Method

Salganik & Levy (2015) propose a Bayesian method for analyzing wikisurvey outcomes. Roughly speaking, this method constructs a posterior distribution of any respondents' "opinion" about any alternative using the *relative* comparison of within-individual opinions implied by observed choices. Alternative scores, available on www.allourideas.com, are computed by sampling from this posterior and taking an appropriate average. Confidence intervals can be obtained by resampling from the posterior, but these are not provided at www.allourideas.com.

We take a slightly different approach. Within respondents, wikisurvey votes are an example of a pairwise-comparison Discrete Choice Experiment (Louvierre, Hensher, & Swait, 2000). Such experiments can be efficiently analyzed with models like the Logit estimated with maximum likelihood (MLE) methods (Train, 2009). Including respondent-specific effects (heterogeneity) in the discrete choice frame can be difficult, however. Generalizations such as the Latent Class or Random Coefficients Logit models capture heterogeneity, but are significantly more complex to estimate.

We use a variant of Logit model that includes both "shared" opinions and respondent-specific effects representing idiosyncratic deviations from those shared opinions. The composition of the shared and idiosyncratic effects is, effectively, Salganik and Levy's (2015) opinion matrix. As is well known in statistics we can estimate models with the *fewest* meaningful respondent-specific effects by including an appropriate regularization term (Tibshirani, 1996). Our treatment is analogous to (but not the same as) how methods like GLMNET (Friedman, Hastie, & Tibshirani, 2010) induce sparsity in Generalized Linear Model coefficients, although we only induce sparsity on a *subset* of coefficients. Estimates depend on the choice of a penalty parameter Λ that determines how much weight is given to optimizing the likelihood versus minimizing the number of respondent-specific effects. Larger penalties imply a focus on shared opinions only, while smaller penalties allow for more respondent-specific effects. Perhaps most importantly this extension of the Logit model has a MLE problem that is *convex* (Boyd & Vandenberghe, 2004). Convexity guarantees that there are unique estimates conditional on the data and, in this case, that computing estimates is quite tractable and possible with freely available software (CVXPy). More information is available at URL TBD.

Additional practical details are as follows. We estimate shared opinions and respondent-specific effects separately for each of the four surveys -- long-text format for both pools, and the short-text format for both pools -- using both alternative-specific coefficient (ASC) coding wherein each alternative has its own associated "opinion" and alternatives coded simply as "I", "E", or "I+E" (see Tables X.1-X.2). Because our estimation strategy involves the choice of a penalty parameter (Λ), we estimate over 50 log-spaced values and focus here on scoring results for three specific values of this parameter. While the numeric values of our three chosen values are not inferentially relevant, they are chosen to span the breadth of influence the respondent-specific effects have in the model fits (cf Figures X.1, X.2). Following Salgnik and Levy (2015) we score each alternative as 'the probability that a randomly chosen respondent would prefer this alternative to a randomly chosen competing alternative.' Because estimation is tractable and reasonably fast we can compute confidence intervals for alternative scores using bootstrapping (Efron and Tibshirani, 1986). We resample the data (with replacement) 500 times to generate data over which we estimate model coefficients and then compute scores to obtain a portrait of estimator variance.

The efficiency of our estimation approach is pivotal to realizing this scale of analysis. Note that we estimate over 200,000 models: one "real" data set and 500 bootstrap resamples, 2 condition, 2 pools, with 2 model specifications, and over 50 penalty parameter values for each. Using simple parallelization tools, we can execute all these model estimates in a few hours on an r4.8xlarge Amazon EC2 instance using freely available python codes. Our own analysis and visualization code is available online at URL TBD.

Results

Figures X.1-X.2 show the optimal likelihoods for the two model types (ASC, IE coding) in each of the two conditions (long/short text) for both pools (N, S). For both conditions and both pools the model based on ASC coding is more descriptive, having higher estimated likelihoods than the IE coding.[1] The IE model does not even fit the data much better than a null model (uniformly random guesses of which alternative wins), even with respondent-specific effects.

Figures X.3-X.4 show scores for ASC coding in each of the two conditions (long/short text) for both pools (N, S). We include our estimated scores, the scores presented at www.allourideas.com, as well as confidence intervals (in the form of box plots) derived from bootstrapping. Our scores line up very well with those from www.allourideas.com, with scores in most cases within the middle 3 quintiles of out bootstrap confidence intervals and in many cases indistinguishable from our actual data estimated scores.

Figure X.5 shows scores for IE coding in each of the two conditions (long/short text) for both pools (N, S). The long text condition results suggest that an "I+E" alternative is perceived as more authentic than either an "I" or "E" alternative, but that "I" and "E" alternatives are probably perceived as equally authentic. The short text condition results are more difficult to interpret. In the N pool, it appears that an "I" alternative is more authentic than an "E" alternative, which is more authentic than an "I+E" alternative. But in the S pool, it appears that an alternative's "I"/"E"/"I+E" classification has no bearing on its authenticity. This inconsistency, along with the relatively poor model fit, suggest that we should draw our inferences from the ASC coded models.

[1] ASC coding has many more parameters, and thus this effect is probably to be expected. Figure ???, however, shows that for the *same* number of parameters the ASC coding can achieve higher likelihoods, suggesting that ASC coding better describes the underlying response processes. Technically, because IE coding is a nested alternative model of the ASC coding, we could us a likelihood ratio test to determine statistically significant difference between these two model forms (Train, 2009).

Figure X.6 shows *class* scores, as in the IE coded model, extracted from the ASC coded model to leverage the higher descriptive power of the ASC coded models. These class scores in the ASC coded model follow the same logic as Salganik & Levy's (2015) scores for the alternatives: 'the probability that a randomly chosen alternative from a particular class is rated as more authentic to a randomly chosen alternative not in that class, for a randomly chosen respondent.' This analysis suggests that "I" alternatives are perceived as the most authentic, while "E" alternatives are perceived as the least authentic, better matching the pattern visually apparent in Figures X.3-X.4 (as one would expect). These features are also more stable across the conditions and pools with ASC-model class scoring, being present with some (conceptual) degree of significance in 3 out of 4 condition/pools.

References

M. J. Salganik & K. E. C. Levy (2015) "Wiki Surveys: Open and Quantifiable Social Data Collection." *PLOS One* **10**(5) e0123483

Stated Choice Methods: Analysis and Applications (2000). J. Louvierre, D. A. Hensher, & J. Swait. Cambridge University Press.

Discrete Choice Methods with Simulation (2009). K. Train. Cambridge University Press.

R. Tibshirani (1996). "Regression Shrinkage and Selection via the LASSO" *Journal of the Royal Statistical Society, Series B (Methodological)*. **58**(1) 267-288

J. Friedman, T. Hastie and R. Tibshirani. (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software*, **33**(1), 1-22

Convex Optimization (2004). S. Boyd & L. Vandenberghe. Cambridge University Press.

CVXPy (2018). *Convex Optimization, for everyone.* http://www.cvxpy.org/

B. Efron and R. Tibshirani (1986). "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy" *Statistical Science* **1**(1) 54-77

Table X.1: Alternatives in the long-text condition with "I"/"E"/"I+E" class labels.

Alt #	Text	Class
1	All of our spirits are blended according to the unique tastes of our master distiller	E
2	All of our spirits are distilled in-house using our own handmade copper still	I
3	All of our spirits are made using ingredients following secret family recipes	E
4	Our barrels are hand-crafted from aged oak trees grown on select properties	E
5	Our distillery features a unique tasting room designed to open your senses	E
6	Our distillery is true grain-to-glass, sourcing only from our own farm	I
7	Our master distiller hails from a family of moonshiners dating back to prohibition	E
8	Our resident still is a direct fire still which caramelizes the mash as it cooks	I+E
9	Our spirits are blended following our master distiller's intuition, rather than a recipe	E
10	Our spirits are distilled in a building once occupied by prohibition moonshiners	E
11	Our spirits are distilled in the basement of an old prohibition-era speakeasy	E
12	Our spirits emanate the special terroir of the locations from where we source the grain	E
13	Our supplier mashes outdoors, allowing microflora to infuse the mash with terroir	E
14	The ingredients we use come only from a family farm that once had its own distillery	E
15	We carefully tend our farm to ensure the grain we use meets our standards	I+E
16	We distill all of our spirits by artisanal methods using our own handmade still	I
17	We distill all our spirits in-house and source grain only from local farmers	I
18	We distill spirits using a copper still made specifically for us by the manufacturer	I+E
19	We make our spirits using grains sourced from secret locations around the globe	E
20	We malt the grain ourselves, ensuring we get the aroma and taste we want	I+E
21	We only use malt processed by hand-raking grain on our special malting floor	E
22	We own every aspect of the spirits production process, from growing to bottling	I
23	We use our own grain, allowing us to make malt originating from our own farm	I

Table X.2: Alternatives in the short-text condition with “I”/”E”/”I+E” class labels.

Alt #	Text	Class
1	Cultivates own farm	I
2	Distills with special partners	E
3	Exotic sources of grain	E
4	Flavor of terroir	E
5	Grows own grain	I
6	Makes own barrels	I
7	Master distiller	E
8	Moonshining ancestors	E
9	Own handmade still	I+E
10	Owns the farm	I
11	Owns the still	I
12	Processes own malt	I
13	Prohibition legacy	E
14	Rakes own malt	I
15	Secret recipes	E
16	Special barrels	E
17	Special blending	E
18	True grain-to-glass	I
19	Uses own copper still	I+E
20	Uses prohibition era recipe	E
21	Uses special third-party still	E

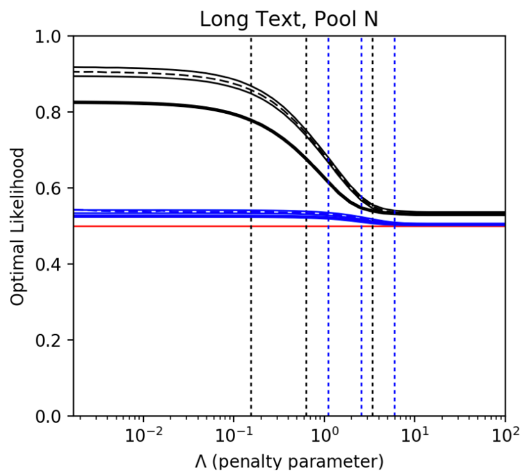

Figure X.1: Optimal likelihood for the long text condition in both pools for various penalty parameters. Black lines show results using ASC coding, blue lines show results using IE coding, the red line shows the likelihood of a null model. Thin solid lines denote the range of likelihoods obtained in 90% (lowest 5% to highest 95%) of the bootstrap samples, and the thin dashed line (if visible) denotes the mean likelihood over the bootstrap samples. The thicker solid line denotes the optimal likelihood on the actual data. The vertical dotted lines correspond to the fixed values of Λ used in our score plots below.

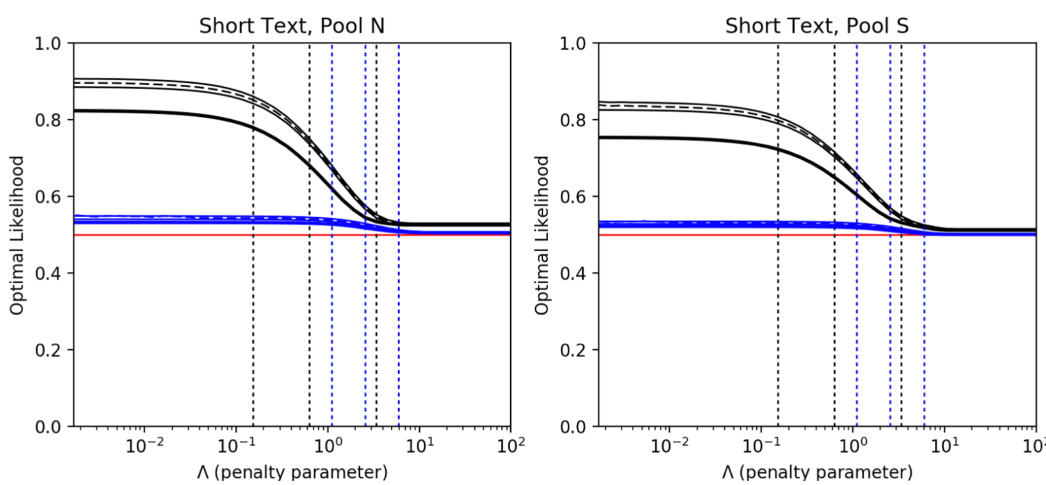

Figure X.2: Optimal likelihood for the short text condition in both pools for various penalty parameters. See Figure X.1 for formatting details.

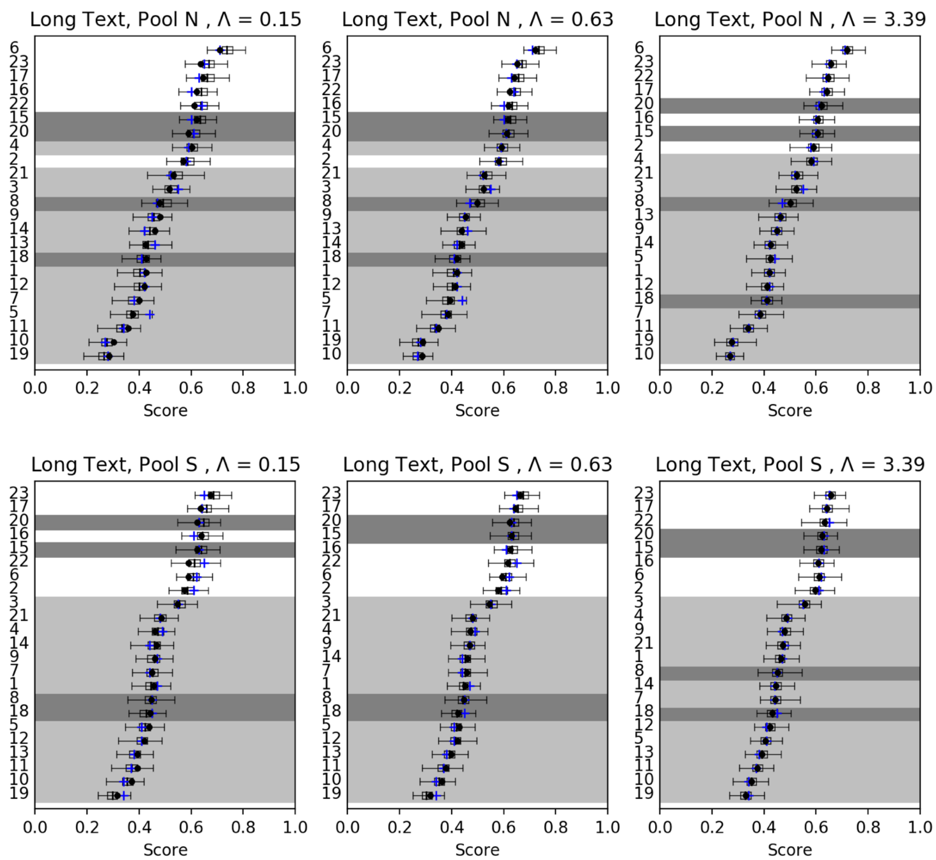

Figure X.3: Scores for the long text condition for three values of the penalty parameter using ASC coding. Box plots show the range of scores estimated with bootstrapping; black dots represent scores estimated on the actual data; and a blue "+" denotes the score estimated and displayed at www.allourideas.com. The background color denotes IE coding: "I" in white, "E" in light grey, and "I+E" in dark grey. See Table X.1 for actual alternative text corresponding to the vertical axis labels.

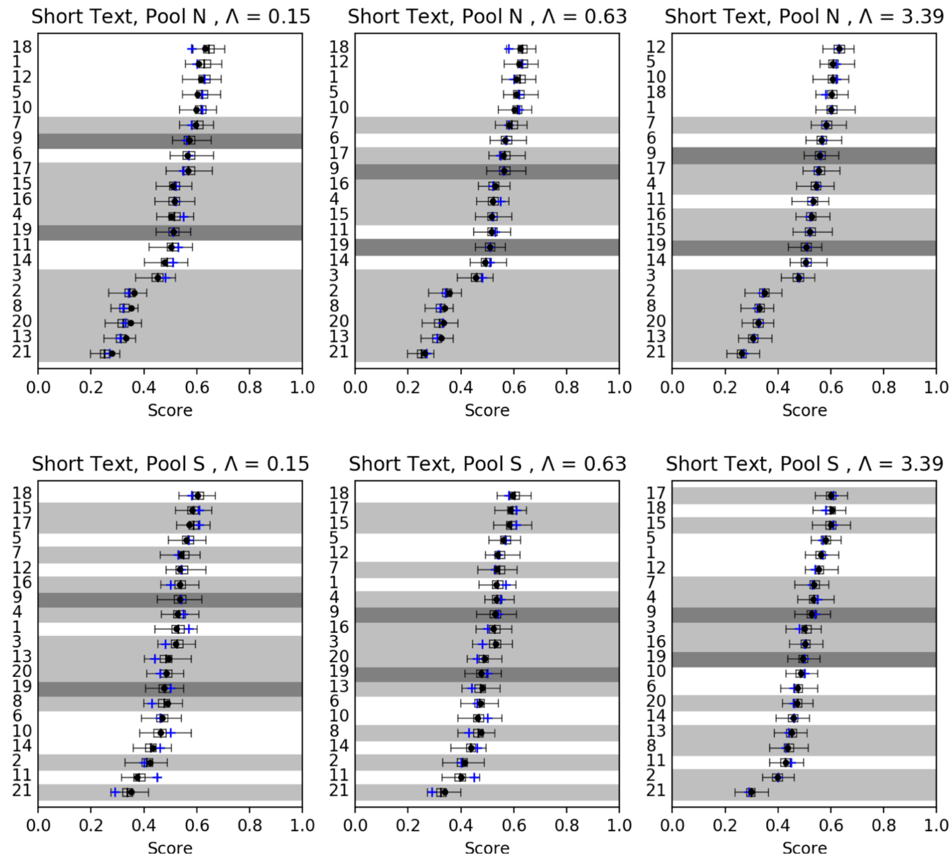

Figure X.4: Scores for the short text condition for three values of the penalty parameter using ASC coding. See Figure X.3 for formatting details, and Table X.1 for actual alternative text corresponding to the vertical

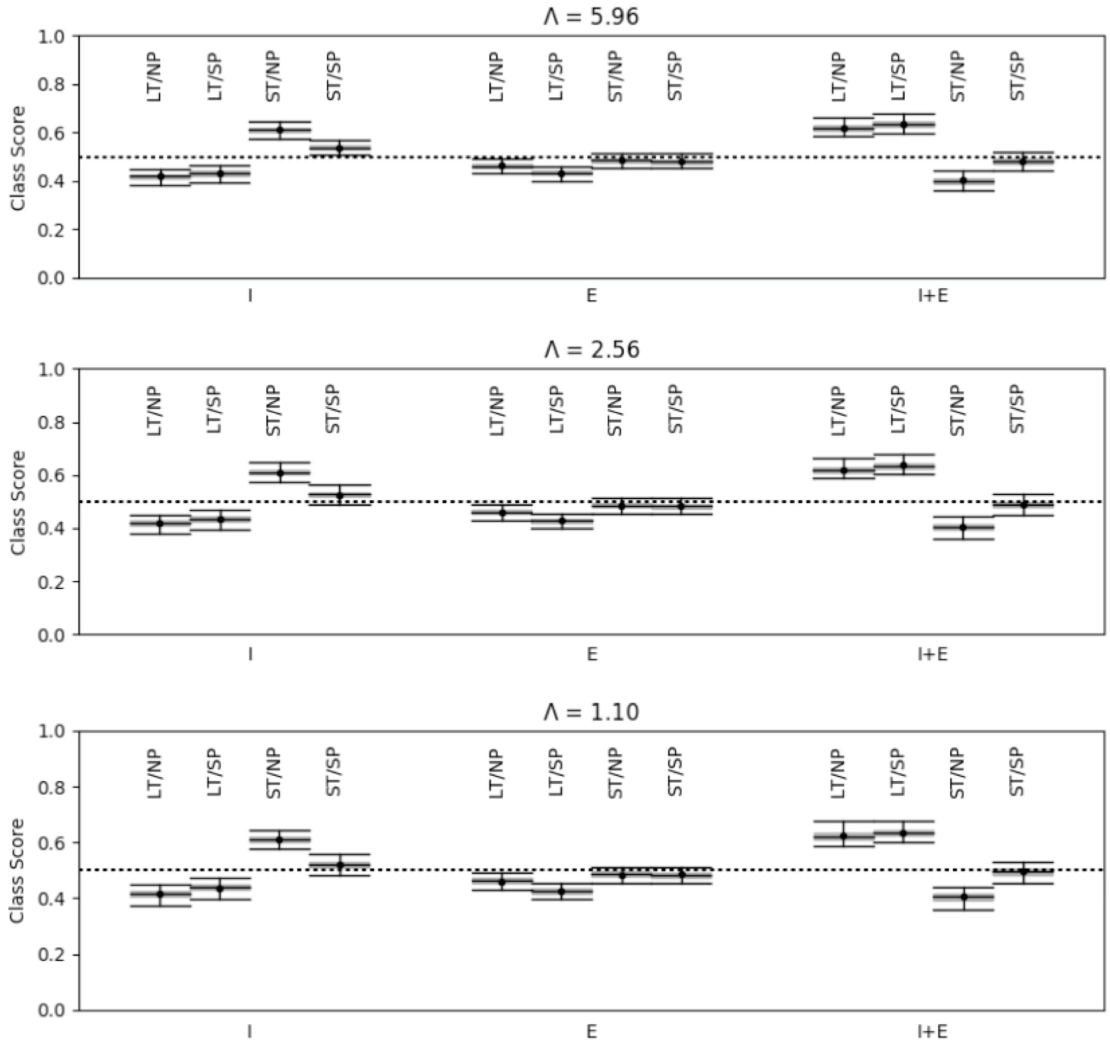

Figure X.5: Scores for both text condition for three values of the penalty parameter using IE coding. Box plots show the range of scores estimated with bootstrapping and black dots represent scores estimated on the actual data. See Table X.2 for alternative classifications. Abbreviations are LT: long text; ST: short text; NP: national pool; SP: Stanford pool.

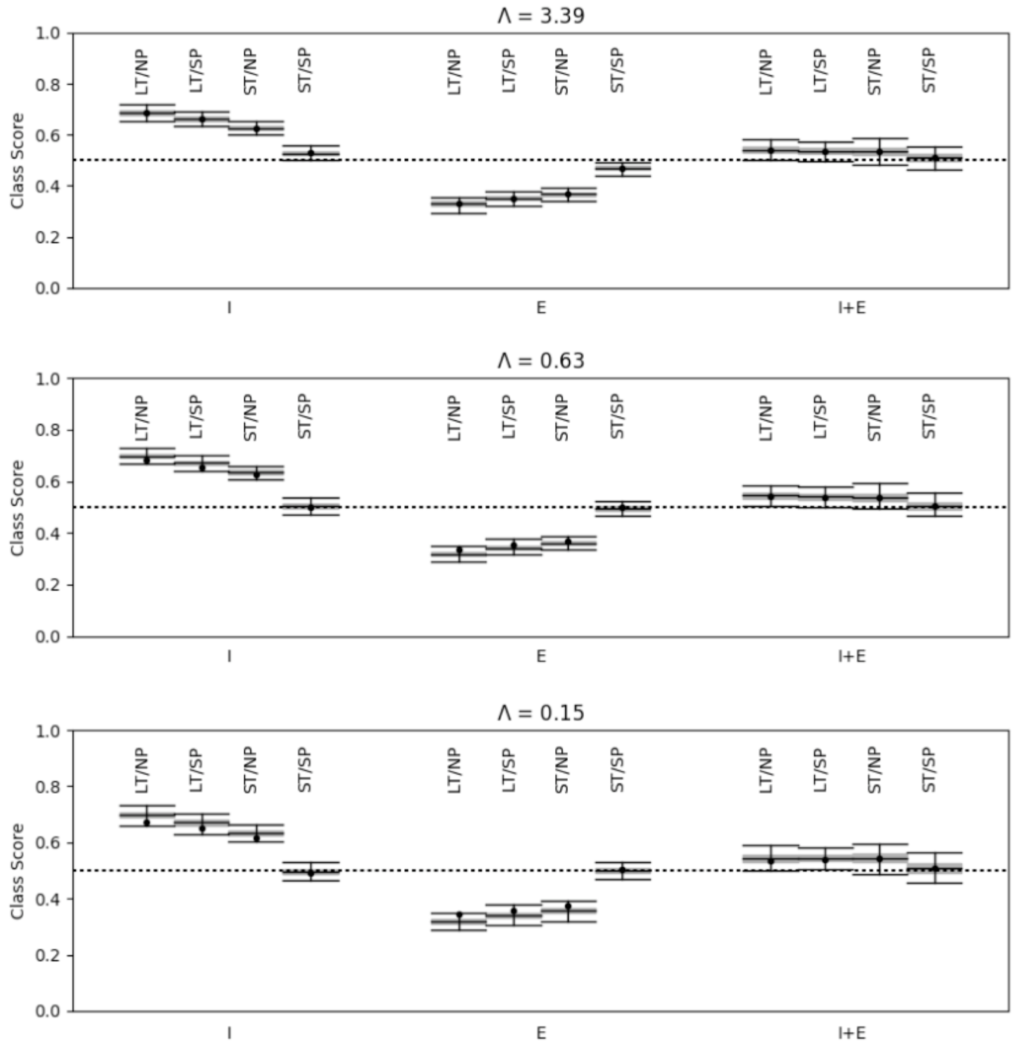

Figure X.6: Class scores for both text conditions for three values of the penalty parameter using the models with ASC coding. Box plots show the range of scores estimated with bootstrapping and black dots represent scores estimated on the actual data. Abbreviations are LT: long text; ST: short text; NP: national pool; SP: Stanford pool.

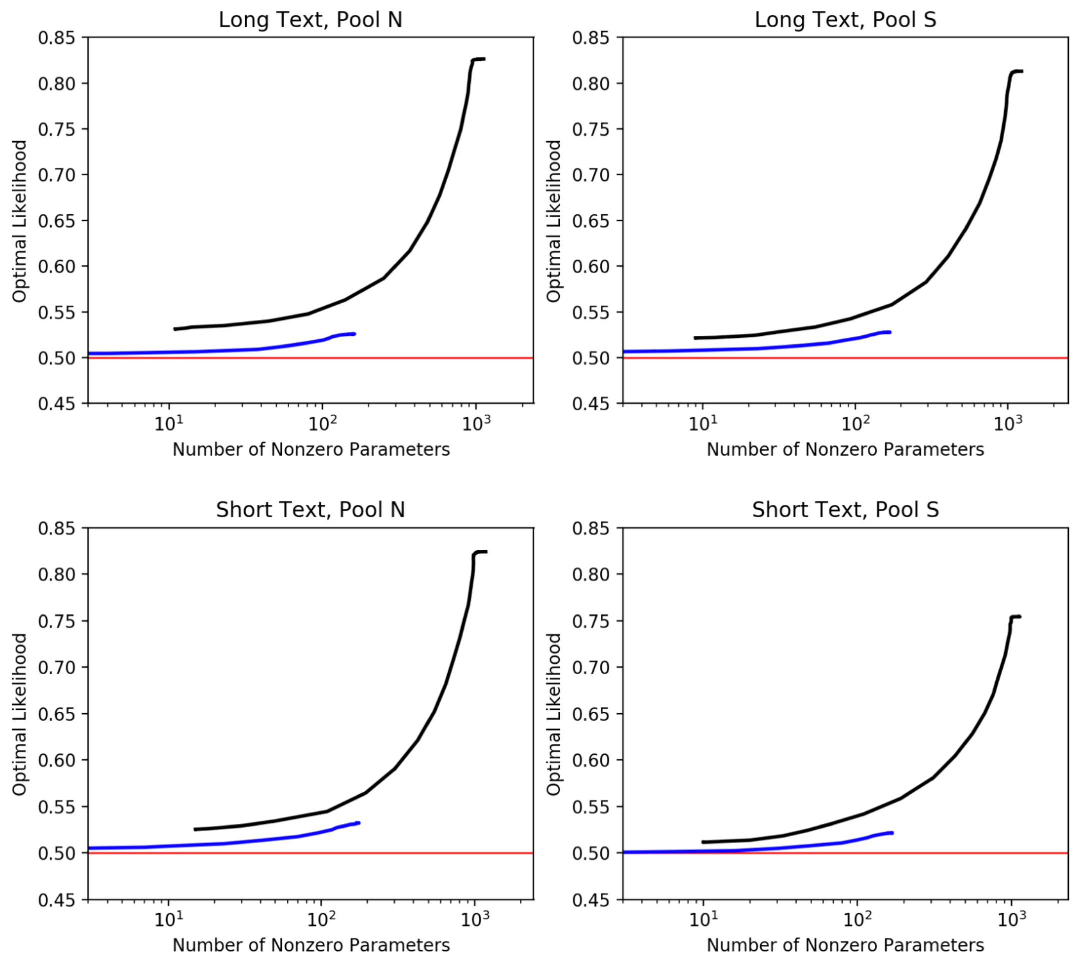

Figure ???: Optimal likelihood as a function of model parameters in all conditions (actual data). Black line: ASC coding; blue line: IE coding; red line: null model (random guessing).