Supplementary Materials for

"Using Conjoint Experiments to Analyze Election Outcomes: The Essential Role of the Average Marginal Component Effect (AMCE)"

Kirk Bansak, Jens Hainmueller, Daniel J. Hopkins, and Teppei Yamamoto

Table of Contents

A Proof of Proposition 1

First, we begin by defining the EBS for the attribute of interest equaling t, $G_i(t)$, as the expected number of times a profile with that attribute is chosen by respondent i against Q profiles randomly drawn from p. That is, with $\mathcal{X}(t)$ denoting the set of all profiles with the lth attribute being t,

$$G_i(t) \quad = \sum_{x_{(mi)} \in \mathcal{X}(t)} g_i(x_{(mi)}; p) \Pr(\tilde{X}_{i1k} = \tilde{x}_{(mi)}).$$

Next, we prove the following lemma, which relates the individual-level average of the potential outcome to the EBS.

Lemma A.1 *The marginal mean potential outcome with respect to profile x_1 for respondent i, $\overline{Y}_i(x_1)$, equals the EBS of x_1 divided by the number of possible unique profiles.*

Proof. Suppose x_1 is the mth-preferred profile by respondent i, such that $x_1 = x_{(mi)}$. Then,

$$\begin{aligned}
\overline{Y}_i(x_1) &= \sum_{x_2 \in \mathcal{X}} Y_i(x_1, x_2) \Pr(X_{i2k} = x_2) \\
&= \sum_{m < m'} Y_i(x_{(mi)}, x_{(m'i)}) \Pr(X_{i2k} = x_{(m'i)}) + \sum_{m \geq m'} Y_i(x_{(mi)}, x_{(m'i)}) \Pr(X_{i2k} = x_{(m'i)}) \\
&= \sum_{m < m'} 1 \cdot \Pr(X_{i2k} = x_{(m'i)}) + \sum_{m > m'} 0 \cdot \Pr(X_{i2k} = x_{(m'i)}) \\
&= \frac{g_i(x_1; p)}{Q}. \qquad \blacksquare
\end{aligned}$$

Now we are ready to prove Proposition 1. First, consider the individual-level AMCE of the lth attribute, $AMCE_{li}(t_1, t_0; p)$, representing the average causal effect of attribute l equaling t_1 as opposed to t_0 on the probability that respondent i would choose the profile against a randomly drawn profile from p. We can show that:

$$\begin{aligned}
AMCE_{li}(t_1, t_0; p) &= \sum_{\tilde{x}_1, x_2} \{Y_i(t_1, \tilde{x}_1, x_2) - Y_i(t_0, \tilde{x}_1, x_2)\} \, p(\tilde{x}_1, x_2) \\
&= \sum_{\tilde{x}_1, x_2} \{Y_i(t_1, \tilde{x}_1, x_2) - Y_i(t_0, \tilde{x}_1, x_2)\} \, p(\tilde{x}_1) p(x_2) \\
&= \sum_{x_1 \in \mathcal{X}(t_1)} \sum_{x_2 \in \mathcal{X}} Y_i(x_1, x_2) \Pr(\tilde{X}_{i1k} = \tilde{x}_1) \Pr(X_{i2k} = x_2) \\
&\quad - \sum_{x_1 \in \mathcal{X}(t_0)} \sum_{x_2 \in \mathcal{X}} Y_i(x_1, x_2) \Pr(\tilde{X}_{i1k} = \tilde{x}_1) \Pr(X_{i2k} = x_2) \\
&= \sum_{x_1 \in \mathcal{X}(t_1)} \Pr(\tilde{X}_{i1k} = \tilde{x}_1) \sum_{x_2 \in \mathcal{X}} Y_i(x_1, x_2) \Pr(X_{i2k} = x_2) \\
&\quad - \sum_{x_1 \in \mathcal{X}(t_0)} \Pr(\tilde{X}_{i1k} = \tilde{x}_1) \sum_{x_2 \in \mathcal{X}} Y_i(x_1, x_2) \Pr(X_{i2k} = x_2) \\
&= \sum_{x_1 \in \mathcal{X}(t_1)} \Pr(\tilde{X}_{i1k} = \tilde{x}_1) \overline{Y}_i(x_1) - \sum_{x_1 \in \mathcal{X}(t_0)} \Pr(\tilde{X}_{i1k} = \tilde{x}_1) \overline{Y}_i(x_1) \\
&= \frac{1}{Q} \left\{ \sum_{x_1 \in \mathcal{X}(t_1)} \Pr(\tilde{X}_{i1k} = \tilde{x}_1) g_i(x_1; p) - \sum_{x_1 \in \mathcal{X}(t_0)} \Pr(\tilde{X}_{i1k} = \tilde{x}_1) g_i(x_1; p) \right\} \\
&= \frac{1}{Q} \{G_i(t_1) - G_i(t_0)\},
\end{aligned}$$

where the first equality follows from the definition of the AMCE, the second from Assumption 1, the sixth from Lemma A.1, and the last from the definition of $G_i(t)$. Since $AMCE_l(t_1, t_0; p) = \frac{1}{N} \sum_{i=1}^{N} AMCE_{li}(t_1, t_0; p)$ by definition, we have

$$AMCE_l(t_1, t_0; p) = \frac{1}{NQ} \sum_{i=1}^{N} \{G_i(t_1) - G_i(t_0)\}. \quad \blacksquare$$

B Details of the Literature Review on Quantities of Interest in Electoral Research

In this section, we provide additional details of the procedure used in our review of the empirical elections literature.

The four journals we collected articles from were *The American Political Science Review*, *The American Journal of Political Science*, *Electoral Studies*, and *Political Behavior*. Our initial parameters identified 279 published articles on voting, and 111 of those included an estimate of the effects of candidate or party characteristics on some electorally relevant outcome. We next removed articles which did not evaluate vote choice specifically and also removed those articles which used a conjoint design, as one of the primary goals of this paper is precisely to clarify the most common implicit quantities of interest in electoral conjoint experiments.

To identify articles that use either aggregate vote shares or their individual-level analogues as a key outcome, we grouped articles whose primary outcome was aggregate vote shares with those that considered their individual-level analog, changes in the individual-level probability of voter support for a party or candidate. We then separately identified articles whose primary outcome was the probability of a candidate/party victory or the number of seats won in a legislature. We also removed articles which did not evaluate vote choice specifically (most commonly because their outcome was instead candidate evaluations), leaving us with 82 articles. We provide the list below.

- Douglas J. Ahler, Jack Citrin, Michael C. Dougal, and Gabriel S. Lenz. 2017. Face Value? Experimental Evidence that Candidate Appearance Influences Electoral Choice. Political Behavior.
- Carlos Algara. 2019. The conditioning role of polarization in US senate election outcomes: A direct-election era & voter-level analysis. Electoral Studies.
- Jennifer Gandhi and Elvin Ong. 2019. Committed or Conditional Democrats? Opposition Dynamics in Electoral Autocracies. American Journal of Political Science.
- Devra Moehler and Jeffrey Conroy-Krutz. 2016. Eyes on the ballot: Priming effects and ethnic voting in the developing world. Electoral Studies.
- Kerri Milita, Elizabeth N. Simas, John Barry Ryan, and Yanna Krupnikov. 2017. The effects of ambiguous rhetoric in congressional elections. Electoral Studies.
- Ignazio De Ferrari. 2017. The accountability effect of endorsements: A survey experiment. Electoral Studies.
- Thomas Gift and Carlos X. Lastra-Anadon. 2018. How voters assess elite-educated politicians: A survey experiment. Electoral Studies.

- Joan E. Cho and Dominika Kruszewska. 2018. Escaping collective responsibiliity in fluid party systems: Evidence from South Korea. Electoral Studies.
- Bethany L. Albertson. 2015. Dog-Whistel Politics: Multivocal Communication and Religious Appeals. Political Behavior.
- Tessa Ditonto. 2017. A High Bar or a Double Standard? Gender, Competence, and Information in Political Campaigns. Political Behavior.
- Ayala Yarkoney Sorek, Kathryn Haglin, and Nehemia Geva. 2018. In Capable Hands: An Experimental Study of the Effects of Competence and Consistency on Leadership Approval. Political Behavior.
- Cheryl Boudreau, Christopher S. Elmendorf and Scott A. MacKenzie. 2019. Roadmaps to Representation: An Experimental Study of How Voter Education Tools Affect Citizen Decision Making. Political Behavior.
- Mara Cecilia Ostfeld. 2019. The New White Flight?: The Effects of Political Appeals to Latinos on White Democrats. Political Behavior.
- Rosario Aguilar, D. Alex Hughes and Micah Gell-Redman. 2019. Phenotypic Preference in Mexican Migrants: Eivdence from a Random Household Survey. Political Behavior.
- Adrian Lucardi and Guillermo Rosas. 2016. Is the incumbent curse the incumbent's fault? Strategic behavior and negative incumbency effects in young democracies. Electoral Studies.
- Lucas Nunez. 2018. Do clientelistic machines affect electoral outcomes? Mayoral incumbency as a proxy for machine prowess. Electoral Studies.
- Daniel Stockemer and Rodrigo Praino. 2019. The Good, the Bad and the Ugly: Do Attractive Politicians Get a "Break" When They are Involved in Scandals?. Political Behavior.
- Santiago Olivella, Kristin Kanthak and Brian F. Crisp. 2017. And keep your enemies closer: Building reputations for facing electoral challenges. Electoral Studies.
- Bruce A. Desmarais, Raymond J. La Raja and Michael S. Kowal. 2015. The Fates of Challengers in U.S. House elections: The Role of Extended Party Networks in Supporting Candidates and Shaping Electoral Outcomes. American Journal of Political Science.
- Mariana Lopes da Fonseca. 2017. Identifying the Source of Incumbency Advantage through a Constitutional Reform. American Journal of Political Science.
- Christian Salas. 2016. Incumbency advantage in multi-member districts: Evidence form congressional elections in Chile. Electoral Studies.

- Kenichi Ariga, Yusaku Horiuchi, Roland Mansilla, and Michio Umeda. 2016. No sorting, no advantage: Regression discontinuity estimates of incumbency advantage in Japan. Electoral Studies.
- Sebatian Dettman, Thomsa B. Pepinsky, Jan H. Pierskalla. 2017. Incumbency advantage and candidate characteristics in open-list proportional representation systems: Evidence from Indonesia. Electoral Studies.
- Woo Chang Kang, Won-ho Park and B.K. Song. 2018. The effect of incumbency in national and local elections: Evidence form South Korea. Electoral Studies.
- Andrew B. Hall and Daniel M. Thompson. 2018. Who Punishes Extremist Nominees? Candidate Ideology and Turning Out the Base in US Elections. American Political Science Review.
- John W. Patty, Constanza F. Schibber, Elizabeth Maggie Penn and Brian F. Crisp. 2019. Valence, Elections, and Legislative Institutions. American Journal of Political Science.
- Marke Klasnja and Rocio Titiunik. 2017. The Incumbency Curse: Weak Parties, Term Limits, and Unfulfilled Accountability. American Political Science Review.
- Jon H. Fiva and Daniel M. Smith. 2018. Political Dynasties and the Incumbency Advantage in Party-Centered Environments. American Political Science Review.
- Jens Olav Dahlgaard. 2016. You just made it: Individual incumbency advantage under Proportional Representation. Electoral Studies.
- Zachary Spicer, Michael McGregor and Christopher Alcantara. 2017. Political opportunity structures and the representation of woment and visible minorities in municipal elections. Electoral Studies.
- Jungho Roh. 2017. The incumbency disadvantage in South Korean National Assembly elections: Evidence from a regression discontinuity approach. Electoral Studies.
- Young-Im Lee. 2019. The leaky pipeline and sacrificial lambs: Gender, candidate nomination, and district assignment in South Korea's national legislative elections. Electoral Studies.
- Steven Rogers. 2017. Electoral Accountability for State Legislative Roll Calls and Ideological Representation. American Political Science Review.
- Gabriele Magni and Andrew Reynolds. 2018. Candidate Sexual Orientation Didn't Matter (in the Way You Might Think) in the 2015 UK General Election. American Political Science Review.

- Lefteris Anastasopoulos. 2016. Estimating the gender penalty in House of Representative elections using a regression discontinuity design. Electoral Studies.
- Michael Jankowski. 2016. Voting for locals: Voters' information processing strategies in open-list PR systems. Electoral Studies.
- Craig Johnson and Alia Middleton. 2016. Junior coalition parties in the British context: Explaining the Liberal Democrat collapse at the 2015 general election. Electoral Studies.
- Boris Heersink and Brenton D. Peterson. 2017. Truman defeats Dewey: The effect of campaign visits on election outcomes. Electoral Studies.
- Seth J. Hill. 2017. Changing votes or changing voters? How candidates and election context swing voters and mobilize the base. Electoral Studies.
- James F. Downes and Matthew Loveless. 2018. Centre right and radical right party competition in Europe: Strategic emphasis on immigration, anti-incumbency, and economic crisis. Electoral Studies.
- R. Urbatsch. 2018. Feminine-sounding names and electoral performance. Electoral Studies.
- David A.M. Peterson. 2018. The dynamic construction of candidate image. Electoral Studies.
- Zeynep Somer-Topcu. 2015. Everything to Everyone: The Elecotral Consequences of the Broad-Appeal Strategy in Europe. American Journal of Political Science.
- Dominic Nyhuis. 2016. Electoral effects of candidate valence. Electoral Studies.
- Marco R. Steenbergen and Thomas Willi. 2019. What consideration sets can teach us about electoral competition: A two-hurdle model. Electoral Studies.
- Chris Tausanovitch and Christopher Warshaw. 2018. Does Ideological Proximity Between Candidates and Voters Affect Voting in the U.S. House Elections?. Political Behavior.
- Jon C. Rogowski and Patrick D. Tucker. 2018. Moderate, extreme, or both? How voters respond to ideologically unpredictable candidates. Electoral Studies.
- Leah C. Stokes. 2016. Electoral Backlash against climate Policy: A Natural Experiment on Retrospective Voting and Local Resistance to Public Policy. American Journal of Political Science.
- Kevin K. Banda and Jason H. Windett. 2016. Negative Advertising and the Dynamics of Candidate Support. Political Behavior.
- Christopher Wlezien and Stuart Soroka. 2019. Mass Media and Electoral Preferences During the 2016 US Presidential Race. Political Behavior.

- Oliver Heath, Gilles Veriers and Sanjay Kumar. 2015. Do Muslim voters prefer Muslim candidates? Co-religiosity and voting behavior in India. Electoral Studies.
- Patrick F.A. van Erkel and Peter Thijssen. 2016. The first one wins: Distilling the primacy effect. Electoral Studies.
- Peter Egge Langsaether, Haakon Gjerlow and Martin G. Soyland. 2019. Is all PR good PR? How the content of media exposure affects candidate popularity. Electoral Studies.
- Lea Portmann and Nenad Stojanovic. 2019. Electoral Discrimination Against Immigrant-Origin Candidates. Political Behavior.
- Aron Kiss. 2015. Identifying strategic voting in two-round elections. Electoral Studies.
- Sjoerdje Charlotte van Heerden and Wouter van der Brug. 2017. Demonisation and electoral support for populist radical right parties: A temporary effect. Electoral Studies.
- Eelco Harteveld. 2016. Winning the losers but losing the winners? The electoral consequences of the radical right moving to the economic left. Electoral Studies.
- Tsung-han Tsai. 2017. A balance between candidate- and party-centric representation under mixed-member systems: The evidence from voter behavior in Taiwan. Electoral Studies.
- Tarik Abou-Chadi and Mark A. Kayser. 2017. It's not easy being green: Why voters punish parties for environmental policies during economic downturns. Electoral Studies.
- Loes Aaldering. 2018. The (ir)rationality of mediated leader effects. Electoral Studies.
- Amanda Bittner. 2018. Leaders always mattered: The persistence of personality in Canadian elections. Electoral Studies.
- Douglas R. Pierce, Richard R. Lau. 2019. Polarization and correct voting in U.S. presidential elections. Electoral Studies.
- Jon Green. 2019. Floating policy voters in the 2016 U.S. presidential election. Electoral Studies.
- Andrew Barclay, Maria Sobolewska, Robert Ford. 2019. Political realignment of British Jews: Testing competing explanations. Electoral Studies.
- Richard Gunther, Paul A. Beck and Erik C. Nisbet. 2019. "Fake news" and the defection of 2012 Obama voters in the 2016 presidential election. Electoral Studies.
- Ron Johnston, Todd Hartman and Charles Pattie. 2019. Feelings about party leaders as a voter's heuristic - what happens when the leaders change?. Electoral Studies.

- Joshua N. Zingher and Michael E. Flynn. 2019. Does polarization affect even the inattentive? Assessing the relationship between political sophistication, policy orientations, and elite cues. Electoral Studies.
- Matthew L. Jacobsmeier. 2015. From Black and White to Left and Right: Race, Perceptions of Candidates' Ideologies, and Voting Behavior in U.S. House Elections. Political Behavior.
- Mollie J. Cohen. 2018. Protesting via the Null Ballot: An Assessment of the Decision to Cast an Invalid Vote in Latin America. Political Behavior.
- Sergio Garcia-Rios, Francisco Pedraza, Bryan Wilcox-Archuleta. 2019. Direct and Indirect Xenophobic Attacks: Unpacking Portfolios of Identity. Political Behavior.
- Benjamin Highton. 2019. Issue Accountability in U.S. House Elections. Political Behavior.
- Benjamin J. Newman and Thomas J. Hayes. 2019. Durable Democracy? Economic Inequality and Democratic Accountability in the New Gilded Age. Political Behavior.
- Harold Clarke, Jason Reifler, Thomas J. Scotto, Maianne C. Stewart, and Paul Whiteley. 2015. Valence politics and voting in the 2012 U.S. presidential election. Electoral Studies.
- Vicki Hesli Claypool, William M. Reisinger, Marina Zaloznaya, Yue Hu, and Jenny Juehring. 2018. Tsar Putin and the "corruption" thorn in his side: The demobilization of votes in a competitive authoritarian regime. Electoral Studies.
- Cheryl Boudreau, Christopher S. Elmendorf and Scott A. MacKenzie. 2019. Racial or Spatial Voting? The Effects of Candidate Ethnicity and Ethnic Group Endorsements in Local Elections. American Journal of Political Science.
- Shigeo Hirano, Gabriel S. Lenz, Maksim Pinkovskiy and James M. Snyder, Jr.. 2015. Voter Learning in State Primary Elections. American Journal of Political Science.
- Zachary David Greene and Matthias Haber. 2015. The consequences of appearing divided: An analysis of party evaluations and vote choice. Electoral Studies.
- M.V. Hood III and Seth C. McKee. 2015. Sunshine State dilemma: Voting for the 2014 governor of Florida. Electoral Studies.
- Eric Arias, Pablo Balan, Horacio Larreguy, John Marshall and Pablo Querubin. 2019. Information Provision, Voter Coordination, and Electoral Accountability: Evidence from Mexican Social Networks. American Political Science Review.
- Taylor C. Boas, F. Daniel Hidalgo and Marcus André Melo. 2019. Norms versus Action: Why Voters Fail to Sanction Malfeasance in Brazil. American Journal of Political Science.

- Lasse Laustsen and Alexander Bor. 2017. The relative weight of character traits in political candidate evaluations: Warmth is more important than competence, leadership and integrity. Electoral Studies.
- Lasse Laustsen. 2017. Choosing the Right Candidate: Observational and Experimental Evidence that Conservatives and Liberals Prefer Powerful and Warm Candidate Personalities, Respectively. Political Behavior.

C Alternative QOIs: Probability of Winning

In this section, we extend our discussion of one alternative quantity of interest, the probability of winning. Replication materials for all results are available in Bansak et al. (2022).

While we focus on two-candidate elections, the discussion in this subsection naturally extends to elections with more than two candidates. In designs with more than two profiles per task, we can define analogous quantities representing seat shares in multiparty plurality elections. For example, the probability of winning greater than some proportion t of the vote share in a J-way single-vote election is a general case of the probability of winning, and the estimation procedure described below can be adapted to accommodate this case by simply including any number J of profiles in the modeling of f and replacing 0.5 with any threshold t of interest in the modeling of M.

C.1 Estimating the Probability of Winning

Here, we provide a sketch of one potential approach to estimating the probability of winning, leaving a comprehensive exposition for future work.

We begin by noting that $\mathbb{E}_{\mathcal{V}}[Y_i([abc],[a'b'c'])] = \mathbb{E}_{\mathcal{V}}[Y_i \mid A = a, B = b, C = c, A' = a', B' = b', C' = c'] = \Pr(Y_i = 1 \mid A = a, B = b, C = c, A' = a', B' = b', C' = c')$ for any (a, b, c, a', b', c') in the support of $\mathcal{A}$ when attributes are randomly assigned. Then, a model-based approach would begin by modeling the following, which we will denote as f for shorthand:

$$f(A, B, C, A', B', C') \equiv \Pr(Y_i = 1|A, B, C, A', B', C').$$

This is a classical discrete choice problem in which the size of the choice set equals two (and hence it easily generalizes to forced choice tasks with more than two profiles), and we can employ a standard modeling strategy for discrete choice outcomes such as the conditional logit model (McFadden, 1974).[1] This is akin to the approach to conjoint survey data traditionally used in marketing research (e.g., McFadden, 1986).

Given the increased dimensionality when including attributes from both profiles in the function, as well as modeling their interactions, it could be useful to additionally employ methods from statistical learning to improve predictive performance in the face of potentially high-dimensional feature spaces. For instance, shrinkage penalties could be layered atop generalized linear models (GLMs) and their multinomial extensions to model f using an elastic net regularized regression framework (e.g., Reid and Tibshirani, 2014; Egami and Imai, 2019). Alternatively, f could be modeled using quasi-parametric learning approaches in place of GLMs, such as random forests,

[1] For paired conjoints, we can also fit a model equivalent to the conditional logit via a binary logit regression of Y_i on the differences of the attributes (i.e., $A - A'$, $B - B'$, etc.)

boosted trees, or neural nets (e.g., Prinzie and Van den Poel, 2008). Employing best practices in supervised learning theory (e.g. model training via cross-validation) is vital, and researchers could allow both theory and cross-validation performance to guide model selection.

Once we obtain a high-performing predictive model f, it is straightforward to estimate the probability of winning. First, given an estimated model $\hat{f}$, one can estimate the vote share for any profile $[abc]$ over any other profile $[a'b'c']$ using $\hat{f}(a, b, c, a', b', c')$. The majority classifier can then be obtained as $\hat{M}([ABC], [A'B'C']) = \mathbf{1}\{\hat{f}(A, B, C, A', B', C') > 0.5\}$, which allows one to predict whether or not $[abc]$ would win a majority over $[a'b'c']$ from the target population of voters, $\mathcal{V}$.

Finally, one can estimate the expectation of M by averaging $\hat{M}$ over the distribution of the attributes corresponding to the target probability of winning. This final step is straightforward since the averaging is with respect to a known sub-distribution of the overall attribute distribution $\mathcal{A}$. To estimate the probability of a female candidate winning against a male candidate (i.e., equation (7)), for example, the following estimator can be used:

$$\sum_{b,c,b',c'} \Pr([ABC] = [abc], [A'B'C'] = [a'b'c'] | A = a, A' = a') \cdot \ \hat{M}([abc], [a'b'c']), \tag{1}$$

where the sum is taken over the possible values of B, C, B', and C' under the target attribute distribution $\mathcal{A}$, conditional on $A = a$ (female) and $A' = a'$ (male).

The procedure outlined above represents a potentially viable approach to estimating the probability of winning with conjoint data. Unlike the estimation of the AMCE, however, the procedure involves the complex problem of modeling a high-dimensional response function, so care must be taken. In particular, validation of the final model is paramount. We remark on the details of the validation procedure in a subsequent SM section.

C.2 Simulated Estimation

To provide a preliminary assessment of the feasibility of using conjoint data to estimate quantities of interest related to the probability of winning, this section reports a series of simulations. The results presented here will focus specifically on the equation (7) target probability of winning estimand. Can our proposed estimation procedure recover the true parameter values in the face of different forms of mis-specification?

We begin by simulating samples of respondents (voters) taking a conjoint experiment in which they evaluate pairs of candidate profiles and select the candidate in each pair that they prefer. We evaluate three separate simulation scenarios with increasing complexity in voters' underlying preference structures. In all scenarios, every candidate profile is comprised of 8 attributes with varying number of levels. Let X_k for $k = 1, ..., 8$ denote each of the attributes, and let the levels of the attributes be represented as integers. The attributes take the following values, for a total of 2,304 unique combinations of values (i.e. 2,304 unique candidates):

$$\begin{array}{cc} X_1 \in \{0,1\} & X_5 \in \{0,1\} \\ X_2 \in \{0,1,2\} & X_6 \in \{0,1,2,3\} \\ X_3 \in \{0,1,2,3\} & X_7 \in \{0,1,2\} \\ X_4 \in \{0,1\} & X_8 \in \{0,1\} \end{array}$$

We draw 1,001 voters in each simulation scenario (i.e. a simulated sample of 1,001 respondents), a respondent sample size that is within the typical range in applied research. In each scenario, we generate for each individual voter her utility over all 2,304 candidate profiles. The way in which these utilities are generated varies by each of the three scenarios, and becomes increasingly more complex over the three scenarios. Specifically, the utilities are generated according to the following, where i denotes the voter, j the candidate, and k the attribute:

$$U_{j(i)} = \sum_{k=1}^{8} \alpha_{k(i)} X_{kj} + \sum_{k=1}^{7} \sum_{k'>k}^{8} \beta_{kk'(i)} X_{kj} X_{k'j} + \sum_{k=1}^{6} \sum_{k'>k}^{7} \sum_{k''>k'}^{8} \gamma_{kk'k''(i)} X_{kj} X_{k'j} X_{k''j} + \varepsilon_{j(i)}$$

The parameters $\alpha_{k(i)}$, $\beta_{kk'(i)}$, and $\gamma_{kk'k''(i)}$ are randomly drawn according to the specifications in Table 1 in each of the three simulation scenarios.

Scenario 1	Scenario 2	Scenario 3
$\alpha_{k(i)} \sim N(\mu_k, 1)$ where μ_k is a sequence of intervals from -1 to 3	$\alpha_{k(i)} \sim N(\mu_k, 1)$ where μ_k is a sequence of intervals from -1 to 3	$\alpha_{k(i)} \sim N(\mu_k, 1)$ where μ_k is a sequence of intervals from -1 to 3
$\beta_{kk'(i)} = 0$ $\forall\ i, k, k'$	$\beta_{kk'(i)} \sim N(\tau_{kk'}, 1)$ where $\beta_{kk'(i)}$ is a sequence of intervals from -3 to 1	$\beta_{kk'(i)} \sim N(\tau_{kk'}, 1)$ where $\beta_{kk'(i)}$ is a sequence of intervals from -3 to 1
$\gamma_{kk'k''(i)} = 0$ $\forall\ i, k, k', k''$	$\gamma_{kk'k''(i)} = 0$ $\forall\ i, k, k', k''$	$\gamma_{kk'k''(i)} \sim N(\omega_{kk'k''}, 1)$ where $\omega_{kk'k''}$ is a sequence of intervals from -1 to 3
$\varepsilon_{j(i)} \sim N(0, 0.01)$	$\varepsilon_{j(i)} \sim N(0, 0.01)$	$\varepsilon_{j(i)} \sim N(0, 0.01)$

Table 1: Simulation Data-Generating Process: Parameters

As can be seen, in all three simulation scenarios, each voter has a randomly drawn utility function over the candidates. In the first scenario, those utility functions are purely linear and additive with respect to the candidate attributes, while the second scenario also incorporates two-way interactions between all attributes. Finally, the third scenario incorporates both two-way and three-way interactions between all attributes.

We further assume that for each individual voter, when confronted with a choice between two candidates, she will always choose the candidate that has the higher utility for her. Therefore, on the basis of this sample of 1,001 voters in each scenario, it is possible to know which candidate would receive a higher sample vote share and hence win for all two-way match-ups between each of the 2,304 candidates. On the basis of this, we then also know for this sample the true probability that a randomly drawn candidate with a particular level of an attribute will win against a randomly drawn candidate with a different level of the same attribute—i.e. the quantity of interest represented by equation (7) in the main text. The question at hand, however, is how well can we estimate this quantity of interest on the basis of observed conjoint data and inevitable model mis-specification?

To evaluate this, in every simulation scenario, we then simulate conjoint data. That is, for each of the 1,001 voters, we generate randomized conjoint pairs of profiles and simulate the voters' choices, again assuming that each voter will choose the candidate with the highest utility in each pair. In keeping within the range of standard practices employed by applied researchers, we generate 10 pairs of profiles for each voter. For simplicity, in all scenarios and for all profiles in all pairs, we also assume independent uniform distributions of the levels, such that all 2,304 candidate profiles are equally likely for any profile, and all possible pairings between any two profiles are equally likely. Furthermore, we generate 1,000 separate simulated conjoint data sets for each simulation scenario to allow later for a better evaluation of the statistical properties of our estimation procedure.

Hence, for each simulation scenario, we have 1,000 simulated conjoint data sets. In each data set, each of the 1,001 voters has evaluated 10 pairs, for a total of 10,010 pairs evaluated in the data set. For any individual data set, we can then use those data to estimate a model for the probability that any given candidate will be chosen over any other given candidate by a randomly drawn voter (equivalent to the predicted vote share) as a function of both candidates' attributes. That is, adapting the notation presented in the previous section and the main text, we seek to estimate the following function f^*:

$$f^*(\vec{X}_j, \vec{X}_{j'}) \equiv Pr(Y_i = 1|\vec{X}_j, \vec{X}_{j'})$$

where $\vec{X}_j$ and $\vec{X}_{j'}$ denote the vectors of attribute values for two candidates j and j', and Y_i takes the value of 1 if a voter i chooses candidate j over j'.

To estimate f^*, we employ a conditional logistic regression to model the choice between the two candidates in each pair, where each choice task between the two candidates in a pair for each voter constitutes one observation. We construct dummy variables for all attribute levels for both candidates in a pair, and we include all two-way interactions in addition to main effects in the model. It is worth noting that because we only include two-way interactions, the model is mis-specified by design for the third simulation scenario and overly flexible for the first simulation scenario. Further, we employ L_2 regularization (i.e. conditional logistic ridge regression) in the estimation, using cross-validation to select the regularization tuning parameter value. This procedure provides an estimate of the function f^* defined above, and we perform this procedure independently for each of the 1,000 data sets in each simulation scenario (i.e. 1,000 separate estimations of f^* in each scenario).

On the basis of an estimated $\hat{f}^*$, we can then, for any hypothetical match-up between any two of the 2,304 candidates, estimate the probability of winning (i.e. vote share) of each candidate. This allows us to predict whether or not any particular candidate will win in a match-up against any other candidate (i.e. estimate the majority classifier denoted by $\hat{M}$ in the previous section and main text). Then, on the basis of aggregating the appropriate two-way match-up predictions, we can estimate the probability that a candidate with a specific level for a particular attribute (e.g. level 1 for the first attribute) will win a contest against a candidate with a different level for the same attribute (e.g. level 0 for the first attribute). This quantity of interest is estimated using an equation analogous to equation (1) above.

For each data set in each of the simulation scenarios, we use the corresponding $\hat{f}^*$ to compute values of this quantity of interest for all contrasts between any two levels of each attribute. We then compare these estimates to the true values in our simulated sample to assess the behavior of the estimation procedure. The results are shown in Figures 1-3. In each figure, the y-axis denotes all of the possible attribute-level contrasts pertaining to the quantity of interest. For instance, the contrast "8-1 vs 8-0" pertains to the probability that a candidate whose eighth attribute is level 1 will prevail over a candidate whose eighth attribute is level 0. The x-axis shows the probability of winning. The true values for each contrast are indicated by the red dots. The estimated values on the basis of the simulated conjoint data are shown by the black dots and bars. The black dots show the mean estimates across all 1000 simulated conjoint data sets, and the thick, medium, and thin bars denote ± 1 standard deviation, ± 2 standard deviations, and the minimum and maximum of the estimates.

As can be seen, the estimation procedure exhibits good and well-behaved performance across all three simulation scenarios, with mean estimates coming extremely close to the truth across all of the contrasts and a reasonably limited amount of variation in the estimates. This performance is particularly notable in the case of the third simulations scenario, where the model (which contains only two-way interactions) is mis-specified by design. As can be seen, in spite of the mis-specification of f^*, the aggregation entailed by the quantities of interest leads to a relatively well-behaved estimation of the probabilities of winning. Table C.2 gives a summary of the results, providing for each scenario the average (across all contrasts) values of the bias, absolute bias, standard deviation, and root mean squared error (RMSE) of the estimates.

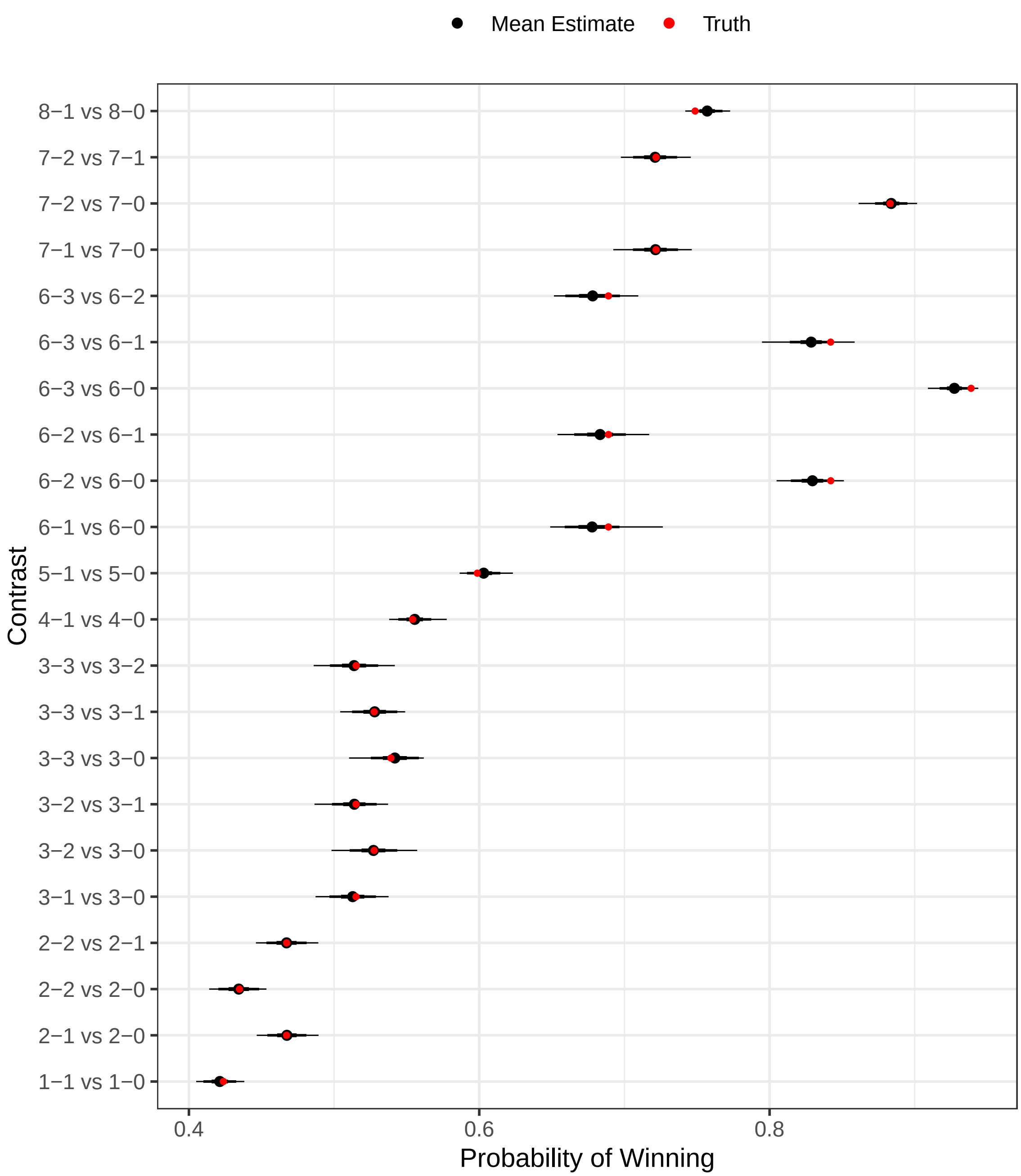

Figure 1: Probability of Winning, Results of Simulation Scenario 1

It is important to note that these are the results of only a single set of simulations. As with all simulations, results can be sensitive to the particular specifications used to define the data-generating process. At the very least, however, this series of simulations does provide baseline

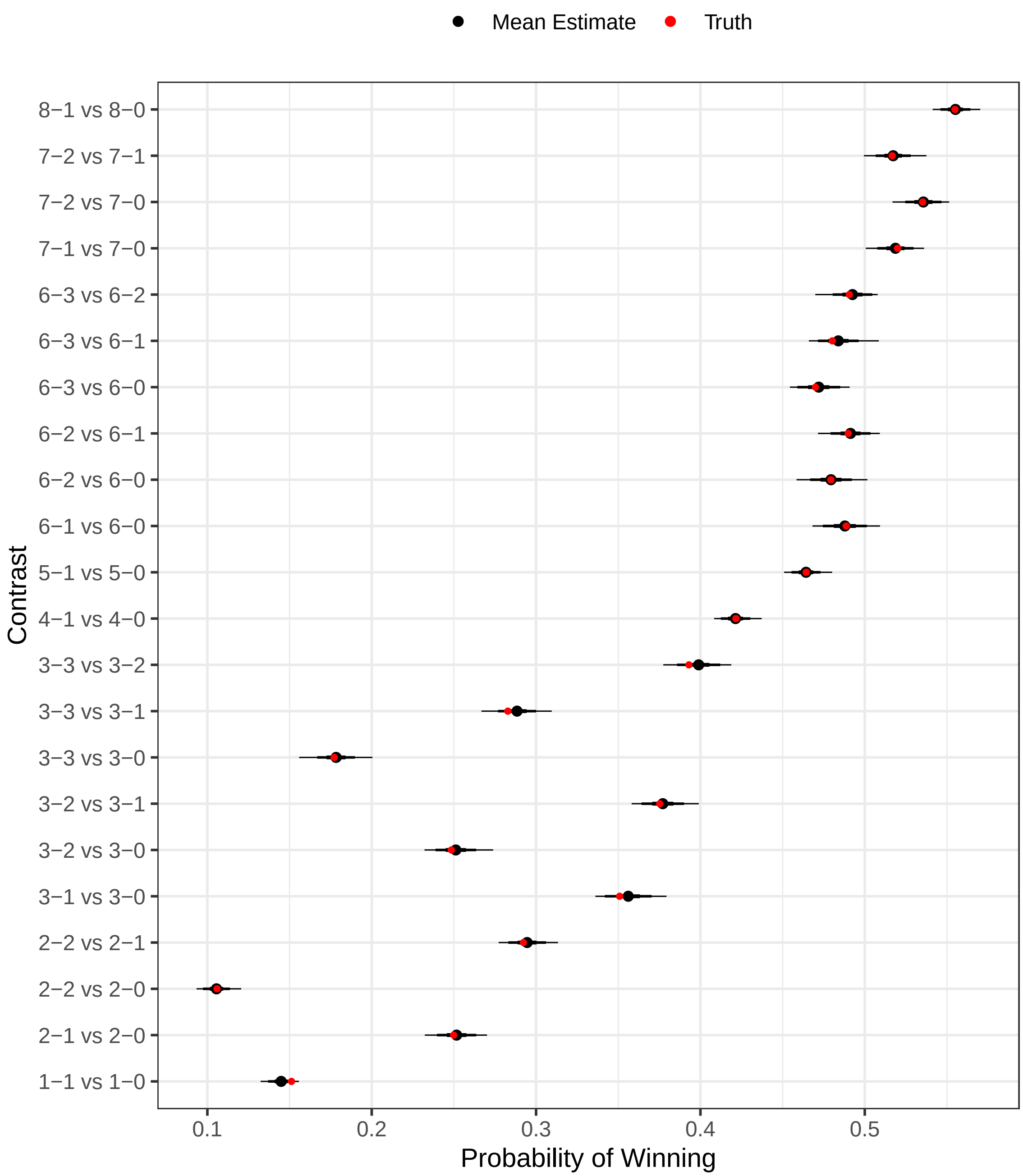

Figure 2: Probability of Winning, Results of Simulation Scenario 2

evidence of the feasibility of using conjoint data to estimate quantities of interest related to the probability of winning. Accordingly, it highlights the value of future research to assess how well this procedure would perform for more complicated data-generating processes, as well as to evaluate the relative performance of other modeling techniques (e.g. random forests, boosted trees, neural nets, etc.) as an alternative to a conditional logistic ridge regression. Recall that each data set in the simulations presented here is comprised of 1,001 respondents, each having evaluated 10 pairs

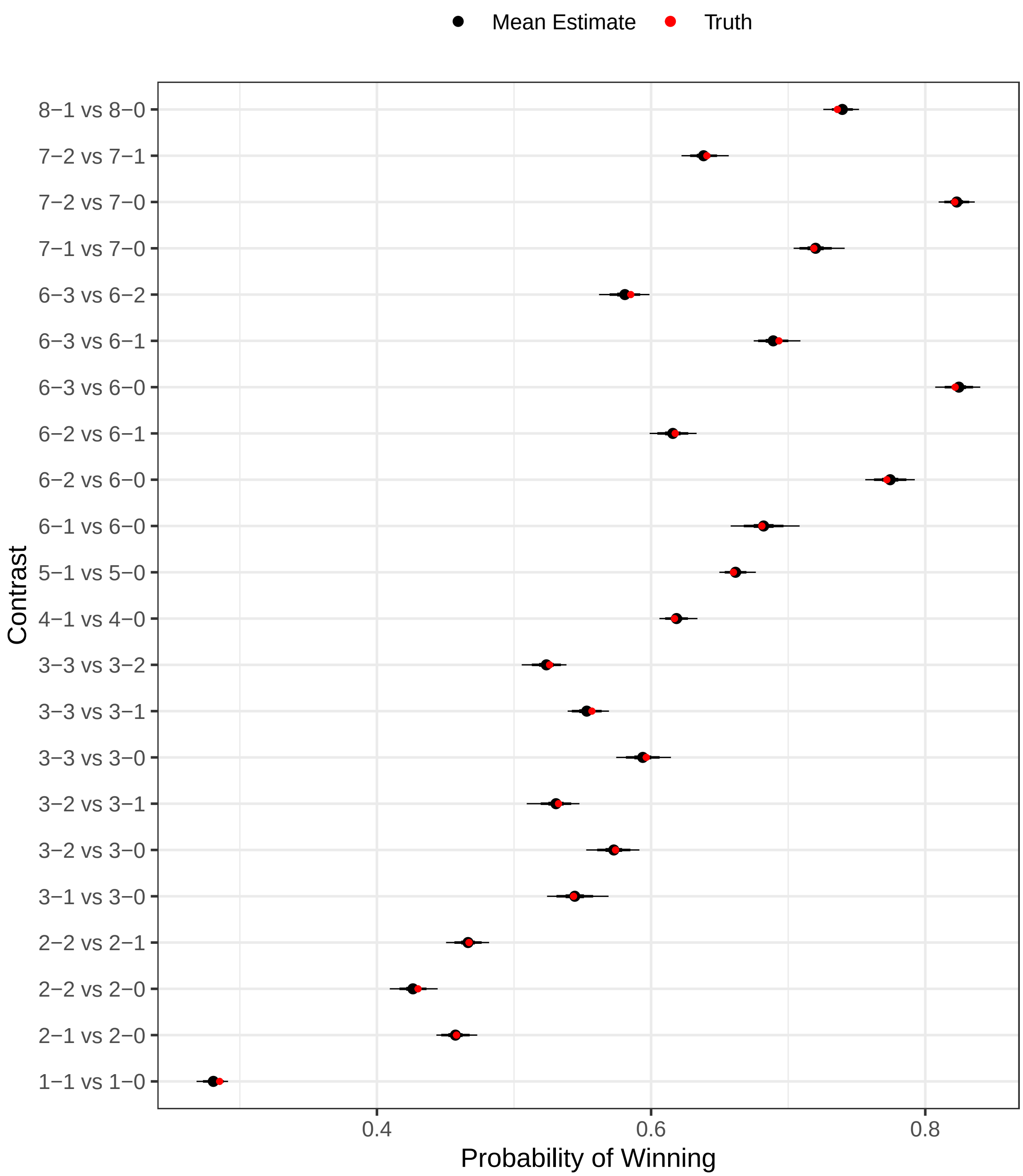

Figure 3: Probability of Winning, Results of Simulation Scenario 3

of candidates that vary along 8 attributes—all values chosen to be within the normal design range of conjoint experiments undertaken by applied researchers. Future research could also evaluate the robustness of these procedures to smaller samples of respondents and/or smaller numbers of tasks.

	Scenario 1	Scenario 2	Scenario 3
Average Bias	-0.00259	0.00126	-0.00080
Average Absolute Bias	0.00423	0.00208	0.00230
Average Standard Deviation	0.00727	0.00570	0.00530
Average RMSE	0.00919	0.00630	0.00590

Table 2: Simulation Summary Results

C.3 Additional Remarks

In an application of the methods above using actual rather than simulated conjoint data, the true values of the quantities of the interest will of course be unknown, and hence other methods of validation are needed. In particular, due to the model dependence of the procedure for estimating the probability of winning described above, validation of the final model is paramount. There is of course no reason to believe, nor do we even need to assume, that the final fitted model perfectly represents the true underlying data-generating process. (Indeed, in the simulations presented above, the models were incorrect by design.) After all, the purpose of these procedures is not to estimate model-specific parameters that themselves are meant to represent particular estimands of interest. Instead, the goal is to learn a model $\hat{f}$ that produces good predictions such that $\hat{f}(a, b, c, a', b', c') \approx f(a, b, c, a', b', c')$. Model validation and evaluation can thus proceed according to standard best practices in machine learning and statistical learning theory, making use of performance metrics that are a function of out-of-sample or cross-validation predictions and the corresponding true outcome values.

Given the focus on estimating the probability of winning, one's first instinct might be to simply compute the out-of-sample or cross-validation classification accuracy of $\hat{M}$. However, while classification accuracy would be informative, it would be insufficient and potentially misleading in terms of the usefulness of the model for predicting the majority-vote outcomes of match-ups. For instance, consider a profile match-up $([abc], [a'b'c'])$ where the true average vote share is 0.55 (i.e. 55% of the population of interest would choose $[abc]$ over $[a'b'c']$). In this case, even if one had perfectly modeled f and had data on this match-up for the entire population, the classification accuracy of $\hat{M}$ at the individual level would be 0.55. This is an underwhelming classification accuracy, but it does not suggest a poorly trained model for our purposes; quite to the contrary, a perfect model would exhibit a classification accuracy of 0.55 if applied to randomly sampled voters' evaluations of this match-up. Classifiers are inherently limited by the signals in the data.

In other words, the focus on predicting the outcome of a match-up at the aggregate level (i.e. which of two candidate profiles would win the majority of votes among a population of interest) means that the classification accuracy of $\hat{M}$ at the individual level (i.e. whether or not $\hat{M}$ accurately predicts a randomly sampled individual's vote $\{0, 1\}$ for a particular match-up) is neither of primary interest nor necessarily even indicative of the quality of the model $\hat{f}$. Since estimates of $M([ABC],[A'B'C'])$ must necessarily happen at some level of aggregation, validation/evaluation of the model must also occur at some level of aggregation. Calibration analysis methods from statistical and machine learning are well-suited for this purpose.

Calibration analysis is a method of assessing the reliability of predicted probabilities. In an ideal world, one would have a perfectly specified and fitted model and hence its predicted probabilities would equal the true probabilities. This is of course not possible in reality, but we may still hope that the predicted probabilities closely approximate the true probabilities. However, empirically assessing this at the individual level is impossible because underlying probabilities are never truly observed. In addition, the true underlying vote share for any matchup $([abc],[a'b'c'])$ is also unobserved given the dimensionality of the feature space and randomization of the attributes. However, the reliability of a model's predicted probabilities can still be (partly) assessed by aggregating the data into bins.

Specifically, for each data point (i.e. each observed matchup evaluation), $\hat{f}$ would be applied to formulate a cross-validated predicted probability, and those predicted probabilities would then be binned into intervals (e.g. 20 intervals of length 0.05 from 0 to 1). Within each bin, the average predicted probability would be computed and compared against the true fraction of 1's in the data points belonging to that bin. Predicted probability averages that are approximately equal to the true fraction of 1's in each bin would be evidence of a well-calibrated model. This would then provide support, albeit not definitive, for the claim that $\hat{f}(A,B,C,A',B',C')$ is meaningfully approximating $\mathbb{E}_{\mathcal{V}}[Y_i([ABC],[A'B'C'])]$, in which case it would then be possible to provide reliable estimates of $M([ABC],[A'B'C'])$.

Figures 4-6 demonstrate this assessment for the use of our conditional logistic ridge regression specification on the simulated conjoint data. Specifically, for each simulation scenario, the figures provide calibration plots along with histograms of the predicted probabilities for one of the simulated data sets (i.e. one data set of 1,001 voters evaluating 10 matchups each). The calibration plot bins are defined based on the ventiles of the predicted probabilities. As can be seen, the conditional logistic ridge regression modeling procedure is able to produce well-calibrated predicted probabilities, evidence that while the model is incorrect by design, it still constitutes a useful approximation for $\mathbb{E}_{\mathcal{V}}[Y_i([ABC],[A'B'C'])]$.

Note also that if one's focus is solely on predicting the ultimate election outcome in a particular matchup, with no additional interest in accurately estimating the vote share, it need only be the case that for any matchup whose vote share is above 0.5, the estimate of the vote share (i.e. predicted probability) is also above 0.5. What that means is one does not necessarily need a model that is calibrated along the entire $[0,1]$ interval. Instead, it would be sufficient to have, for instance, a model's whose calibration curve hits the identity line at 0.5 and is otherwise

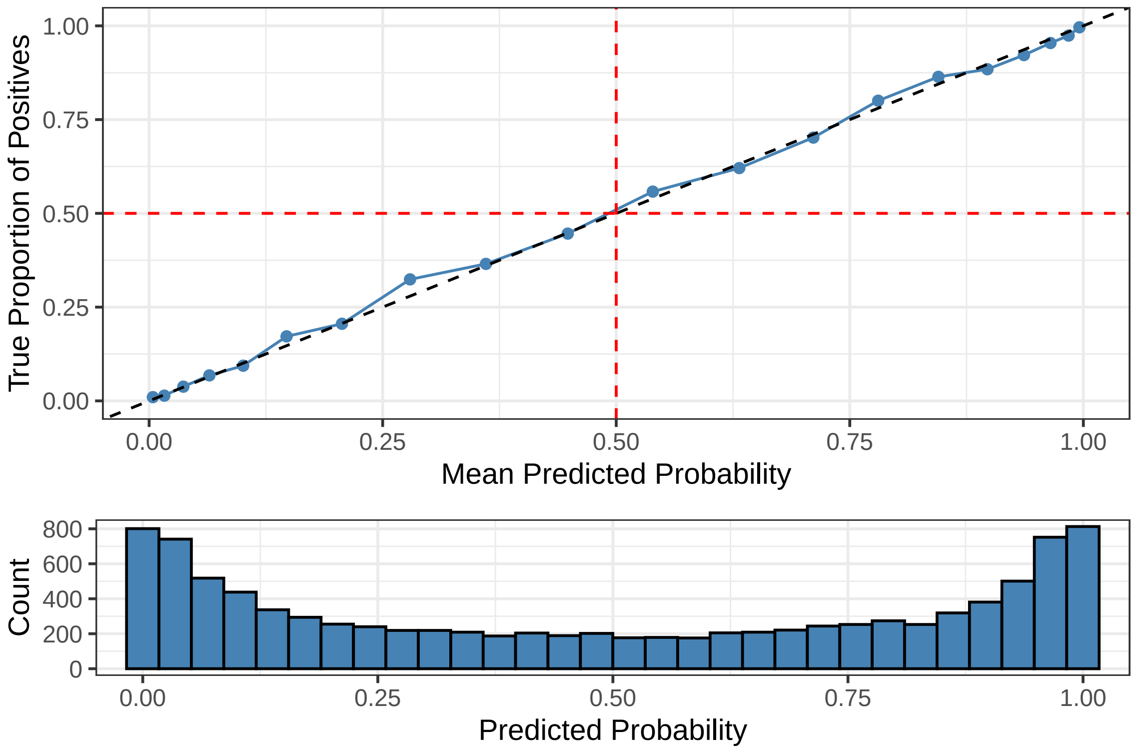

Figure 4: Calibration Plot with Histogram, Simulation Scenario 1

monotonically increasing, which is a strictly easier condition for classification models to satisfy.

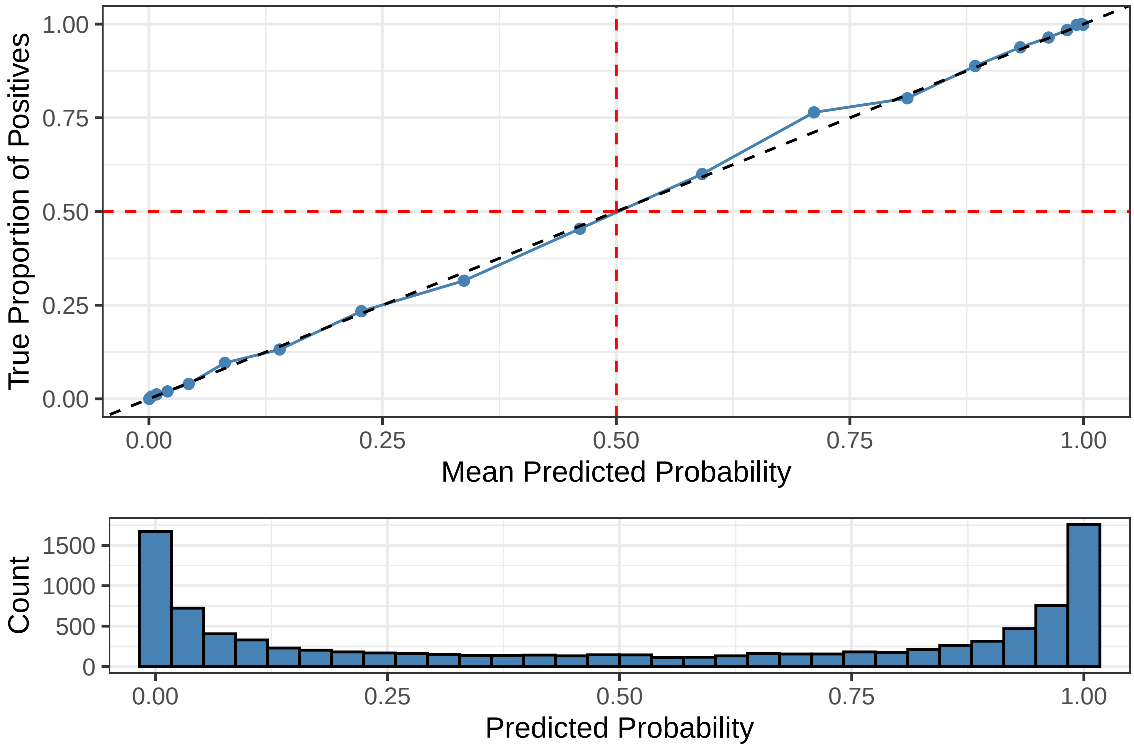

Figure 5: Calibration Plot with Histogram, Simulation Scenario 2

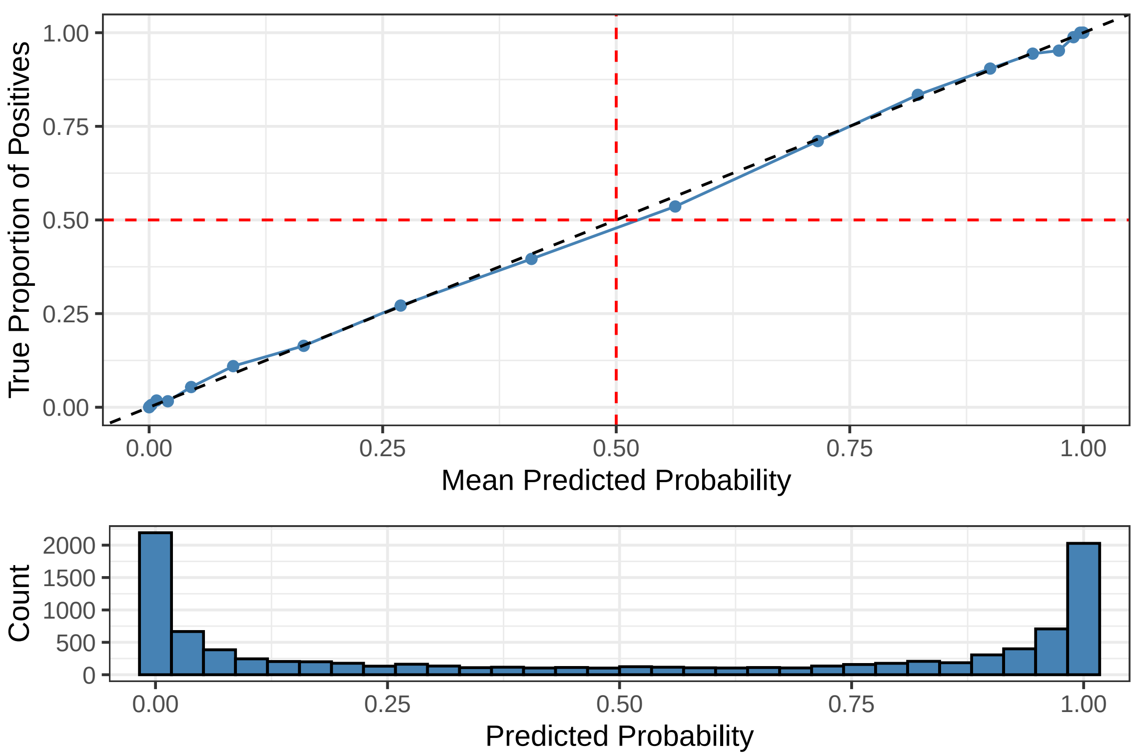

Figure 6: Calibration Plot with Histogram, Simulation Scenario 3

D Alternative QOIs: Fraction of Voters Preferring an Attribute

As discussed in the main text, other possible quantities of interest pertain to the fraction of voters preferring attribute $A = a$ over $A = a'$ and whether that fraction constitutes a majority. For instance, as one target quantity of interest, we can define the fraction of voters preferring $A = a$ over $A = a'$, which is equation (8) in the main text:

$$\mathbb{E}_{\mathcal{V}}\left[\mathbf{1}\{\mathbb{E}_{\mathcal{A}}[Y_i([aBC],[a'B'C'])] > 0.5\}\right]$$

This quantity amounts to first classifying all voters into those preferring $A = a$ over $A = a'$, and then calculating the proportion of voters who prefer $A = a$.

As explained in the main text, one should expect major difficulties in reliably and accurately estimating this (and other related) quantities of interest due to a sparsity problem. Note that by the form of this expression, one first needs to estimate the inner expectation term, $\mathbb{E}_{\mathcal{A}}[Y_i([aBC],[a'B'C'])]$, which equals the proportion of profile comparisons in which *a specific voter i* would choose a profile containing $A = a$ versus another profile containing $A = a'$. Given that respondents are asked to evaluate a limited number of profile pairs (e.g. 3-10 pairs) in standard conjoint experimental designs in political science, this means that for any particular comparison a vs. a' only a handful of observations will be available to estimate that inner expectation for each voter. This is in contrast to the probability of winning quantities of interest, in which the inner expectation is taken with respect to all voters, and hence can be modeled and estimated on the basis of an entire conjoint data set.

As a result, in the present case of the fraction-preferring quantity of interest (and in contrast to the probability of winning quantities), an extremely noisy inner expectation estimator must thus be passed through the indicator function. Poor performance in estimating the fraction-preferring quantity can hence be expected in this situation, as the application of the indicator function to a noisy quantity will result in severe misclassification, and misclassification is always negatively correlated with the true value. Hence, a systematically biased indicator estimator will then be passed into the outer expectation, resulting in a biased outer expectation aggregation. These properties are highlighted by the simulation results presented below.

D.1 Simulated Estimation

On the basis of the same simulated data described above in Section C.2, we also estimated the fraction-preferring quantity of interest corresponding to equation (8) but adapted to the eight-attribute simulation data context. As with the previous simulation results, we estimate this quantity of interest for all possible attribute-level contrasts independently for each of the 1,000 simulated conjoint datasets and each of the three simulation scenarios.

Due to the within-voter sparsity (i.e. each voter only had 10 pairs of profiles), estimation of the inner expectation (analogous to $\mathbb{E}_{\mathcal{A}}[Y_i([aBC],[a'B'C'])]$) for each voter for each contrast was limited to the computation of a simple mean (proportion) on all relevant choice tasks—i.e. profile pairs corresponding to the contrast of interest. Some voters by chance had no relevant observations for any given contrast, in which case they were dropped from the estimation for that particular contrast. In addition, in situations in which the estimate of a particular voter's inner expectation equaled 0.5 (and hence could not be properly classified as preferring one attribute-level or the other), the voter was also dropped from the estimation for that particular contrast. (Note that an alternative way of dealing with this situation would be to randomly choose a classification or to replace the classification with the value of 0.5; this was also assessed, and it leads to worse performance than that of dropping these voters.) For all voters that were not dropped for any given contrast, the mean of their classifications (i.e. the mean of the application of the indicator to their estimated inner expectations) then represents the estimate of the fraction of voters preferring the one attribute-level over the other attribute-level in the contrast.

We can then compare these estimates to the true values in our simulated sample in order to assess the feasibility of using conjoint data to estimate the fraction-preferring quantities of interest. The results are shown in Figures 7-9. As before, the y-axis denotes all of the possible attribute-level contrasts pertaining to the quantity of interest, while now the x-axis shows the fraction of voters who prefer the first attribute-level over the second in the contrast. The true values for each contrast are indicated by the red dots. The estimated values on the basis of the simulated conjoint data are shown by the black dots and bars. The black dots show the mean estimates across all $1,000$ simulated conjoint datasets, and the thick, medium, and thin bars denote ± 1 standard deviation, ± 2 standard deviations, and the minimum and maximum of the estimates. As can be seen, the estimates are severely biased, and systematically so toward 0.5. This behavior is unsurprising; as discussed above, application of the indicator function to a noisy quantity results in severe misclassification, and misclassification is always negatively correlated with the true value. This results in the final estimate of the outer expectation being biased toward 0.5.

As these results demonstrate, conjoint data sets (at least those with the typical number of observations per respondent found in the social sciences) are not well-suited to estimating quantities of interest related to the fraction of voters who prefer a particular attribute-level. It is recommended that researchers interested in such quantities undertake alternative research designs.

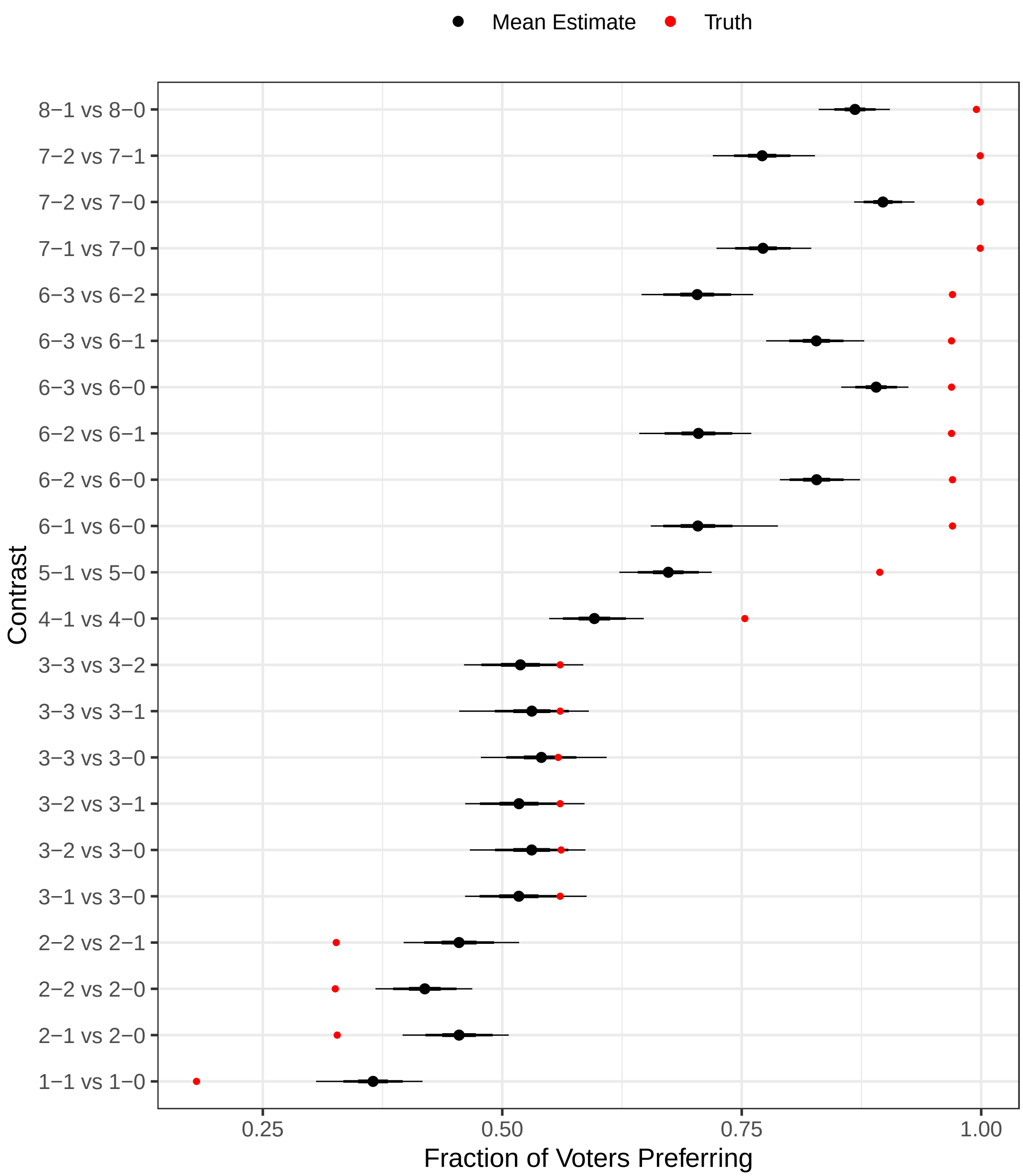

Figure 7: Fraction of Voters Preferring, Results of Simulation Scenario 1

D.2 Additional Remarks

As discussed above and highlighted by the simulation results, the underlying complication with reliably and accurately estimating the fraction-preferred quantities of interest is one of sparsity: estimation of the inner expectation is constrained by the number profile pairs each individual respondent evaluates. One potential strategy to mitigate this problem would be to leverage re-

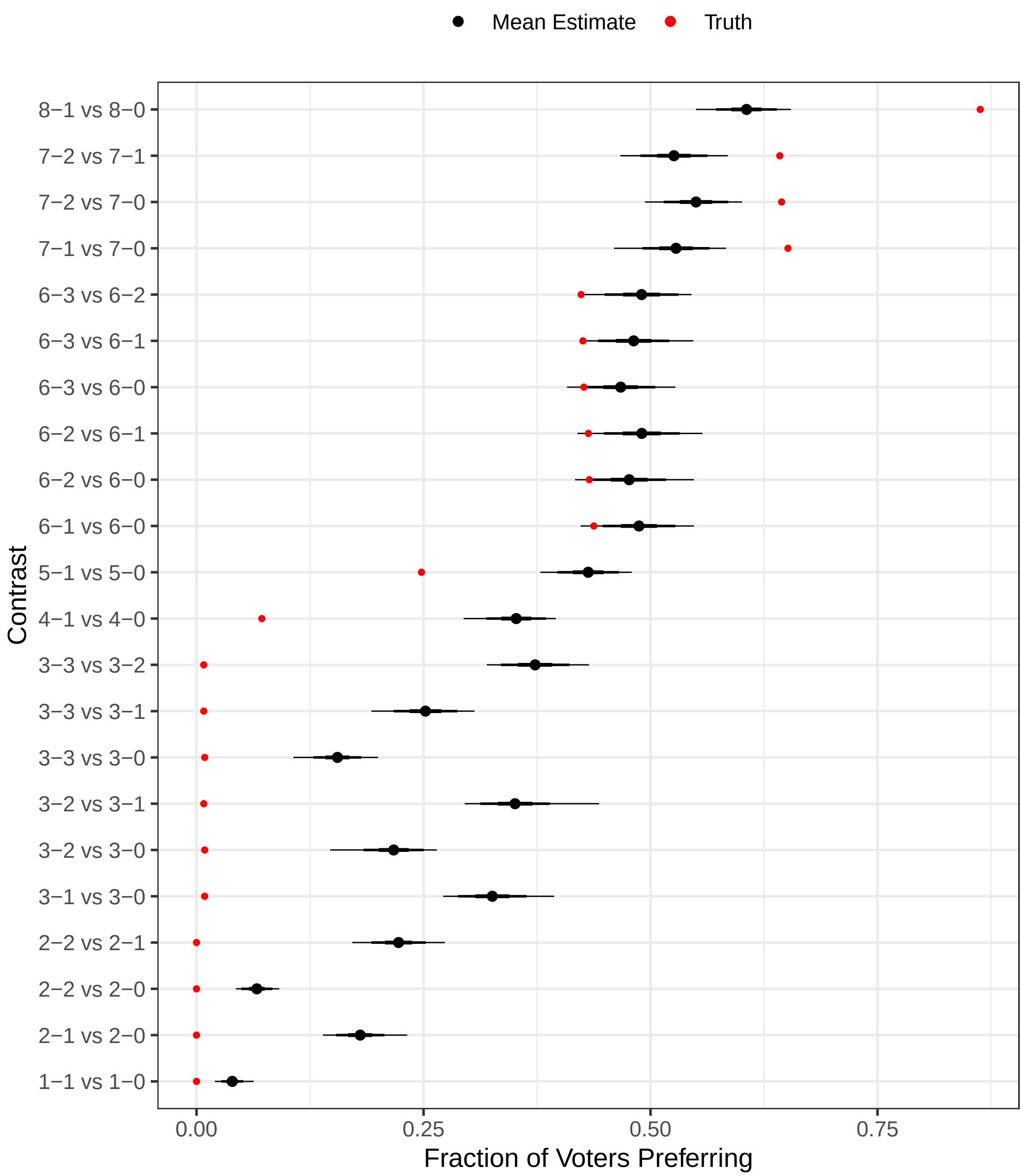

Figure 8: Fraction of Voters Preferring, Results of Simulation Scenario 2

spondent covariates to model the inner expectation as a function of those covariates and hence pool more information together. At its most intuitive level, this would amount to defining strata (or types) of voters on the basis of covariate values and then aggregating data within those strata to estimate the inner expectation at the level of strata rather than individuals. However, regardless of the richness of the stratifying covariates, this strategy would still require the extremely strong assumption of complete homogeneity within strata. In other words, it would not be sufficient for

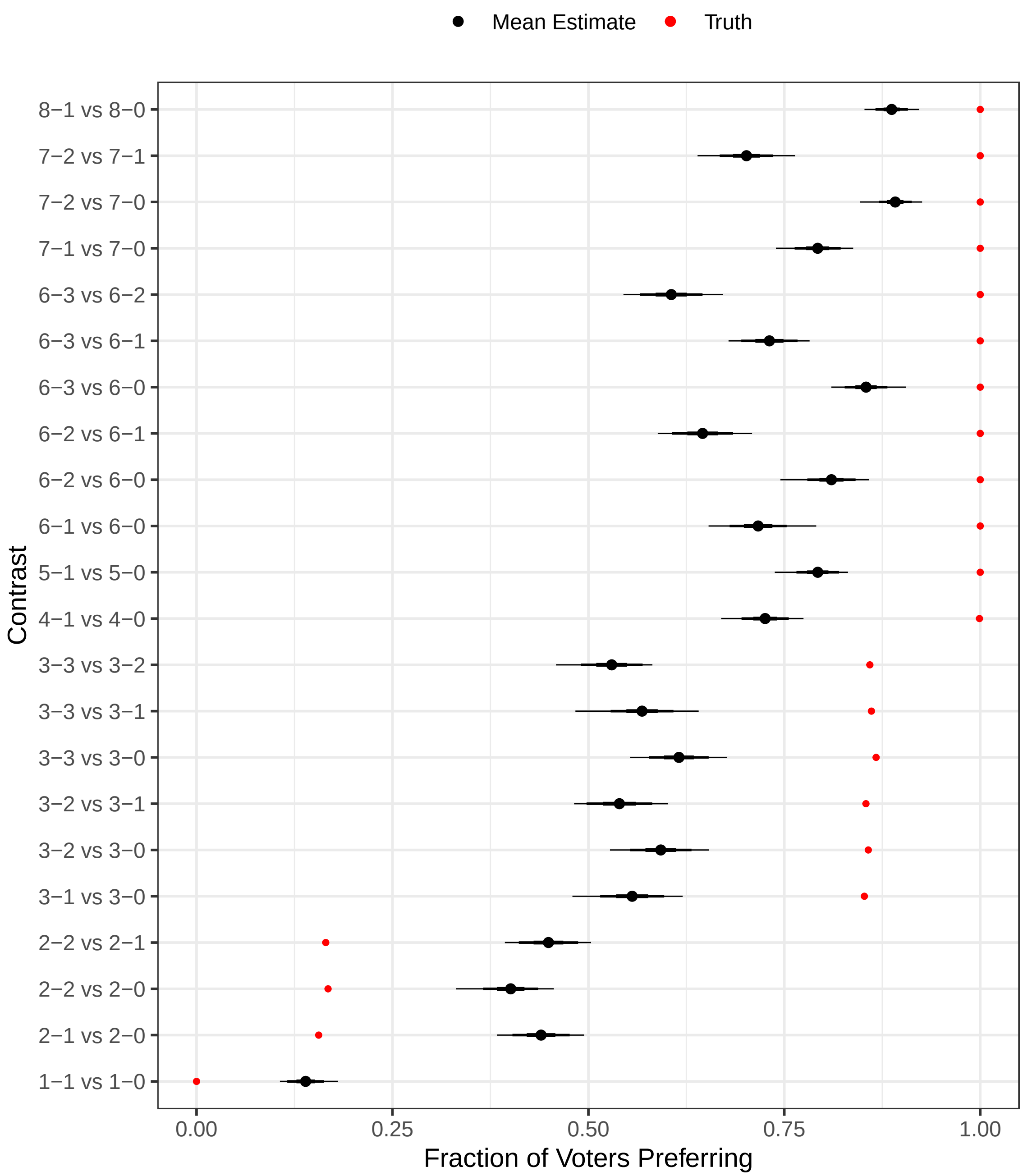

Figure 9: Fraction of Voters Preferring, Results of Simulation Scenario 3

voters within a stratum to have profile rankings that share a common central tendency or are drawn from a common distribution; instead, it would need to be the case that voters within a stratum have the exact same rankings across all possible candidate profiles—an assumption that is far more stringent than standard exchangeability assumptions.[2] Indeed, this is already evident

[2]Although hierarchical Bayes approaches (e.g. Lenk et al., 1996) are often used in marketing to model individual preference heterogeneity, accurate prediction of choice behavior for each individual respondent is still regarded as

from the simulation results above, where all of the simulated voters have rankings that are drawn from a common distribution.

It is also worth noting that the concerns raised about the applicability of conjoint data for studying the fraction-preferring quantity of interest as defined by equation (8) apply as well to a restricted version of this quantity of interest that has been proposed in other work. Specifically, Abramson et al. (2021) propose defining an individual attribute preference for attribute $A = a$ over $A = a'$ as $\mathbf{1}\{\mathbb{E}_{\mathcal{A}}[Y_i([aBC], [a'BC])] > 0.5\}$, thereby considering only *ceteris paribus* comparisons (i.e. B and C are equal across the two profiles). Under this definition, the fraction of voters preferring $A = a$ over $A = a'$ simplifies to

$$\mathbb{E}_{\mathcal{V}}\left[\mathbf{1}\{\mathbb{E}_{\mathcal{A}}[Y_i([aBC], [a'BC])] > 0.5\}\right]. \tag{2}$$

This definition differs from that provided in equation (8) in that it is a function of preference relations only between pairs of profiles identical on all but one attribute. For example, consider the profile [111] in the case of three binary attributes. This version only allows the profile to be compared against three out of the other seven possible profiles, i.e., [011], [101] and [110], which are each identical to the original profile in all but one attribute. Preferences over other profiles—[100], [010], [001] and [000]—are ignored.[3]

In other words, this restricted definition of individual attribute preferences leaves a large number of profile pairs incomparable. This significantly limits the applicability of conjoint survey experiments, not only because this further exacerbates the sparsity problems that plague estimation of the unrestricted version of this quantity, but also because it deviates from the underlying motivation of conjoint experiments, which is to analyze choices about profiles that vary across multiple attributes simultaneously. To see the gravity of the problem, consider our toy example of a conjoint experiment with three binary attributes. Assuming the uniform independent randomization of the attributes (and disregarding the exact ties), the probability that a randomly generated pair results in a *ceteris paribus* comparison in which all attributes are equal save one is $3/7 \simeq .43$. That is, the expected proportion of conjoint tasks that provide *any* information about respondents' preferences per this restricted definition is only 43%, with the remaining 57% of the data contributing nothing. Moreover, for a given attribute, only one out of seven comparisons ($\simeq$ 14%) is considered informative about respondents' preferences. The signal-to-noise ratio continues to decline rapidly as the number of attributes increases to more realistic settings, rendering most

a challenging inferential problem for these models.

[3]The limitation of focusing on *ceteris paribus* comparisons is not readily apparent in the framework Abramson et al. (2021) initially uses to prove its main results, since the framework rules out any interaction between attributes by construction. Under the no-interaction assumption, if $\exists\, b, c$ such that $[1bc] \succ [0bc]$ then $[1b'c'] \succ [0b'c']$ for any $b' \in \{0, 1\}$ and $c' \in \{0, 1\}$, making consideration of all but one *ceteris paribus* comparison per attribute redundant. Although analytically convenient, this no-interaction assumption is unrealistically restrictive as a framework for voter preferences and therefore of limited utility to empirical election scholars.

of the actual choice data "uninformative" by definition. With ten binary attributes, for example, only 10 out of 1,023 pairs ($\leq 1\%$) are *ceteris paribus* and thus informative about respondents' preferences per this restricted definition.

Defining preferences based exclusively on *ceteris paribus* comparisons is not only problematic for comparing profiles, but also for understanding individual attribute preferences. To illustrate, consider a voter choosing a male white Democratic candidate (e.g., [000]) over both a female White Republican candidate ([101]) and a male Black Republican candidate ([011]). According to the restricted definition of individual preferences, these two choice outcomes contain *no* information about the voter's preference between a Democratic candidate and a Republican, since neither is a *ceteris paribus* comparison with respect to party affiliation. In the real world, virtually no elections are about *ceteris paribus* contests between candidates; no two candidates for office differ in just one way. Hence, based on the restricted definition of individual attribute preferences presented in Abramson et al. (2021), individual votes in almost all actual elections reveal nothing about the voters' preferences about candidates attributes such as partisanship, race, or gender.

References

Abramson, S. F., K. Koçak, and A. Magazinnik (2021). What do we learn about voter preferences from conjoint experiments? *American Journal of Political Science (forthcoming)*.

Bansak, K., J. Hainmueller, D. J. Hopkins, and T. Yamamoto (2022). Replication Materials for: Using Conjoint Experiments to Analyze Election Outcomes: The Essential Role of the Average Marginal Component Effect (AMCE). `https://doi.org/10.7910/DVN/NBRBHO`, Harvard Dataverse, V1.

Egami, N. and K. Imai (2019). Causal interaction in factorial experiments. *Journal of the American Statistical Association 114*(526), 529–540.

Lenk, P. J., W. S. DeSarbo, P. E. Green, and M. R. Young (1996). Hierarchical bayes conjoint analysis. *Marketing Science 15*(2), 173–191.

McFadden, D. (1986). The choice theory approach to market research. *Marketing science 5*(4), 275–297.

McFadden, D. L. (1974). Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, pp. 105–142. New York: Academic Press.

Prinzie, A. and D. Van den Poel (2008). Random forests for multiclass classification. *Expert systems with Applications 34*(3), 1721–1732.

Reid, S. and R. Tibshirani (2014). Regularization paths for conditional logistic regression. *Journal of statistical software 58*(12).