"Measuring Subgroup Preferences in Conjoint Experiments: Supplemental Information"

Thomas J. Leeper, Sara B. Hobolt, and James Tilley
2019-05-24

A Definition of Quantities of Interest **2**

B Impact of Reference Category Choice on AMCEs **7**

C Re-analysis of 'Political Experience' Feature from Teele et al. (2018) **8**

D Hainmueller et al. (2014) Candidate Experiment **9**
D.1 Replication using AMCEs and MMs . . . 9
D.2 Numerical Results: AMCEs . . . 10
D.3 Numerical Results: MMs . . . 11

E Hainmueller et al. (2014) Immigration Experiment **12**
E.1 Replication using AMCEs . . . 12
E.2 Numerical Results: AMCEs . . . 13
E.3 Replication using MMs . . . 14
E.4 Numerical Results: MMs . . . 15
E.5 Subgroup Analysis using AMCEs . . . 16
E.6 Subgroup Analysis using MMs . . . 17
E.7 Nested Model Comparison . . . 18

F Teele et al. (2018) Candidate Experiment **19**
F.1 Replication using AMCEs and MMs . . . 19
F.2 Numerical Results: AMCEs . . . 20
F.3 Numerical Results: MMs . . . 21
F.4 Subgroup Analysis using AMCEs . . . 22
F.5 Subgroup Analysis using MMs . . . 22
F.6 Nested Model Comparison: Male/Female Voters . . . 23
F.7 Nested Model Comparison: Democratic/Republican Voters . . . 23
F.8 Comparison of Alternative Reference Categories . . . 24

G Bechtel and Scheve (2013) Climate Agreement Experiment **25**
G.1 Replication using AMCEs and MMs . . . 25
G.2 Subgroup Analysis using AMCEs: Country . . . 26
G.3 Subgroup Analysis using MMs: Country . . . 27
G.4 Subgroup Analysis using AMCEs: Environmentalism . . . 28
G.5 Subgroup Analysis using MMs: Environmentalism . . . 29
G.6 Subgroup Analysis using AMCEs: Reciprocity . . . 30
G.7 Subgroup Analysis using MMs: Reciprocity . . . 31
G.8 Nested Model Comparison: Country . . . 32
G.9 Nested Model Comparison: Environmentalism . . . 32
G.10 Nested Model Comparison: Reciprocity . . . 33
G.11 Comparison of Alternative Reference Categories . . . 34

A Definition of Quantities of Interest

A conjoint experiment serves two purposes: (1) description of the conditional distribution of favorability over variations in multiple features, and (2) leveraging the random observation of combinations of features (so-called "profiles") to infer that any differences in favorability over features are causally attributable to the features as opposed to something else. The quantities of interest are therefore functions of the features being randomized as in any factorial experiment. But additionally, conjoints typically involve within-subjects research designs (i.e., multiple, different profile observations per participant) thus necessitating some additional notation to account for the *survey implementation* of the conjoint in addition to the definition of the descriptive and causal parameters of interest.

Ultimately, a conjoint since Hainmueller, Hopkins, and Yamamoto (2014) is a complex survey-experimental design involving multiple observations across a high-dimension factorial experimental space. Specifically, I respondents ($i \in \{1, \ldots, I\}$) are presented with K rating or forced choice decision tasks, each involving J (typically 2) alternative profiles of, for example, candidates or policies. Each profile consists of a vector of F (typically discrete) features or attributes that describe the profile (e.g., age, sex), each composed of D_f alternative levels, a number which can vary across features. The experiment thus generates a dataset with $N = I \times J \times K$ observations of some rating scale or discrete choice outcome, Y, from a random sample of profiles drawn from the $C = \prod_{f=1}^{F} D_f$ population of experimental *cells* in the F-dimension feature space.

The survey implementation of the conjoint therefore generates N observations that can be indexed by i, j, k, forming an $N \times (L + 4)$ dimensional data matrix $\mathbf{M}$ with each row representing the vector of feature levels $\vec{F}$ in each profile j of respondent i's task k, with indicators for i, j, k, and the corresponding outcome $Y_{i,j,k}$.[11]

With no loss of information, we can think of each row in this matrix equivalently as an observation of $Y_{i,\vec{F}}$. This is because Hainmueller, Hopkins, and Yamamoto (2014) make several important assumptions that allow us to interpret these data in a different way than the survey implementation implies. First, they assume no carryover effects (Assumption 1), such that multiple observations from the same respondent can be treated as independent of one another. Second, they assume no profile order effects within-task (Assumption 2), such that profiles within a task can be treated as independent of each other. Assumptions 1 and 2 imply that the survey implementation indices for task, k, and profile-within-task, j, can be ignored. They have no bearing on any quantity of interest, by assumption.

The analyst is therefore left with a dataset of N observations, grouped into i participants, each providing into $Y_{\vec{F}}$. All quantities of interest must therefore be specified over as features of the distribution of Y over the F-dimensional feature space. In what follows, we therefore focus on the experimental features being randomized rather than the survey design factors being assumed away. Hainmueller, Hopkins, and Yamamoto (2014) make a third assumption that profiles are randomly constituted (Assumption 3), which in a fully randomized design, has the effect of meaning that features and feature combinations are randomly sampled for observation. If this randomization is uniform

[11] In typical paired designs (where $J = 2$), this means each task generates two data points: $Y_{i,1,k}$ and $Y_{i,2,k}$. Note, too, that in fully randomized designs, these two profiles can be identical. Furthermore in fully randomized, forced-choice designs this can yield the additional curiosity that $Y_{i,\mathbf{c}} \neq Y_{i,\mathbf{c}}$ for a given respondent, i, and profile, $\mathbf{c}$.

(which it almost always is in applied examples) this means we can additionally ignore the probability of observing any given combination (as all profiles are equally likely to be observed). This is a point we return to in a moment.

The most basic thing that can be estimated about the distribution of Y is the expected value, $E[Y]$, or *grand mean* (in the parlance of factorial experiments). We can think of this quantity in terms of the survey implementation process (namely, respondents, tasks, and profiles) or as a simple function of the resulting data:

$$\bar{Y} = \frac{1}{I \times J \times K} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} Y_{i,j,k} = \frac{1}{N} \sum_{n=1}^{N} Y_n \tag{4}$$

The nested summation over i, j, k could be stated explicitly but is unnecessary as the grand mean is simply the mean of all observed Y. A useful check on intuition is that in a forced choice design, where a respondent must choose only one profile, j, of all those presented in each task k, then by design $\bar{Y} = \frac{1}{J}$. For common, two-alternative, forced choice designs, $\bar{Y}$ therefore always equals 0.5. By contrast, in rating scale designs, $\bar{Y}$ can take any value between the lower and upper bounds of the rating scale.

In a *full factorial* experiment where $N > C$ (the number of observations is larger than the number of cells) due to a large sample, or few factors, or levels of each factor, or both (or both of these design characteristics), a sensible next quantity of interest is the *cell mean*: $E[Y|\vec{X} = \vec{x}]$, which in a conjoint simply measures the mean favorability toward a particular profile, $\vec{x}$. An effort to actually estimate this quantity will, however, become obviously intractable when one recognizes that the number of observations in a typical conjoint is much lower the number of feasible profiles ($N \ll C$). The cell mean can be unobserved for many or perhaps most experimental cells.

Therefore quantities of interest that derive from it — such as pairwise differences of means between cells — cannot be estimated for any of the arbitrary $\binom{C}{2}$ pairs of cells. As an example, in the Hainmueller, Hopkins, and Yamamoto (2014) candidate experiment, $C = 6^6 * 2^2 = 186,624$ and $N = 3466$, so less than 2% of experimental cells were observable and a minuscule fraction of the 17.4 billion pairwise cell combinations could have generated estimable effects.

It is at this point that the quantities of interest in a conjoint can become confusing. In a typical experiment where $N > C$, these pairwise differences of means are the standard estimator for a causal effect (the estimand being the causal effect on favorability of changing from one profile to another). For example, we might be interested in the effect on Y of changing the value of one feature to another theoretically interesting value of that feature, holding all other feature values in the profile constant:

$$\tau = E[Y|X_1 = x_1, X_2 = x_2, \ldots, X_f = x_f] - E[Y|X_1 = \neg x_1, X_2 = x_2, \ldots, X_f = x_f] \tag{5}$$

but we have no guarantee that both or, in fact, either of those particular cells are observed. If even this minimal causal quantity cannot be guaranteed to be estimable by design, questions about higher-order interactions across features are even more difficult to estimate as they require observing four or more specific cells, any of which may be missing. Even if we were interested in such quantities, we would be unlikely to be able to estimate them.

Conjoint designs therefore ask us to think about completely different quantities of

interest from typical sentiment measurement or experimentation. Consequently, what quantities might we care about that can be estimated from an L-dimension factorial experimental with considerable sparsity other than the grand mean?

Even though $N \ll C$ in most applied conjoints, $N > F$. This means that even if we probably cannot learn about particular high-dimensional *combinations of features*, we can learn about favorability toward particular features alone. That is, we can learn about conditional expectations over each feature dimension, $E[Y|X_f = x_f]$. In the factorial experiments literature, this conditional mean is called the *marginal mean* (as it lies at the margins of a tabular presentation cell means for the complete design). For example, the following 2x3 factorial design contains 6 cell means ($2 * 3$), 1 grand mean, and five marginal means ($2 + 3$, one for each level of each factor):

	$A = 1$	$A = 2$	
$B = 1$	$\bar{Y}_{A=1,B=1}$	$\bar{Y}_{A=2,B=1}$	$E[Y\|B = 1]$
$B = 2$	$\bar{Y}_{A=1,B=2}$	$\bar{Y}_{A=2,B=2}$	$E[Y\|B = 2]$
$B = 3$	$\bar{Y}_{A=1,B=3}$	$\bar{Y}_{A=2,B=3}$	$E[Y\|B = 3]$
	$E[Y\|A = 1]$	$E[Y\|A = 2]$	$E[Y]$

The uniform sampling of cells in the design means that this is quantity can be estimated by the simple mean of $Y \forall X_f = x_f$.[12]

Were a constrained conjoint design used where some feature combinations were impossible, the marginal means would only be intelligible in the fractions of the design where all cells are observed.[13]

To clarify this point, consider the constrained 2x3 design below where one cell is unobserved by design:

	$A = 1$	$A = 2$	
$B = 1$	$\bar{Y}_{A=1,B=1}$	$\bar{Y}_{A=2,B=1}$	$E[Y\|B = 1]$
$B = 2$	$\bar{Y}_{A=1,B=2}$	$\bar{Y}_{A=2,B=2}$	$E[Y\|B = 2]$
$B = 3$	$\bar{Y}_{A=1,B=3}$	–	$E[Y\|B = 3]$
	$E[Y\|A = 1]$	$E[Y\|A = 2]$	$E[Y]$

Were the lower-right cell ($A = 2, B = 3$) observable by design, then a direct comparison of the marginal means, $E[Y|A = 1]$ and $E[Y|A = 2]$, in the lower table margin would provide direct insight into the relative favorability of respondents to profiles

[12]In unbalanced designs where the probability of being in a given cell is not uniform across cells, there is sometimes a distinction made between *descriptive* marginal means that equally weight observations and *design* marginal means that equally weight cells in the design. Given conjoint designs generally do not allow for the observation of cell means, the distinction is not relevant and we refer to *descriptive* marginal means simply as "marginal means."

[13]Practically, the random sampling of cells does not need to be uniform; over- and under-representation of cells is possible. We focus here on fully randomized designs that draw profiles from the full space with equal probability. A nuance in Hainmueller, Hopkins, and Yamamoto's notation is that their quantities of interest are conditioned on an arbitrary joint distribution of features rather than the particular joint distribution of features that was used to construct design or the joint distribution of features that happens to emerge empirically. In other words, they weight cells by an arbitrary joint probability mass function.

with features $A = 1$ and $A = 2$, marginalized over B. But because this cell is unobserved, these marginal means marginalize over different subsets of the possible values of B making them not obviously comparable. By contrast, the first and second marginal means at the top-right of the table — $E[Y|B = 1]$ and $E[Y|B = 2]$ — provide insight into the favorability of participants toward profiles with features $B = 1$ and with feature $B = 2$ marginalizing over the two possible values of A. A researcher could safely conclude that participants are more (less) favorable toward profiles with feature $B = 1$ than $B = 2$ from this information alone. But they would not be able to so for feature A without either (a) an explicit caveat that the comparison is of dissimilar subsets of profiles along dimension B or (b) calculating marginal means over only the completely observable[14] portion of the feature space due to the curse of dimensionality.

For the common *descriptive* use of conjoint designs to measure preferences over multi-dimensional objects, these marginal means alone are of direct interest. They express favorability on the scale of the outcome over alternative values of each feature independent of the features in the design.[15]

For the *causal* interpretation of conjoint designs, comparisons of these marginal means is required. Comparisons between them provide causal inferences about the effect of changing a focal feature, marginalizing across the distribution of other features. Because feature combinations (i.e., the profiles) are randomly constructed and randomly observed from all possible combinations, the distribution of other non-focal features is, in expectation, is independent of the focal feature, thus identical across all levels of the focal feature, and therefore ignorable.

A typical causal effect of interest is therefore the difference in marginal means across two levels of a feature (i.e., the marginal effect of a change in a feature's levels). For an unconstrained design, this difference is the *average marginal component effect* (AMCE) defined by Hainmueller, Hopkins, and Yamamoto (2014). In this way, an AMCE is simply a marginal effect of the factorial design: the difference of two marginal means.

Unfortunately, this is not a perfectly complete definition, but it covers the vast majority of applied cases. The exceptions are few. First, Hainmueller, Hopkins, and Yamamoto allow the joint distribution of features used in calculating the difference of marginal means to be arbitrary. This is meant to accommodate the weighting of marginal means to reflect the real-world distribution of feature combinations (e.g., down-weighting African American Republican political candidates given their rarity in real-world politics). Their definition of an AMCE allows for arbitrary weighting, but in practice this is uncommon.

Second, in constrained designs where some cells are unobservable, care needs to be taken in both defining and estimating AMCEs. Take, for example, the trivial example

[14]Note that what matters here is *observability*, not whether any given cell is actually observed. We know from above that most cells will be unobserved even in a uniformly sampled, unconstrained design.

[15]They do not necessarily convey favorability in an absolute sense. A high marginal mean for a given feature does not imply that the sample prefers that feature in an absolute sense. Instead, favorability has to be understood in light of the features presented to respondents. This is the innovation in conjoints; rather than asking respondents whether they will support a Mormon candidate (for example), we can infer their favorability toward a Mormon candidate in light of other candidate characteristics they may consider. Still our design may not contain all such features, so caution is needed in drawing typical public opinion inferences from these marginal means.

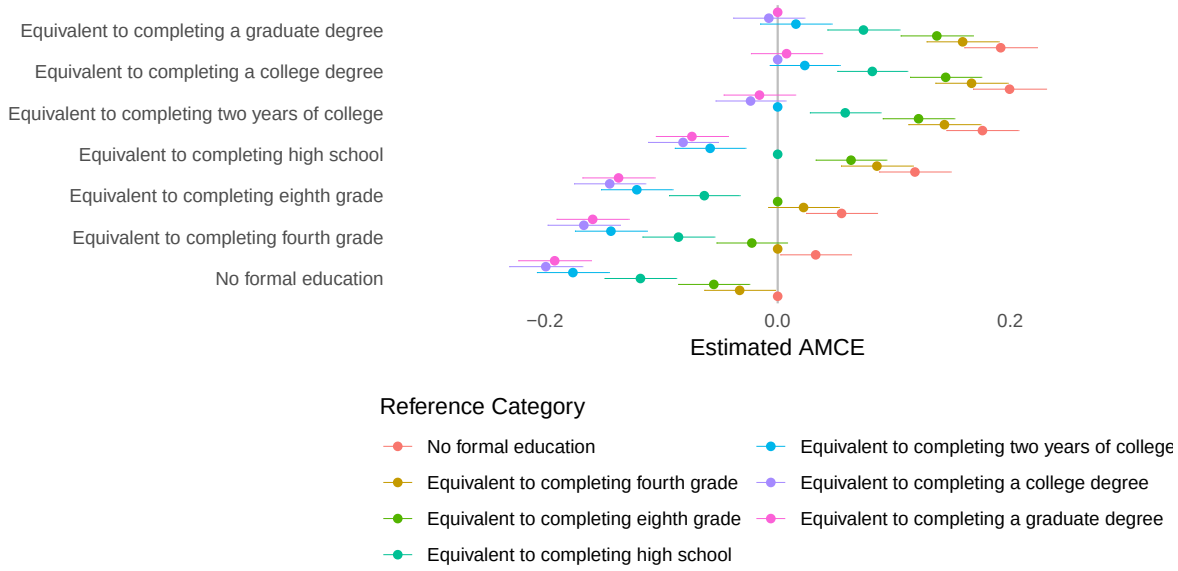
just above. The difference $E[Y|B = 2] - E[Y|B = 1]$ marginalizes over the full set of levels of A but $E[Y|B = 3] - E[Y|B = 1]$ marginalizes only over case where $A = 1$. Thus these two marginal effects reflect different subsets of the data and different combinations of values of A.

Hainmueller, Hopkins, and Yamamoto allow for these two differences to be presented as the AMCE despite the fact that the quantities marginalize over distinct subsets of the design. Indeed, their definition of AMCE for constrained designs diverges from the intuitive marginal effect to instead define the AMCEs for levels of B as an average of marginal effects of B over subsets of A and the AMCEs for levels of A as averages of the marginal effects of A over subsets of B (again, weighting these marginal effects arbitrarily). For example, if feature A is race $Caucasian, AfricanAmerican$ and feature B is religion $Evangelical, Catholic, Jewish$. In Hainmueller, Hopkins, and Yamamoto's notation, the AMCE of a candidate being Jewish relative to being Evangelical Christian is defined only for Caucasian candidates, while the AMCE of being Catholic is defined for both African American and Caucasian candidates. They present these subset marginal effects as the sample AMCEs even though they are not defined for the whole sample. There is nothing inherently problematic about that but, as noted earlier, it requires either being clear about what features are being marginalized over for each AMCE or an analysis of only the complete and comparable subset of the design (i.e., partitioning the design to form two complete, overlapping experimental designs). So, the researcher in this example may prefer to not present the AMCE of being Jewish together with the other results as it does not draw upon the complete set of feature combinations used in other portions of the analysis.

B Impact of Reference Category Choice on AMCEs

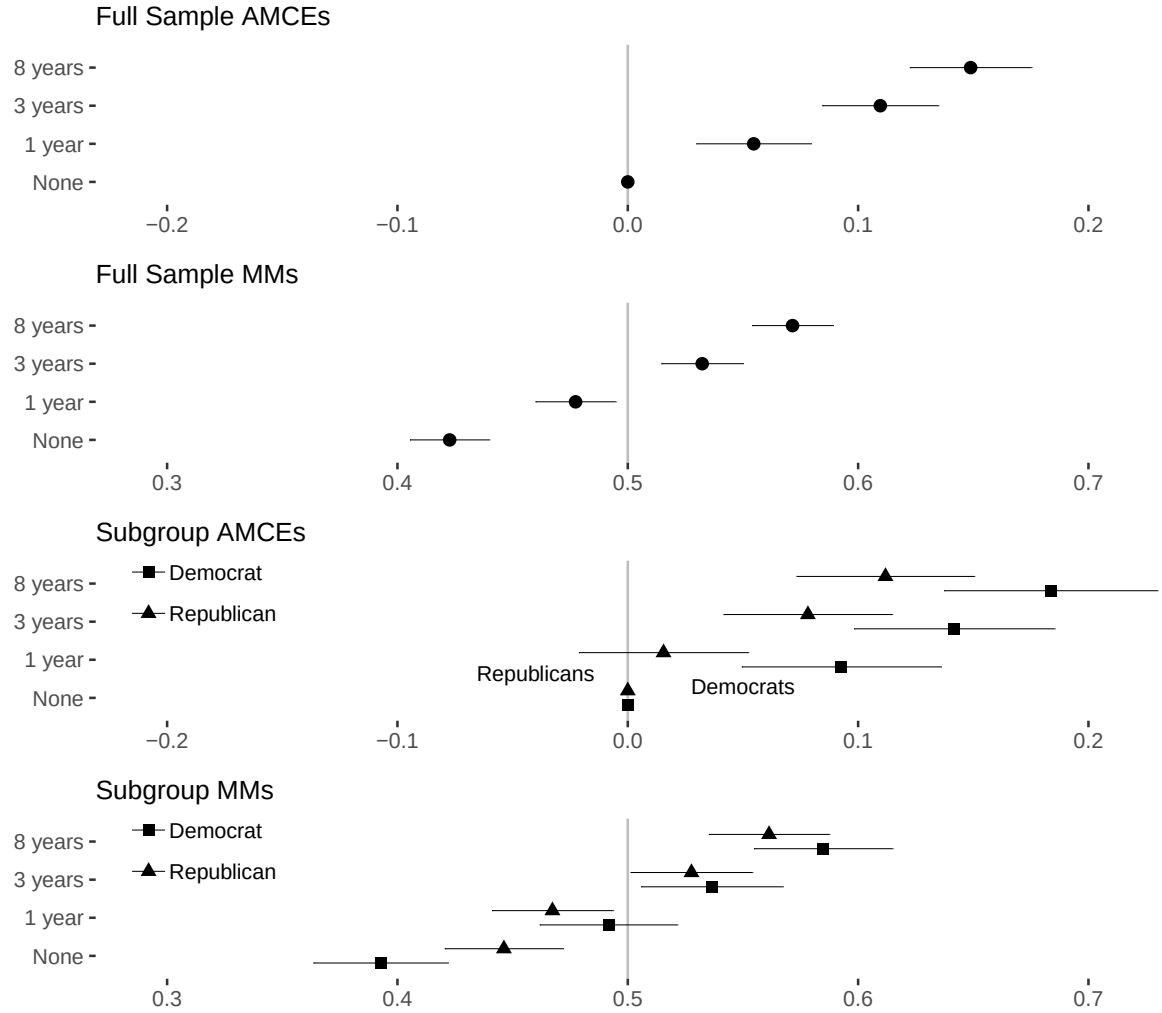
Though seemingly arbitrary, the choice of reference category for estimating AMCEs can be quite consequential. For example, in Hainmueller, Hopkins, and Yamamoto's candidate experiment (again, see 1), the least liked education level ("no formal education") is chosen as a reference category, but the authors could have presented the results using any of the categories as the baseline.

The figure below shows how the estimated AMCEs for each level of the education feature would have differed depending on that choice. Selecting a reference category that receives middling support (i.e., more favorability than some other feature levels but less favorability than others), makes some AMCEs positive and others negative but all AMCEs can be made positive (or negative) simply by choosing a different baseline. The results would be numerically equivalent — the alternative linear models used to the estimate the AMCEs have a mathematical equivalence — but the choice has sizeable consequences for the interpretation of conjoint analyses, as we discuss below.



In *constrained* conjoint designs, the choice of reference category is even more important. Consider, for example, the design of Hainmueller, Hopkins, and Yamamoto's immigration experiment, which constrains the "Country of Origin" feature so that levels 'India,' 'Germany,' 'France,' 'Mexico,' 'Philippines,' and 'Poland' cannot co-occur with the 'Escape Persecution' level of the "Reason for Application" feature. Consequently, the AMCE for the "Escape Persecution" level (relative to the "Reunite with family" reference category) is only defined for the subset of the design involving countries 'China,' 'Sudan,' 'Somalia,' and 'Iraq.' The AMCEs for those four countries (relative to India as a baseline) marginalize across all reasons for application, but the AMCEs for the first six countries marginalize only across the latter two reasons. Thus the interpretation of AMCEs — and the basic ability to estimate them in constrained designs — depends entirely upon the selection of a reference category where all feature levels can co-occur. In a design where *all* features are constrained, then AMCEs are undefined for the design as a whole and only estimable for subsets of the design that are *conditionally* unconstrained.

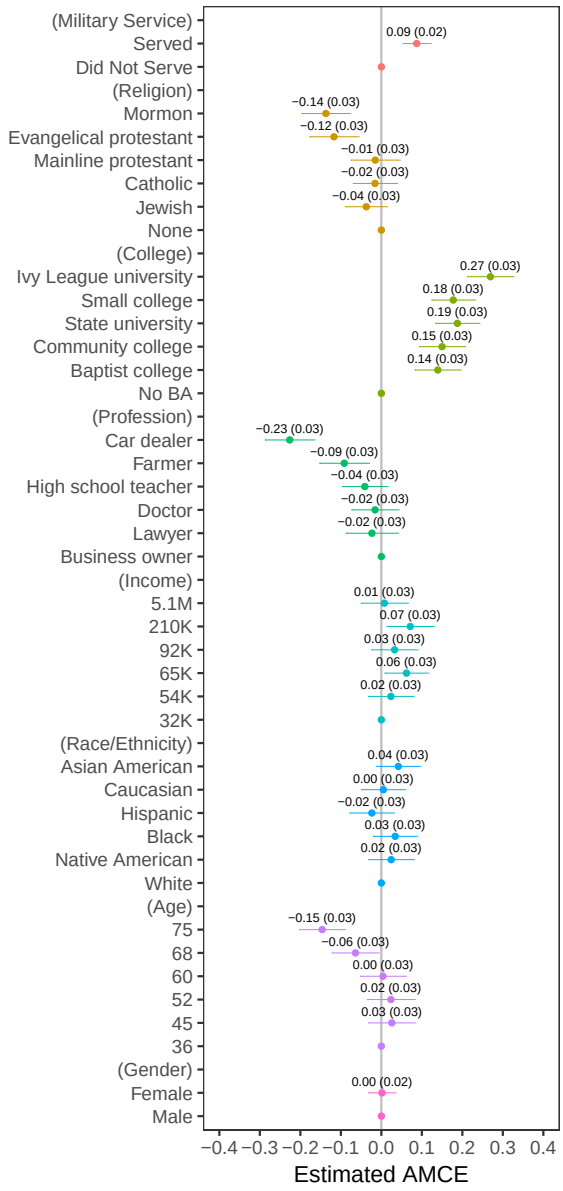
C Re-analysis of 'Political Experience' Feature from Teele et al. (2018)

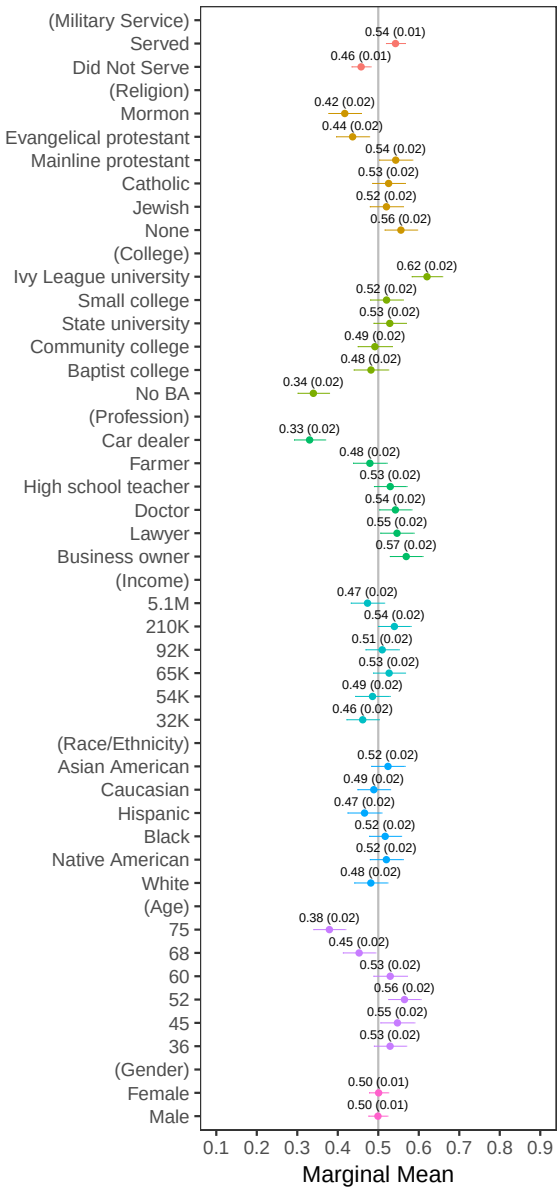


Conditional AMCEs in this experiment (see 3rd panel, above) correctly convey that both Democrats and Republicans are more likely to favor experienced than inexperienced candidates. Reading the AMCEs descriptively, however, would suggest that Democratic voters are more favorable toward candidates with all levels of experience compared to Republican voters (i.e., Republicans and Democrats differ in their preferences over experienced candidates). Yet the conditional marginal means (4th panel, above) reveal that Democrats and Republicans have very similar preferences toward candidates with 1 or 3 years of experience, but differ dramatically in their preferences over candidates with no experience (the reference category) and those with 8 years experience. Democrats are much more sensitive to experience than are Republicans and important differences in preferences between the groups are apparent for very high and very low experience, but the conditional AMCEs suggest that preferences differ at all levels of experience, when in reality they do not.

D Hainmueller et al. (2014) Candidate Experiment

D.1 Replication using AMCEs and MMs





D.2 Numerical Results: AMCEs

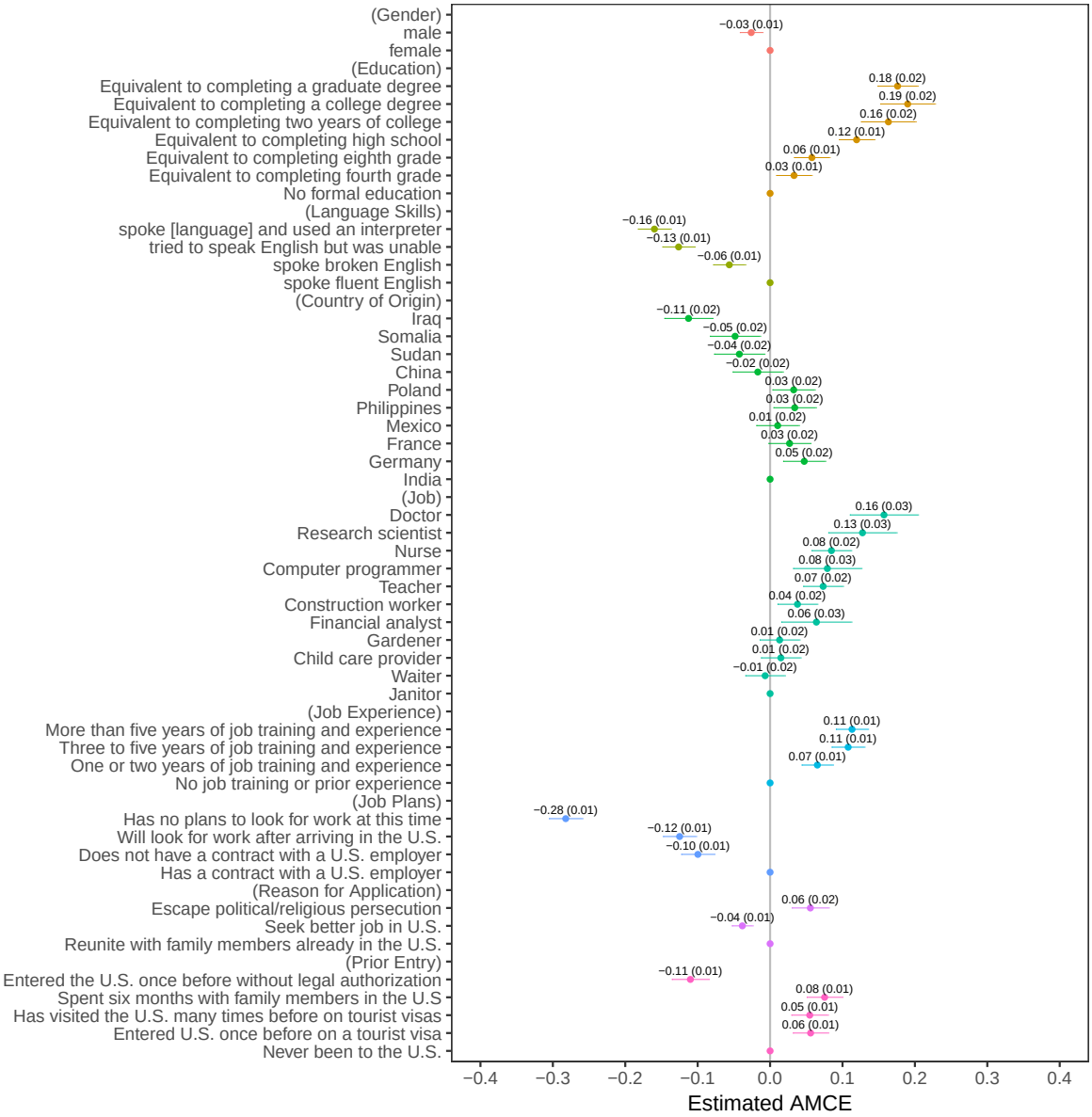
feature	level	estimate	std.error	z
Military Service	Did Not Serve	0.00		
Military Service	Served	0.09	0.02	4.95
Religion	None	0.00		
Religion	Jewish	-0.04	0.03	-1.42
Religion	Catholic	-0.02	0.03	-0.56
Religion	Mainline protestant	-0.01	0.03	-0.48
Religion	Evangelical protestant	-0.12	0.03	-3.78
Religion	Mormon	-0.14	0.03	-4.46
College	No BA	0.00		
College	Baptist college	0.14	0.03	4.82
College	Community college	0.15	0.03	5.17
College	State university	0.19	0.03	6.77
College	Small college	0.18	0.03	6.50
College	Ivy League university	0.27	0.03	9.26
Profession	Business owner	0.00		
Profession	Lawyer	-0.02	0.03	-0.71
Profession	Doctor	-0.02	0.03	-0.53
Profession	High school teacher	-0.04	0.03	-1.42
Profession	Farmer	-0.09	0.03	-2.94
Profession	Car dealer	-0.23	0.03	-7.24
Income	32K	0.00		
Income	54K	0.02	0.03	0.82
Income	65K	0.06	0.03	2.26
Income	92K	0.03	0.03	1.12
Income	210K	0.07	0.03	2.41
Income	5.1M	0.01	0.03	0.25
Race/Ethnicity	White	0.00		
Race/Ethnicity	Native American	0.02	0.03	0.85
Race/Ethnicity	Black	0.03	0.03	1.22
Race/Ethnicity	Hispanic	-0.02	0.03	-0.84
Race/Ethnicity	Caucasian	0.00	0.03	0.18
Race/Ethnicity	Asian American	0.04	0.03	1.51
Age	36	0.00		
Age	45	0.03	0.03	0.88
Age	52	0.02	0.03	0.78
Age	60	0.00	0.03	0.14
Age	68	-0.06	0.03	-2.17
Age	75	-0.15	0.03	-5.06
Gender	Male	0.00		
Gender	Female	0.00	0.02	0.09

D.3 Numerical Results: MMs

feature	level	estimate	std.error	z
Military Service	Did Not Serve	0.46	0.01	-3.54
Military Service	Served	0.54	0.01	3.55
Religion	None	0.56	0.02	2.73
Religion	Jewish	0.52	0.02	0.96
Religion	Catholic	0.53	0.02	1.24
Religion	Mainline protestant	0.54	0.02	2.06
Religion	Evangelical protestant	0.44	0.02	-3.05
Religion	Mormon	0.42	0.02	-4.04
College	No BA	0.34	0.02	-8.11
College	Baptist college	0.48	0.02	-0.83
College	Community college	0.49	0.02	-0.39
College	State university	0.53	0.02	1.39
College	Small college	0.52	0.02	0.99
College	Ivy League university	0.62	0.02	6.27
Profession	Business owner	0.57	0.02	3.35
Profession	Lawyer	0.55	0.02	2.20
Profession	Doctor	0.54	0.02	2.08
Profession	High school teacher	0.53	0.02	1.44
Profession	Farmer	0.48	0.02	-0.98
Profession	Car dealer	0.33	0.02	-8.64
Income	32K	0.46	0.02	-1.89
Income	54K	0.49	0.02	-0.65
Income	65K	0.53	0.02	1.33
Income	92K	0.51	0.02	0.46
Income	210K	0.54	0.02	1.94
Income	5.1M	0.47	0.02	-1.26
Race/Ethnicity	White	0.48	0.02	-0.88
Race/Ethnicity	Native American	0.52	0.02	0.96
Race/Ethnicity	Black	0.52	0.02	0.85
Race/Ethnicity	Hispanic	0.47	0.02	-1.59
Race/Ethnicity	Caucasian	0.49	0.02	-0.53
Race/Ethnicity	Asian American	0.52	0.02	1.14
Age	36	0.53	0.02	1.43
Age	45	0.55	0.02	2.21
Age	52	0.56	0.02	3.18
Age	60	0.53	0.02	1.40
Age	68	0.45	0.02	-2.31
Age	75	0.38	0.02	-5.99
Gender	Male	0.50	0.01	-0.07
Gender	Female	0.50	0.01	0.07

E Hainmueller et al. (2014) Immigration Experiment

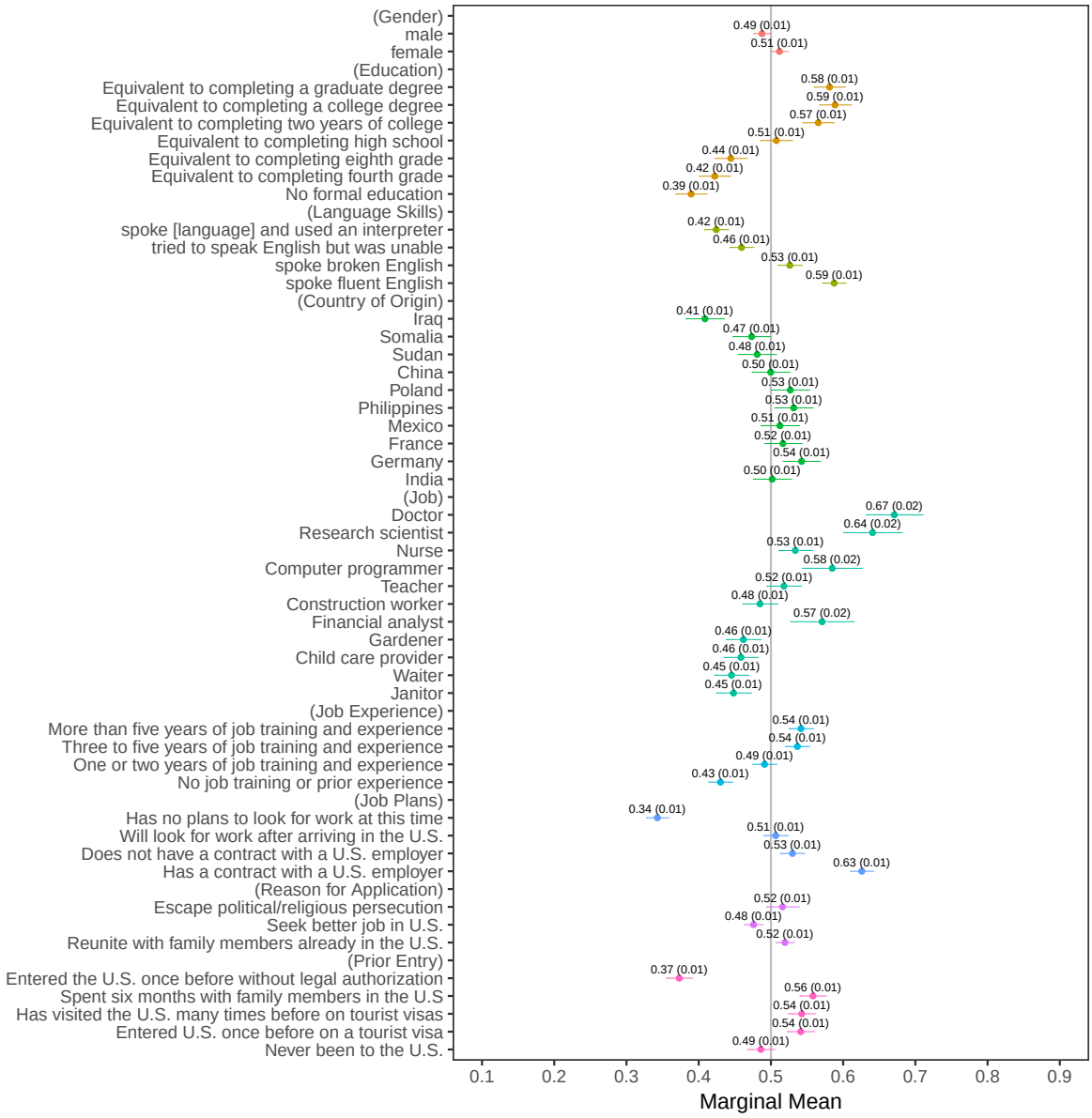
E.1 Replication using AMCEs



E.2 Numerical Results: AMCEs

feature	level	estimate	std.error	z
Gender	female	0.00		
Gender	male	-0.03	0.01	-3.25
Education	No formal education	0.00		
Education	Equivalent to completing fourth grade	0.03	0.01	2.22
Education	Equivalent to completing eighth grade	0.06	0.01	3.86
Education	Equivalent to completing high school	0.12	0.01	7.98
Education	Equivalent to completing two years of college	0.16	0.02	7.12
Education	Equivalent to completing a college degree	0.19	0.02	8.26
Education	Equivalent to completing a graduate degree	0.18	0.02	10.41
Language Skills	spoke fluent English	0.00		
Language Skills	spoke broken English	-0.06	0.01	-4.98
Language Skills	tried to speak English but was unable	-0.13	0.01	-11.11
Language Skills	spoke [language] and used an interpreter	-0.16	0.01	-13.78
Country of Origin	India	0.00		
Country of Origin	Germany	0.05	0.02	2.66
Country of Origin	France	0.03	0.02	1.53
Country of Origin	Mexico	0.01	0.02	0.59
Country of Origin	Philippines	0.03	0.02	1.91
Country of Origin	Poland	0.03	0.02	1.83
Country of Origin	China	-0.02	0.02	-0.81
Country of Origin	Sudan	-0.04	0.02	-2.01
Country of Origin	Somalia	-0.05	0.02	-2.29
Country of Origin	Iraq	-0.11	0.02	-5.56
Job	Janitor	0.00		
Job	Waiter	-0.01	0.02	-0.41
Job	Child care provider	0.01	0.02	0.89
Job	Gardener	0.01	0.02	0.78
Job	Financial analyst	0.06	0.03	2.17
Job	Construction worker	0.04	0.02	2.26
Job	Teacher	0.07	0.02	4.39
Job	Computer programmer	0.08	0.03	2.76
Job	Nurse	0.08	0.02	5.08
Job	Research scientist	0.13	0.03	4.44
Job	Doctor	0.16	0.03	5.49
Job Experience	No job training or prior experience	0.00		
Job Experience	One or two years of job training and experience	0.07	0.01	5.92
Job Experience	Three to five years of job training and experience	0.11	0.01	9.32
Job Experience	More than five years of job training and experience	0.11	0.01	9.96
Job Plans	Has a contract with a U.S. employer	0.00		
Job Plans	Does not have a contract with a U.S. employer	-0.10	0.01	-8.50
Job Plans	Will look for work after arriving in the U.S.	-0.12	0.01	-10.69
Job Plans	Has no plans to look for work at this time	-0.28	0.01	-23.91
Reason for Application	Reunite with family members already in the U.S.	0.00		
Reason for Application	Seek better job in U.S.	-0.04	0.01	-4.37
Reason for Application	Escape political/religious persecution	0.06	0.02	3.58
Prior Entry	Never been to the U.S.	0.00		
Prior Entry	Entered U.S. once before on a tourist visa	0.06	0.01	4.49
Prior Entry	Has visited the U.S. many times before on tourist visas	0.05	0.01	4.24
Prior Entry	Spent six months with family members in the U.S	0.08	0.01	5.98
Prior Entry	Entered the U.S. once before without legal authorization	-0.11	0.01	-8.45

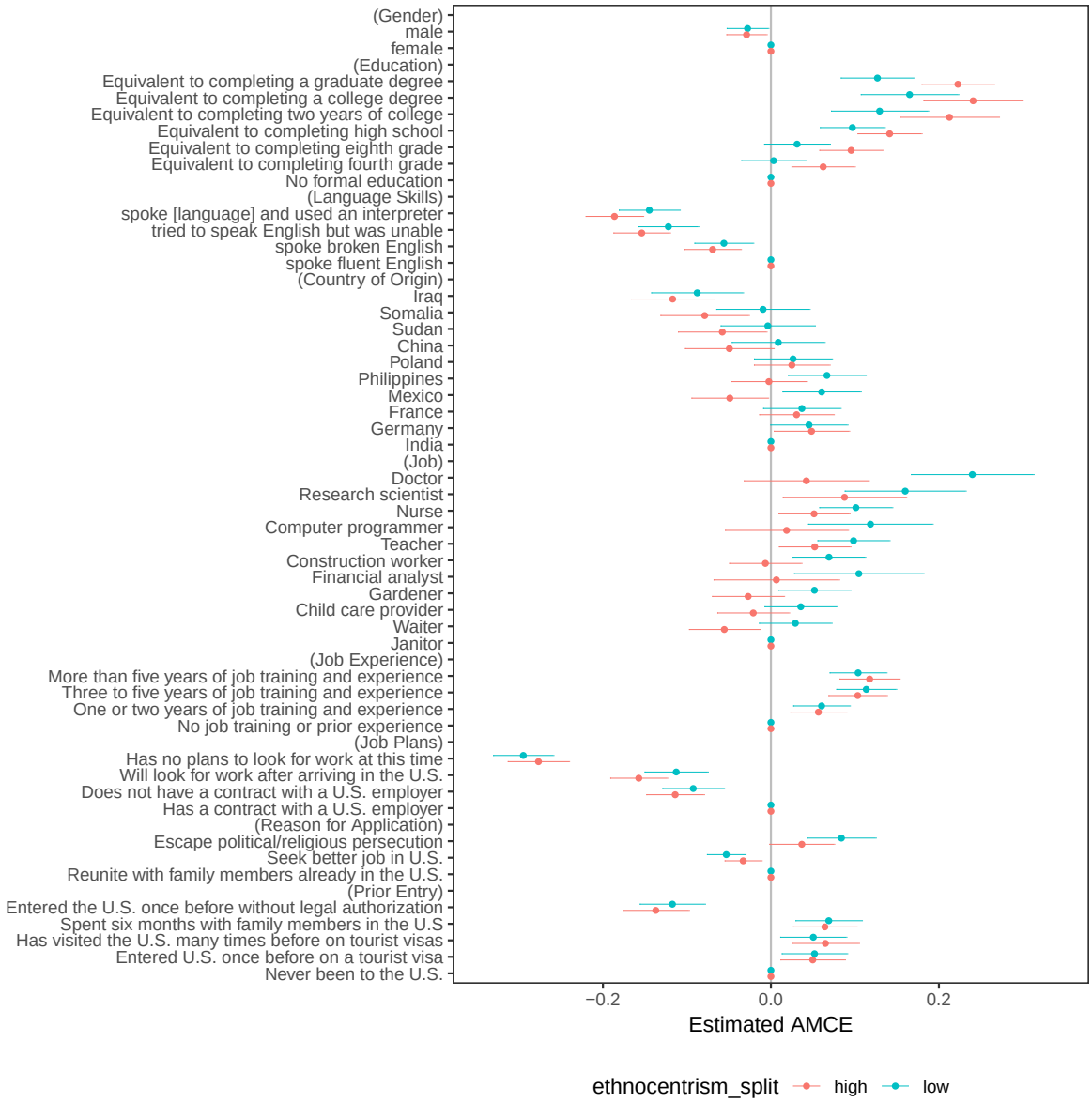
E.3 Replication using MMs



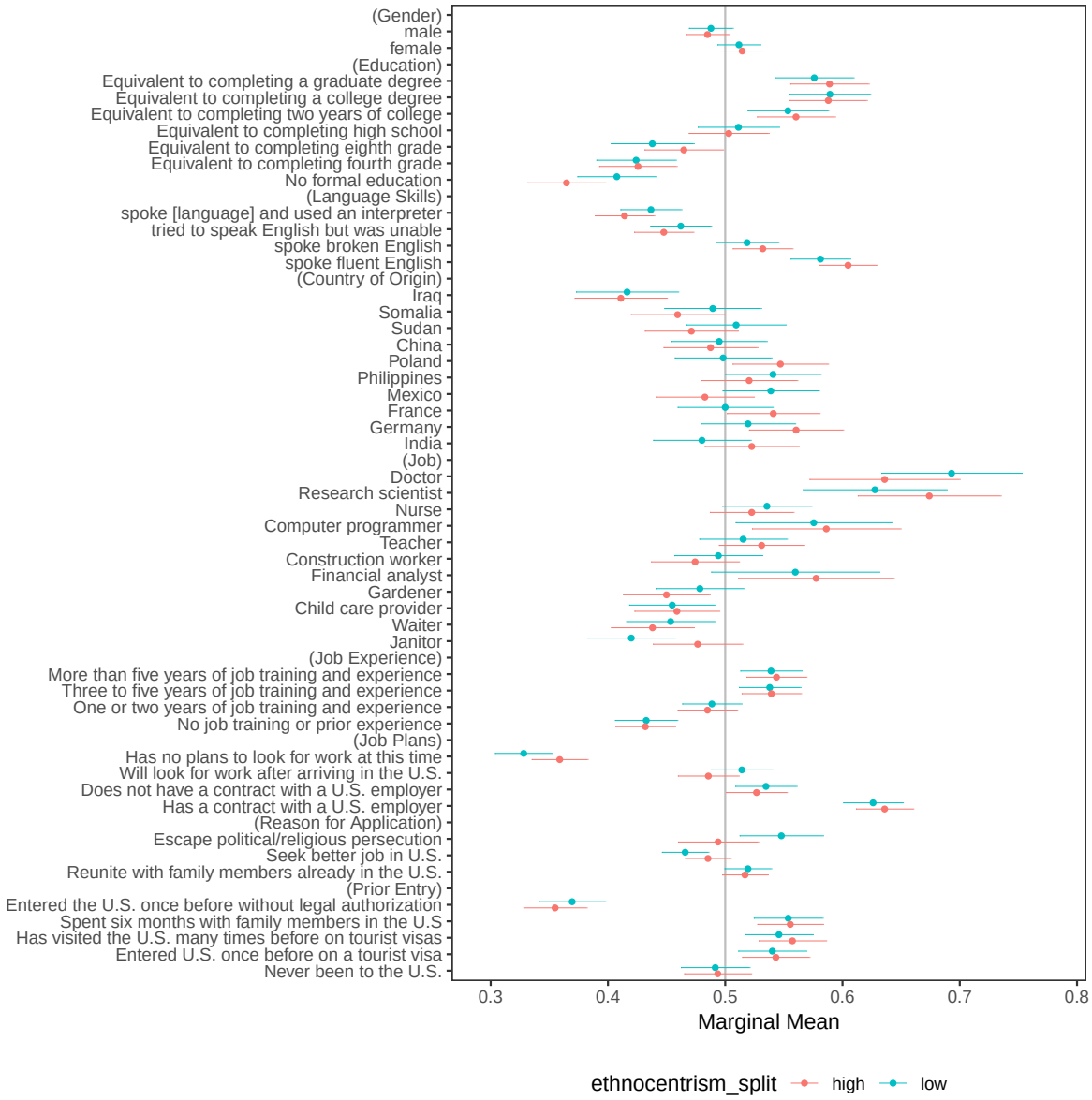
E.4 Numerical Results: MMs

feature	level	estimate	std.error	z
Gender	female	0.51	0.01	1.99
Gender	male	0.49	0.01	-2.03
Education	No formal education	0.39	0.01	-10.04
Education	Equivalent to completing fourth grade	0.42	0.01	-7.08
Education	Equivalent to completing eighth grade	0.44	0.01	-5.00
Education	Equivalent to completing high school	0.51	0.01	0.67
Education	Equivalent to completing two years of college	0.57	0.01	5.92
Education	Equivalent to completing a college degree	0.59	0.01	8.00
Education	Equivalent to completing a graduate degree	0.58	0.01	7.40
Language Skills	spoke fluent English	0.59	0.01	10.63
Language Skills	spoke broken English	0.53	0.01	3.07
Language Skills	tried to speak English but was unable	0.46	0.01	-4.83
Language Skills	spoke [language] and used an interpreter	0.42	0.01	-8.98
Country of Origin	India	0.50	0.01	0.13
Country of Origin	Germany	0.54	0.01	3.22
Country of Origin	France	0.52	0.01	1.26
Country of Origin	Mexico	0.51	0.01	0.92
Country of Origin	Philippines	0.53	0.01	2.36
Country of Origin	Poland	0.53	0.01	2.01
Country of Origin	China	0.50	0.01	-0.03
Country of Origin	Sudan	0.48	0.01	-1.42
Country of Origin	Somalia	0.47	0.01	-2.01
Country of Origin	Iraq	0.41	0.01	-6.76
Job	Janitor	0.45	0.01	-4.20
Job	Waiter	0.45	0.01	-4.56
Job	Child care provider	0.46	0.01	-3.50
Job	Gardener	0.46	0.01	-3.11
Job	Financial analyst	0.57	0.02	3.16
Job	Construction worker	0.48	0.01	-1.23
Job	Teacher	0.52	0.01	1.49
Job	Computer programmer	0.58	0.02	4.01
Job	Nurse	0.53	0.01	2.82
Job	Research scientist	0.64	0.02	6.82
Job	Doctor	0.67	0.02	8.53
Job Experience	No job training or prior experience	0.43	0.01	-8.27
Job Experience	One or two years of job training and experience	0.49	0.01	-1.05
Job Experience	Three to five years of job training and experience	0.54	0.01	4.33
Job Experience	More than five years of job training and experience	0.54	0.01	4.92
Job Plans	Has a contract with a U.S. employer	0.63	0.01	15.40
Job Plans	Does not have a contract with a U.S. employer	0.53	0.01	3.47
Job Plans	Will look for work after arriving in the U.S.	0.51	0.01	0.78
Job Plans	Has no plans to look for work at this time	0.34	0.01	-19.86
Reason for Application	Reunite with family members already in the U.S.	0.52	0.01	3.00
Reason for Application	Seek better job in U.S.	0.48	0.01	-3.76
Reason for Application	Escape political/religious persecution	0.52	0.01	1.40
Prior Entry	Never been to the U.S.	0.49	0.01	-1.47
Prior Entry	Entered U.S. once before on a tourist visa	0.54	0.01	4.37
Prior Entry	Has visited the U.S. many times before on tourist visas	0.54	0.01	4.50
Prior Entry	Spent six months with family members in the U.S	0.56	0.01	6.24
Prior Entry	Entered the U.S. once before without legal authorization	0.37	0.01	-13.96

E.5 Subgroup Analysis using AMCEs



E.6 Subgroup Analysis using MMs

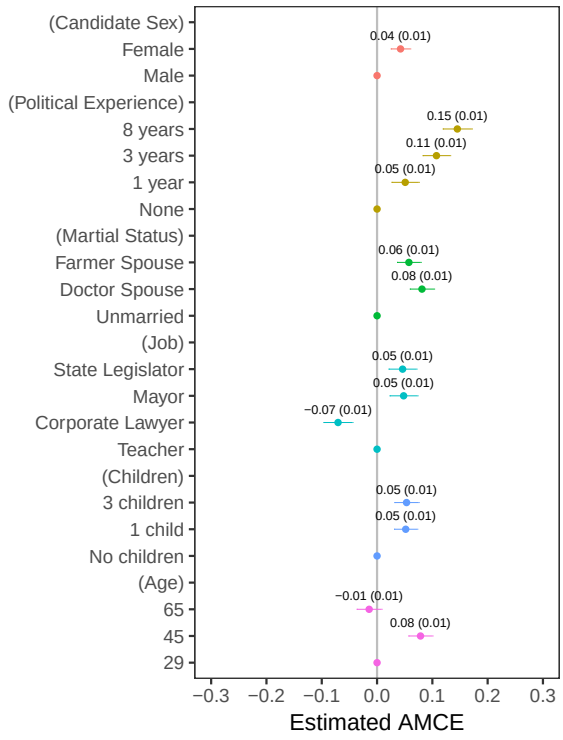


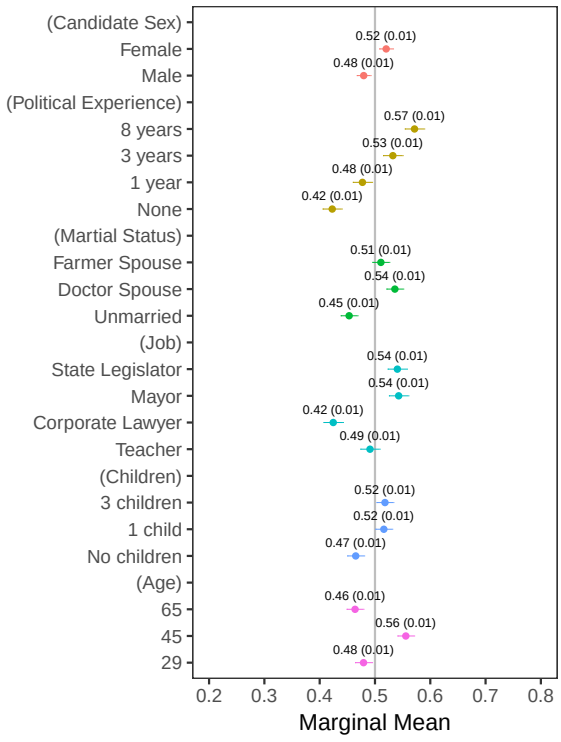
E.7 Nested Model Comparison

```
## Analysis of Deviance Table
##
## Model 1: ChosenImmigrant ~ Gender + Education + LanguageSkills + CountryOfOrigin +
##     Job + JobExperience + JobPlans + ReasonForApplication + PriorEntry +
##     Education:Job + CountryOfOrigin:ReasonForApplication
## Model 2: ChosenImmigrant ~ Gender + Education + LanguageSkills + CountryOfOrigin +
##     Job + JobExperience + JobPlans + ReasonForApplication + PriorEntry +
##     ethnocentrism_split + Education:Job + CountryOfOrigin:ReasonForApplication +
##     Gender:ethnocentrism_split + Education:ethnocentrism_split +
##     LanguageSkills:ethnocentrism_split + CountryOfOrigin:ethnocentrism_split +
##     Job:ethnocentrism_split + JobExperience:ethnocentrism_split +
##     JobPlans:ethnocentrism_split + ReasonForApplication:ethnocentrism_split +
##     PriorEntry:ethnocentrism_split + Education:Job:ethnocentrism_split +
##     CountryOfOrigin:ReasonForApplication:ethnocentrism_split
##   Resid. Df Resid. Dev Df Deviance     F Pr(>F)
## 1     11402     2500.7
## 2     11304     2475.8 98   24.834 1.157 0.1384
```

F Teele et al. (2018) Candidate Experiment

F.1 Replication using AMCEs and MMs





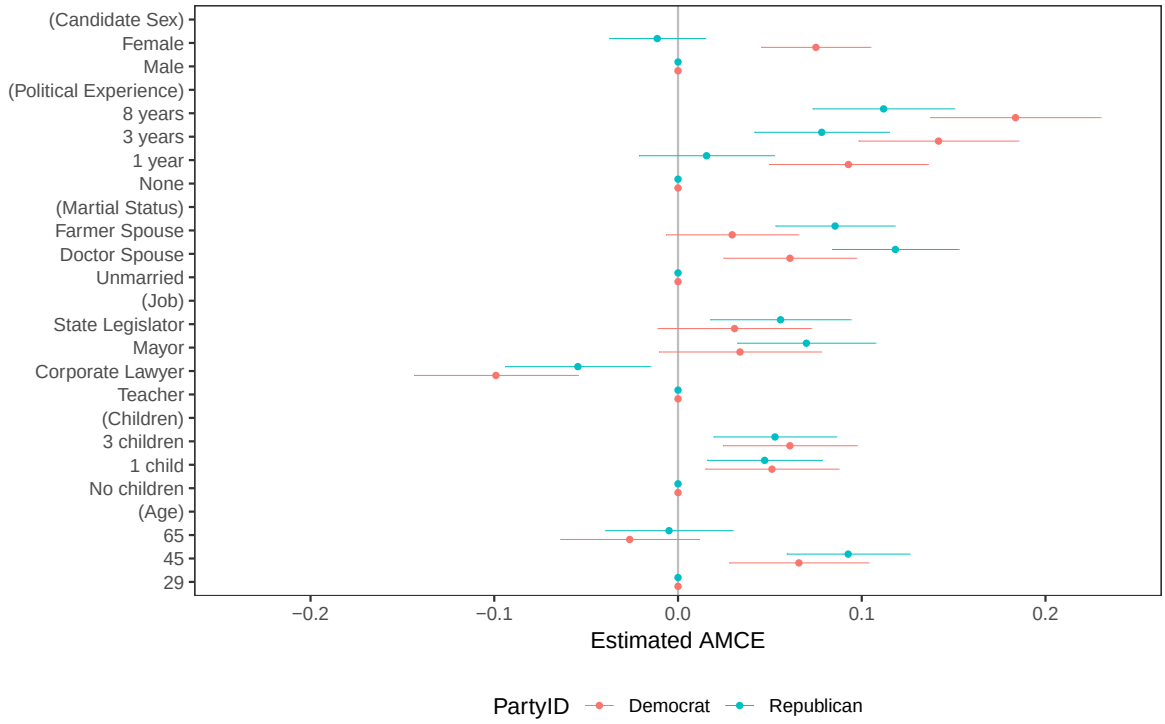
F.2 Numerical Results: AMCEs

feature	level	estimate	std.error	z
Candidate Sex	Male	0.00		
Candidate Sex	Female	0.04	0.01	4.74
Political Experience	None	0.00		
Political Experience	1 year	0.05	0.01	4.06
Political Experience	3 years	0.11	0.01	8.47
Political Experience	8 years	0.15	0.01	10.83
Martial Status	Unmarried	0.00		
Martial Status	Doctor Spouse	0.08	0.01	7.25
Martial Status	Farmer Spouse	0.06	0.01	5.26
Job	Teacher	0.00		
Job	Corporate Lawyer	-0.07	0.01	-5.29
Job	Mayor	0.05	0.01	3.74
Job	State Legislator	0.05	0.01	3.59
Children	No children	0.00		
Children	1 child	0.05	0.01	4.84
Children	3 children	0.05	0.01	4.77
Age	29	0.00		
Age	45	0.08	0.01	7.07
Age	65	-0.01	0.01	-1.24

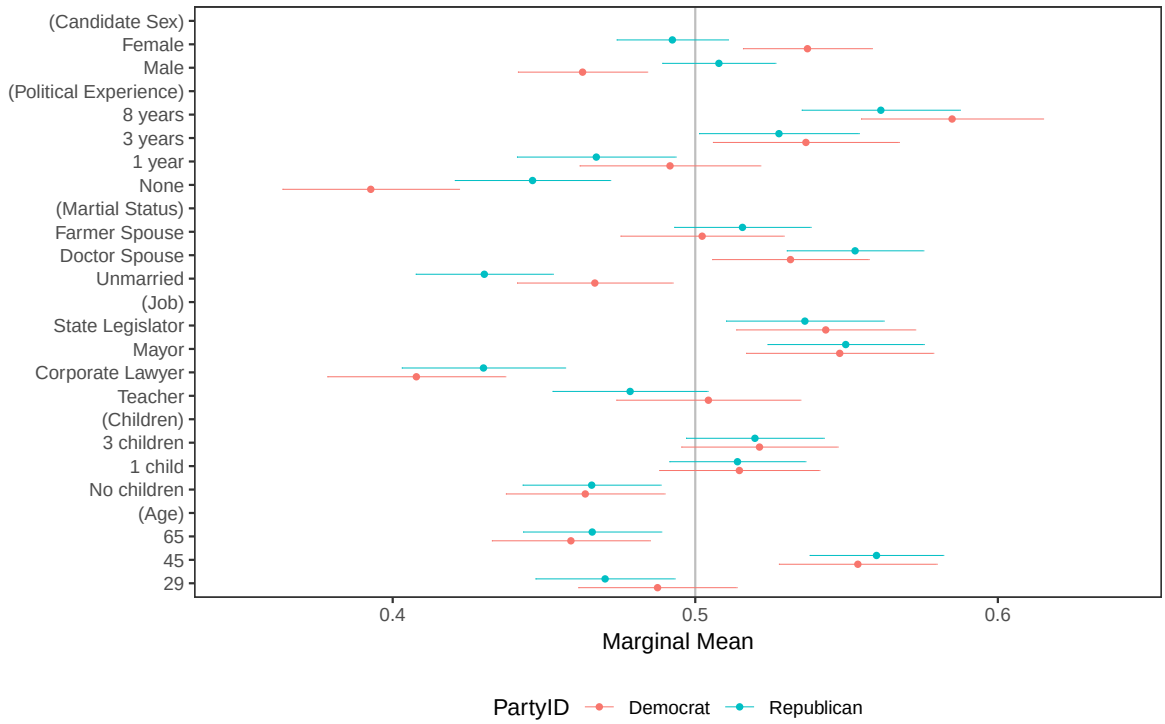
F.3 Numerical Results: MMs

feature	level	estimate	std.error	z
Candidate Sex	Male	0.48	0.01	-3.22
Candidate Sex	Female	0.52	0.01	3.20
Political Experience	None	0.42	0.01	-8.81
Political Experience	1 year	0.48	0.01	-2.56
Political Experience	3 years	0.53	0.01	3.58
Political Experience	8 years	0.57	0.01	7.99
Martial Status	Unmarried	0.45	0.01	-6.05
Martial Status	Doctor Spouse	0.54	0.01	4.66
Martial Status	Farmer Spouse	0.51	0.01	1.37
Job	Teacher	0.49	0.01	-1.01
Job	Corporate Lawyer	0.42	0.01	-8.44
Job	Mayor	0.54	0.01	4.77
Job	State Legislator	0.54	0.01	4.55
Children	No children	0.47	0.01	-4.47
Children	1 child	0.52	0.01	2.07
Children	3 children	0.52	0.01	2.34
Age	29	0.48	0.01	-2.65
Age	45	0.56	0.01	7.28
Age	65	0.46	0.01	-4.66

F.4 Subgroup Analysis using AMCEs



F.5 Subgroup Analysis using MMs



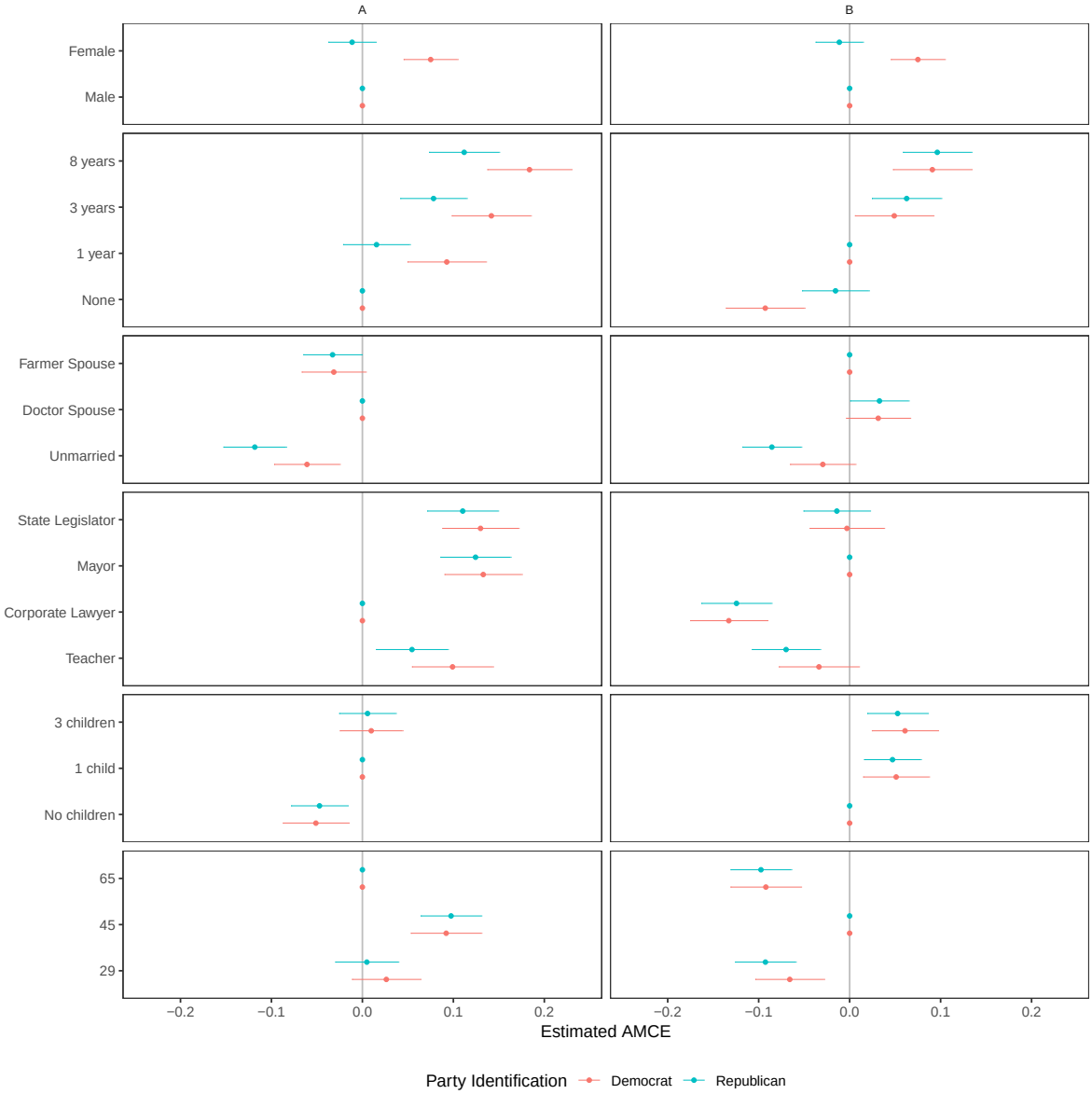
F.6 Nested Model Comparison: Male/Female Voters

```
## Analysis of Deviance Table
##
## Model 1: winner ~ feature_sex + feature_experience + feature_marital +
##     feature_job + feature_children + feature_age
## Model 2: winner ~ feature_sex + feature_experience + feature_marital +
##     feature_job + feature_children + feature_age + Sex + feature_sex:Sex +
##     feature_experience:Sex + feature_marital:Sex + feature_job:Sex +
##     feature_children:Sex + feature_age:Sex
##   Resid. Df Resid. Dev Df Deviance      F Pr(>F)
## 1     12436     2997.2
## 2     12422     2992.2 14   5.0738 1.5046 0.1002
```

F.7 Nested Model Comparison: Democratic/Republican Voters

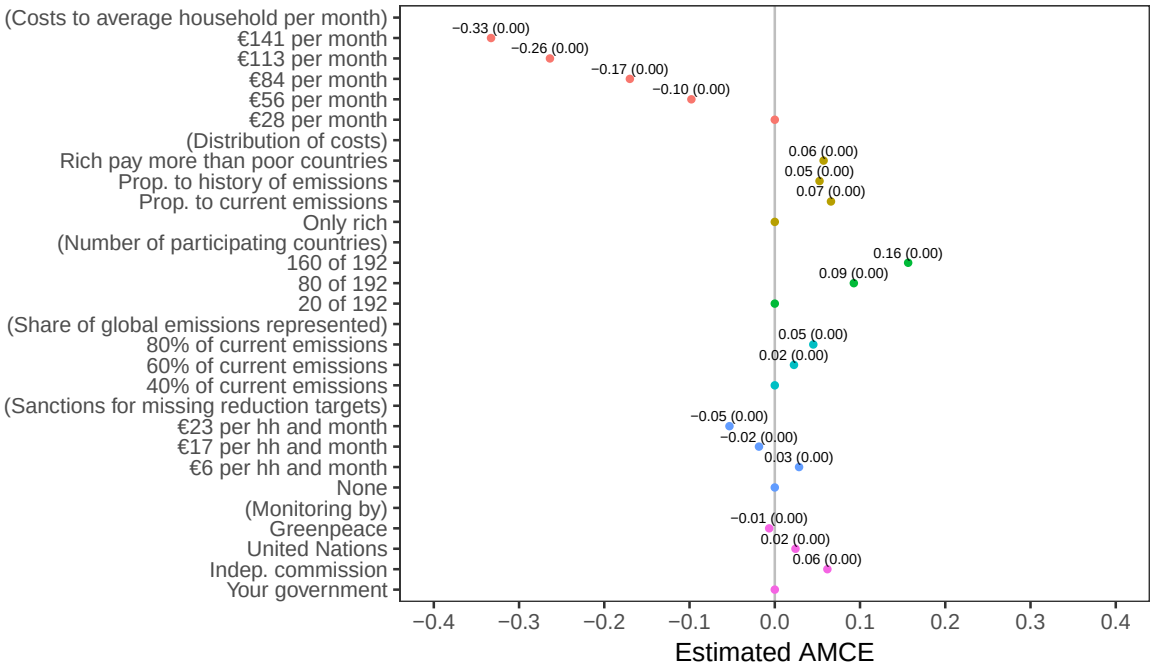
```
## Analysis of Deviance Table
##
## Model 1: winner ~ feature_sex + feature_experience + feature_marital +
##     feature_job + feature_children + feature_age
## Model 2: winner ~ feature_sex + feature_experience + feature_marital +
##     feature_job + feature_children + feature_age + PartyID +
##     feature_sex:PartyID + feature_experience:PartyID + feature_marital:PartyID +
##     feature_job:PartyID + feature_children:PartyID + feature_age:PartyID
##   Resid. Df Resid. Dev Df Deviance      F    Pr(>F)
## 1      9776     2352.8
## 2      9762     2342.9 14   9.8697 2.9373 0.0001739 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

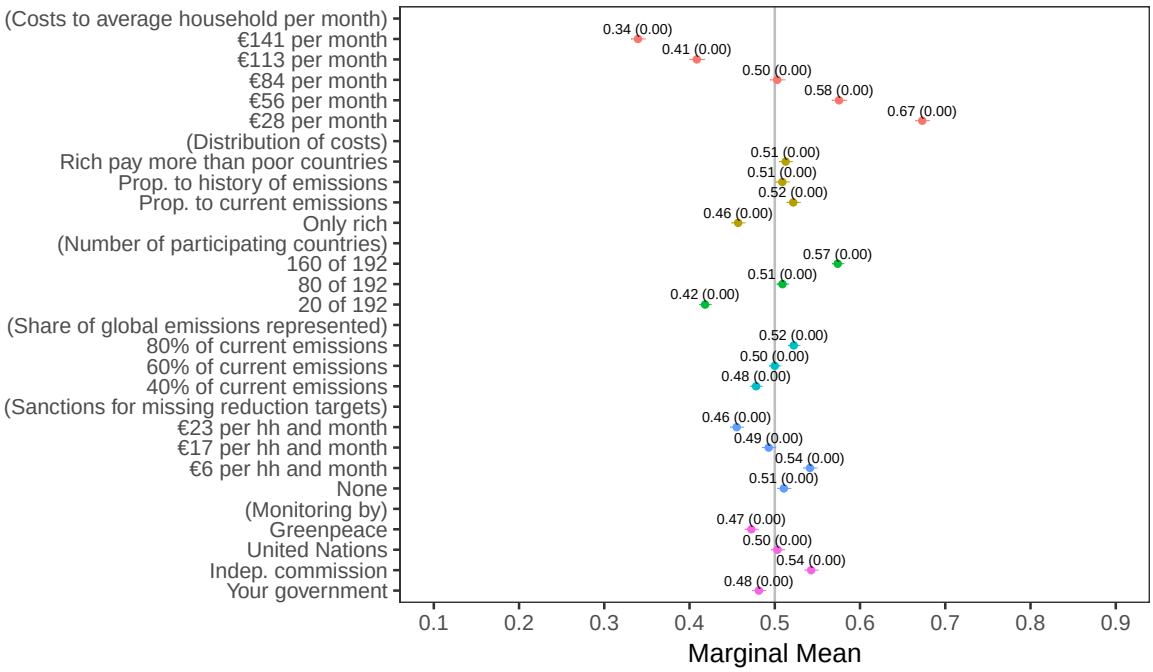
F.8 Comparison of Alternative Reference Categories



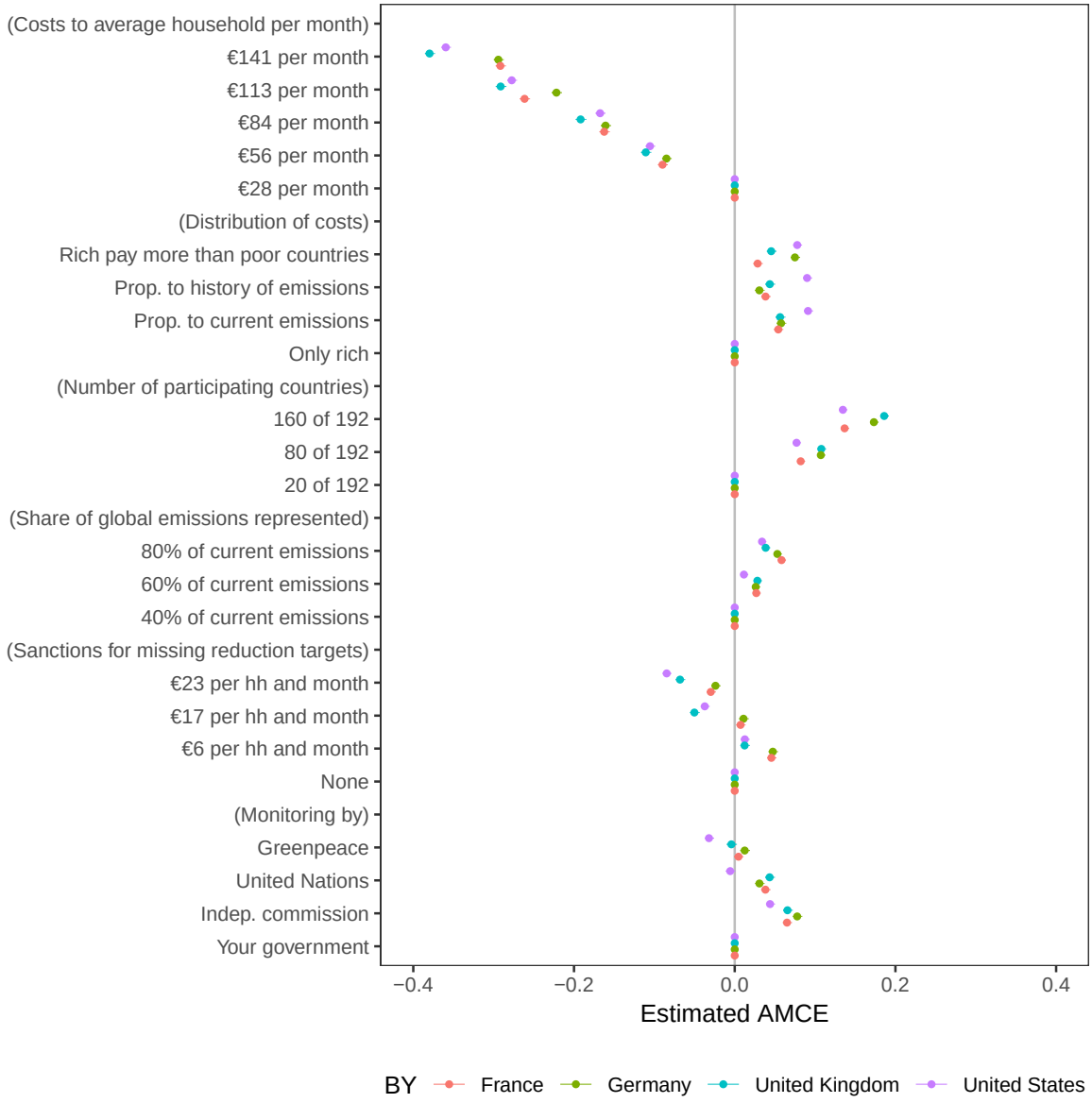
G Bechtel and Scheve (2013) Climate Agreement Experiment

G.1 Replication using AMCEs and MMs

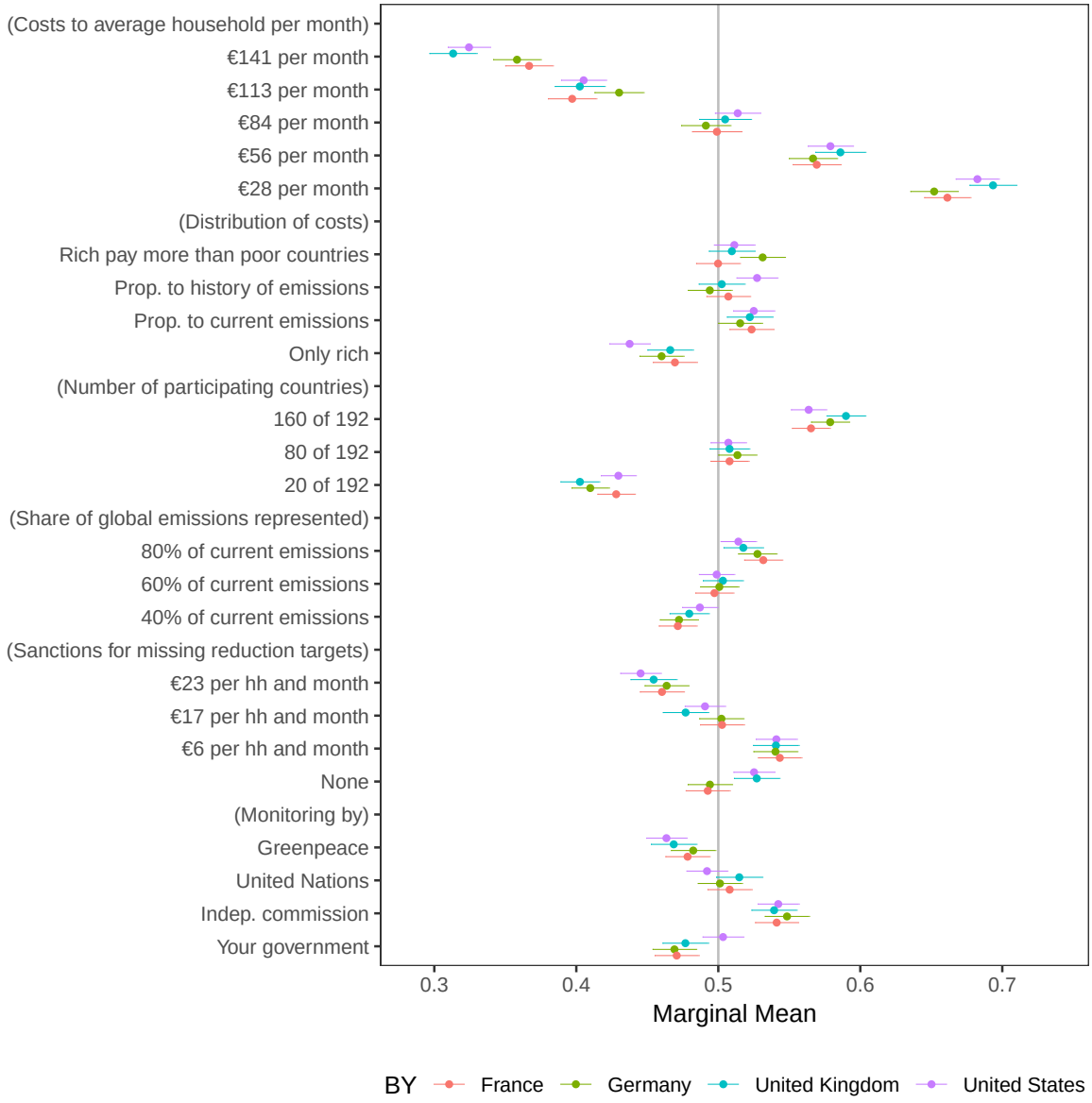




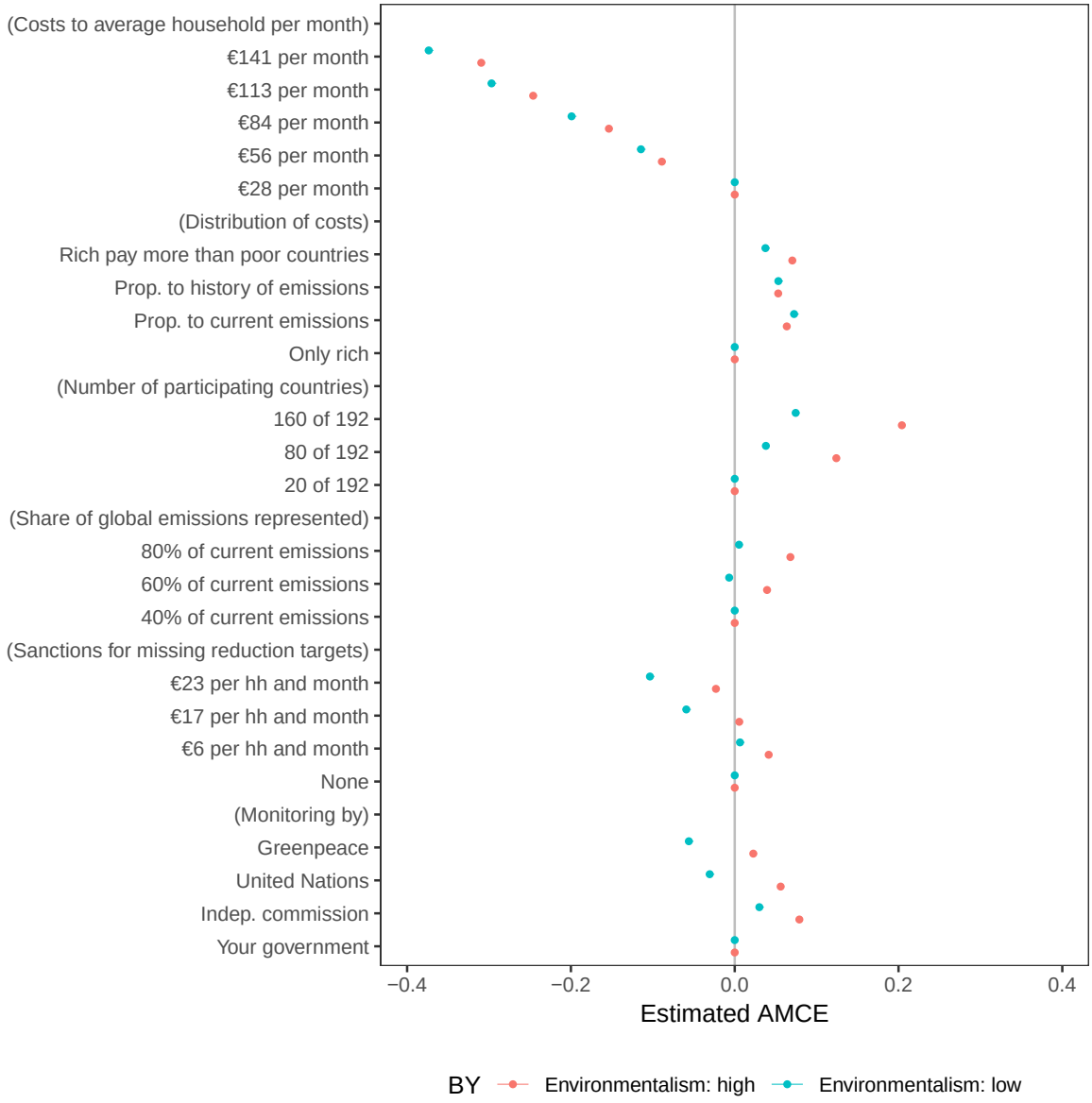
G.2 Subgroup Analysis using AMCEs: Country



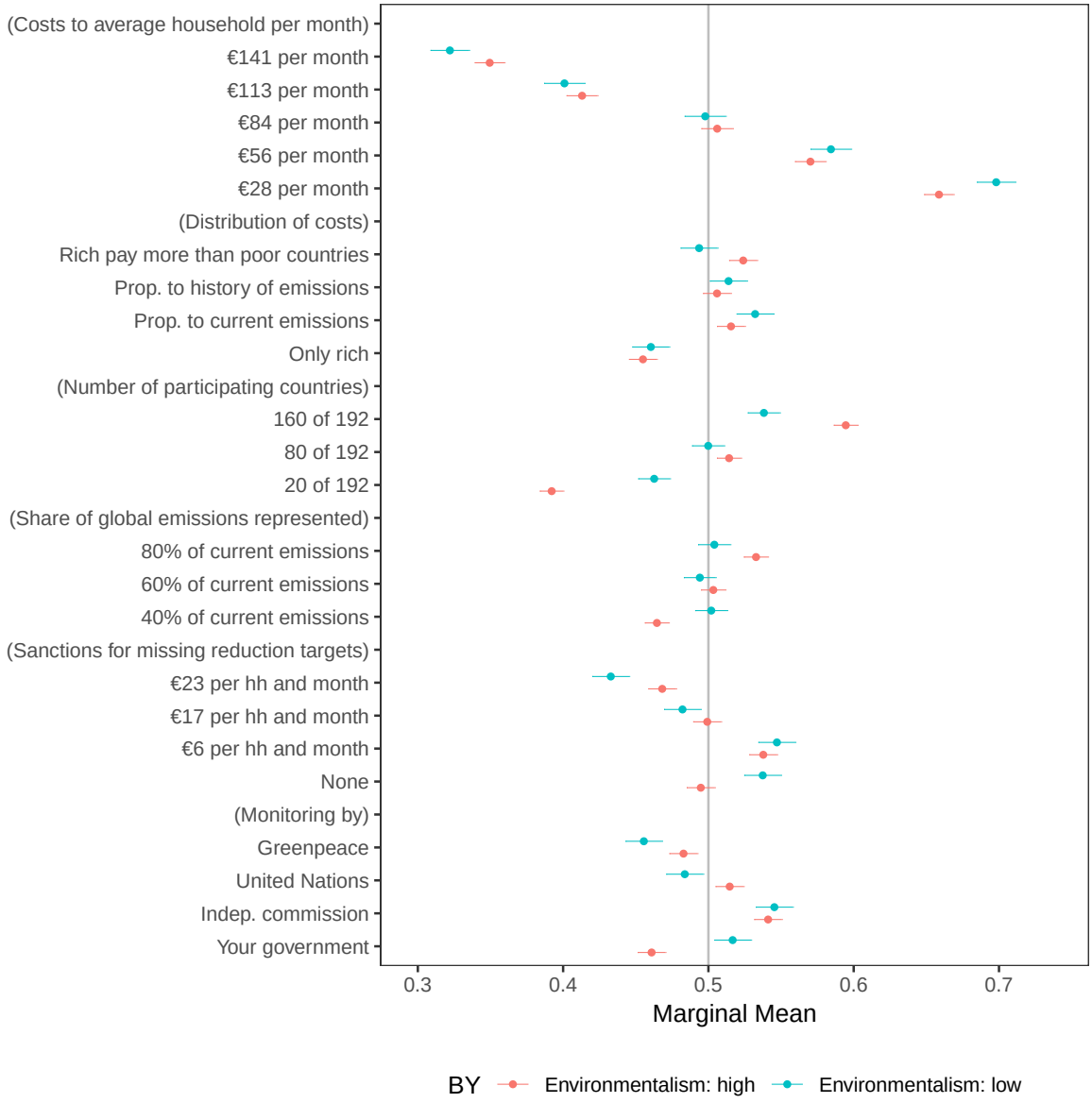
G.3 Subgroup Analysis using MMs: Country



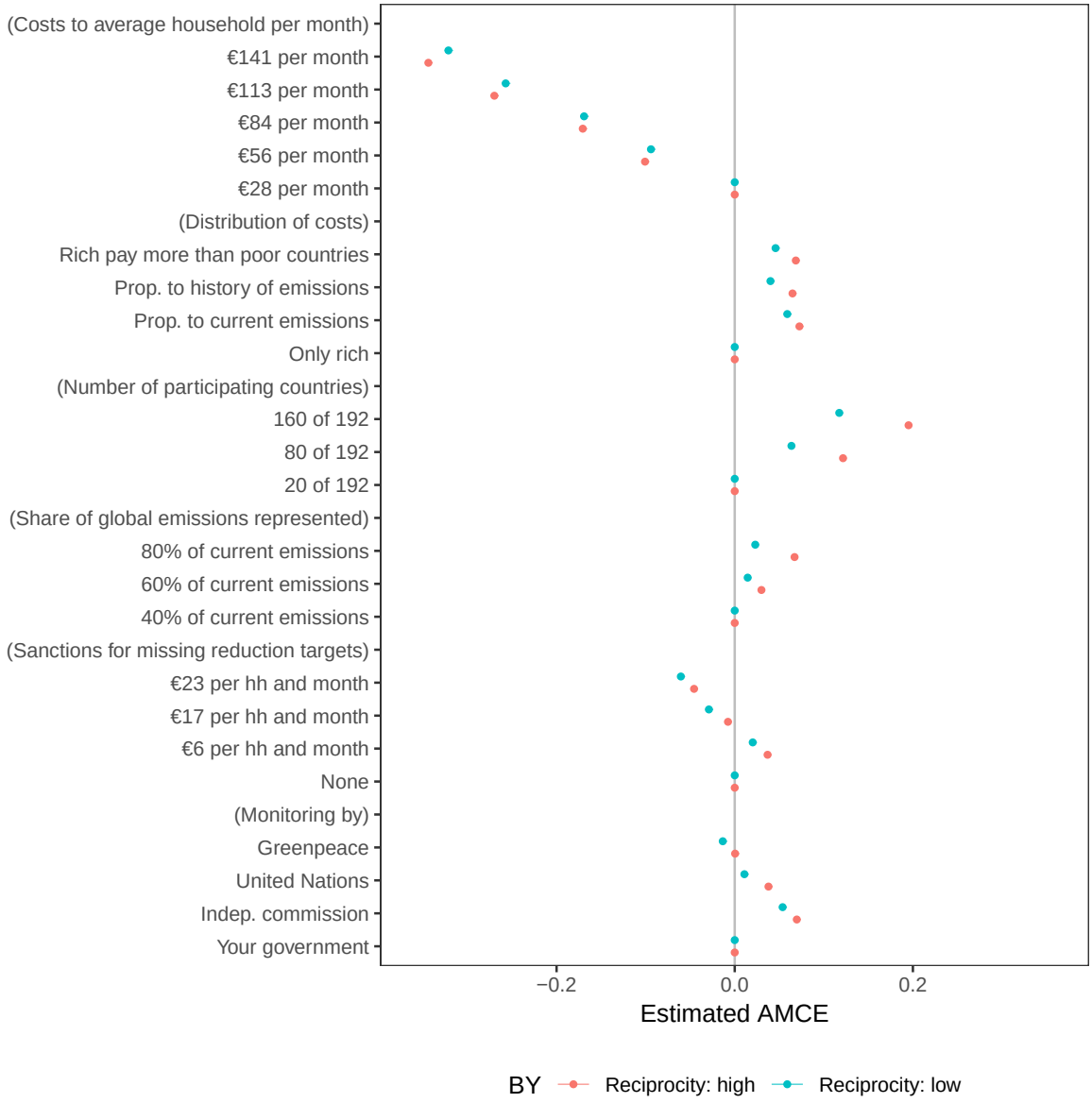
G.4 Subgroup Analysis using AMCEs: Environmentalism



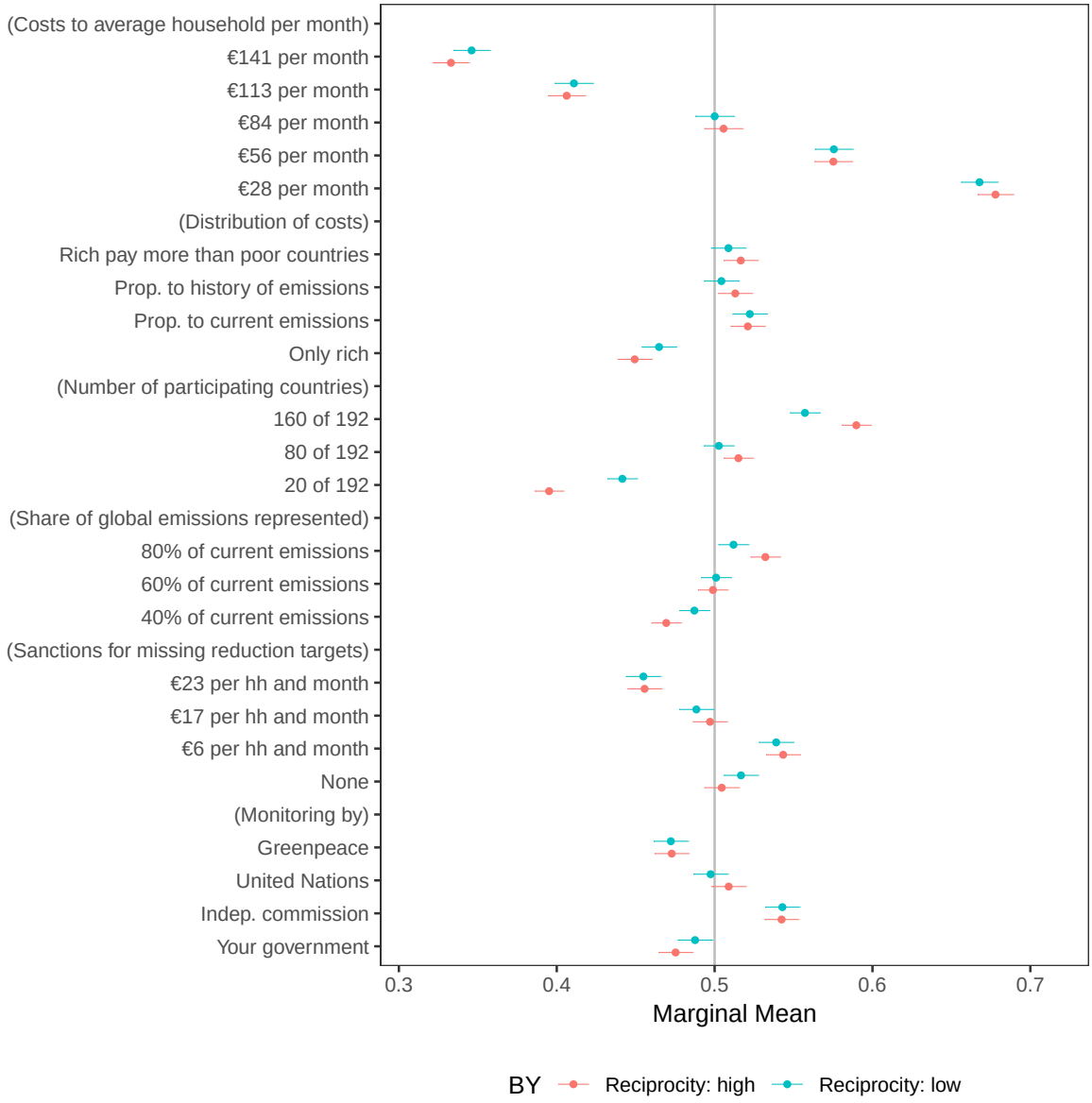
G.5 Subgroup Analysis using MMs: Environmentalism



G.6 Subgroup Analysis using AMCEs: Reciprocity



G.7 Subgroup Analysis using MMs: Reciprocity



G.8 Nested Model Comparison: Country

```
## Analysis of Deviance Table
##
## Model 1: choice_cj ~ cost_cj + distrib_cj + ctries_cj + emissions_cj +
##     sanctions_cj + monitoring_cj
## Model 2: choice_cj ~ cost_cj + distrib_cj + ctries_cj + emissions_cj +
##     sanctions_cj + monitoring_cj + country + cost_cj:country +
##     distrib_cj:country + ctries_cj:country + emissions_cj:country +
##     sanctions_cj:country + monitoring_cj:country
##   Resid. Df Resid. Dev Df Deviance      F    Pr(>F)
## 1     67982      15601
## 2     67928      15555 54   45.983 3.7187 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

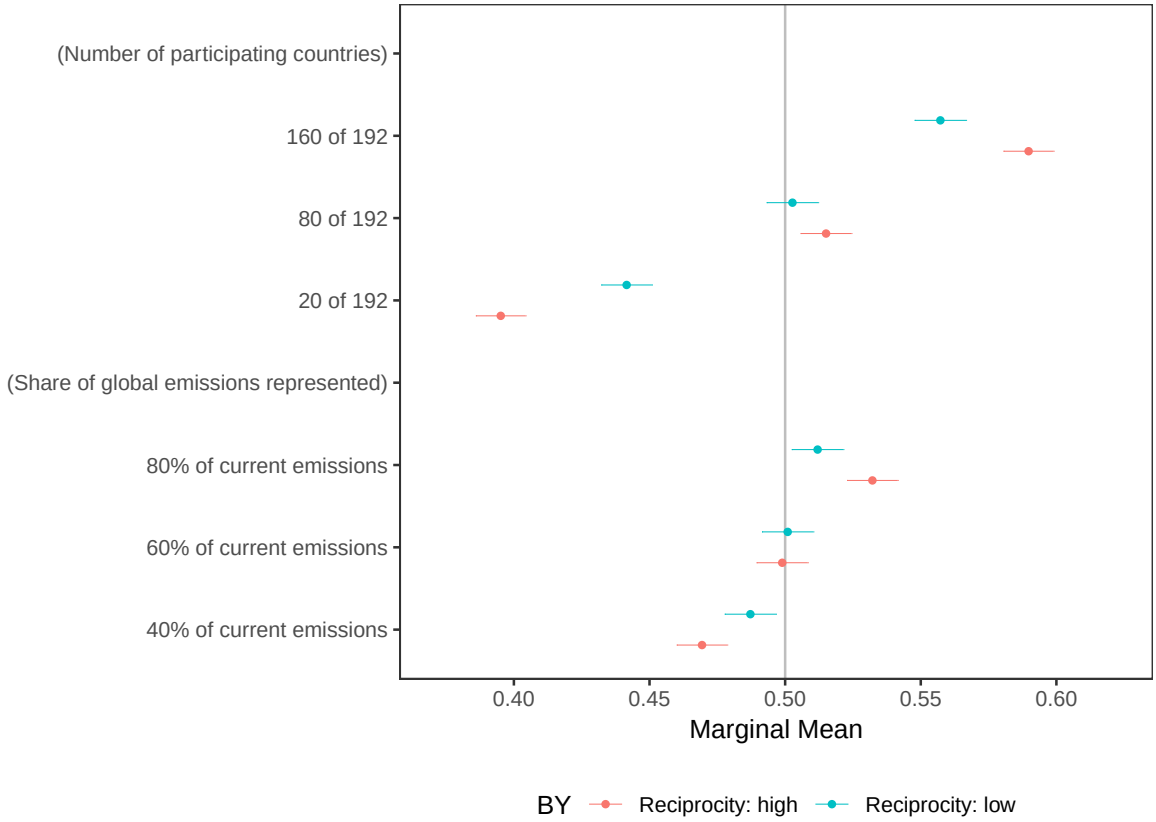
G.9 Nested Model Comparison: Environmentalism

```
## Analysis of Deviance Table
##
## Model 1: choice_cj ~ cost_cj + distrib_cj + ctries_cj + emissions_cj +
##     sanctions_cj + monitoring_cj
## Model 2: choice_cj ~ cost_cj + distrib_cj + ctries_cj + emissions_cj +
##     sanctions_cj + monitoring_cj + environmentalism + cost_cj:environmentalism +
##     distrib_cj:environmentalism + ctries_cj:environmentalism +
##     emissions_cj:environmentalism + sanctions_cj:environmentalism +
##     monitoring_cj:environmentalism
##   Resid. Df Resid. Dev Df Deviance      F    Pr(>F)
## 1     67974      15599
## 2     67956      15491 18   107.83 26.279 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

G.10 Nested Model Comparison: Reciprocity

```
## Analysis of Deviance Table
##
## Model 1: choice_cj ~ cost_cj + distrib_cj + ctries_cj + emissions_cj +
##     sanctions_cj + monitoring_cj
## Model 2: choice_cj ~ cost_cj + distrib_cj + ctries_cj + emissions_cj +
##     sanctions_cj + monitoring_cj + reciprocity + cost_cj:reciprocity +
##     distrib_cj:reciprocity + ctries_cj:reciprocity + emissions_cj:reciprocity +
##     sanctions_cj:reciprocity + monitoring_cj:reciprocity
##   Resid. Df Resid. Dev Df Deviance      F    Pr(>F)
## 1     67982      15601
## 2     67964      15570 18   30.831 7.4767 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

G.11 Comparison of Alternative Reference Categories



This paper was built using `knitr::knit2pdf()` under the following environment:

```
## R version 3.6.0 (2019-04-26)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 7 x64 (build 7601) Service Pack 1
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United Kingdom.1252
## [2] LC_CTYPE=English_United Kingdom.1252
## [3] LC_MONETARY=English_United Kingdom.1252
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United Kingdom.1252
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] gridExtra_2.3 ggplot2_3.1.1 cregg_0.3.0   rio_0.5.16
##
## loaded via a namespace (and not attached):
##  [1] zip_2.0.1         Rcpp_1.0.1        cellranger_1.1.0
##  [4] pillar_1.3.1      compiler_3.6.0    plyr_1.8.4
##  [7] forcats_0.4.0     tools_3.6.0       digest_0.6.18
## [10] lattice_0.20-38   ggstance_0.3.1    evaluate_0.13
## [13] tibble_2.1.1      gtable_0.3.0      pkgconfig_2.0.2
## [16] rlang_0.3.4       Matrix_1.2-17     openxlsx_4.1.0
## [19] DBI_1.0.0         curl_3.3          haven_2.1.0
## [22] xfun_0.6          withr_2.1.2       stringr_1.4.0
## [25] dplyr_0.8.0.1     knitr_1.22        mitools_2.4
## [28] hms_0.4.2         lmtest_0.9-36     grid_3.6.0
## [31] tidyselect_0.2.5  glue_1.3.1        data.table_1.12.2
## [34] R6_2.4.0          survival_2.44-1.1 readxl_1.3.1
## [37] foreign_0.8-71    purrr_0.3.2       magrittr_1.5
## [40] splines_3.6.0     scales_1.0.0      assertthat_0.2.1
## [43] xtable_1.8-4      colorspace_1.4-1  sandwich_2.5-1
## [46] survey_3.36       stringi_1.4.3     lazyeval_0.2.2
## [49] munsell_0.5.0     crayon_1.3.4      zoo_1.8-5
```