Supplementary Information for

Eliminating unintended bias in personalized policies using Bias Eliminating Adapted Trees (BEAT)

Eva Ascarza and Ayelet Israeli

E-mail: eascarza@hbs.edu or aisraeli@hbs.edu.

This PDF file includes:
- Supplementary text
- Fig. S1 (not allowed for Brief Reports)
- Tables S1 to S5 (not allowed for Brief Reports)
Supporting Information Text

1. Simulation: additional results

Table S1 presents the full set of results, including two new benchmarks for targeting (that use plugin estimators instead of causal forests), two additional de-biasing approaches (an additional pre-processing method where both X and Y are de-biased with respect to Z, and a post-processing method where the (predicted) individual treatment effect ($\hat{\tau}_i$) is de-biased with respect to Z), and two additional benchmark policies (random and uniform).

We present the full set of results for 8 scenarios varying the strength of the relationship between protected (Z) and unprotected (X) attributes, the relationship between Z, X and $\tau$, and the type of protected attribute Z (either binary or continuous). Scenarios 1–4 correspond to those presented in the main manuscript. Scenarios 5–8 are simulated as follows.

- Scenario 5 corresponds to low correlation between Z and X, $\tau = f(X, Z)$, and Z now is a continuous variable.
- Scenario 6 corresponds to high correlation between Z and X, $\tau = f(X, Z)$ when Z is binary.
- Scenario 7 corresponds to high correlation between Z and X, $\tau = f(X, Z)$, and Z being a continuous variable.
- Scenario 8 corresponds to low correlation between Z and X, $\tau = f(X, -Z)$, and Z being a continuous variable.
# Table S1. Simulated scenarios: Comparing outcomes across methods

| Scenario 1 | Scenario 2 |
|------------|------------|
| **Method** | **Efficiency** | **Imbalance** | **ΔPolicy** | **ΔRanking** | **Efficiency** | **Imbalance** | **ΔPolicy** | **ΔRanking** |
| CF-FD      | 81.2%       | 100.0%       | 82.8%       | 2356        | 78.7%       | 100.0%       | 82.6%       | 2373        |
| Plug-RF    | 81.3%       | 96.7%        | 81.7%       | 2338        | 79.1%       | 93.6%        | 80.2%       | 2344        |
| Plug-XGB   | 81.1%       | 98.3%        | 81.5%       | 2299        | 79.1%       | 96.1%        | 80.5%       | 2329        |
| CF-NP      | 71.1%       | 60.8%        | 0.0%        | 0           | 50.8%       | 5.6%         | 0.0%        | 0           |
| De-biased X| 45.8%       | 0.3%         | 13.5%       | 495         | 43.0%       | 0.1%         | 8.9%        | 304         |
| De-biased XY| 45.2%      | 0.3%         | 17.0%       | 531         | 43.4%       | 0.1%         | 9.0%        | 301         |
| De-biased Tau| 44.0%     | 0.0%         | 4.3%        | 155         | 43.4%       | 0.0%         | 5.5%        | 162         |
| BEAT       | 44.5%       | 0.1%         | 0.0%        | 0           | 42.1%       | 0.1%         | 0.0%        | 0           |
| Uniform    | -4.5%       | 0.0%         | 0.0%        | 0           | -1.1%       | 0.0%         | 0.0%        | 0           |

| Scenario 3 | Scenario 4 |
|------------|------------|
| **Method** | **Efficiency** | **Imbalance** | **ΔPolicy** | **ΔRanking** | **Efficiency** | **Imbalance** | **ΔPolicy** | **ΔRanking** |
| CF-FD      | 74.6%       | 100.0%       | 28.8%       | 950         | 76.3%       | 100.0%       | 1.6%        | 30          |
| Plug-RF    | 74.4%       | 96.1%        | 27.4%       | 927         | 75.8%       | 94.5%        | 2.1%        | 36          |
| Plug-XGB   | 75.3%       | 95.7%        | 28.1%       | 901         | 77.3%       | 67.2%        | 5.6%        | 206         |
| CF-NP      | 71.8%       | 30.3%        | 0.0%        | 0           | 76.7%       | 112.0%       | 0.0%        | 0           |
| De-biased X| 59.9%       | 0.6%         | 24.3%       | 872         | 75.0%       | 0.8%         | 26.7%       | 976         |
| De-biased XY| 59.9%      | 0.4%         | 24.2%       | 867         | 75.3%       | 1.0%         | 27.1%       | 980         |
| De-biased Tau| 61.0%     | 0.3%         | 4.4%        | 168         | 75.2%       | 0.4%         | 0.8%        | 31          |
| BEAT       | 44.8%       | 0.1%         | 0.0%        | 0           | 45.4%       | 0.5%         | 0.0%        | 0           |
| Uniform    | -3.4%       | 0.0%         | 0.0%        | 0           | -3.3%       | 0.0%         | 0.0%        | 0           |

| Scenario 5 | Scenario 6 |
|------------|------------|
| **Method** | **Efficiency** | **Imbalance** | **ΔPolicy** | **ΔRanking** | **Efficiency** | **Imbalance** | **ΔPolicy** | **ΔRanking** |
| CF-FD      | 76.3%       | 100.0%       | 32.3%       | 1066        | 77.4%       | 100.0%       | 53.5%       | 1548        |
| Plug-RF    | 76.1%       | 99.0%        | 32.8%       | 1078        | 77.5%       | 99.0%        | 52.0%       | 1473        |
| Plug-XGB   | 77.6%       | 93.2%        | 30.6%       | 978         | 77.6%       | 95.4%        | 28.6%       | 853         |
| CF-NP      | 65.9%       | 15.6%        | 0.0%        | 0           | 75.7%       | 74.7%        | 0.0%        | 0           |
| De-biased X| 48.9%       | 0.1%         | 8.1%        | 281         | 51.0%       | 0.1%         | 38.4%       | 1621        |
| De-biased XY| 49.1%      | 0.1%         | 7.8%        | 294         | 49.9%       | 0.1%         | 38.4%       | 1623        |
| De-biased Tau| 51.8%     | 0.3%         | 6.5%        | 252         | 48.5%       | 0.1%         | 14.9%       | 535         |
| BEAT       | 45.4%       | 0.1%         | 0.0%        | 0           | 45.3%       | 0.1%         | 0.0%        | 0           |
| Uniform    | -1.1%       | 0.0%         | 0.0%        | 0           | -3.0%       | 0.0%         | 0.0%        | 0           |

| Scenario 7 | Scenario 8 |
|------------|------------|
| **Method** | **Efficiency** | **Imbalance** | **ΔPolicy** | **ΔRanking** | **Efficiency** | **Imbalance** | **ΔPolicy** | **ΔRanking** |
| CF-FD      | 79.0%       | 100.0%       | 28.5%       | 916         | 73.8%       | 100.0%       | 42.4%       | 1352        |
| Plug-RF    | 79.1%       | 100.0%       | 29.3%       | 910         | 73.3%       | 98.5%        | 43.0%       | 1377        |
| Plug-XGB   | 79.6%       | 91.4%        | 23.6%       | 764         | 75.7%       | 89.2%        | 42.3%       | 1359        |
| CF-NP      | 74.8%       | 48.6%        | 0.0%        | 0           | 43.4%       | 0.1%         | 0.0%        | 0           |
| De-biased X| 50.4%       | 0.1%         | 13.6%       | 488         | 51.4%       | 0.4%         | 11.1%       | 373         |
| De-biased XY| 49.5%      | 0.1%         | 13.8%       | 496         | 51.4%       | 0.3%         | 11.5%       | 373         |
| De-biased Tau| 52.6%     | 0.6%         | 9.8%        | 389         | 55.0%       | 0.4%         | 9.1%        | 337         |
| BEAT       | 47.4%       | 0.1%         | 0.0%        | 0           | 41.8%       | 0.1%         | 0.0%        | 0           |
| Uniform    | -0.5%       | 0.0%         | 0.0%        | 0           | -2.1%       | 0.0%         | 0.0%        | 0           |

Reporting results comparing: CF-FD = CF with full data; Plug-RF = Regression forest plugin estimator; Plug-XGB = Xgboost plugin estimator; CF-NP = CF without protected attributes; De-biased uses regression forests to de-bias the data; de-bias just the X variables, both the X and the Y variables, or the predicted \( \gamma \); BEAT is denoted with * to highlight that scalar \( \gamma \) is set to a large number, yielding the most conservative estimates for efficiency and imbalance; Random is a random policy that targets 50% of customers in the test data at random; and Uniform is the optimal uniform policy that either targets everyone or no one, depending on which of the two policies yields a higher outcome.

**Efficiency** is measured as the percent increase in outcome over random allocation. **Imbalance** is normalized to 100% for the imbalance obtained with CF-FD (e.g., Column 2 in Scenario 1 should be read: CF-NP generates 60.8% of the imbalance obtained when using the full data). **ΔPolicy** measures the percent of individuals for which the outcome would change if their protected attributes were different. **ΔRanking** indicates the average change in ranking for identical individuals with different protected attributes.
2. Experiment: additional details and results

A. Experimental design. Consider Domino’s Pizza as an example of a firm which could leverage their customer data to increase the profitability of their marketing campaigns via personalization. Our ideal data would have replicated as closely as possible what Domino’s Pizza would do in this case, i.e., set up an experiment where coupons are randomly sent to individuals, measure and compare revenues and profits among those who received a coupon and those who did not, and design a targeting policy accordingly. However, due to the request to have access to detailed data including protected attributes, a collaboration with a firm was not possible. Instead, we set up an online experiment. The experiment consisted of a simple choice manipulation task akin to what companies call an “A/B Test.” Individuals in the control condition were offered a choice between $5 gift cards to Panera Bread and to Domino’s Pizza. In the treatment condition, participants were given a similar choice, but the Domino’s Pizza gift card had a value of $10 (mimicking a $5 coupon). To ensure truth-telling, the experiment was incentive-compatible—participants were informed that one of every 100 participants will receive the gift card of their choice.

In addition to the main experimental choice manipulation, we collected detailed information from participants about their brand preferences and consumption behaviors, similar to data a company may collect (e.g., via browsing cookies) or infer about its consumers: (i) preferences for brands in four categories (sportswear, apparel, chain restaurants, and technology and social media companies) which were rated on a three-point scale of dislike, indifferent, and like; (ii) consumption behaviors such as coupon usage, Facebook usage, exercise frequency, and other daily behaviors, rated on a six-point scale between “never” and “on a daily basis;” (iii) other general information such as subscription services usage, operating system, mobile service provider, dietary restrictions, and mTurk usage frequency; and (iv) other personal information such as monthly budget of various activities, pet ownership, price sensitivity, environmentally friendliness scale, education level, and monthly income. Those variables are used in the analysis as the unprotected attributes. Some of these may turn out to be “guilty” (i.e., related to the protected attributes) while others are “innocent” (unrelated to the protected attributes). We also collected information from the Qualtrics survey session (each participant’s latitude and longitude, and the duration of time they spent responding to the survey).

Finally, we also collected protected demographic attributes including gender, race, and age. We collected three different types of protected attributes in order to maximize the potential relationships between those and the unprotected attributes. The gift card task appeared at the end of the survey: after participants learned about the lottery, the next page presented the experiment stimuli and asked them to make a choice, and the final page asked them to enter their email to receive the gift card. Section 4 in the SI Appendix provides the full details of the study.

The back story for the study was that we were conducting research to better understand consumer preferences and choices, and the gift cards were presented as additional compensation for the study. Via Cloud Research, we recruited 3,205 participants who completed the survey. In addition to the gift card of their choice that 32 of the participants received, each participant was compensated $1.30 for their participation in the seven to eight minutes survey, a standard market rate at the time the study was conducted. Seven participants did not complete the gift card selection task and an additional 52 did not fill out their demographic information. Therefore, the final sample included 3,146 participants.

B. Experimental results. In this subsection, we first present summary statistics of the data. We then calculate the effect of the $10 gift card treatment in our experiment using a simple regression to demonstrate the average treatment effect in the data. We also use linear regressions to explore interaction effects of the protected attributes and the treatment to get a sense of the data. Finally, we present details for the forest-based analyses.

B.1. Summary statistics. The protected attributes in our data are based on three variables we collected from participants: age, gender, and race. Summary statistics for these variables are reported in Table S2. We highlight below summary statistics for the protected attributes in our data. With regard to gender: 56% of participants identified as female, 43.5% identified as male, 0.6% as non-binary, and the remaining participants identified as other (7 total). The reported age range was 18 to 92, with a median of 39 and average of 41.5. With regard to race, participants could indicate any race or ethnicity that applied. The reported race range was 18 to 92, with a median of 39 and average of 41.5. With regard to race, participants could indicate any race or ethnicity that applied. The resulting rates in the data are (by frequency): White (77%), Black or African American (10%), Asian (9%), Hispanic or Latino (6%), American Indian or Alaska Native (1%), Middle Eastern (.7%), Other (.7%), Native Hawaiian or Pacific Islander (.4%), and North African (.1%).

* Any firm that collects individual-level customer data and that has the capability to personalize its marketing promotional activity would be a valid candidate.
† For details, see: Litman, L., Robinson, J., and Abberbock, T. (2017). TurkPrime.com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. Behavior Research Methods, 49(2), 433-442.
‡ To calculate the required sample size, we first ran a pretest with 100 participants in each condition. Using a linear regression, we observed a statistically significant main effect of an increase Domino’s Pizza choice. We also saw some (not statistically significant) interactions between some of the protected variables and the treatment. We thus recruited 16 times the sample size to identify interaction effects. See https://statmodeling.stat.columbia.edu/2018/03/15/need-16-times-sample-size-estimate-interaction-estimate-main-effect/ for more details.
Table S2. Experiment variables summary Statistics

| Item                  | Average | SD   |
|-----------------------|---------|------|
| **Sportswear (rated 1-3)** |         |      |
| Nike                  | 2.47    | 0.748|
| Adidas                | 2.61    | 0.566|
| Reebok                | 2.34    | 0.638|
| Puma                  | 2.30    | 0.620|
| New Balance           | 2.44    | 0.625|
| Under Armor           | 2.48    | 0.617|
| Lululemon             | 1.97    | 0.591|
| Fabletics             | 2.00    | 0.480|
| Athleta               | 2.06    | 0.448|
| **Apparel (rated 1-3)** |         |      |
| Gucci                 | 2.03    | 0.689|
| Zara                  | 2.04    | 0.508|
| Dior                  | 2.08    | 0.694|
| Prada                 | 2.02    | 0.629|
| Calvin Klein          | 2.30    | 0.661|
| Ann Taylor            | 2.08    | 0.563|
| Theory                | 1.96    | 0.399|
| H&M                   | 2.20    | 0.643|
| Forever 21           | 1.99    | 0.662|
| GAP                   | 2.30    | 0.675|
| Old Navy              | 2.43    | 0.690|
| Banana Republic       | 2.22    | 0.649|
| **Restaurants (rated 1-3)** |         |      |
| Starbucks             | 2.31    | 0.801|
| Dunkin’ Donuts        | 2.47    | 0.682|
| Sweetgreen            | 1.99    | 0.374|
| Così                  | 1.96    | 0.375|
| Panera Bread          | 2.49    | 0.658|
| Pizza Hut             | 2.33    | 0.770|
| Domino’s Pizza        | 2.30    | 0.777|
| Burger King           | 2.25    | 0.798|
| McDonald’s            | 2.30    | 0.802|
| Tasty Burger          | 1.95    | 0.374|
| Shake Shack           | 2.20    | 0.569|
| Chopt                 | 1.97    | 0.363|
| Papa John’s           | 2.16    | 0.811|
| Little Caesars        | 2.13    | 0.789|
| **Technology (rated 1-3)** |         |      |
| Apple                 | 2.32    | 0.775|
| Google                | 2.63    | 0.649|
| Facebook              | 2.04    | 0.886|
| Amazon                | 2.69    | 0.614|
| Twitter               | 2.07    | 0.792|
| Instagram             | 2.27    | 0.741|
| Pinterest             | 2.26    | 0.722|
| TikTok                | 1.83    | 0.747|
| Snapchat              | 1.87    | 0.686|
| Ebay                  | 2.46    | 0.627|
| **Activity frequency (rated 1-6)** |         |      |
| Facebook              | 4.41    | 1.96 |
| Twitter               | 3.33    | 2.09 |
| Social media          | 5.31    | 1.32 |
| Take selfies          | 2.36    | 1.41 |
| Cook meals            | 5.23    | 1.17 |
| Use meal service      | 1.93    | 1.31 |
| Food delivery/pickup  | 3.04    | 1.40 |
| Shop online           | 4.01    | 1.23 |
| Return items          | 2.05    | 0.97 |
| TV                    | 4.55    | 1.75 |
| Streaming             | 4.82    | 1.51 |
| Grocery store         | 4.27    | 1.12 |
| Online groceries      | 2.41    | 1.53 |
| Exercise              | 4.40    | 1.59 |
| Talk to friend        | 4.51    | 1.56 |
| Dating app            | 1.40    | 1.07 |
| Drive a car           | 4.80    | 1.64 |
| Coupon Usage          | 3.43    | 1.31 |
| **Subscription services (Yes/No)** |         |      |
| Dollar Shave Club     | 0.121   | 0.326|
| Chewy                 | 0.220   | 0.415|
| Amazon Subcribe and Save | 0.217   | 0.412|
| Blue Apron            | 0.100   | 0.299|
| Hello Chef            | 0.066   | 0.249|
| Dropbox               | 0.481   | 0.500|
| **Average spending ($ amount)** |         |      |
| Delivery              | 43.0    | 72.2 |

Reporting average and standard deviation (SD) for variables collected from 3,146 participants in the experiment and used in subsequent analyses. In the interest of brevity, although used in the analyses the table does not report: i) Qualtrics’ inferred longitude and latitude; ii) mobile companies that were reported by less than 5% of respondents; and iii) sixteen dietary restriction categories.
B.2. Linear regression analyses. Using a simple linear regression, we find that the treatment increases the Domino’s gift card choice by 19.5 percentage points ($p$-value < 0.01), from a baseline of 46.6%. We then analyze the data using linear regressions with interactions of the form:

$$Dominos_i = \alpha + \beta Treatment_i + \gamma ProtectedChar_i + \delta Treatment \times ProtectedChar + \epsilon_i$$  \[1\]

where $Dominos_i$ indicates whether participant $i$ selected Domino’s in the choice task, $Treatment_i$ indicates whether that participant was in the treatment group, and $ProtectedChar_i$ indicates whether they belong to a specific protected attribute group. For simplicity and interpretability, we use indicator variables indicating whether gender is non-male, race is non-white, and age is greater than 39.\(^9\) We investigate the interactions with race, gender, and age. We find that non-males and whites are less likely to select the Domino’s gift card, by 15% and 10.6%, respectively. However, the interaction effects were indistinguishable from zero. For age, we do find a meaningful interaction. We find that older participants ($Age > 39$) were 10.9% less likely to choose Domino’s Pizza, and that the interaction effect between the age indicator and being in the treatment group corresponded with an additional 6.7% decrease in likelihood to choose Domino’s. These results remain when we use age as a continuous variable; there is a main effect of age such that every additional year is associated with a statistically significant 0.3% decrease in propensity to choose Domino’s, and an interaction effect of an additional decrease of 0.4% per year for those in the treatment group. Regression results are reported in Table S3.

Table S3. Experiment Results

| Specification | Main Effect | Non-Male | Non-White | 1 (Age>39) | Age |
|---------------|-------------|----------|-----------|------------|-----|
| Treatment     | 0.195***    | 0.205*** | 0.193***  | 0.246***   | 0.346*** |
|               | (0.017)     | (0.026)  | (0.020)   | (0.024)    | (0.057)  |
| ProtectedChar | -0.154***   | 0.106*** | -0.067*** | -0.004***  |       |
|               | (0.024)     | (0.029)  | (0.024)   | (0.001)    |       |
| Treatment $\times$ ProtectedChar | -0.012 | 0.007 | -0.109*** | -0.003*** |
|               | (0.035)     | (0.041)  | (0.034)   | (0.001)    |       |
| Constant      | 0.466***    | 0.551*** | 0.442***  | 0.498***   | 0.598*** |
|               | (0.012)     | (0.018)  | (0.014)   | (0.017)    | (0.041)  |

Significance level: 10% (*); 5% (**); 1% (***)

The specifications indicate which protected characteristic was used in regression Eq. (1). All regressions include 3,146 observations.

B.3. Forest-based analyses. Finally, we use all of the variables we collected in the data to compare the performance of the causal forest with the full data (CF-FD), the causal forest excluding the demographics data (CF-NP), the de-biased method, and BEAT. For these analyses, we used all the variables reported in Table S2, where the protected attributes are age, gender, and race; the assignment into treatment and control is defined by the experimental conditions; and the outcome variable is as before, as an indicator of whether participants chose the Domino’s gift card. In our analyses we used a 90/10 split for train and test samples, running a random split 1,000 times. The main results are reported in the main paper.

Table S4 reports the complete set of results for the variety of benchmarks we calculated. In addition to $\Delta$Policy, we report $\Delta$Ranking which indicates the average change in ranking for identical individuals with different protected attributes. The chosen protected attributes for which we calculated this method was Age. The total number of individuals in the test data is 315, and 50% are chosen to be targeted in each policy. The plugin XGboost estimator was first tuned using cross-validation and then we used the optimal parameters for the estimation. The uniform policy yields the highest outcome but the cost is also highest because this policy targets 100% of the population, whereas the other policies all target 50% of the population.

\(^9\)For age, we use the U.S.’s Age Discrimination in Employment Act (ADEA) as a guide, and indicate those who are 40 or older as protected. Coincidentally, the median age in our data is 39.
Table S4. Experiment results: Comparing outcomes across methods

| Method          | Efficiency | Imbalance | ∆Policy | ∆Ranking |
|-----------------|------------|-----------|---------|----------|
| CF-FD           | 0.578      | 0.157     | 16.7%   | 36.2     |
| (0.050)         | (0.064)    | (4.0%)    | (7.3)   |          |
| Plug-RF         | 0.594      | 0.073     | 10.0%   | 20.3     |
| (0.046)         | (0.03)     | (2.2%)    | (3.2)   |          |
| Plug-XGB        | 0.586      | 0.107     | 11%     | 25.4     |
| (0.049)         | (0.043)    | (2.3%)    | (3.6)   |          |
| CF-NP           | 0.574      | 0.057     | 0%      | 0        |
| (0.050)         | (0.026)    | –         | –       |          |
| De-biased X     | 0.569      | 0.240     | 42.3%   | 82.9     |
| (0.047)         | (0.088)    | (3.9%)    | (7.1)   |          |
| De-biased X,Y   | 0.569      | 0.224     | 41.0%   | 81.2     |
| (0.047)         | (0.09)     | (3.9%)    | (7.3)   |          |
| De-biased τ     | 0.573      | 0.034     | 8.7%    | 19.4     |
| (0.051)         | (0.012)    | (2.1%)    | (3.3)   |          |
| BEAT (γ = 3)    | 0.575      | 0.042     | 0%      | 0%       |
| (0.050)         | (0.018)    | –         | –       |          |
| BEAT (γ = 5)    | 0.573      | 0.042     | 0%      | 0%       |
| (0.049)         | (0.018)    | –         | –       |          |
| BEAT (γ = 8)    | 0.582      | 0.041     | 0%      | 0%       |
| (0.049)         | (0.017)    | –         | –       |          |
| Random          | 0.564      | 0.039     | 0%      | 0%       |
| (0.049)         | (0.017)    | –         | –       |          |
| Uniform         | 0.661      | 0.000     | 0%      | 0%       |
| (0.036)         | (0)        | –         | –       |          |

Reporting results comparing: CF-FD = CF with full data; Plug-RF = Regression forest plugin estimator; Plug-XGB = Xgboost plugin estimator; CF-NP = CF without protected attributes; De-biased uses regression forests to de-bias the data: de-bias just the X variables, both the X and the Y variables, or the predicted ŷ; BEAT estimated with moderate and high penalties for imbalance; Random is a random policy that targets 50% of customers; and Uniform is the optimal uniform policy that, in this case, targets everyone. Efficiency measures the proportion of users choosing the discounted product (i.e., market share). ∆Policy measures the percent of individuals for which the outcome would change if their protected attributes were different. Imbalance calculates the average distance between standardized protected attributes of targeted and non-targeted individuals in the test data. The Table reports the average outcomes of 1,000 runs of each method, and the standard deviation is reported in parentheses. ∆Ranking indicates the average change in ranking for identical individuals with different protected attributes.
To get a better understanding of which variables are playing the biggest role in these policy allocations, Figure S1 below reports the top 20 variables in terms of variable importance for each of the methods. As in the previous analysis, age turns out to be an important attribute for personalization. The methods that remove age from the attributes leveraged for personalization still generate some imbalance because they use unprotected attributes that are related to age and other protected attributes. Finally, BEAT removes most of the imbalance (but also all of the benefit from personalization), suggesting that in this case it eliminates the unprotected attributes that drive the incremental benefit over a uniform policy because they are also related to the protected attributes.

Fig. S1. Experiment results: Variable importance across methods.
The variable importance of the top 20 variables for each method.
3. Applying BEAT to prediction tasks using regression forests (instead of causal forests)

In this section we demonstrate how BEAT can be used in cases where the decision maker allocates the policy based on a prediction algorithm that does not require running an experiment. For example, let us suppose that, instead of running the experiment, the company uses historical data to predict which customers are most likely to respond to the coupon. Once that prediction model is calibrated, the firm would predict customer responses for a new sample of customers and allocate the coupon accordingly. They would either send coupons to all individuals whose responsiveness is above some threshold (e.g. above 0) or based on coupon costs. For illustration, and for consistency with the analyses in the manuscript, we assume that the company would send coupons to 50% of the new sample; in particular, to those for whom the model predicts the highest response.

To demonstrate this use case, we leverage the data collected in the experiment but recreate a situation in which the focal firm does not run any experiment but rather uses historical data for the prediction task. We use the control data to train a model that predicts which types of customers are likely to choose the Domino’s gift card (which, for this illustrative example, we take as a proxy for their responsiveness to the offer). We set optimal policies based on the model predictions for out-of-sample customers, and evaluate the policies using the outcomes of the treatment data. As with other analyses, we evaluate the policies using efficiency and imbalance metrics. We compute imbalance as before, using the Euclidean distance between the average value of the protected attributes (Z) between targeted and non-targeted units of each policy. For efficiency, we use the outcome generated by each allocation policy in the test data (i.e., proportion of targeted customers who would have chosen the focal brand).

Similar to the main analysis, we evaluate three policy methods: using regression forest with the full data (RF-FD), regression forest without using the protected attributes (RF-NP), and BEAT for prediction tasks. For each policy, we replicate this procedure 1,000 times (using different randomization seeds) and report the summary statistics in Table S5.

| Policy          | Imbalance | Outcome |
|-----------------|-----------|---------|
| RF-FD           | 0.083     | 0.436   |
| (0.005)         | (0.001)   |         |
| RF-NP           | 0.055     | 0.435   |
| (0.004)         | (0.001)   |         |
| BEAT (γ = 5)    | 0.009     | 0.374   |
| (0.001)         | (0.002)   |         |

Reporting results comparing: RF-FD = RF with full data; RF-NP = RF without protected attributes; BEAT estimated with penalty=5. The Table reports the average outcomes of 1,000 runs of each method, and the standard deviation is reported in parentheses. Imbalance calculates the average distance between standardized protected attributes of targeted and non-targeted individuals in the test data. Outcome reports the predicted coupon choice in the test sample among those who were targeted (i.e., market share).

The policy that uses the prediction task with the full data (RF-FD) is highly imbalanced, with differences across targeted and non-targeted units of 0.083. While removing the protected attributes (RF-NP) reduces imbalance (from 0.083 down to 0.055), the outcomes of such policy are still disproportionate with respect to the protected attributes. Interestingly, both efficiency metrics for the RF with and without demographics are not statistically different from each other (43.6% and 43.5%), suggesting that, in this case, one could achieve the full value of personalization from unprotected attributes only. However, as indicated by the imbalance metric, doing so would not allocate resources equally between protected and unprotected groups.

Importantly, BEAT reduces the imbalance significantly, providing the lowest level of imbalance corresponding to 89.2% below RF-FD and 83.6%, below RF-NP. As expected, BEAT results in a lower level of efficiency (37.4% vs. 43.6% and 43.5%), which corresponds to the “cost” of using fair policies.
### 4. Experiment Survey

#### Brands

Please indicate your preference for the following sportswear brands:

| Brands       | Dislike | Indifferent | Like |
|--------------|---------|-------------|------|
| New Balance  |         |             |      |
| Fabletics    |         |             |      |
| Reebok       |         |             |      |
| Adidas       |         |             |      |
| Puma         |         |             |      |
| Athleta      |         |             |      |
| Nike         |         |             |      |
| Under Armor  |         |             |      |
| Lululemon    |         |             |      |

Please indicate your preference for the following Apparel brands:

| Brands   | Dislike | Indifferent | Like |
|----------|---------|-------------|------|
| Dior     |         |             |      |
| Calvin Klein |      |             |      |
| GAP      |         |             |      |
| Gucci    |         |             |      |
| H&M      |         |             |      |
| Old Navy |         |             |      |
| Banana Republic | |             |      |
| Theory  |         |             |      |
| Forever 21 |        |             |      |
| Prada   |         |             |      |
| Zara    |         |             |      |
| Anne Taylor |        |             |      |
Please indicate your preference for the following restaurants and cafes:

|               | Dislike | Indifferent | Like |
|---------------|---------|-------------|------|
| Cosi          | ○       |             | ○    |
| Little Caesars| ○       |             | ○    |
| Domino’s Pizza| ○       |             | ○    |
| Sweetgreen    | ○       |             | ○    |
| Pizza Hut     | ○       |             | ○    |
| Panera Bread  | ○       |             | ○    |
| Tasty Burger  | ○       |             | ○    |
| Shake Shack   | ○       |             | ○    |
| Chopt         | ○       |             | ○    |
| Starbucks     | ○       |             | ○    |
| Papa John’s   | ○       |             | ○    |
| Dunkin Donuts | ○       |             | ○    |
| McDonalds     | ○       |             | ○    |
| Burger King   | ○       |             | ○    |

Please indicate your preference for the following technology companies and social networks:

|               | Dislike | Indifferent | Like |
|---------------|---------|-------------|------|
| Tik Tok       | ○       |             | ○    |
| Pinterest     | ○       |             | ○    |
| Amazon        | ○       |             | ○    |
| Google        | ○       |             | ○    |
| Apple         | ○       |             | ○    |
| Facebook      | ○       |             | ○    |
| Snapchat      | ○       |             | ○    |
| Twitter       | ○       |             | ○    |
| Instagram     | ○       |             | ○    |
| Etsy          | ○       |             | ○    |
# Consumer Behavior Information

How frequently do you engage in the following activities?

| Activity                                      | Never | 0-1 times per month | About half the time | 0-1 times per week | On a weekly basis | On a daily basis |
|-----------------------------------------------|-------|---------------------|---------------------|--------------------|-------------------|-----------------|
| Use Facebook                                  | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Use Twitter                                   | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Use any social media                          | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Take selfies                                  | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Cook your meals                               | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Use a meal delivery service                   | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Order meal delivery/pickup from a restaurant  | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Shop online                                   | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Return items you previously bought            | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Watch a show/movie on tv                      | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Watch a show/movie on a streaming service (e.g. netflix, disney+) | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Go to a grocery store                         | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Order your groceries online                   | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Exercise                                      | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Talk to your best friend                      | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Use a dating app                              | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Drive a car                                   | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
| Use a coupon or a promo code                  | ○     | ○                   | ○                   | ○                  | ○                 | ○               |
Which of these subscription services have you used? (Please select all that apply)

- Blue Apron
- Dollar Shave Club
- Dropbox
- Hello Chef
- Amazon Subscribe and Save
- Chewy

Please estimate your average monthly spend (in dollars) on:

- Delivery
- Grocery
- Exercise/Sports
- Entertainment/ Sports
- Activity
- Online Shopping
- Offline Shopping

Who is your mobile phone service provider (e.g. AT&T, Verizon, Cricket, Boost Mobile)?

Demographics

Do you have any dietary restriction?

- Yes
- No

What is your restriction? (please select all that apply)

- Gluten-free or celiac
- Milk/dairy allergy
- Kosher diet
- Counting calories
- Do not currently consume milk
- Pescatarian
- Vegan
- Vegetarian
- Lactose-intolerance
- Other
Do you own any pets?

Yes
No

Please estimate how many studies you have completed on Mturk:

Today
In the last 7 days
In the last month

Please indicate the extent to which you agree or disagree with the following statements:

| Statement                                                                 | Strongly Disagree | Mildly Disagree | Unsure | Mildly Agree | Strongly Agree |
|--------------------------------------------------------------------------|-------------------|-----------------|--------|--------------|----------------|
| I would describe myself as environmentally responsible.                 | ○                 | ○               | ○      | ○            | ○              |
| My purchase habits are affected by my concern for our environment.       | ○                 | ○               | ○      | ○            | ○              |
| It is important to me that the products I use do not harm the environment.| ○                 | ○               | ○      | ○            | ○              |
| I am willing to be inconvenienced in order to take actions that are more environmentally friendly. | ○                 | ○               | ○      | ○            | ○              |
| I consider the potential environmental impact of my actions when making many of my decisions. | ○                 | ○               | ○      | ○            | ○              |
| I am concerned about wasting the resources of our planet.                | ○                 | ○               | ○      | ○            | ○              |
Please indicate the extent to which you agree or disagree with the following statements:

| Statement                                                                 | Strongly Disagree | Disagree | Somewhat disagree | Neither agree nor disagree | Somewhat agree | Agree | Strongly agree |
|---------------------------------------------------------------------------|--------------------|----------|-------------------|----------------------------|----------------|-------|----------------|
| The money saved by finding low prices is usually not worth the time and effort. | ○                  | ○        | ○                 | ○                          | ○              | ○     | ○              |
| The time it takes to find low prices is usually not worth the effort.      | ○                  | ○        | ○                 | ○                          | ○              | ○     | ○              |
| I typically seek out cheap retail outlets to buy products for the house.   | ○                  | ○        | ○                 | ○                          | ○              | ○     | ○              |
| I would never shop at more than one store to find low prices.              | ○                  | ○        | ○                 | ○                          | ○              | ○     | ○              |
| I will grocery shop at more than one store to take advantage of low prices.| ○                  | ○        | ○                 | ○                          | ○              | ○     | ○              |
| I am not willing to go to extra effort to find lower prices.               | ○                  | ○        | ○                 | ○                          | ○              | ○     | ○              |

What is your current operating system (OS)?

PC
Mac
Other

Please select your gender below:

Male
Female
Non-Binary
Other
Choose one or more races or ethnicities that you consider yourself to be:

American Indian or Alaska Native
Asian
Black or African American
Hispanic or Latino
Native Hawaiian or Pacific Islander
Middle Eastern
North African
White
Other

Please input your age (in years) below:

What is the highest level of school you have completed or the highest degree you have received?

Less than high school degree
High school graduate (high school diploma or equivalent including GED)
Some college but no degree
Associate degree in college (2-year)
Bachelor’s degree in college (4-year)
Master’s degree
Doctoral degree
Professional degree (JD, MD)

Information about income is very important to understand. Would you please give your best guess?
Please indicate the answer that includes your entire household income in (previous year) before taxes.

Less than $10,000
$10,000 to $19,999
$20,000 to $29,999
$30,000 to $39,999
$40,000 to $49,999
$50,000 to $59,999
$60,000 to $69,999
$70,000 to $79,999
$80,000 to $89,999
$90,000 to $99,999
$100,000 to $149,999
$150,000 or more
Gift Cards

To thank you for your participation, we are going to enter you into a lottery where one out of every hundred participants will receive a gift card to one of two restaurants.

**Control:**
If selected to win the lottery, which of the following restaurants would you like a gift-card to?

- Panera Bread
- Domino's

**Treatment:**
If selected to win the lottery, which of the following restaurants would you like a gift-card to?

- Panera Bread
- Domino's

Please enter your email address below where you would like to receive the gift card if selected to win the lottery:

[Email field]