Smiles in Profiles: Improving Fairness and Efficiency Using Estimates of User Preferences in Online Marketplaces

Pre-analysis Plan Experiment 2

Susan Athey* Dean Karlan† Emil Palikot‡ Yuan Yuan§

November 13, 2024

Abstract

Online platforms often face the challenge of being both fair (i.e., non-discriminatory) and efficient (i.e., maximizing revenue). Using computer vision algorithms and observational data from a micro-lending marketplace, we find that the choices that online borrowers make when creating online profiles impact both of these objectives. We further support this finding with a web-based randomized survey experiment. In the experiment, we create profile images using Generative Adversarial Networks that differ in a specific feature and estimate the impact of the feature on lender demand. We then evaluate counterfactual platform policies based on the changeable profile features, and identify approaches that can ameliorate the fairness-efficiency tension.

*Stanford Graduate School of Business
†Kellogg School of Management, Northwestern University
‡Stanford Graduate School of Business
§Carnegie Mellon University; Stanford Graduate School of Business

1 Introduction

This document outlines the experimental design intended to evaluate the effect of specific borrower profile image features on the probability of obtaining a loan on the microfinance platform Kiva.org. Based on the analysis of the historical data on lending outcomes on the Kiva platform, we selected the following profile image characteristics, which are hypothesized to influence lender decision-making:

- *Age* – a binary variable classifying an individual's facial appearance in an image as aged (coded as 1) or youthful (coded as 0),
- *Sunglasses* – a binary variable indicating the presence of sunglasses in the image (coded as 1 if present, 0 otherwise),
- *Glasses* – a binary variable indicating the presence of regular glasses (coded as 1 if present, 0 otherwise),
- *Dark Hair* – a binary variable indicating whether the person in the image has dark hair (coded as 1 if present, 0 otherwise).

Experimental Design

To assess the impact of these features, we conduct an experiment in which subjects are presented with pairs of borrower profiles and asked to select the profile they prefer. Each subject is assigned to a protocol consisting of eight choice situations. Seven out of the eight pairs present images of hypothetical borrowers, while the final pair presents real borrowers who are actively seeking funding on Kiva.org.

The images of the hypothetical borrowers are generated using Generative Adversarial Networks (GANs) to introduce variation in the experimental features. We begin by selecting a set of 19 baseline images from Shutterstock.com. All baseline images are carefully chosen to resemble typical borrower images on Kiva.org. In pilot testing, we confirmed that subjects could not differentiate between the real borrower images and the stock images.

Each baseline image is systematically modified to introduce variation across the experimental features. Specifically, each image is altered to reflect combinations of glasses (no glasses, regular glasses, or sunglasses), age (old or young), and hair color (dark or light), resulting in 12 distinct versions for each baseline image. These image variations are then randomly assigned across the protocols. We employ stratified randomization to ensure that in each protocol a baseline image appears only once. Specifically, we create strata at the baseline image level and select randomly from these strata.

In the final choice pair, subjects are shown profiles of real borrowers currently fundraising on Kiva.org. These profiles are identified using Kiva's API[1]. From the images obtained from Kiva, we select those that show only one person in the main image; additionally, the information available on Kiva.org is simplified to fit the protocol format.

Incentive Structure

To incentivize participation and ensure subject engagement, we provide a real monetary stake in the experiment. Specifically, for each subject, we will loan \$10 to one of the actual borrowers that they select during the experiment. This ensures that the selection decisions made by subjects have real-world consequences.

2 Samples used in the experiment

We recruit 400 experimental subjects on Prolific.com. Subjects are recruited from the pool of users from the USA or UK, who are fluent in English, and have made a charitable donation in the previous year.

3 Hypotheses to be tested

Let $P(\text{Selected})$ denote the probability that a subject selects a lender

$$P(\text{Selected}) = f(Age, Sunglasses, Glasses, Dark\ Hair)$$

We test the following hypotheses for each feature:

[1]See here for more details about the API: https://www.kiva.org/build/docs

$$H_1 : \frac{\partial f}{\partial Age} = 0 \qquad \text{(Age has no effect)}$$
$$H_2 : \frac{\partial f}{\partial Sunglasses} = 0 \qquad \text{(Sunglasses have no effect)}$$
$$H_3 : \frac{\partial f}{\partial Glasses} = 0 \qquad \text{(Glasses have no effect)}$$
$$H_4 : \frac{\partial f}{\partial Dark\ Hair} = 0 \qquad \text{(Dark Hair has no effect)}$$

4 Variable construction

Each choice option has the following covariates: age, sunglasses, glasses, and dark hair (binary indicators), and an indicator of the baseline image.

The outcome of being selected takes the value of 0 or 1. Finally, we collect the following sociodemographic information about the experimental subjects: age, amount of charitable giving in the previous year, gender, ethnicity, country of birth, country of residence, nationality, student status, and employment status.

5 Treatment effects equations

We estimate treatment effects using the conditional logit model, where we account for both the characteristics of the image that was selected and the characteristics of the non-selected image. We adjust the estimates using the image characteristics.

1. Conditional Logit Model

The conditional logit model with image fixed effects:

$$P(\text{Selected} = 1 \mid Age, Sunglasses, Glasses, Dark\ Hair, \text{Image Fixed Effects}) = \frac{\exp\left(\beta_1 \cdot Age + \beta_2 \cdot Sunglasses + \beta_3 \cdot Glasses + \beta_4 \cdot Dark\ Hair + \lambda_i\right)}{\sum_{j=1}^{J} \exp\left(\beta_1 \cdot Age_j + \beta_2 \cdot Sunglasses_j + \beta_3 \cdot Glasses_j + \beta_4 \cdot Dark\ Hair_j + \lambda_i\right)}$$

Where: - λ_i represents image fixed effects.

Additional we include a specification with subject characteristics.

2. Logit Model

We also consider the logit model with image fixed effects:

$$P(\text{Selected} = 1 \mid \textit{Age}, \textit{Sunglasses}, \textit{Glasses}, \textit{Dark Hair}, \text{Image Fixed Effects}) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1 \cdot \textit{Age} + \beta_2 \cdot \textit{Sunglasses} + \beta_3 \cdot \textit{Glasses} + \beta_4 \cdot \textit{Dark Hair} + \lambda_i\right)\right)}$$

Where: - λ_i represents image fixed effects.

Additional specification with subject characteristics X:

$$P(\text{Selected} = 1 \mid \textit{Age}, \textit{Sunglasses}, \textit{Glasses}, \textit{Dark Hair}, X, \text{Image Fixed Effects}) = \frac{1}{1 + \exp\left(-\left(\beta_0 + \beta_1 \cdot \textit{Age} + \beta_2 \cdot \textit{Sunglasses} + \beta_3 \cdot \textit{Glasses} + \beta_4 \cdot \textit{Dark Hair} + \gamma \cdot X + \lambda_i\right)\right)}$$

Where: - X represents subject characteristics.

3. Linear Regression Model

Finally, we also use the linear regression model with image fixed effects:

$$P(\text{Selected}) = \beta_0 + \beta_1 \cdot \textit{Age} + \beta_2 \cdot \textit{Sunglasses} + \beta_3 \cdot \textit{Glasses} + \beta_4 \cdot \textit{Dark Hair} + \lambda_i + \epsilon$$

Where: - λ_i represents image fixed effects.

Additional specification with subject characteristics X:

$$P(\text{Selected}) = \beta_0 + \beta_1 \cdot \textit{Age} + \beta_2 \cdot \textit{Sunglasses} + \beta_3 \cdot \textit{Glasses} + \beta_4 \cdot \textit{Dark Hair} + \gamma \cdot X + \lambda_i + \epsilon$$

Where: - X represents subject characteristics.

5.1 Minimum Detectable Effect

Each experimental subject observes 16 profiles. Thus, 400 subjects generate 6400 observations. For each feature, the treated group is a profile for which the feature takes the value of 1 and the control group when the value is 0. The size of treatment and control groups is 3200, and the baseline probability of being selected 0.5. Using the two-sample proportion test, we find that the minimum detectable effect is a 2.47 percentage point difference between treatment and control in the probability of being selected. We have the same minimum detectable effect for all features of interest.

6 Experiment attrition

Subjects who did not consent to participate in the experiment are removed from the analysis. Subjects who failed to complete the protocol are removed from the analysis.

In the protocol, we include the attention check, in which we ask for a purpose of the loan presented in the previous slide. Based on the responses we construct an indicator variable taking the value of 1 when the subject has correctly responded to the question and 0 otherwise. We carry out a robustness check of this analysis in which we remove subjects that did not respond correctly.