Supporting Information (Online)

The Limitations of Using Forced Choice in Electoral Conjoint Experiments

Giancarlo Visconti* Yang Yang†

January 29, 2025

Contents

*Assistant Professor, Department of Government & Politics, University of Maryland, College Park; gvis@umd.edu.

†Ph.D. Candidate, Department of Political Science and Center for Social Data Analytics, Pennsylvania State University;yky5272@psu.edu

A List of Published Articles Using Conjoint Analysis in Electoral Studies

Paper	Journal	Context of Study	Forced-Choice Design
Thomas et al (2024)	AJPS	India	Y
Greene (2022)	AJPS	Mexico	Y
Spater (2022)	AJPS	India	Y
Costa (2021)	AJPS	US	Y
Frederiksen (2022)	APSR	Many	Y
Teele, Kalla and Rosenbluth (2018)	APSR	US	Y
Carnes and Lupu (2016)	APSR	US, UK, and Argentina	Y
Rains and Wibbels (2023)	BJPS	India	Y
Loewen and Rheault (2021)	BJPS	US and Canada	Y
Campbell et al (2019)	BJPS	UK	Y
Schuler (Forthcoming)	CPS	Vietnam	Y
Laterzo (2024)	CPS	Argentina and Brazil	Y
Ventura, Ley and Cantú (2023)	CPS	Mexico	Y
Hankla et al (2023)	CPS	India	Y
Singh (2022)	CPS	Argentina	Y
Chou et al (2021)	CPS	Germany	Y
Carter (2021)	CPS	Peru	Y
Clayton et al (2019)	CPS	Malawi	Y
Agerberg (2019)	CPS	Spain	N
Arceneaux and Wielen (2023)	Elect. Stud.	US	Y
Busby (2022)	Elect. Stud.	US	Y
Kim (2021)	Elect. Stud.	Kenya	Y
Kang et al (2021)	Elect. Stud.	Australia	Y
Foulon and Reyes-Housholder (2021)	Elect. Stud.	Uruguay, Argentina and Chile	Y
Shockley and Gengler (2020)	Elect. Stud.	Qatar	Y
Vivyan et al (2020)	Elect. Stud.	Austria, Germany and UK	Y
Badas and Stauffer (2019)	Elect. Stud.	US	Y
Arnesen, Duell and Johannesson (2019)	Elect. Stud.	Norway	Y
Gift and Lastra-Anadón (2018)	Elect. Stud.	US	Y
Matsuo and Lee (2018)	Elect. Stud.	UK	Y
Aguilar, Cunow and Desposato (2015)	Elect. Stud.	Brazil	Y
Frederiksen (2024)	JOP	US, UK, Czech, Mexico, and South Korea	Y
Driscoll and Nelson (2023)	JOP	US	Y
Portmann (2022)	JOP	Switzerland	Y
Henderson et al (2022)	JOP	US	Y
Eshima and Smith (2022)	JOP	Japan	Y
Weaver (2021)	JOP	Peru	Y

Continued on next page

Paper	Journal	Context of Study	Forced-Choice Design
Magni and Reynolds (2021)	JOP	US, UK, and New Zealand	Y
Bakker, Schumacher and Rooduijn (2021)	JOP	US	Y
Schneider (2020)	JOP	Germany	Y
Chauchard, Klašnja and Harish (2019)	JOP	India	Y
Campbell et al (2019)	JOP	UK	Y
Peterson and Simonovits (2018)	JOP	US	Y
Eggers, Vivyan and Wagner (2018)	JOP	UK	N
Peterson (2017)	JOP	US	Y
Robinson (2023)	PolBeh	US	Y
Wood (2023)	PolBeh	US	Y
Visconti (2022)	PolBeh	Chile	Y
Magni and Reynolds (2022)	PolBeh	US, UK, and New Zealand	Y
Manento and Testa (2022)	PolBeh	US	Y
Rehmert (2022)	PolBeh	Germany	Y
Funck and McCabe (2022)	PolBeh	US	Y
Neuner and Wratil (2022)	PolBeh	Germany	Y
Saha and Weeks (2022)	PolBeh	US and UK	Y
Martin and Blinder (2021)	PolBeh	UK	Y
Rosenzweig (2021)	PolBeh	Kenya	Y
Blackman and Jackson (2021)	PolBeh	Tusnia	Y
Mummolo, Peterson and Westwood (2021)	PolBeh	US	Y
Crowder-Meyer et al (2020)	PolBeh	US	Y
Leeper and Robison (2020)	PolBeh	US	Y
Ono and Burden (2019)	PolBeh	US	Y
Kirkland and Coppock (2018)	PolBeh	US	Y
Sances (2018)	PolBeh	US	Y
DeMora et al (2022)	POQ	US	Y
Lehrer, Stöckle and Juhl (2024)	PSRM	Germany	N
Dai and Kustov (2023)	PSRM	US	Y
Erlich and Beauvais (2023)	PSRM	Ukraine	N
Ono and Yamada (2020)	PSRM	Japan	Y
Mares and Visconti (2020)	PSRM	Romania	Y
Horiuchi, Smith and Yamamoto (2020)	PSRM	Japan	Y
Franchino and Zucchini (2015)	PSRM	Italy	Y
Carlson (2015)	WP	Uganda	Y

B Simulating Respondents' Voting Behavior in Forced-Choice Conjoint Analyses

In this section, we simulate plausible voting behaviors of respondents under both the typical and proposed designs, evaluating the potential of our proposed design to improve how we conduct electoral conjoint experiments.

B.1 Forced-Choice Design and Forced Voting Decisions

When an electoral conjoint experiment compels respondents to choose between two candidate profiles, they may approach the task strategically, particularly when they do not have a strong preference for either profile. Consider a nationally representative sample that includes respondents who have never participated in real-world elections and those who occasionally abstain from voting despite being eligible. What plausible decision-making process might these respondents follow if researchers deprive them of the option to abstain or cast a protest (blank or null) vote, forcing them to choose between two candidates?

For respondents who are eligible voters but consistently abstain from real-world voting, one possible coping approach when forced to choose is to appear *cooperative*. They might carefully evaluate each profile based on certain criteria in a way they would not in real-world elections. This behavior would not bias inference as long as, on average, they exhibit voting preferences similar to those of regular voters. However, existing studies have shown that the political preferences of non-voters and regular voters are distinct. If such respondents answer in a socially desirable way by heavily valuing or devaluing certain attributes, such as being more or less educated, experienced, younger or older, or holding certain policy positions, this could distort the aggregated preferences in an unpredictable manner. Consequently, the measurement error bias can either be upward or downward.

They might also approach the task by selecting between two candidates in a completely *random* manner, similar to flipping a coin. At first glance, this might seem inconsequential: if respondents whose true preference is to abstain are forced to randomly select between two candidates, their random choices would not cause an estimation problem when we are interested in the difference in aggregated preferences between the two profiles, given that measurement errors introduced by random choices can be canceled out by the law of large numbers. However, since the estimators of interest, AMCEs, are calculated by the weighted average of differences in means, we are essentially interested in the average preference, that is, which candidates with certain attributes are more likely to be selected on average. Forcing respondents who would prefer to abstain to make random voting choices introduces downward biases. This is because, in a forced-choice conjoint analysis, the difference in aggregated preferences between candidates is averaged over an entire sample that includes everyone, even though in reality, non-voters should never be counted.

Likewise, although regular voters actively engage in elections, this does not necessarily mean they always vote exclusively for one candidate or another. In reality, they might prefer to cast a protest vote (blank or null) or abstain as needed. However, under a forced-choice design, when they do not have a strong preference for either profile and are compelled to choose between the two, they might make a *random* choice or a *trade-off* choice by considering some second-order attributes to cope with the conjoint experiments.[1] This type of decision-making is likely when they find both candidates unattractive or indifferent. The estimates can be biased if respondents who would prefer to abstain are forced to make either a random or trade-off choice, or if respondents who would prefer to cast a protest vote are forced to make a random or a trade-off choice.

[1]By secondary attributes, we mean that if respondents do not have a strong preference for either profile and are forced to make a choice, they might select certain attributes they care about more as the main criteria to decide which candidate they slightly prefer in the second order.

B.2 Bootstrapping Simulation Studies

We now present bootstrapping simulation studies to demonstrate how the typical forced-choice conjoint design can introduce biases while the proposed unforced-choice design improves estimation by reducing design-induced biases. To do this, we first assume that each respondent (voter) has a randomly drawn utility function, which is linear and additive. This function aggregates multiple dimensions into a composite criterion using the information provided by each candidate profile, defined by eight attributes with varying number of levels:[2]

Religion: l_1 ={Catholic, Protestant, None}

Education: l_2 ={State university, Small college, Community college, Ivy League college, No college}

Profession: l_3 ={Lawyer, High school teacher, Business owner, Farmer, Doctor, Car dealer}

Income: l_4 ={32K, 54K, 65K, 92K, 210K, 5.1M}

Race: l_5 ={Black, Asian American, Native American, Hispanic, Caucasian}

Gender: l_6 ={Male, Female}

Military Service: l_7 ={Served, Did not serve}

Age: l_8 ={45, 52, 60, 68, 75}

This produces 54,000 unique combinations of attribute levels, representing distinct candidates. Let each simulated respondent be indexed by $i \in \{1, ..., N\}$. As defined, each conjoint profile is composed of eight attributes represented by the corresponding $\{l_1, ..., l_8\} \in L$ factors, where each factor l has a total of D_l levels. For example, $D_1 = 3$ and $D_2 = 5$ represent the first and second attributes, *Religion* and *Education*, varying with three and five levels respectively. Formally, our latent utility function of respondent i for profile j is defined as follows:

$$U_{j(i)} = \sum_{l=1}^{L} \alpha_{L(i)} * \mathbf{X}_j + \varepsilon_{j(i)} \tag{1}$$

where $\mathbf{X}_j$ is a vector of D_l dummy variables for the levels of attributes contained in profile j. The coefficient $\alpha_{L(i)}$ denotes the specific utility value respondent i earns from any profiles containing a certain level of factor l. All the coefficients are randomly drawn at the individual level from a normally distributed population with predefined means and standard deviations, detailed in Table A3 in Appendix B.2.1.

We assume there are two types of eligible voters (respondents) in the experiment: regular voters and non-voters. These groups have distinct utility functions with different population means,

[2] These attributes and their values do not imply any substantive meaning in the simulation studies; they are used solely to aid in readability. We utilize the primary attributes and values from Hainmueller, Hopkins and Yamamoto (2014). However, we have omitted a few values due to memory constraints and the size limitations of the RStudio environment we are using.

reflecting their unique political preferences, as detailed in Section B.1. However, we do not assume that each respondent always chooses profile j when its latent utility is higher than that of the other profile j' in comparison. In some cases, some respondents might not be interested in politics or might find both candidates unattractive. As a result, they could be indifferent to both candidates and might choose randomly or based on other criteria.

Respondents are complex decision-makers, and a forced-choice design can compel them to make voting decisions that deviate from their unobserved true preferences in various ways. It is challenging to determine the degree and direction of misclassification or external validity bias introduced by a specific type of deviation when various forced decisions are mixed together. However, simulation tools allow us to fully model and customize respondent behaviors. Therefore, we analyze each type of forced voting decision separately as an ideal case to gain a deeper understanding of how a forced-choice design introduces biases. Table A2 summarizes the scenarios in which a forced-choice conjoint design might introduce biases.

Table A2: Simulated Conjoint Scenarios

Scenarios	Type of Voters	True Preference	Forced-Choice Design	
			Observed Choice	Decision-Making
1	Non-Voters	Abstention	Candidate A or B	Random
				Cooperative
	Regular Voters	Candidate A	Candidate A	=
		Candidate B	Candidate B	=
2	Regular Voters	Candidate A	Candidate A	=
		Candidate B	Candidate B	=
		Abstention	Candidate A or B	Random
				Trade-Off
		Protest Vote	Candidate A or B	Random
				Trade-Off

We use a simulated sample of 1,001 respondents (eligible voters) for each simulation scenario. To evaluate the difference between forced-choice and unforced-choice electoral conjoint designs in estimates, we generate randomized conjoint pairs of profiles and simulate the choices of each of the 1,001 voters under both designs for every scenario. Following 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 assume no interactions between attributes and independent uniform distributions of all the levels, ensuring that all 54,000 candidate profiles are equally likely, and all possible pairings between any two profiles are equally likely. Finally, using bootstrapping method, we iterate the simulated conjoint analysis for each simulation scenario under both designs 100 times to allow for a more robust evaluation of the statistical properties of our estimation procedure.

B.2.1 Simulation Set-Up

Table A3: Simulation Data-Generating Process: Parameters

	Regular Function	**Socially Desirable Function**
Religion-Catholic	$\alpha_{11(i)} \sim N(\mu_{11}, 1)$	same
-Protestant	$\alpha_{12(i)} \sim N(\mu_{12}, 1)$	same
-None	$\alpha_{13(i)} \sim N(\mu_{13}, 1)$	same
Education-No college	$\alpha_{21(i)} \sim N(\mu_{21}, 1)$	$\alpha_{21(i)} \sim N(\mu_2 \times -2, 1)$
-State university	$\alpha_{22(i)} \sim N(\mu_{22}, 1)$	same
-Small college	$\alpha_{23(i)} \sim N(\mu_{23}, 1)$	same
-Community college	$\alpha_{24(i)} \sim N(\mu_{24}, 1)$	$\alpha_{24(i)} \sim N(\mu_{24} \times 1.5, 1)$
-Ivy League college	$\alpha_{25(i)} \sim N(\mu_{25}, 1)$	$\alpha_{25(i)} \sim N(\mu_{25} \times 2, 1)$
Profession-Lawyer	$\alpha_{31(i)} \sim N(\mu_{31}, 1)$	same
-High school teacher	$\alpha_{32(i)} \sim N(\mu_{32}, 1)$	same
-Business owner	$\alpha_{33(i)} \sim N(\mu_{33}, 1)$	same
-Farmer	$\alpha_{34(i)} \sim N(\mu_{34}, 1)$	same
-Doctor	$\alpha_{35(i)} \sim N(\mu_{35}, 1)$	same
-Car dealer	$\alpha_{36(i)} \sim N(\mu_{36}, 1)$	same
Income-32K	$\alpha_{41(i)} \sim N(\mu_{41}, 1)$	same
-54K	$\alpha_{42(i)} \sim N(\mu_{42}, 1)$	same
-65K	$\alpha_{43(i)} \sim N(\mu_{43}, 1)$	same
-92K	$\alpha_{44(i)} \sim N(\mu_{44}, 1)$	same
-210K	$\alpha_{45(i)} \sim N(\mu_{45}, 1)$	same
-5.1M	$\alpha_{46(i)} \sim N(\mu_{46}, 1)$	same
Race-Black	$\alpha_{51(i)} \sim N(\mu_{51}, 1)$	same
-Asian American	$\alpha_{52(i)} \sim N(\mu_{52}, 1)$	same
-Native American	$\alpha_{53(i)} \sim N(\mu_{53}, 1)$	same
-Hispanic	$\alpha_{54(i)} \sim N(\mu_{54}, 1)$	same
-Caucasian	$\alpha_{55(i)} \sim N(\mu_{55}, 1)$	same
Gender-Male	$\alpha_{61(i)} \sim N(\mu_{61}, 1)$	same
-Female	$\alpha_{62(i)} \sim N(\mu_{62}, 1)$	same
Military Service-Did not Serve	$\alpha_{71(i)} \sim N(\mu_{71}, 1)$	same
-Served	$\alpha_{72(i)} \sim N(\mu_{72}, 1)$	$\alpha_{72(i)} \sim N(\mu_{72} \times 2, 1)$
Age-45	$\alpha_{81(i)} \sim N(\mu_{81}, 1)$	same
-52	$\alpha_{82(i)} \sim N(\mu_{82}, 1)$	same
-60	$\alpha_{83(i)} \sim N(\mu_{83}, 1)$	$\alpha_{83(i)} \sim N(\mu_{83} \times 0, 1)$
-68	$\alpha_{84(i)} \sim N(\mu_{84}, 1)$	$\alpha_{84(i)} \sim N(\mu_{84} \times -1, 1)$
-75	$\alpha_{85(i)} \sim N(\mu_{85}, 1)$	same
	$\varepsilon_{j(1)} \sim N(0, 1)$	same

where, the sequences μ_{1i} through μ_{8i} are defined as follows: μ_{1i} is a sequence of intervals ranging from -3 to 8; μ_{2i} spans intervals from -5 to 8; μ_{3i} ranges from 0 to 6; μ_{4i} covers intervals from -4 to 5; and μ_{5i} ranges from 0 to 8. Additionally, μ_{6i} consists of two identical random numbers drawn from integers between 2 and 10. The sequence μ_{7i} spans intervals from 0 to 5, while μ_{8i} represents a sequence of decreasing intervals from -8 to 8.

B.2.2 Simulation Study 1

In the first scenario, we assume a nationally representative sample where regular voters constitute 70% and non-voters make up 30%. We further assume that regular voters have clear preferences between each pair of candidates, always choosing the one with the highest utility. In contrast, non-voters either choose randomly or select a candidate in a socially desirable manner by heavily valuing or devaluing certain attributes, such as education, age, and military experience, under the forced-choice design. However, under an unforced-choice conjoint design, these non-voters would choose to abstain. To examine how forcing non-voters to vote introduces measurement error bias, we first assume that all non-voters only choose randomly, and then we consider the scenario where they select a candidate only in the socially desirable way.

We generate 100 random samples from bootstrapping for each decision-making processes non-voters and regular voters might follow in either forced- or unforced-choice conjoint experiments and compute the sample means and sampling distributions. The sampling distributions that we computed provide valuable insights into estimating average preference about which candidates with certain attributes are more likely to be selected. Since the AMCEs are unbiased estimators, the sampling distributions are centered around the true average preference of the population (voters). The spread of the sampling distribution indicates the amount of variability induced by sampling 100 simulated conjoint analyses.

In Figure A1, the left panel compares the mean estimates based on the simulated conjoint data in which non-voters who would prefer to abstain randomly select between candidates with those based on the unforced-choice data where abstention is allowed. Similarly, the right panel compares the mean estimates based on the simulated conjoint data in which non-voters who would prefer to abstain are forced to choose candidates in a socially desirable manner with those based on the unforced-choice data where abstention is available.

As can be seen from Figure A1, if respondents whose true preference is to abstain are forced to randomly select between two candidates, the forced-choice conjoint design tends to generate parameter estimates with smaller magnitudes — downward biases. In contrast, if respondents whose true preference is to abstain are compelled to choose candidates in a socially desirable manner, the forced-choice conjoint design tends to introduce either downward or upward biases, depending on which attributes they might value or devalue more. For example, in our simulation, we specifically assume and set up the socially desirable voting criteria such that, when forced to choose candidates, cooperative abstainers devalue candidates aged 52 and 68, while they largely value candidates educated at Ivy League colleges and those with military service. The results from the right panel confirm that using a forced-choice design causes upward and downward biases in the estimates of those attribute levels, to different degrees, where these biased selection criteria are assumed.

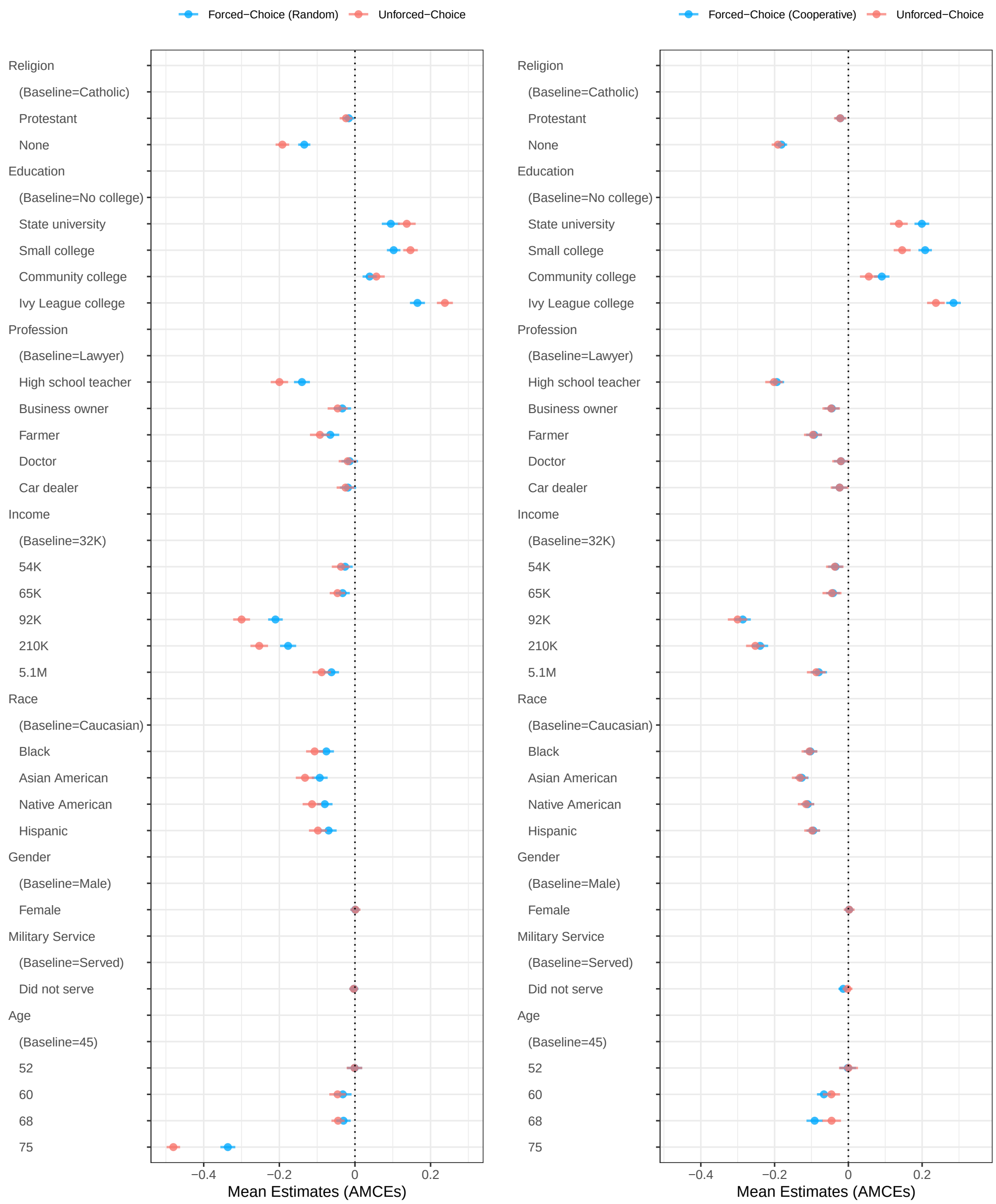

Figure A1: Simulated Choice-Based Mean Estimates in Scenario One. The estimated values based on the simulated forced-choice and unforced-choice conjoint data are represented by the blue and red dots and bars, respectively. All the dots show the mean estimates across all 100 simulated conjoint data sets from bootstrapping, and the bars denote ±1.96 standard deviations and the minimum and maximum of the estimates.

B.2.3 Simulation Study 2

In the second scenario, we assume there are no voters who consistently abstain from voting; instead, all are regular voters. We further assume that a half of these regular voters possess clear preferences, always choosing the candidate with the highest utility and avoiding null votes or abstention under any design. The remaining half might prefer to abstain or cast a protest vote if such options are available when presented with a pair of candidates delivering utilities lower than the median candidate according to their own utility functions. We use this scenario to mirror the real-world cases where some people do not vote for the "lesser of two evils." We assume that when forced to make a decision under the forced-choice design, they might choose to vote randomly or make a trade-off decision by selecting the candidate with the highest utility based on the secondary attributes they care about most, such as profession and age.

In the following simulations, we first examine scenarios where half of the regular voters' true preferences are to abstain, but they must vote either completely randomly or in line with their secondary attribute-based preferences when both profiles have low utilities. We then explore cases where the same group's true preference is to cast a protest vote. Similar to Section B.2.2, we generate 100 random samples and compute the sample means and sampling distributions for each case. Figures A2 and A3 present the results for each case, respectively.

As shown in the left panel of Figure A2, when half of the regular voters' true preferences are to abstain but they vote randomly when confronted with both profiles having low utilities, it results in downward bias for the AMCE estimates. This bias is particularly evident for those attribute levels where respondents have a stronger preference compared to the reference level. However, if voters base their decisions on secondary attributes when both profiles have low utilities, the estimates for these secondary attributes and for those attribute levels where respondents do not have a strong preference compared to the reference level are less likely to be biased, while other estimates show varying degrees of bias.

In the right panel of Figure A2, we simulate a scenario where half of the regular voters' true preference is to abstain, but they end up voting based on second-order attributes, such as profession and age, when faced with profiles offering low utility. Given our assumption that some respondents prioritize profession and age as second-order attributes, we observe that the estimates of AMCEs for these attributes remain largely unbiased. However, for other attributes, especially those where respondents exhibit a stronger preference compared to the reference level, there is a noticeable bias.

In the left panel of Figure A3, we consider a scenario in which half of the regular voters genuinely prefer to cast a protest vote, but they end up voting randomly when both profiles present low utilities. This randomness leads to a downward bias in the AMCE estimates, especially for attribute levels where respondents demonstrate stronger preferences compared to the baseline. Similar to the abstention scenario, this bias is more significant for attributes that voters prioritize heavily in their decision-making process.

In the right panel of Figure A3, we simulate a situation where half of the voters intend to cast a protest vote but instead rely on secondary attributes, such as profession and age, when neither profile provides significant utility. In this case, the estimates of AMCEs for these secondary attributes are mostly unbiased, reflecting the voters' focus on these specific factors. However, for attributes where respondents hold more pronounced preferences compared to the baseline, we observe varying degrees of bias, with some estimates deviating significantly from the true preferences, either downward or upward.

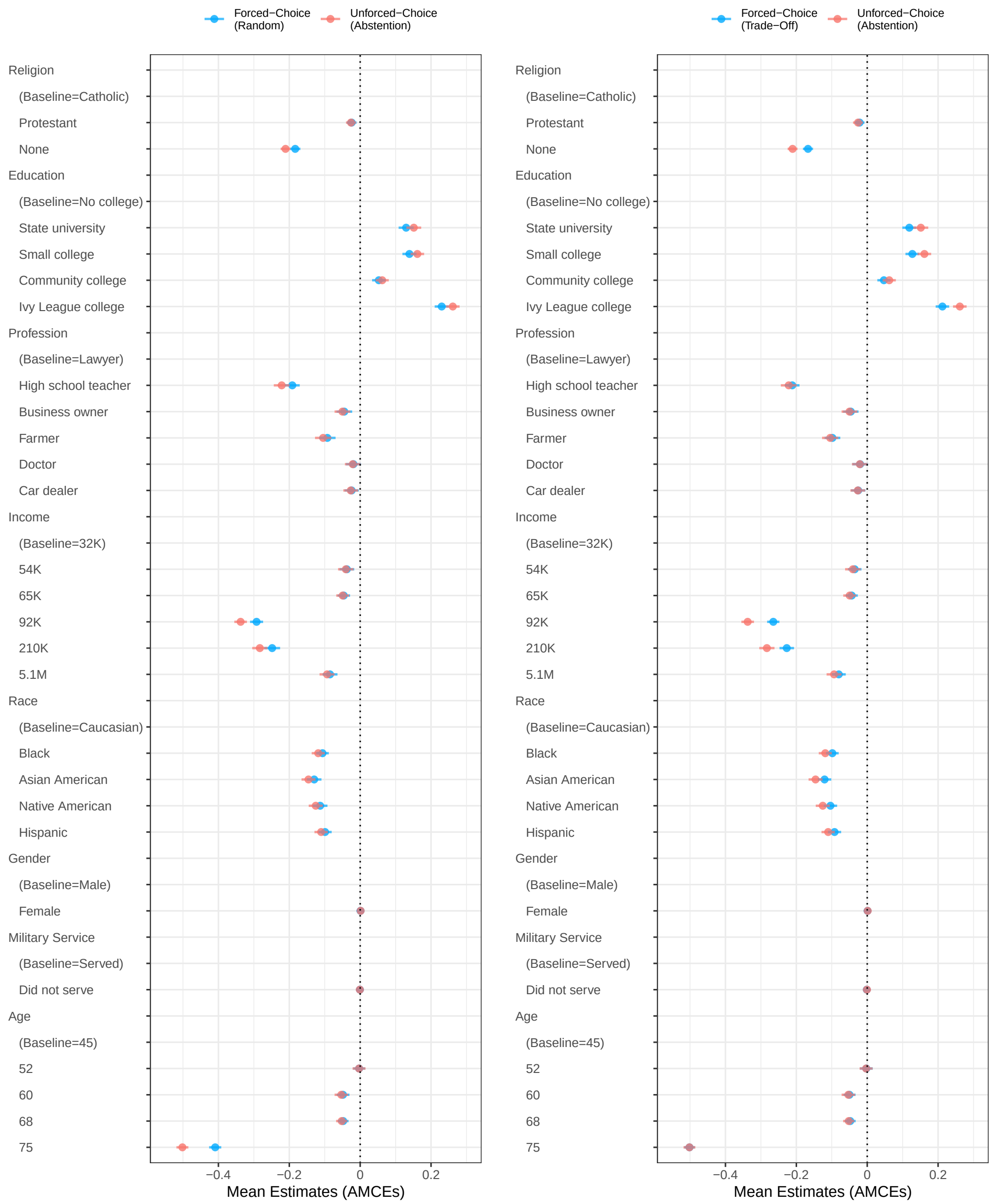

Figure A2: Simulated choice-based mean estimates in Scenario Two, where half of the respondents' true voting preference is to abstain. The estimated values based on the simulated forced-choice and unforced-choice conjoint data are represented by the blue and red dots and bars, respectively. All the dots show the mean estimates across all 100 simulated conjoint data sets from bootstrapping, and the bars denote ± 1.96 standard deviations and the minimum and maximum of the estimates.

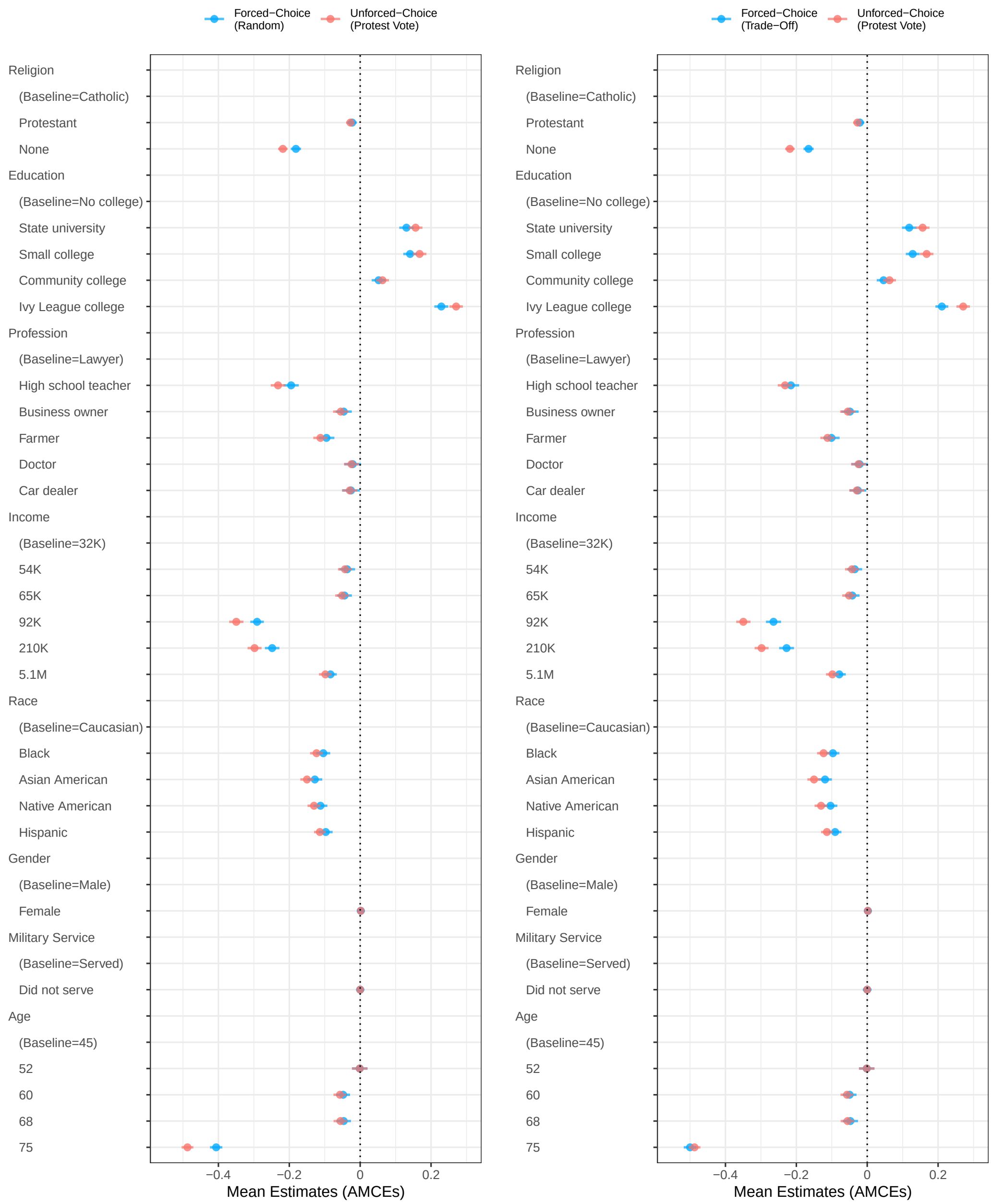

Figure A3: Simulated choice-based mean estimates in Scenario Two, where half of the respondents' true voting preference is to to cast a protest vote. The estimated values based on the simulated forced-choice and unforced-choice conjoint data are represented by the blue and red dots and bars, respectively. All the dots show the mean estimates across all 100 simulated conjoint data sets, and the bars denote ±1.96 standard deviations and the minimum and maximum of the estimates.

C Additional Methods and Results

C.1 Two-Step Heckmane Selection

We first employ the Two-Step Heckman Selection Approach, which is designed to account for selection bias arising from non-random missingness. This method allows us to explicitly model the selection process—specifically, whether individuals choose to abstain or turn out—in estimating the relationship between candidate attributes and respondent preferences.

The Heckman approach consists of two stages. The first stage, known as the selection equation, models the likelihood that an observation is included in the sample, which in this case corresponds to whether the respondent chooses to turn out ($T_i = 1$). We use a probit model for this stage, predicting the probability of turnout based on both profile attributes and respondent-level pre-treatment covariates. The predictors in the selection equation include candidate or policy features presented in the conjoint task, as well as respondent characteristics such as past voting history, propensity to vote in the upcoming election, age, education, income, gender, and race. Mathematically, the selection equation is expressed as:

$$\Pr(T_i = 1) = \Phi(Z_i \beta) \tag{2}$$

where $T_i = 1$ indicates the respondent turns out, Z_i includes profile attributes and respondent covariates, β represents the coefficients, and Φ is the cumulative distribution function of the probit model.

From this stage, we compute the inverse Mills ratio (IMR) for each observation, which captures the likelihood of inclusion based on the selection process. The IMR is given by:

$$\text{IMR}_i = \frac{\phi(Z_i \beta)}{\Phi(Z_i \beta)} \tag{3}$$

where ϕ and Φ are the probability density function and cumulative density function of the normal distribution, respectively. This ratio is later used in the second stage to adjust for the selection bias.

The second stage, known as the outcome equation, models the relationship between candidate attributes (treatment variables) and the outcome variable (Y_i) for the subset of respondents who chose to turn out. This stage is a linear regression that resembles that in a conjoint analysis but includes the inverse Mills ratio as an additional explanatory variable to control for selection bias. The outcome equation is expressed as:

$$Y_i = X_i \gamma + \lambda \cdot \text{IMR}_i + \epsilon_i, \tag{4}$$

where Y_i is the choice-based outcome variable (e.g., vote choice), X_i represents the treatment variables (candidate attributes), γ are the coefficients for the treatment effects, λ is the coefficient for the inverse Mills ratio, and ϵ_i is the error term. The inclusion of the IMR ensures that the estimates of γ are unbiased, even in the presence of selection bias.

The coefficients in the outcome equation represent the effects of candidate attributes on the outcome variable, conditional on respondents being included in the sample (i.e., having chosen to turn out). The selection equation accounts for the factors influencing the decision to abstain, thereby correcting for bias introduced by non-random missingness. By explicitly modeling the selection process, the Two-Step Heckman Selection Approach provides robust estimates of the causal effects of candidate attributes, even when abstention behavior leads to deviations from complete randomization. This approach ensures that our analysis remains valid and that the observed treatment effects are accurately estimated, as shown in Figures A4 and A5.

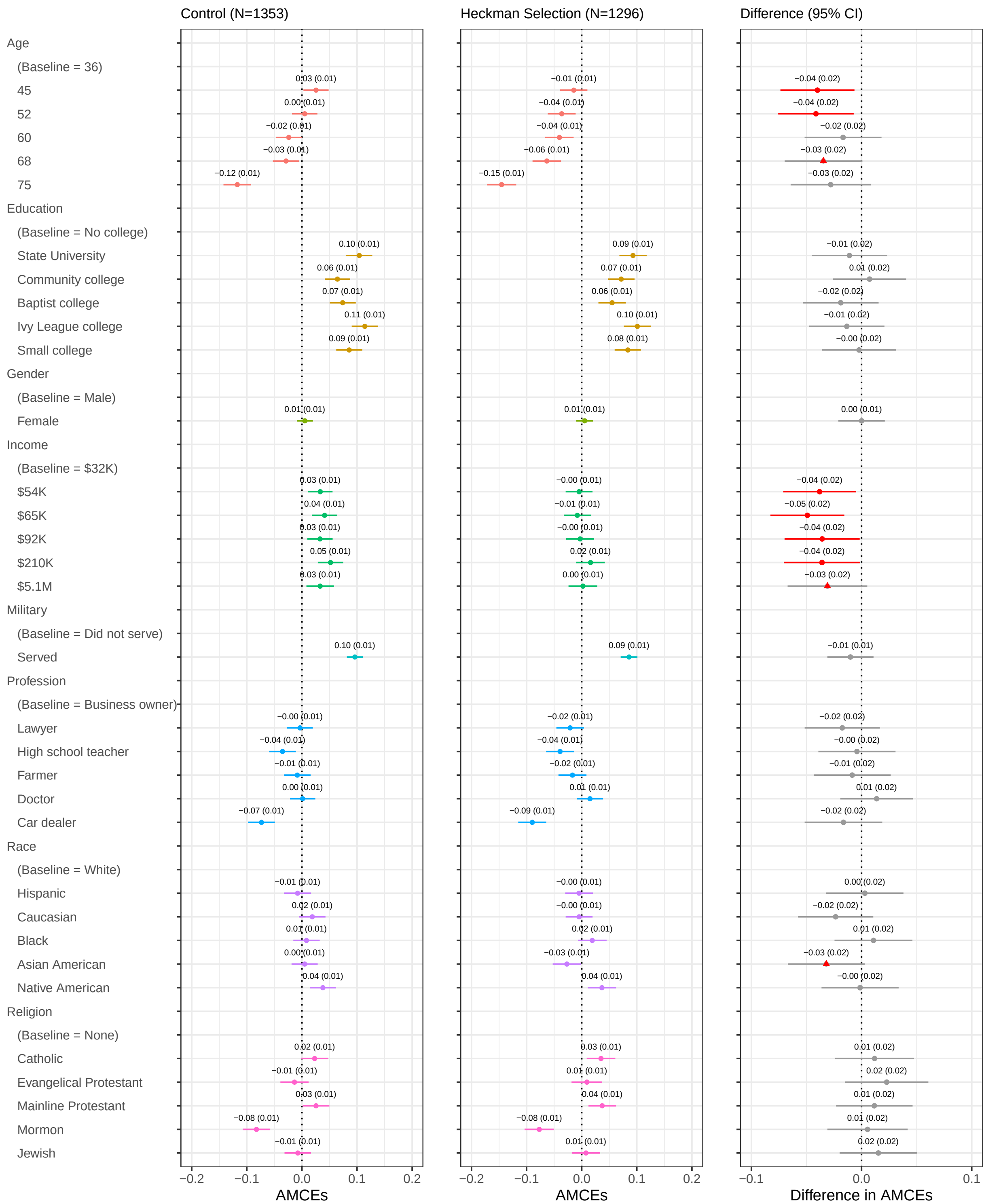

Figure A4: AMCEs of Presidential Candidate Attributes by Design: The left and middle panels present the AMCE results for Presidential candidates under the control and treatment designs. The treatment design estimates were obtained using the 2-step Heckman selection approach to account for potential selection bias. The rightmost panel displays the differences in AMCEs for each attribute level, with horizontal bars representing 95% confidence intervals robust to clustering at the respondent level. Red dots with red bars denote significance at the 5% level, while red triangles with gray bars indicate significance only at the 10% level. Gray dots with gray bars represent attribute levels that are not significant at either conventional level.

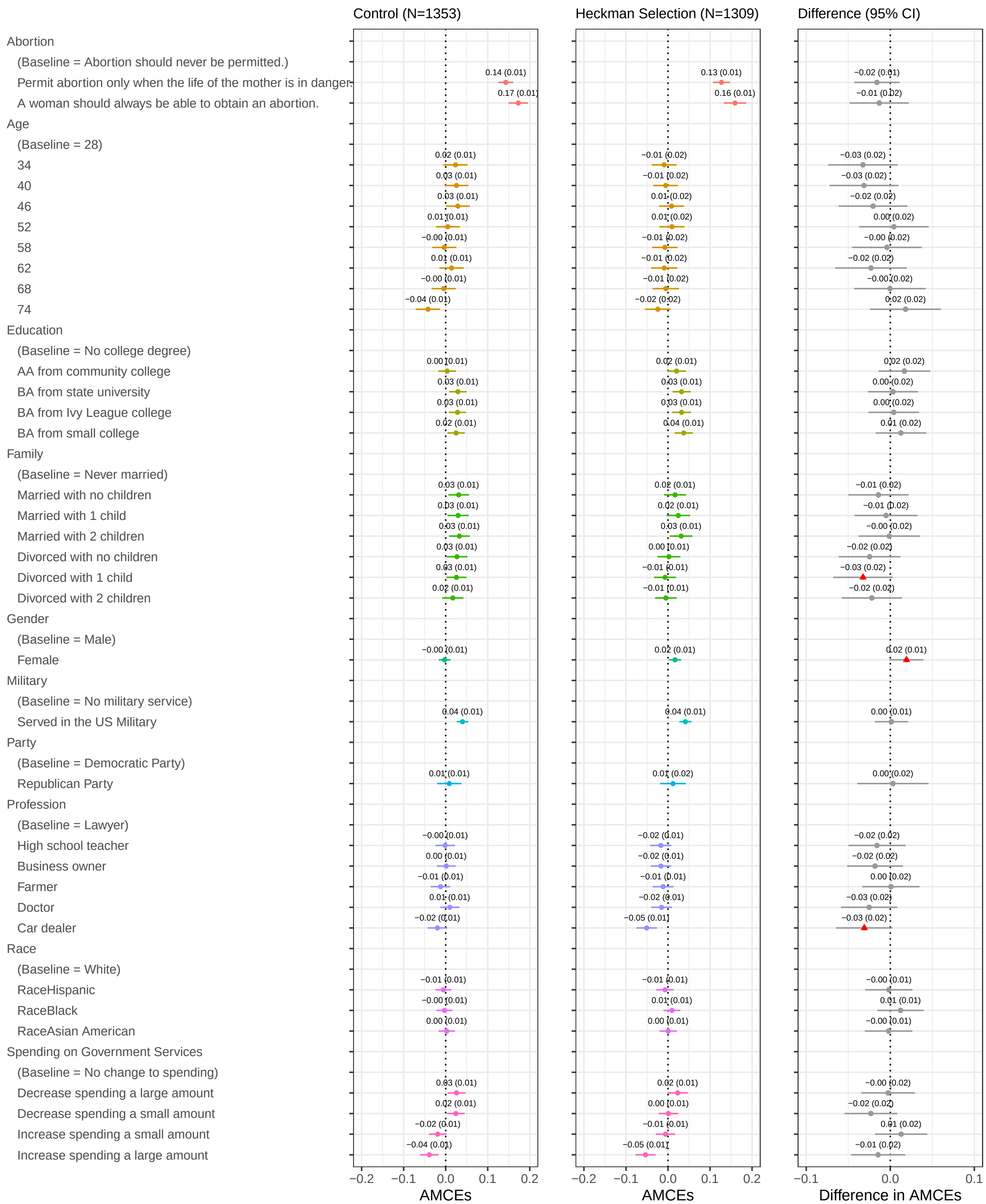

Figure A5: AMCEs of Congressional Candidate Attributes by Design: The left and middle panels present the AMCE results for Congressional candidates under the control and treatment designs. The treatment design estimates were obtained using the 2-step Heckman selection approach to address potential selection bias. The rightmost panel displays the differences in AMCEs for each attribute level, with horizontal bars representing 95% confidence intervals robust to clustering at the respondent level. Red dots with red bars indicate significance at the 5% level, while red triangles with gray bars denote significance only at the 10% level. Gray dots with gray bars represent attribute levels that are not significant at either conventional level.

C.2 Inverse Probability Weighting

We also apply Inverse Probability Weighting (IPW) as an alternative approach to mitigate potential selection bias. Unlike the Two-Step Heckman Selection model, which explicitly models the selection process, IPW adjusts for bias by reweighting the observed data to approximate the distribution of the target population.

The IPW approach begins by estimating the probability of each respondent turning out ($T_i = 1$), conditional on both candidate attributes and respondent-level pre-treatment covariates. A logistic regression model is used to estimate the turnout probability:

$$\Pr(T_i = 1 | D_j^{(k)}, X_i) = \Phi(Z_i \beta) \tag{5}$$

where Z_i includes the profile attributes ($D_j^{(k)}$) and individual covariates (X_i), and Φ represents the cumulative distribution function of the logistic model. Once the probabilities are estimated, weights are assigned to each observation as the inverse of their estimated probability of turnout:

$$w_i = \frac{1}{\Pr(T_i = 1 | D_j^{(k)}, X_i)} \tag{6}$$

These weights are inversely proportional to the likelihood of being observed in the sample. Observations with a lower probability of turnout (based on their characteristics and the presented profiles) are assigned higher weights, ensuring that the weighted sample better represents the overall target population.

In the outcome analysis, these weights are incorporated directly into the estimation of the Average Marginal Component Effects (AMCEs). The weighted AMCE is calculated as:

$$\text{AMCE}j^{(kk\prime)} = \frac{\sum_{i=1}^{N} w_i \cdot (Y_i | D_j^{(k)} = 1)}{\sum_{i=1}^{N} w_i} - \frac{\sum_{i=1}^{N} w_i \cdot (Y_i | D_j^{(k\prime)} = 1)}{\sum_{i=1}^{N} w_i} \tag{7}$$

This weighted estimator adjusts for the non-random selection process introduced by abstention, ensuring that the estimated treatment effects are robust to selection bias.

The key advantage of IPW lies in its flexibility. By reweighting the observed data, it addresses differences in the likelihood of turnout across subgroups without requiring explicit modeling of the selection process in the outcome equation. Additionally, using both pre-treatment covariates and candidate attributes in the weighting model ensures that the estimated probabilities capture the primary factors influencing abstention, reducing bias and aligning the weighted study population with the intended target population. Figures A6 and A7 illustrate the robustness of our results when using this approach.

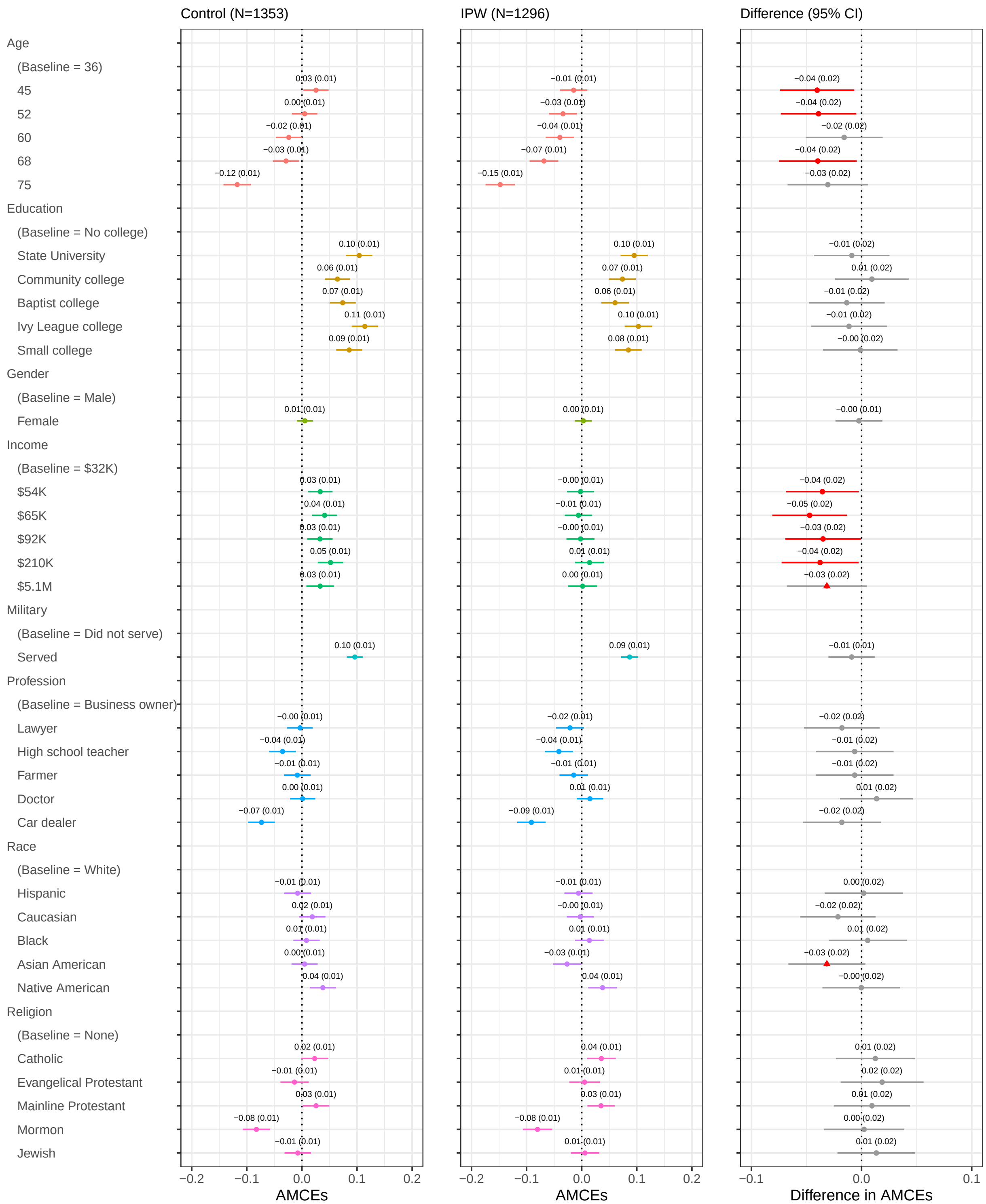

Figure A6: AMCEs of Presidential Candidate Attributes by Design: The left and middle panels present the AMCE results for Presidential candidates under the control and treatment designs. The treatment design estimates were obtained using the Inverse Probability Weighting approach to account for potential selection bias. The rightmost panel displays the differences in AMCEs for each attribute level, with horizontal bars representing 95% confidence intervals robust to clustering at the respondent level. Red dots with red bars denote significance at the 5% level, while red triangles with gray bars indicate significance only at the 10% level. Gray dots with gray bars represent attribute levels that are not significant at either conventional level.

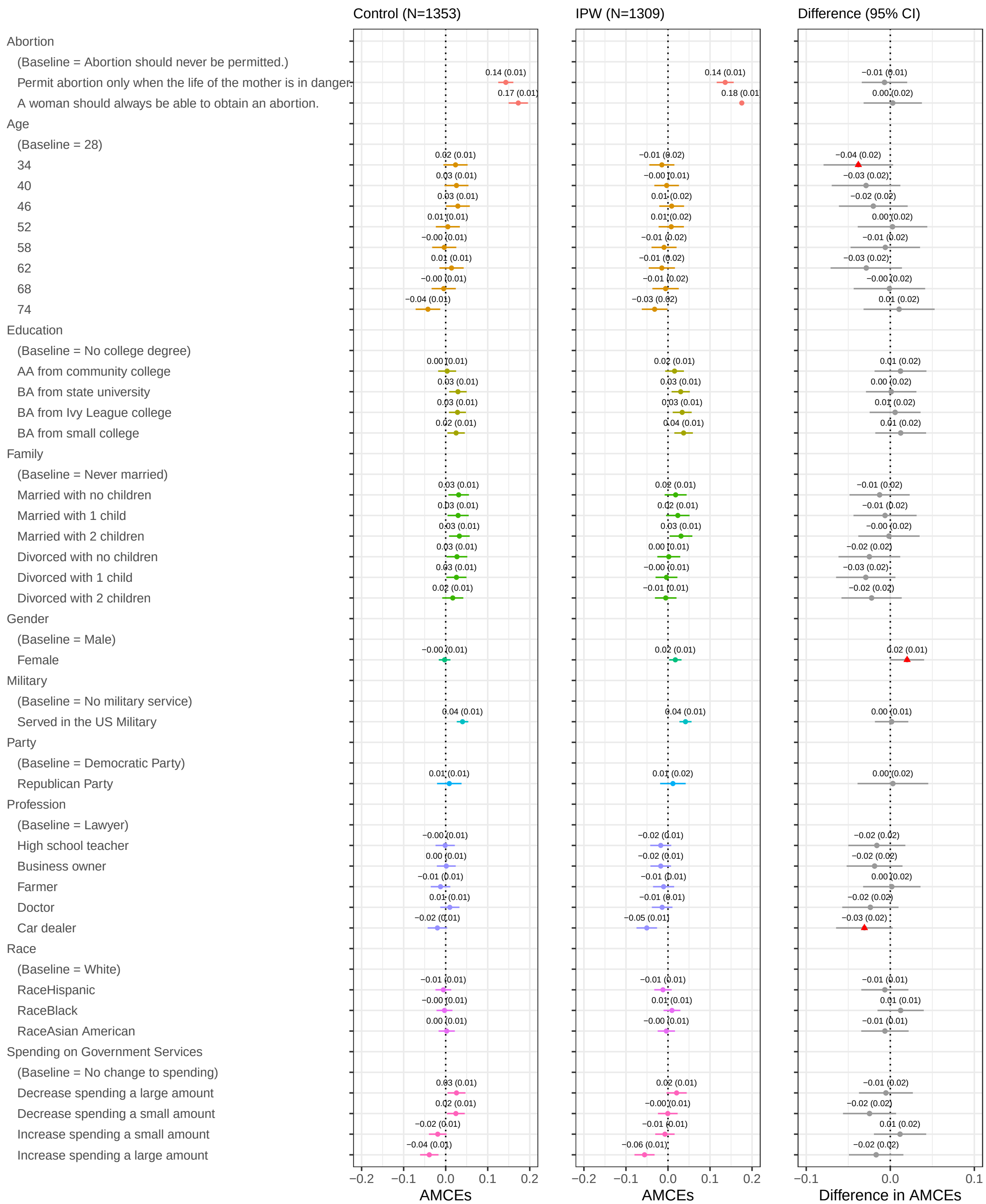

Figure A7: AMCEs of Congressional Candidate Attributes by Design: The left and middle panels present the AMCE results for Congressional candidates under the control and treatment designs. The treatment design estimates were obtained using the Inverse Probability Weighting approach to address potential selection bias. The rightmost panel displays the differences in AMCEs for each attribute level, with horizontal bars representing 95% confidence intervals robust to clustering at the respondent level. Red dots with red bars indicate significance at the 5% level, while red triangles with gray bars denote significance only at the 10% level. Gray dots with gray bars represent attribute levels that are not significant at either conventional level.