheme using pilot coding experiments, to see if results can be reproduced – since observed texts are converted to data through a stochastic coding process
- Focus on a new scaling method for left-right positions from coded units
- (not done yet) Test whether unitization would work with less subjective units (e.g. natural instead of "quasi-" sentences

Estimating uncertainty through simulation: Bootstrapping the texts

- We can simulate stochastic manifesto generation by bootstrapping from quasi-sentence database and reconstructing all CMP scores

  (*bootstrapping* refers to repeated resampling of sentences with replacement)

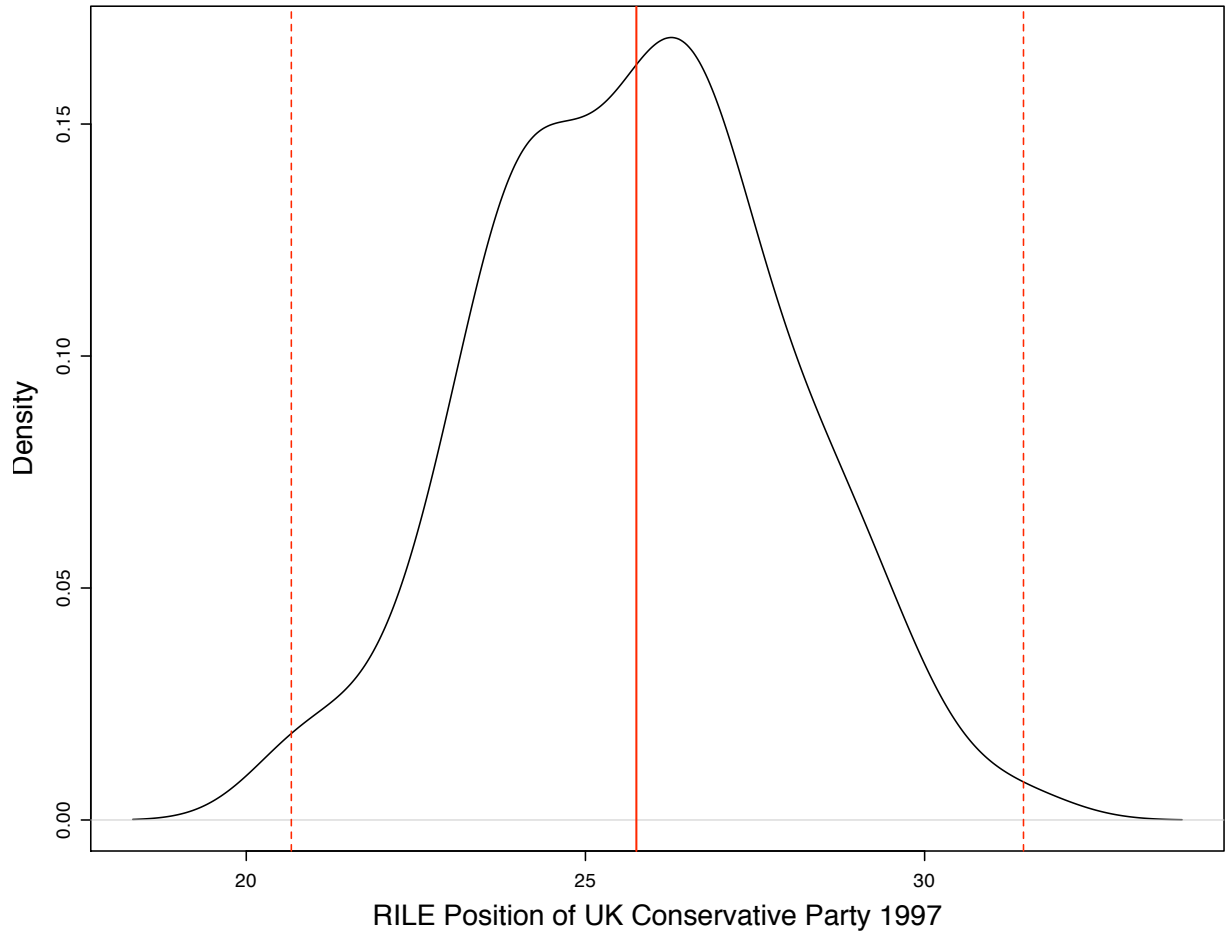
- Robust in the absence of parametric assumptions

- A very flexible approach that allows for the introduction of additional stochastic elements – such as variable text length

- Allows direct simulation of error for additive scales - unlike analytical solution

Estimating error result: Benoit, Laver and Mikhaylov (2009)

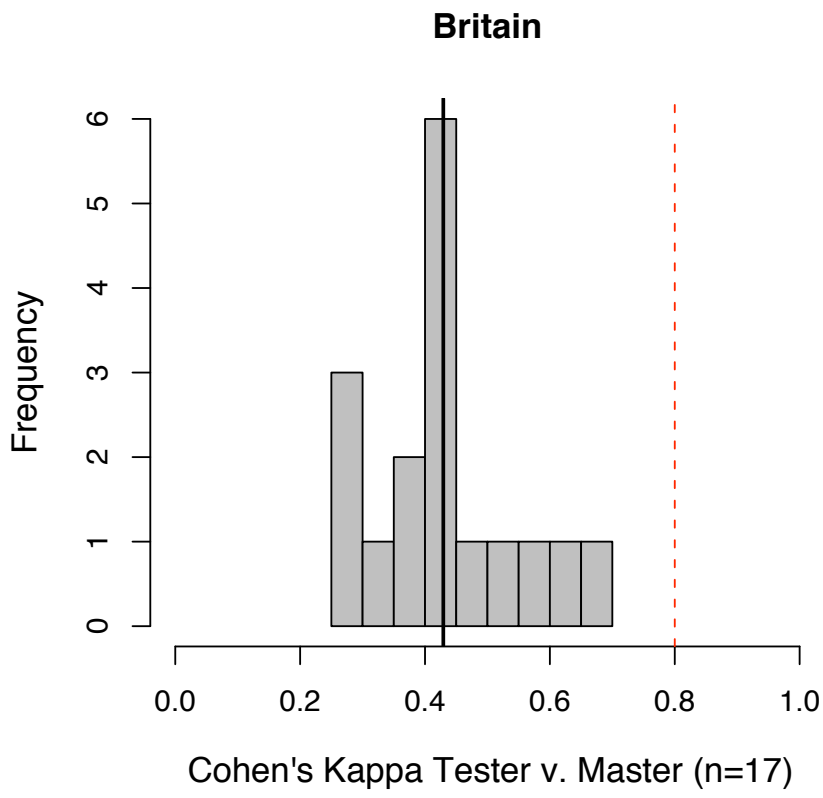
Example: British Convervative party: The CMP reported "Rile" value is 25.7, but 95% confidence interval is [20.7, 31.4]

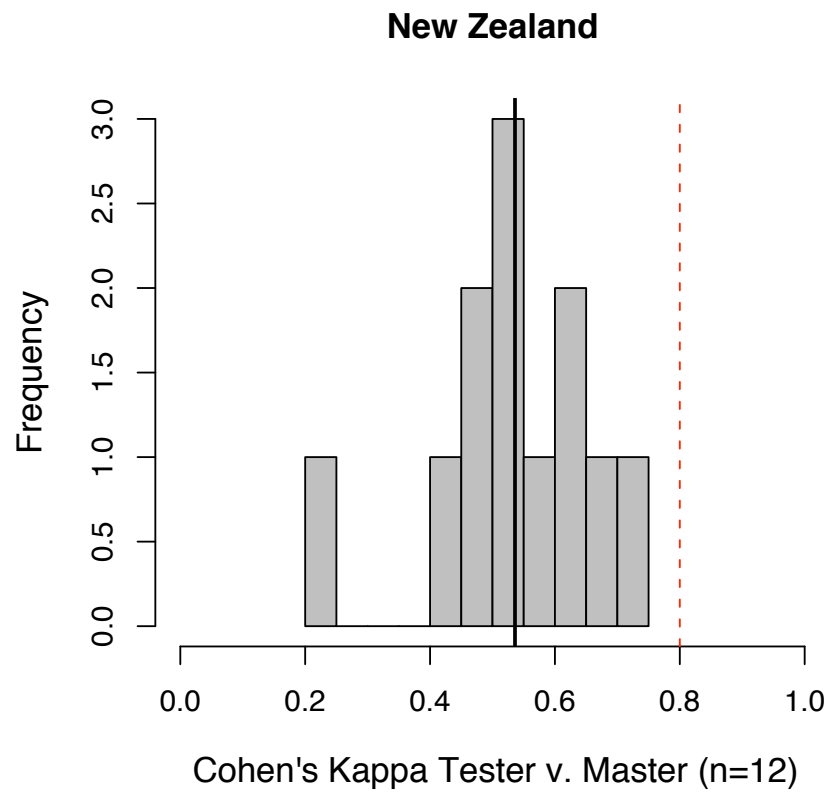


Testing coder reliability: an experiment

- Experiment: Asked coders to assign categories to two pre-unitized manifestos
  - UK Liberal/SDP Alliance 1983
  - New Zealand National Party 1972
- Both were coded "officially" in the CMP training manual
- Used measures of agreement and misclassification to assess reliability
- We recruited trained and experienced coders to take the test
- We discarded the bottom worst set of results to make the test as fair as possible

Reliability Results





Scaling analysis: overview

- Target: Estimating policy positions of parties on multiple, separable dimensions of policy (e.g. economic, social, environment, EU, etc.)
- Data source: Hand-coded manifestos where texts are divided into sub-units and each sub-unit is classified into a category
  - Hand-coded text remains the simplest and most common form of textual data on political actors
  - The process also generalizes to other non-ordered (automated) classification methods
  - Specific source: the Comparative Manifesto Project (CMP)
- Issue: How to construct continuous *scales* of policy positions from non-ordered category counts
- Solution: Fix the scaling method, demonstrate how ours is better, and apply it to the existing data on $> 3,000$ manifestos

Scaling Results for Dummies

- The king of all policy position datasets, the Comparative Manifesto Project, scales policy positions as absolute porportional difference, measured by proportion of "Right" mentions less proportion of "Left" mentions: $\frac{(R-L)}{N}$
- This scale works OK for "Rile" but poorly for everything else
- The alternative is to scale $\frac{(R-L)}{(R+L)}$ but this is *even worse* for everything except Rile
- Our better mousetrap is to scale position as $\log \frac{R}{L}$ and this works great for everything!
- By applying to confrontational category pairs from the CMP, we can provide about 14 new specific dimensions of policy never before really used from the dataset

The payoff: More scales than ever before

- Old scales will perform better:
  - CMP's left-right "Rile" scale (CMP)
  - Planned v. market economy (CMP)
  - Welfare and social security (CMP)
  - Social liberalism (Benoit and Laver 2007)
  - Pro-/Anti-EU (CMP)
- A dozen "new" confrontational single-pair scales:
  - Foreign alliances ($L = 101, R = 102$)
  - Militarism ($L = 101, R = 102$)
  - Internationalism ($L = 107, R = 109$)
  - Constitionalism ($L = 203, R = 204$)
  - Decentralisation ($L = 301, R = 302$)
  - Protectionism ($L = 406, R = 407$)
  - Keynesian Policy ($L = 409, R = 414$)
  - Nationalism ($L = 601, R = 602$)
  - Traditional Morality ($L = 603, R = 604$)
  - Multiculturalism ($L = 607, R = 608$)
  - Labour policy ($L = 701, R = 702$)
- And some brand spanking new ones we propose in the paper
  - Environmental protection v. growth economy
  - Free-market economy
  - State provision of social services

CMP's "saliency" scaling of position from category counts

- Doctrinaire "saliency theory" (Budge 1994) states that parties will only take one side of any issue, and distinguish positions through differing relative emphasis
- Strictly interpreted, this means we only need one-sided issue categories for most issues, such as "501: Environmental Protection: Positive" since no party will advocate trashing the environment
- But since not even the CMP group believes this, many scales are in fact opposite pairs. Example: "406 Protectionism: Positive" and "407 Protectionism: Negative"
- The CMP suggests taking the %603 - %604 to measure position – what we refer to as the absolute proportional emphasis approach
- But this makes the scale sensitive to non-protectionism-related text, and insenstive to relative changes when total mentions of either category is small. Also we never get near the theoretical (-100, 100) endpoints

Alternative scale: The relative proportional approach

- Condition the difference on the sum of the categories (Kim an Fording 2002; Laver and Garry 2000): $\frac{R-L}{R+L}$
- This makes the scale insensitive to irrelevant content, and also uses much more of the scale including the endpoints
- The problem is that this is hyper-sensitive in the middle range
- Like the saliency scaling approach, this also imples a linear change in position with additional content – an assumption that lacks practical and linguistic justification
  - This is because position does not increase linearly with repetition
  - This is because position does not increase linearly with repetition
  - This is because position does not increase linearly with repetition
  - This is because position does not increase linearly with repetition
  - This is because position does not increase linearly with repetition
  - This is because position does not increase linearly with repetition
  - This is because position does not increase linearly with repetition
  - This is because position does not increase linearly with repetition

What, you're not conviced yet?

- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition
- Position does not increase linearly with repetition

Our better scale: empirical logit

- Scale defined as: $P^{(L)} = \log \frac{R}{L}$ where R and L are counts of "right" and "left" text units, respectively
- This focuses attention the *balance* or true *relative* emphasis of L and R
- Since the counts on a given issue must be either R or L, the counts have a natural log ratio form
- Has a basis in linguistic theory (see previous slide!)
- Has a basis in psychophysical theory (Weber-Fechner Law)
- Consistent with "text as data" parameterizations of unobservable policy position θ generating counts of text (e.g. Elff 2008, Monroe, Colaresi and Quinn 2008, Slapin and Proksch 2008)

More precisely

Absolute proportion (saliency)

$$P^{(S)} = \frac{R - L}{N}. \tag{1}$$

Relative proportion ("ratio")

$$P^{(R)} = \frac{R - L}{R + L} \tag{2}$$

logit

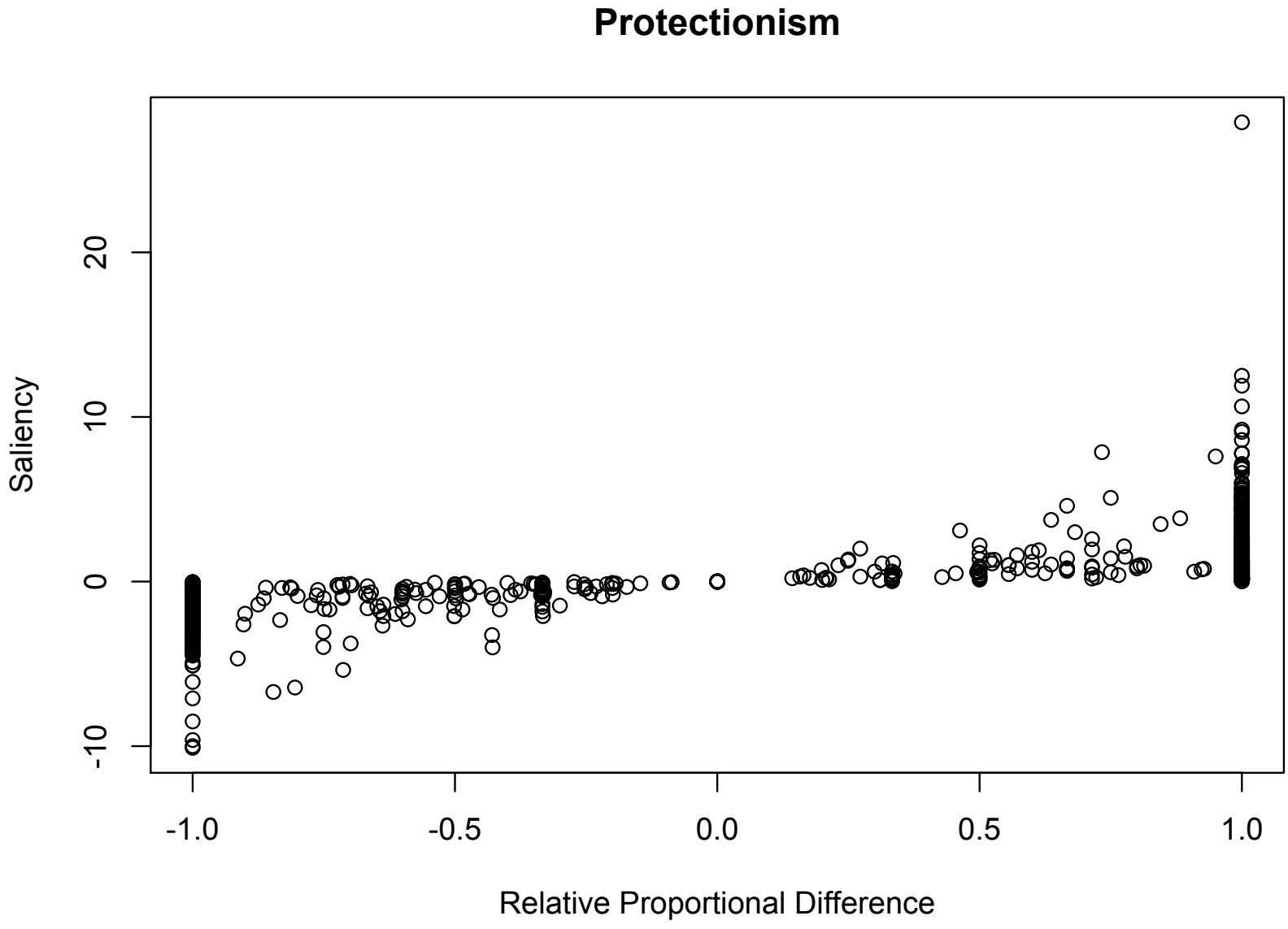
$$P^{(R)} = \log \frac{R + .5}{L + .5} \tag{3}$$

logit for indexes

$$P^{(L)}_{\text{index}} = \log \left[\sum_{j=1}^{J} R_j \Big/ \sum_{k=1}^{K} L_k \right]$$

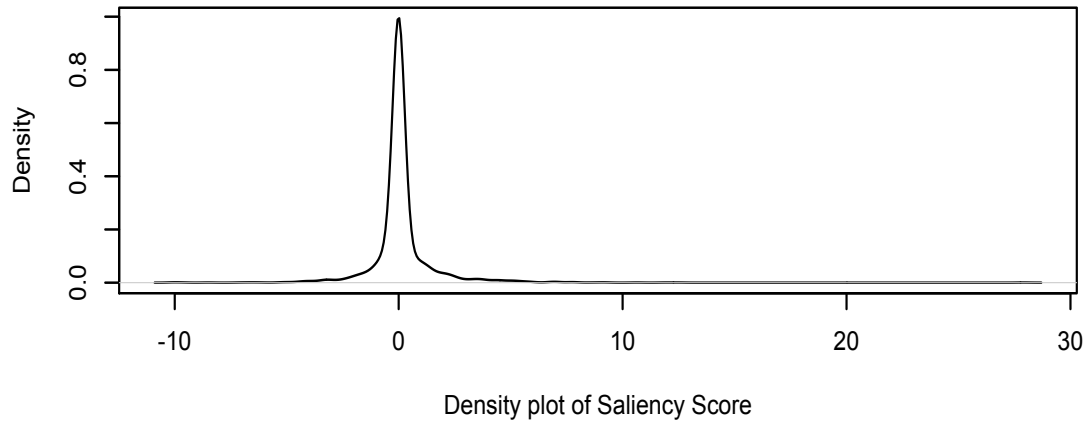
and now for some great pictures from scaling results

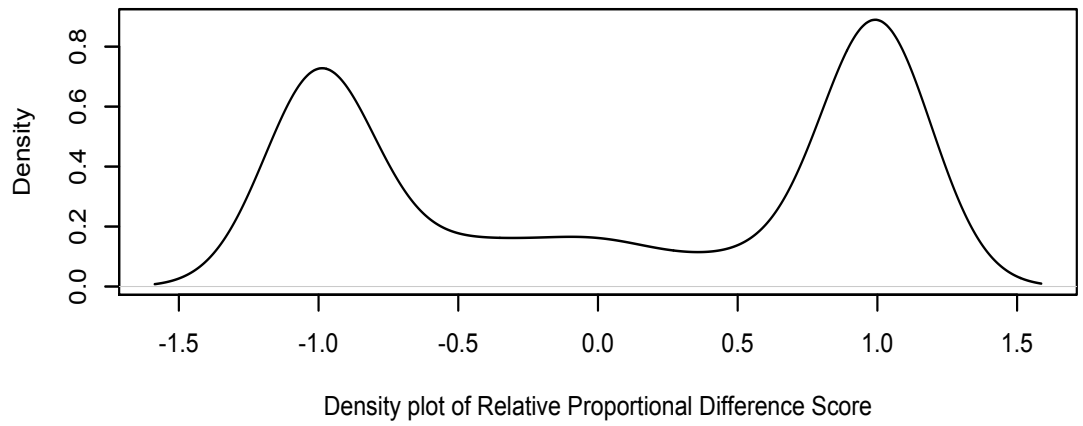
Comparing scales: $P^{(S)}$ v. $P^{(R)}$

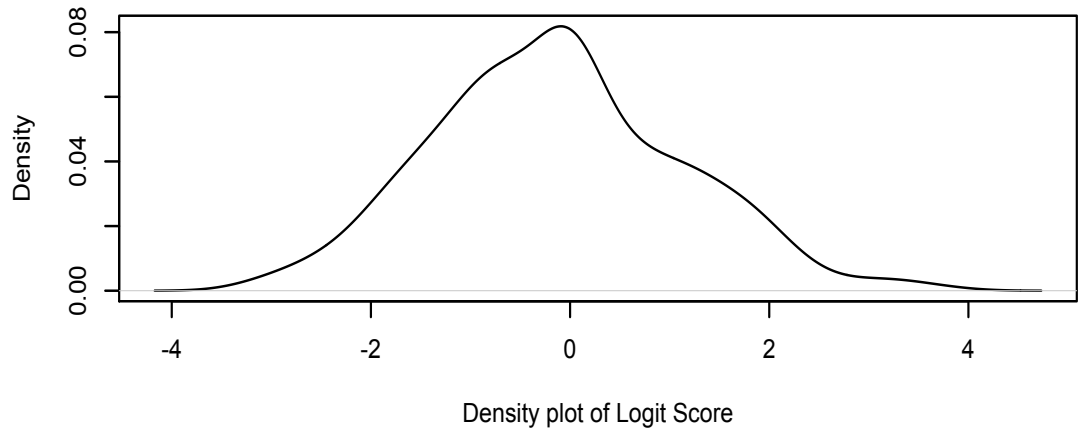


Comparing scales

Protectionism distributions





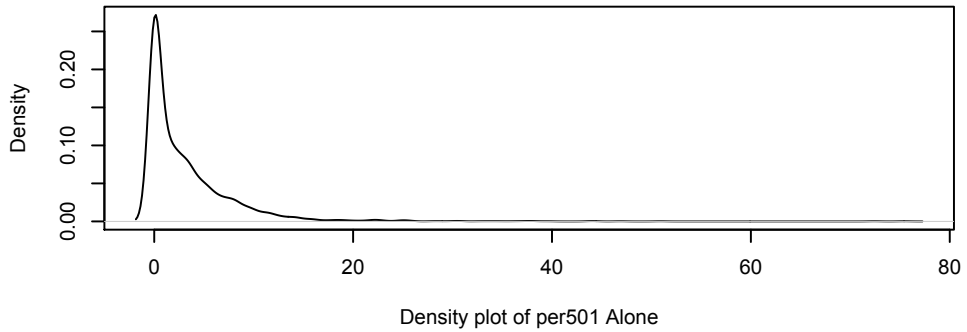


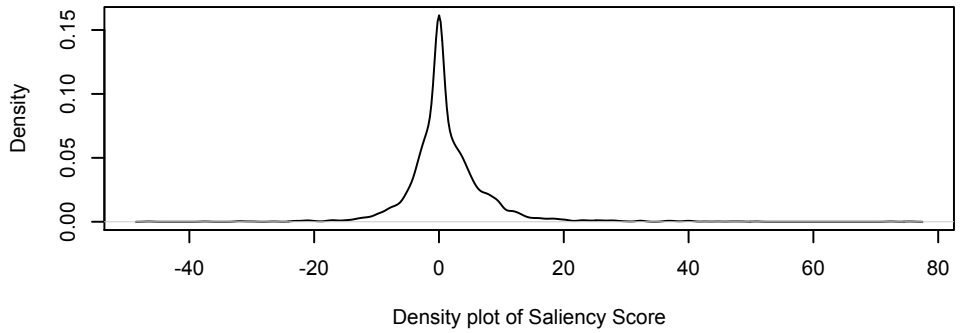
Comparing scales

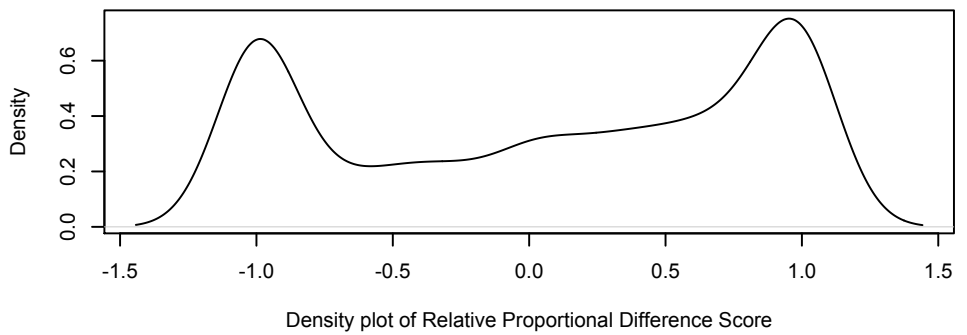
Environmental distributions

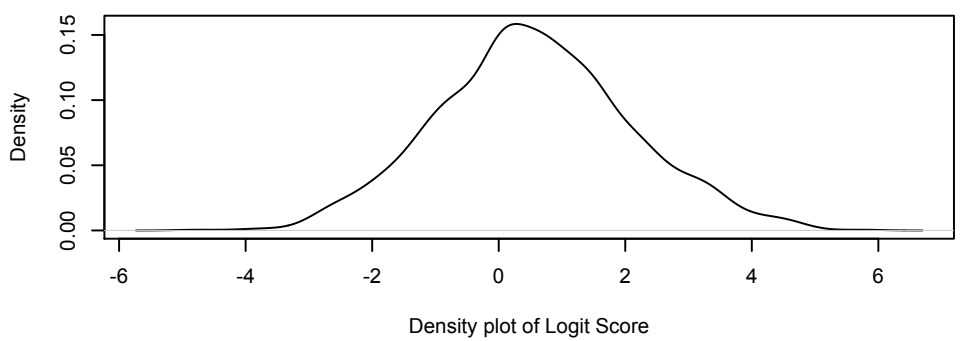
$L_{Env} = N_{401}+$ (Env'l Protection: $+$)
N_{416} (Anti-Growth: $+$)

$L_{Env} = N_{410}$ (Productivity: Positive)



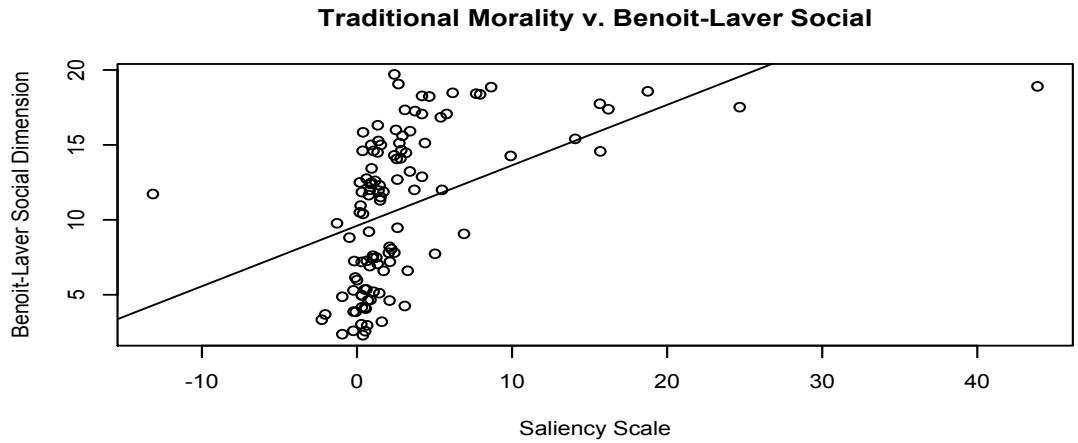


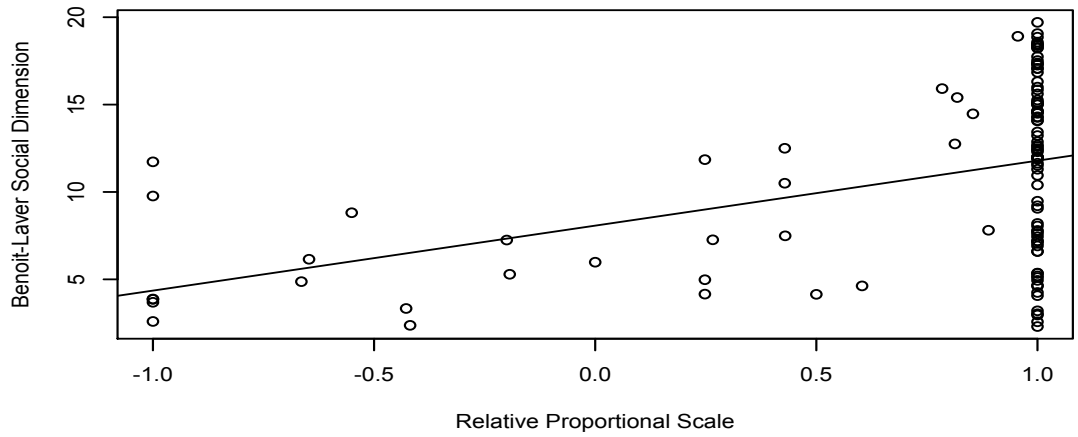


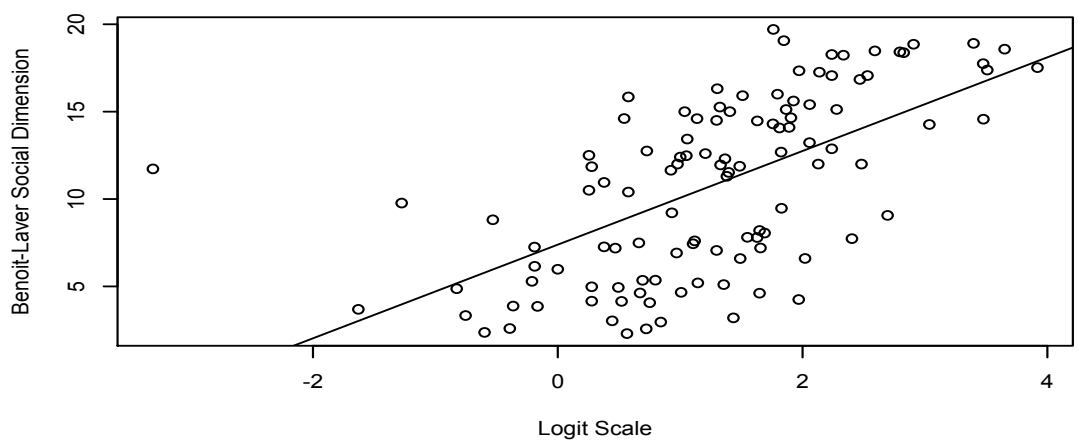


Comparison w/Expert Surveys

Social Liberalism

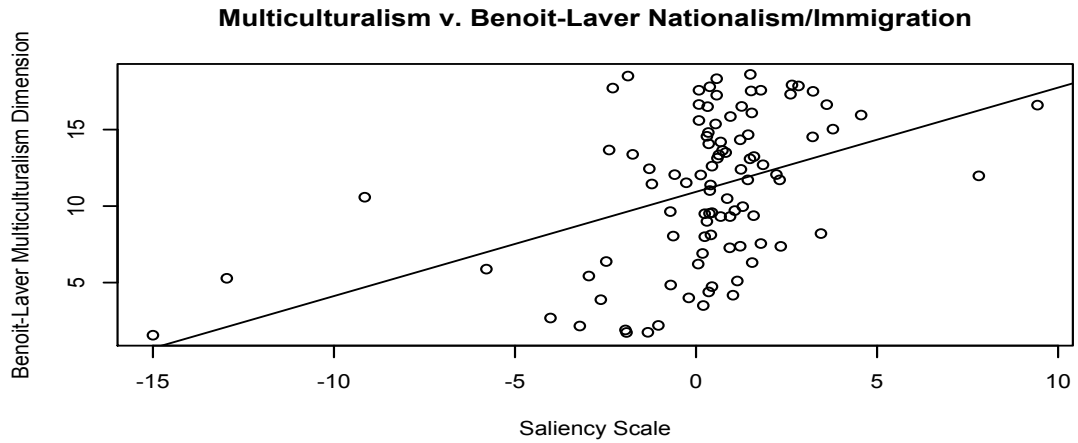


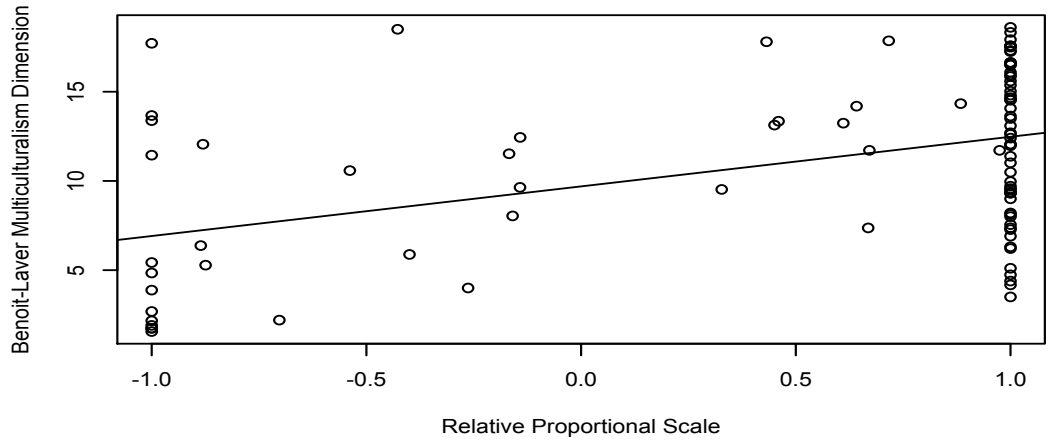


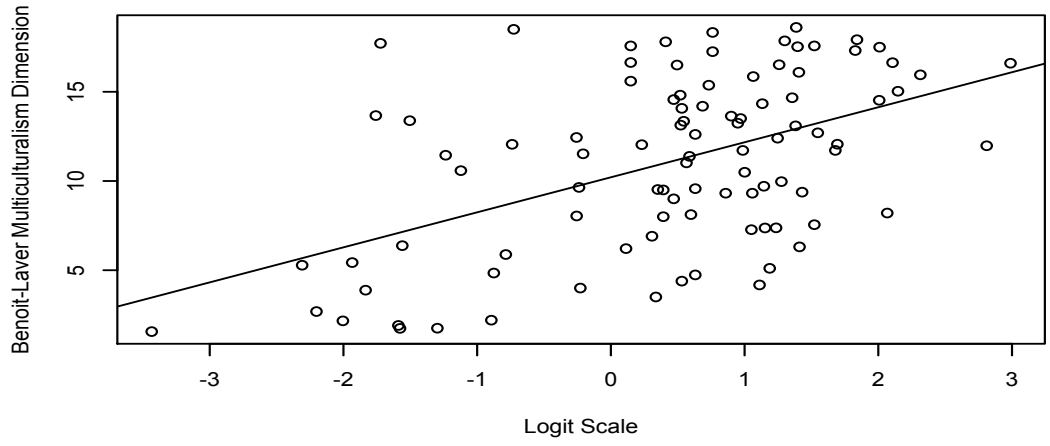


Comparison w/Expert Surveys

Multiculturalism

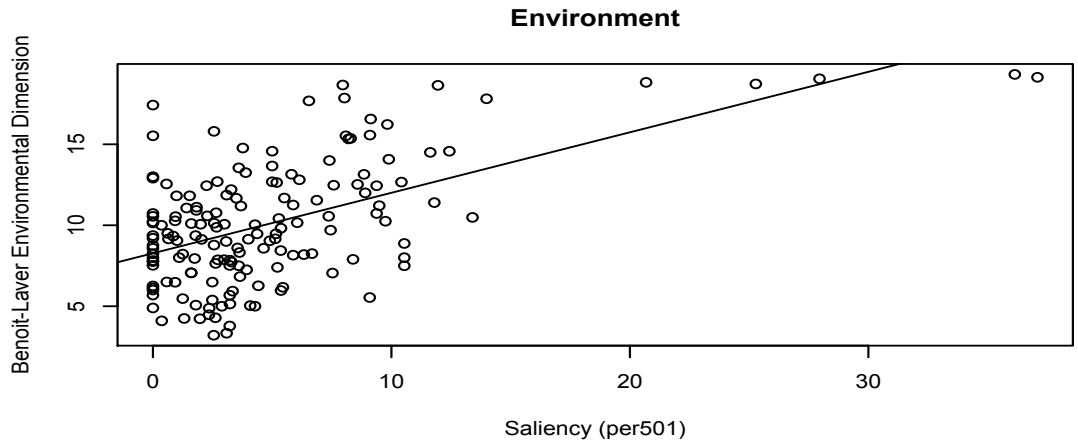


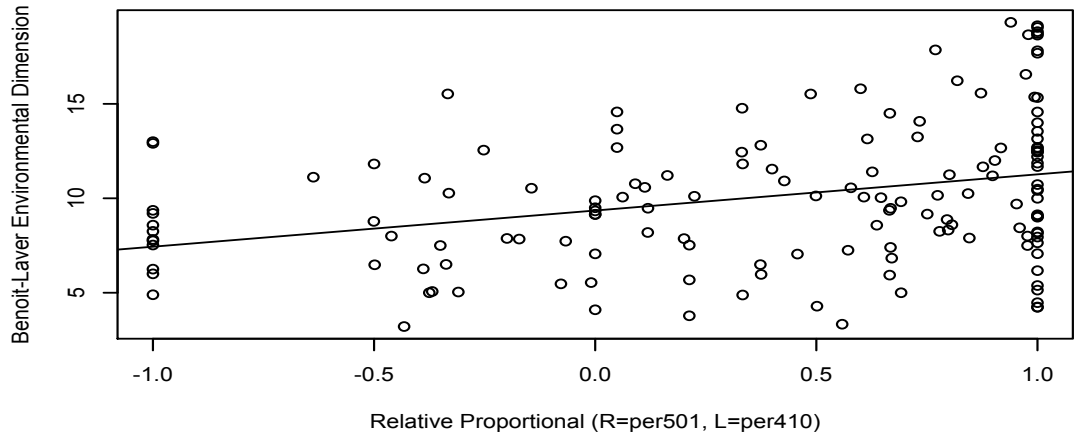


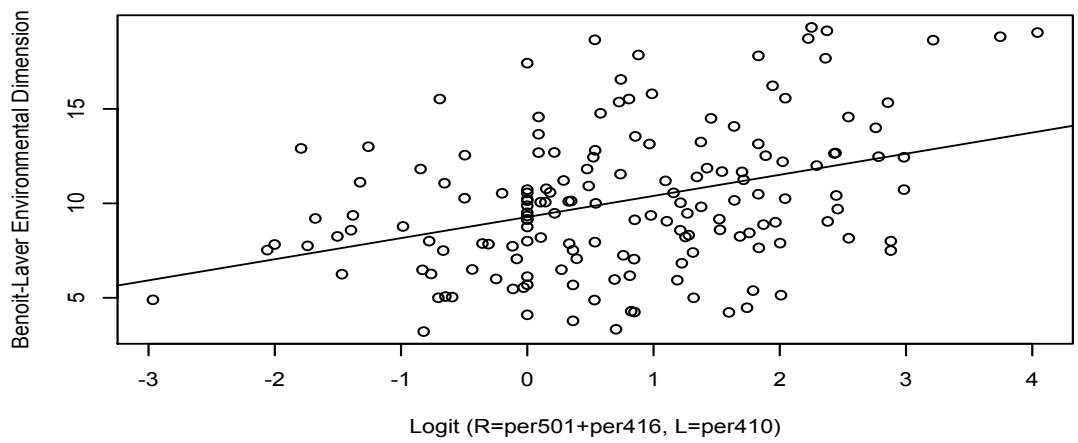


Comparison w/Expert Surveys

Environment







Next project: unitization testing

- Unitization varies +/- 10% in reliability tests on the training document
- The notion of a *quasi-sentence* is extremely subjective
- What if: natural sentences were just as good in the aggregate?

A Test: Count the Quasi-Sentences!

We believe that continued double-figure inflation will destroy the basis of the New Zealand economy and cause untold misery. The fight against increases in the cost of living is the most important single issue in economic management.
People without jobs represent waste of productive effort: National supports a policy of full employment and the dignity of labour. We do not accept unemployment as a balancing factor in economic management.
Finally, the National Development Council will be restored and consultation resumed between Government departments, academic specialists and private industry, including farming and organised labour.

A Test: How many of you said seven?

We believe that continued double-figure inflation will destroy the basis of the New Zealand economy and cause untold misery. / The fight against increases in the cost of living is the most important single issue in economic management. / People without jobs represent waste of productive effort: / National supports a policy of full employment / and the dignity of labour. / We do not accept unemployment as a balancing factor in economic management. / Finally, the National Development Council will be restored and consultation resumed between Government departments, academic specialists and private industry, including farming and organised labour.