Mitigating Bias in Online Microfinance Platforms: A Case Study on Kiva.org

Soumajyoti Sarkar(✉) and Hamidreza Alvari

Arizona State University, Tempe, USA
{ssarka18,halvari}@asu.edu

Abstract. Over the last couple of decades in the lending industry, financial disintermediation has occurred on a global scale. Traditionally, even for small supply of funds, banks would act as the conduit between the funds and the borrowers. It has now been possible to overcome some of the obstacles associated with such supply of funds with the advent of online platforms like Kiva, Prosper, LendingClub. Kiva, in particular, allows lenders to fund projects in different sectors through group or individual funding. Traditional research studies have investigated various factors behind lender preferences purely from the perspective of loan attributes and only until recently have some cross-country cultural preferences been investigated. In this paper, we investigate lender perceptions of economic factors of the borrower countries in relation to their preferences towards loans associated with different sectors. We find that the influence from economic factors and loan attributes can have substantially different roles to play for different sectors in achieving faster funding. We formally investigate and quantify the hidden biases prevalent in different loan sectors using recent tools from causal inference and regression models that rely on Bayesian variable selection methods. We then extend these models to incorporate fairness constraints based on our empirical analysis.

Keywords: Linear regression · Causal inference · Machine learning · Online lending

1 Introduction

Online lending in recent years has been considered to be an important contributor to financial restructuring in developing and underdeveloped nations by way of opening access to alternate sources of funding for them [4]. Online platforms that enable such peer-to-peer transactions whereby certain groups of people invest in projects from poor entrepreneurs, have become very popular. There exist different types of microlending services including for-profit lending services like LendingClub, Prosper and the pro-social platforms like Kiva[1] where the lenders offer interest-free money to the borrowers. Platforms like Kiva are beneficial to

[1] http://www.kiva.org.

I. Koprinska et al. (Eds.): ECML PKDD 2020 Workshops, CCIS 1323, pp. 75–91, 2020.
https://doi.org/10.1007/978-3-030-65965-3_6

borrowers, since lenders typically are risk-free indicating they do not expect any interest returns for the loan and hence can select their portfolio being less biased. Additionally, such pro-social platforms overcome the biases in loan disbursement through auctions in online platforms which is unfavorably inclined towards the credit-trustworthy users and undermines new users.

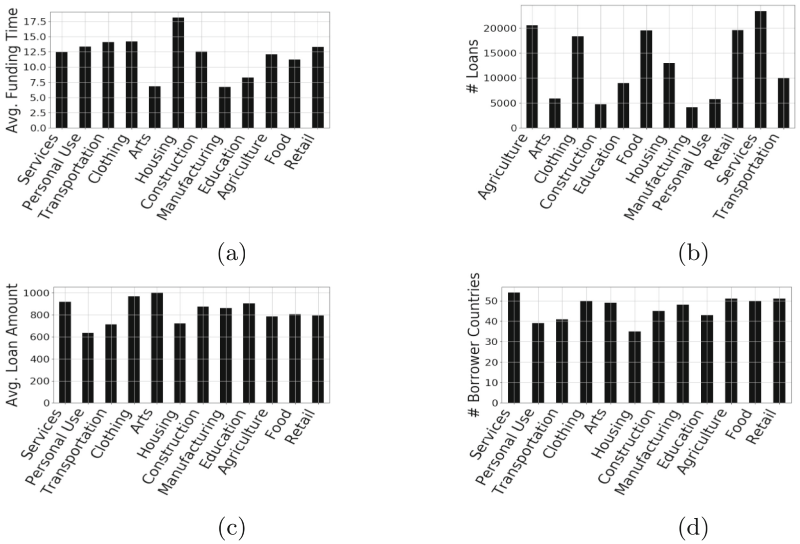
Broadly, there have been a few groups of research studies conducted on understanding and promoting microfinance lending on such platforms. (1) Investigating biases: previous studies have focused on understanding and predicting bilateral trade transactions based on migration and GDP differences between country pairs [23]. (2) Borrower and lender features: past studies include understanding various platform-external lender and borrower personal and regional characteristics that facilitate the transactions between countries [8] and the role of matching characteristics. However, the loan attribute concerning the loan sector is often overlooked especially to its connections to philanthropic and pro-social motivations of investors [16], (3) Fairness aware lending: recent studies have acknowledged the existence of bias in lending models and the need to diversify the distribution of donations to reduce the inequality of loans [12], and (4) Social networks: the role of networks have been studied from the perspective of facilitating bidding behavior in platforms [14].

What is often overlooked is the impact of external factors pertaining to the borrower countries that influence lender preferences and which cannot be directly observed from the platform data. Furthermore, there has been substantial evidence in the recent past that supports Lucas paradox, which indicates that, counter-intuitively the liberalization of international capital regimes using the internet platforms has not produced an open club, rather a rich club, a group of countries that exhibit the country-pair bias [1]. Since recommendation models typically do not consider such external data while building their models [21], such latent biases arising from external factors including lender perceptions of countries[2] can be quite detrimental for certain projects especially ones from specific countries.

To this end, we investigate the factors behind the funding speed of loans using the dataset available from Kiva. The goal is to see whether the lenders fall for region specific economic factors that they expect would help them avoid loan defaults from borrowers and whether that affects funding projects in certain sectors. We compare the effect of different sectors on project funding times when the economic external factors form part of the models in consideration. Using data from 143,856 loans over a period of 4 years and economic indicators from World Bank Data, we make the following contributions:

- We gather data from Kiva loans and heterogeneous data sources and build regression models to estimate the impact of such factors on the funding speed. We observe the role of the project or loan sector as a sensitive attribute in the models especially when its correlation with the funding speed differs for different sectors.

[2] https://bit.ly/2LF9Mpp.



(a) (b)

(c) (d)

Fig. 1. Distribution of (a) average funding times, (b) number of loans, (c) average loan amount by loan sectors and (d) number of borrower countries by each sector.

- We use recent causal inference and machine learning tools to estimate the effects the sector attribute on funding times. We specifically find that loans catering to Retail are funded 4 days slower relative to the other sectors on aggregate and loans for Arts are funded 6 days faster - all these when considering the economic factors of the location of the borrowers and the loan attributes. This is in contrast to observations from data that do not reveal such hidden discrepancies when excluding external factors.
- Following this, we incorporate fairness driven constraints to mitigate some of the biases arising from these loan specific attributes for particular sectors of loans. Our results suggest that even with such fairness constraints, the model performances are not too far-off from baselines, thus giving hope for future systems that take into account such constraints.

We note that this is the first work in attempting to understand the existing biases from loan attributes when external factors are also considered to be the contributors to such decisive disparities. Throughout our work, we mainly focus on linear regression models, however we adopt the models to use Bayesian variable selection techniques.

2 Data and Modeling Issues

Kiva is a non-profit micro-financial organization and its lending model is based on crowdfunding in which any individual can fund a particular loan by contributing to a loan individually or as a part of a lender team. The choice behind this

Table 1. Basic statistics for loans used in our study

# Loans	**# Lender Countries**	**# Borrower Countries**
143856	216	57
# Languages	**Avg. Loan Amount (USD)**	**Avg. Funding Time (std)**
7	836.18	12.58 (14.6)

platform is driven by the motivation to test a few hypotheses in this research - we want to be able to understand the presence or absence of behavioral and social bias that could create preferences for certain projects. Since public perceptions of societies can elicit biases towards countries with specific geographical, cultural or political fabric and that can affect funding in such online platforms, we set out to test the interplay of economic externalities and the loan specific attributes in such settings.

The publicly available Kiva dataset[3] contains various entities: (1) the data for the loans that contains various attributes associated with the borrowers, (2) the lenders' information containing various attributes regarding a lender's history of funding projects (3) the borrowers' information containing various attributes regarding a borrower's project and repayment history, (4) field partner which acts as the mediator and allocates loans from the lenders to the borrowers. Since our objective in this study is to understand the role of developmental factors when paired with the sector that receives the most funding, we use the following attributes that are associated with a loan in Kiva's platform from January 2010 to December 2014: (1) sector: categorical attribute denoting the sector of loan activity. The sectors considered in our study after removing sparse data is shown in Fig. 1. Note that the set of sector tags are fixed for all loans and are not randomly generated. (2) currency policy - binary attribute to reduce risk of currency fluctuation[4], (3) language - the language of the loan description - since 70% of the loans we considered were in English, we converted this to a binary attribute by considering all non-English languages as one category, (4) loan amount - numerical attribute denoting the amount of loan requested for the project, (5) borrower gender - binary attribute denoting the gender of the borrower, and (6) funding time - this is a derived numerical attribute calculated as the difference between the time of the loan request and the time when it was fully funded. We use this attribute for measuring the preference of the investors towards particular projects and our models are based on understanding what attributes account for lesser funding times. We plot the distribution of the funding times and the number of loans by sectors in our dataset in Fig. 1. However, unlike similar analyses, we do not found any substantial difference in the average loan amounts by sectors that get funded. Apart from entertainment, most of the sectors have similar funding requests that ultimately get funded.

[3] https://www.kiva.org/build/data-snapshots.

[4] https://pages.kiva.org/blog/new-kiva-feature-currency-risk-protection.

As a first task, we try to investigate the causal effects of borrower-lender differences arising from lender perceptions of the borrower countries as well as implicit economic and cultural variations. We try to measure the extent of impact it has on the funding times when considered alongside the sector of the loans. To this end, for each loan, we gather the following data from the world bank metrics dating back to 2010 [5]. We gather the following attributes: (1) ease of business: an ordinal attribute denoting the rank of the borrower country for ease of business, (2) loan access - numerical attribute denoting the ease of access to loan in the borrower country through formal financial institutions, (3) women ratio - numerical attribute measuring the ratio of women in labor force compared to men, (4) affordability - numerical attribute pertaining to the costs associated with using services, including both interest rates and fees, (5) VC financing - ordinal attribute that indicates how easy it is for the borrower to seek capital locally or otherwise in their country, (6) capacity innovation - ordinal attribute denoting the capacity of people in the borrower country to innovate and (7) internet penetration - numerical attribute denoting the percentage of people in the borrower country using the internet. To measure the cross-cultural similarities, we proceed as done in [23] to use the following features: (8) colonization - binary attribute denoting whether the borrower country was colonized by lender country, (9) distance - geographical distance between borrower and lender countries obtained from [17], (10) migrants - numerical attribute that measures the number of people or borrower country origin living in the lender country obtained from world bank data and (11) GDP difference - numerical attribute denoting the GDP Difference between borrower and lender countries obtained from world bank data. For the derived attributes which were calculated based on borrower and lender countries, we used the following method: for numerical attributes, for a specific loan we took the average of all the borrower-lender pairs for that loan. For categorical or binary attributes like colonization, we randomly picked one of the borrower-lender pairs for that loan and used that for the loan feature. This however introduces some approximation into the feature measurements. For all numerical attributes, we performed standardization for the regression models which would be described henceforth.

The dataset is publicly available for download[5]. After merging the data from these heterogeneous sources, we list the basic statistics of the loans used in this study (see Table 1). We find that while the average funding time is 12.5 days, the standard deviation is 14.6 days, which demands further investigation behind the variations.

3 Preliminary Analysis of Potential Disparities

We begin with a simple linear regression model to investigate the importance that these economic indicators capturing the borrower's nations, have on the funding time and how they play a role compared to the loan sector. When we regress the variables of the economic factors and the loan attributes barring the

[5] https://bit.ly/2TnqhL7.

Table 2. Table: OLS Regression estimates on funding time for a project loan. For model **M1**, we do not include the Sector attribute and for model **M2**, the attribute Sector (categorical) is used as the dummy variable.

OLS estimates	M1	M2 (Sector)	M3 (Services)	M4 (Agri.)	M5 (Retail)
Intercept	21.0346	20.2312	22.3927	20.4209	19.9626
Sector			−1.5268	−0.5921	2.3072
Currency policy [T.shared]	−1.0349	−1.1697	−3.2858	−0.894	−0.2923
Language	−2.5225	−2.0611	−1.3739	−2.1189	−2.3513
Ease of business	−0.0313	−0.027	−0.0309	−0.0282	−0.0281
Colonization	−1.6048	−2.0484	1.5085	−3.7422	−0.8401
Borrower gender [T.female]	−4.2708	−4.609	−4.4875	−4.1959	−4.7179
Loan amount	3.8751	4.1365	3.6231	5.1564	3.6773
Distance	−0.7032	−0.6359	−0.6739	−0.8225	−0.3331
Migrants	−2.1397	−2.4934	−1.9945	−2.3797	−2.4833
GDP difference	−0.3162	−0.2086	−0.0415	−0.2439	0.0622
Loan access	−1.2037	−0.526	1.619	−1.582	0.2476
Women ratio	−2.9224	−3.2598	−2.1065	−2.0469	−3.3524
Affordability	−2.2799	−2.2284	2.4288	−2.2366	−2.7703
VC finance	2.2952	2.0675	0.4367	2.3493	1.042
Capacity innov.	−0.0378	−0.0177	0.5504	0.5695	−0.0052
Internet pen.	−0.2639	0.4767	0.7634	−1.3048	0.9828

loan sector, on the funding time denoted by model **M1**, we find from Table 2 instantly that there are some borrower-lender attributes that have a larger role to play - the distance, and GDP difference have a negligible impact on the funding time whereas the feature measuring the migrants of borrower country in lender country has a negative correlation with funding times - it suggests that the cultural similarities arising from cross-border migration results towards faster funding for borrowers with such cultural advantages. Similarly, the women ratio factor has a significant negative correlation on the funding time along with the borrower gender in line with previous research [1]. This indicates that the perception about the role of women in such economies is a significant driver towards deciding whether the project would receive faster funding.

Next, we include the project sector attribute in the regression model as a dummy variable with the corresponding one-hot encodings denoted by model **M2**. We observe that while there are minor changes in the magnitude of the coefficients, the correlations of these variables do not change in the presence of the sector variable. This tempts us to conclude that when recommendation systems rely on these attributes to predict the best projects in terms of having better chances of funding, they can make a pretty fair classification for all projects based on these attributes. However, upon closer analyses, we look into the effect that these attributes have when considering each sector at a time. Taking each loan sector category s as the main category in contention, we build sector specific models. For a model on sector s, we consider all loans belonging to s as one category and all other categories as a unified dummy sector category separate from the sector s. We build regression models for all 12 sectors in a similar fashion. We show four of the models corresponding to four different sectors in Table 2 - in model **M3** we convert the multi-category sector attribute into a binary category by considering the services sector and loans belonging to that category as one cluster and all the other loans belonging to other sectors as the other cluster. When we compare M3 with model **M4** catering to the agriculture sector, we can observe that not only do some of the attributes differ in the magnitude significantly, but the influence from the attributes also reverses in some cases. Particularly, we find that the influence from the attribute colonization has opposite effects for the two sectors and similarly for attributes like affordability and loan access.

When comparing the role of sectors, we observe that the coefficient magnitude demonstrates that the relative number of days by which each sector gets funded faster or slower relative to the other sectors. However, the fact that these results are also heavily affected by the varying sparsity of the data. This leads us to turn our attention to recent literature on more robust causal reasoning tools that allow for explaining the effect of sectors on funding speed in the presence of such externalities [2,20].

4 Causal Inference

Note that the treatment of interest here is the loan sector assignment for the loan requested and we are interested in estimating the effects of loan sector relative to the economic, cultural and other loan characteristics, on the funding time. Following the work done in [20], we would use the Robin Causal Model (RCM) or the Potential Outcome Framework to estimate the treatment effects. We use the RCM model in this study due to the principled framework on which it applies - the treatment of unit i (loan in our case) only affects i and that the treatment is homogeneous across the units.

4.1 Treatment Effects Indicators

We describe in this section how we measure the causal impact metrics for each sector s. We estimate the treatment effects of sector on loan funding time considering separate models for each sector and treating the whole batch of data separately for each sector. Let the features be denoted by X, which in our case are all the attributes except the project sector s. Let Y be the outcome of interest, in our case the funding time of the loans. For each sector s, we consider W to be the binary treatment variable (whether a loan belongs to s or not). Following this, for each sector s, we represent the dataset in the form $(Y_i, X_i, W_i)_{i=1}^{n}$, where W_i denotes whether the sector for loan i is s or not ($W_1 = 1$ when loan belongs to s), n denoting the number of loans in the data. Note that W_i would be different for loan i when considering different sectors since the observational data gives us the actual loan sector. We will drop the subscripts from W when generalizing the inference settings for all loans. We will also refrain from attaching s as sub/super-scripts to notations since we perform all the following steps and estimate models in the same was irrespective of the sectors. We are interested in estimating the average treatment effects (ATE) of W on Y for each sector s and this is given by:

$$\tau = \mathbb{E}[Y(1) - Y(0)] \tag{1}$$

where $Y(1)$ is the potential outcome of a loan that belongs to s while $Y(0)$ is the one that does not belong to s. However, in the data, only one of them is observed for each loan when considering models for a specific sector. The three assumptions that are made during this estimation procedure are: (1) ($SUTVA$) - The apriori assumption that the value of Y_i when instance i is exposed to treatment W_i will be the same, no matter what mechanism is used to assign the treatment to i and no matter what treatments others receive, (2) the probability of outcome Y_i is independent of the features X_i given W_i - it means that the features X_i do not simultaneously affect W_i and Y_i. In our case this is more intuitive since firstly the external economic factors in itself have no bearing on the choice of the loan sectors and secondly, the loan sector also has little in relation to other loan features like gender, loan amount, and (3) both treatment and control groups have has at least one instance assigned to them (see [20] for more details on these assumptions).

4.2 Estimating Treatment Effects

With recent advances in machine learning to create estimators for ATE [3,9], we use the Doubly Robust Estimator (DRE) [11,22] to measure τ. We briefly lay out the steps for estimating τ using DRE for our data - note we follow these steps for all sectors individually:

1. **Outcome Model** - For loan sector s, we consider the loans i belonging to s as having $W_i = 1$ and all other loans as $W_i = 0$. Then we use the treated data $\{i : W_i = 1\}$ to estimate $\mu(1, x) = \mathbb{E}[Y(1)|X = x]$ with estimator $\hat{\mu}(1, x)$ and use control data $\{i : W_i = 0\}$ to estimate $\mu(0, x) = \mathbb{E}[Y(0)|X = x]$ with estimator $\hat{\mu}(0, x)$.
2. **Propensity Score Model**: We then estimate the propensity score model - use all loans data to estimate $e(x) = \mathbb{P}(W = 1|X = x)$ with estimator $\hat{e}(x)$.
3. The DRE $\hat{\tau}_{DRE}$ is given by $\hat{\tau}_{DRE} = \frac{1}{n}\sum_{i=1}^{n}\Big[W_i \times \frac{Y_i - \hat{\mu}(1,X_i)}{\hat{e}(X_i)} - (1 - W_i) \times \frac{Y_i - \hat{\mu}(0,X_i)}{1-\hat{e}(X_i)} - \hat{\mu}(1, X_i) - \hat{\mu}(0, X_i)\Big]$.
4. The standard error is then estimated following [15] by using an empirical sandwich estimator. For each instance/loan i, we have $IC_i = W_i \times \frac{Y_i - \hat{\mu}(1,X_i)}{\hat{e}(X_i)} - (1 - W_i) \times \frac{Y_i - \hat{\mu}(0,X_i)}{1-\hat{e}(X_i)} + \hat{\mu}(1, X_i) - \hat{\mu}(0, X_i) - \hat{\tau}_{DRE}$ and $\sigma^2 = \frac{1}{n}\sum_{i=1}^{n} IC_i^2$. The standard error is estimated as $\frac{\sigma}{\sqrt{n}}$.

The DRE has the double robustness property: given that either the outcome model or the propensity score model or both are correctly specified, the estimator is consistent.

4.3 Learning Outcome Models

In order to estimate $\hat{\mu}(1, x)$ and $\hat{\mu}(0, x)$ for each sector s, we use regression models, however we observe from Table 2 that not all variables are equally important when measuring their outcome on funding times and these differ substantially among the sectors. To this end, we adopt some variable selection techniques while building separate regression models for $\hat{\mu}(1, x)$ and $\hat{\mu}(0, x)$ for a sector s. We specifically adopt Bayesian methods where sparsity can be favored by assuming sparsity-enforcing priors on the model coefficients. These types of priors are characterized by density functions that are peaked at zero and also have a large probability mass in a wide range of non-zero values. Ideally, the posterior mean of truly zero coefficients should be shrunk towards zero and the posterior mean of non-zero coefficients should remain unaffected by the assumed prior. We use spike-and-slab priors which have some advantages when compared to other sparsity enforcing priors like Laplace and Student's t priors [19]. We briefly review the spike-and-slab model [10] as the regression model in choice and we learn separate models for $\hat{\mu}(1, x)$ and $\hat{\mu}(0, x)$ for a specific sector.

Let $\mathbf{y} \in \mathbb{R}^{n \times 1}$ be an n-dimensional row vector denoting the target variable and $\mathbf{X} \in \mathbb{R}^{n \times p}$ denote the design matrix, p denoting the number of attributes in our model except the sector attribute. Briefly, the spike-and-slab model specifies the prior hierarchy in the following way:

$$\begin{aligned} y_i &\sim N(\beta x_i, \sigma^2) \\ \beta_i &\sim (1 - \pi_i)\delta_0 + \pi_i N(0, \sigma^2 \tau^2) \\ \tau^2 &\sim \text{Inverse-Gamma}(\frac{1}{2}, \frac{s^2}{2}) \\ \pi_i &\sim \text{Bern}(\theta) \\ \theta &\sim \text{Beta}(a, b) \\ \sigma^2 &\sim \text{Inverse-Gamma}(\alpha_1, \alpha_2) \end{aligned} \tag{2}$$

where $i \in [1, p]$ indexes the features in the regression model, β denotes the coefficients in the regression model. The first equation defines a regression model where the response y_i follows a normal distribution conditioned on $\mathbf{x}_i$ and the parameters β. The second equation models the way in which sparsity is enforced on the model coefficients. The sparsity of β can be favored by assuming a spike-and-slab prior for the components of this vector - the slab $N(0, \sigma^2 \tau^2)$ is a zero mean broad Gaussian whose variance τ^2 is large and the scale σ^2 is multiplied so that the prior scales with outcome. The spike δ_0 is a Dirac Delta function (point probability mass) centered at 0 and this component is responsible for deciding whether the posterior for these coefficients would be zeroed out. $\pi \in [0, 1]$ is a mixture weight between the spike-and-slab components in the prior. The rest of the equations denote the hierarchical structure of the parameters σ^2, τ^2 and π. Note that τ^2 and θ are common to all predictors. We briefly describe how we sample the parameters for Gibbs sampling, with details added to Appendix[6].

[6] Appendix to the manuscript can be accessed here.

Sampling θ: The parameter is sampled from the conditional posterior using $\theta|\pi \sim Beta\Big(a + \sum_{i=1}^{p} \pi_i, b + \sum_{i=1}^{n}(1-\pi_i)\Big)$.

Sampling τ^2: The conditional posterior of τ^2 can be derived from the probability $p(\tau^2|\mathbf{y}, \beta, \pi, \theta, \sigma^2) = p(\tau^2|\pi, \beta)$. Here since π can assume values 0 or 1 we tackle each case independently and derive the following. We sample from the prior if all π_i's are zero. Let $\pi = \{\pi_1, \ldots, \pi_p\}$ be the vector of mixture weights and let $\mathbf{0}$ be a vector of zeros of length p. Following this, we have

$$\begin{gathered} p(\tau^2|\pi, \beta) = \frac{1}{Z} p(\beta|\tau^2, \pi) p(\pi) p(\tau^2) \\ = \frac{1}{Z} \prod_{i=1}^{p} \pi_i (2\pi\sigma^2\tau^2)^{-\frac{1}{2}} exp\Big(-\frac{1}{2\sigma^2\tau^2}\beta^T\beta\Big) \frac{\Big(\frac{s^2}{2}\Big)^{\frac{1}{2}}}{\Gamma\Big(\frac{1}{2}\Big)} (\tau^2)^{-\frac{1}{2}-1} exp\Big(-\frac{\frac{s^2}{2}}{\tau^2}\Big) \end{gathered} \tag{3}$$

which is a Gamma distribution and therefore we sample $\tau^2|\beta, \pi \sim$ Inverse-Gamma$(\frac{1}{2} + \frac{\sum_{i=1}^{p}\pi_i}{2}, \frac{s^2}{2} + \frac{\beta^T\beta}{2\sigma^2})$. On the other hand, when $\pi = 0$, the β_i's are 0 and we simply sample from the prior $\tau^2|\beta, \pi \sim$ Inverse-Gamma$(\frac{1}{2}, \frac{s^2}{2})$.

Sampling σ^2: The conditional posterior of τ^2 can be derived in a similar manner as above from the probability $p(\sigma^2|\mathbf{y}, \beta, \pi, \theta, \sigma^2)= p(\sigma^2|\mathbf{y}, \beta)$. Proceeding as before, we can derive the sampling as follows: $\sigma^2|\mathbf{y}, \beta \sim$ Gamma$\Big(\alpha_1 + \frac{n}{2}, \alpha_2 + \frac{(\mathbf{y}-\mathbf{X}\beta)^T(\mathbf{y}-\mathbf{X}\beta)}{2}\Big)$.

Sampling β: Proceeding as before, when all π_i's are zero, the corresponding β_i's are all sampled from the Dirac Delta function δ_o resulting in all zeros. We can now sample all β_i's as follows

$$\beta_i|\mathbf{y}, \pi_i, \sigma^2, \tau^2 \sim \begin{cases} \delta_0, & \pi_i = 0 \\ N\Big(\Big(\mathbf{X}^T\mathbf{X}\frac{1}{\sigma^2} + \mathbf{I}\frac{1}{\sigma^2\tau^2}\Big)^{-1} \mathbf{X}^T\mathbf{y}\frac{1}{\sigma^2}, & \\ \Big(\mathbf{X}^T\mathbf{X}\frac{1}{\sigma^2} + \mathbf{I}\frac{1}{\sigma^2\tau^2}\Big)^{-1}\Big), & \pi_i = 1 \end{cases} \tag{4}$$

Sampling $\boldsymbol{\pi}$

The individual π_j's are conditionally independent given θ. We compare two cases: one when the j^{th} element of β is zero or π_j is zero and the other when $\pi_j = 1$. We denote by π_{-j} the state of the variables barring j. Let $\pi_j = 1|\mathbf{y}, \beta_{-j}, \pi_{-j}, \sigma^2, \tau^2, \theta \sim$ Bern(ζ_j). Let $a = p(\pi_j = 1|\mathbf{y}, \beta_{-j}, \pi_{-j}, \sigma^2, \tau^2, \theta)$ and $b = \pi_j = 1|\mathbf{y}, \beta_{-j}, \pi_{-j}, \sigma^2, \tau^2, \theta$. Then $\zeta_j = \frac{a}{a+b}$. We then draw π_j from a Bernoulli with a chance parameter ζ_j and we repeat this for all predictors β_j. For the case when $\pi_j = 0$,

$$\begin{gathered} p(\pi_j = 0|\mathbf{y}, \beta_{-j}, \pi_{-j}, \sigma^2, \tau^2, \theta) \\ = \frac{1}{Z} exp\Big(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}_{-j}\beta_{-j})^T(\mathbf{y} - \mathbf{X}_{-j}\beta_{-j})\Big)(1-\theta) \end{gathered} \tag{5}$$

Table 3. Result comparison on the test set. For p-score estimation, we use F1 score and accuracy (the higher, the better); for outcome estimations, we use RMSE (the lower the better).

Sector name	Treatment (RMSE)		Control (RMSE)		p - score	
	SSR	LR	SSR	LR	F1	Acc. %
Manufacturing	12.34	12.39	5.44	5.38	0.68	64.84
Transportation	12.68	12.83	13.98	14.17	0.69	61.76
Clothing	10.01	10.01	4.89	5.02	0.64	62.52
Personal use	9.84	10	10.71	10.84	0.65	64.32
Housing	12.1	12.23	11.79	11.93	0.68	60
Food	11.42	11.61	10.17	10.36	0.69	66.96
Arts	11.11	11.21	12.03	12.11	0.7	59.69
Retail	11.35	11.69	11.32	11.45	0.74	71.82
Construction	10.15	10.21	10.15	10.21	0.7	69.33
Agriculture	10.66	10.75	11.39	11.55	0.71	62.89
Services	12.72	12.94	12.83	13	0.74	68.35
Education	12.57	12.67	6.18	6.36	0.66	61.05

where we have absorbed all the irrelevant terms into Z, the normalizing constant. The expression for $\pi_j = 1$ can be written similarly except that it would require integration over β_j. Defining $\mathbf{z} = \mathbf{y} - \mathbf{X}_{-j}\beta_{-j}$, we have

$$p(\pi_j = 1|\mathbf{y}, \beta_{-j}, \pi_{-j}, \sigma^2, \tau^2, \theta)$$
$$= \frac{1}{Z}\theta(2\pi\sigma^2\tau^2)^{-\frac{1}{2}} exp\Big(-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{X}_{-j}\beta_{-j})^T(\mathbf{y} - \mathbf{X}_{-j}\beta_{-j})\Big) exp\Big(\frac{(\sum_{i=1}^n x_i z_i)^2}{2\sigma^2(\sum_{i=1}^n x_i^2 + \frac{1}{\tau^2})}\Big) \tag{6}$$

The conditional posterior of $\pi = 0$ is therefore a Bernoulli distribution with chance parameter $1 - \zeta_j = \frac{1-\theta}{(\sigma^2\tau^2)^{-\frac{1}{2}} exp(K)\Big(\frac{\sigma^2}{(\sum_{i=1}^n x_i^2 + \frac{1}{\tau^2})}\Big)^{\frac{1}{2}}\theta + (1-\theta)}$, where $K = \frac{(\sum_{i=1}^n x_i z_i)^2}{2\sigma^2(\sum_{i=1}^n x_i^2 + \frac{1}{tau^2})}$ and where z_j changes depending on which β_j we sample.

5 Experiments and Results

In this section, we first start by evaluating the effectiveness of the learning methods in modeling individual estimators that form the components of $\hat{\tau}_{DRE}$. The outcome models through spike-and-slab Bayesian variable selection models have been described in the previous sections. For estimating the propensity score $e(x) = \mathbb{P}(W = 1|X = x)$ with estimator $\hat{e}(x)$ in step 2 outlined in Sect. 4.2, we use a logistic regression model with the same attributes as the outcome model. We further experimented with Random Forests, but did not observe any substantial

Table 4. Summary of ATE Estimation for different sectors comparing models. Numbers marked in asterisk indicate substantial differences in the estimates from the regression coefficients estimated in Table 2.

Sector name	Naive		Baseline		DRE (SSR)	
	ATE	std	ATE	std	ATE	std
Construction	1.04	0.28	−0.09	0.27	0.51	0.29
Clothing	1.85	0.16	3.12	0.15	2.63	0.29
Retail	0.59	0.15	3.25	0.15	4.81*	0.18
Education	−4.86	0.2	−5.48	0.19	−5.37	0.19
Services	−0.52	0.15	−0.92	0.14	−0.85	0.15
Manufacturing	−5.16	0.25	−5.53	0.23	−5.38	0.24
Transportation	1.34	0.22	1.6	0.2	1.05	0.22
Agriculture	−0.61	0.16	−0.28	0.15	−0.6	0.15
Housing	6.34	0.19	6.77	0.18	7.9*	0.22
Arts	−5.46	0.23	−5.4	0.22	−5.61	0.26
Personal use	1.61	0.27	1.35	0.25	−2.26*	0.71
Food	−1.43	0.15	0.34	0.14	−0.4	0.17

difference in the results. For the Gibbs sampling procedure, we set the following hyper-parameter values: $a = b = 1$, $a_1 = a_2 = 0.01$, $\theta = 0.5$ and $s = 1/2$ for all the models. We use a burn-in of 1000 samples for the procedure and use 4000 samples for the sampling procedure. We use these posterior estimates as the coefficient estimates in the spike-and-slab regression model for predictive purposes.

As mentioned before, for each sector, we consider treated and control groups considering that sector and evaluate the outcome models for treatment and control and the propensity score (p-score) models. We use Root Mean Squared Error (RMSE) for the outcome regression models and F1 sore and Accuracy for the p-score model using logistic regression. For each model we split the data into 70%–30% train-test and evaluate the models using these metrics on the held-out test set. The results shown for all sectors in Table 3 compares Linear Regression (LR) without any regularization with the Spike and Slab (SSR) model. We find that while for most sectors the models fare comparably for both treated and control groups, for 3 sectors namely Manufacturing, Clothing and Education where the regression models for Treatment are an order of magnitude worse than control groups evidenced by their RMSE scores. This can be attributed to the relatively low number of projects in these areas shown in Fig. 1. We also find that the SSR model outperforms the LR model in most cases in terms of lower RMSE scores for the SSR model. For the p-score model, we find that the logistic regression model performs similar for most sectors showing lesser disparity among the several models used for the purpose.

Next, we compare the ATE for different sectors against a model where the ATE is estimated with just the target variable - the funding time. We compare 3 models for measuring the Average Treatment Effect (ATE):

1. **Naive** - ATE is calculated using the differences in means of Y for treatment and control groups, and the standard deviation is calculated using the group standard deviations.
2. **Baseline** - Here we use the Linear Regression (LR) model as discussed above to estimate 2 relations: (1) $Y(1) = \mathbf{X}\beta_1$ with estimate $\hat{\beta}_1$ using the treated data and $Y(0) = \mathbf{X}\beta_0$ with estimate $\hat{\beta}_0$ using the control data. The estimator $\hat{\tau} = \frac{1}{n}\sum_{i=1}^{n}(\hat{Y}_{1,i} - \hat{Y}_{0,i})$. The standard error is then calculated as $\sqrt{\frac{var(Y_i - \hat{Y}_{1,i}|i:W_i=1)}{n_t - 1} + \frac{var(Y_i - \hat{Y}_{0,i}|i:W_i=0)}{n_t - 1}}$.
3. **DRE (SSR)** - Here we use the SSR models for the estimators $\hat{Y}_1$ and $\hat{Y}_0$ from the treated and control data and $\hat{\tau}_{DRE}$ and teh standard errro is calculated as described in Sect. 4.2.

The results for the model is shown in Table 4. From the table, we find that the four sectors where the ATE from the DRE estimator is substantially different from the naive estimator are Retail, Housing, Arts and Personal Use (we keep the 3 sectors, Manufacturing, clothing and education out of our discussion since the SSR models for the treated data in these 3 sectors were substantially worse than control data). In fact, we find that the funding time for Arts loans have almost 6 days (ATE $= -5.61$) faster funding when compared to all other sectors using our DRE (SSR) model, whereas the naive estimator suggests a slower funding. This suggests that when we combine these economic factors along with the loan attributes for these specific sectors, the effect of this loan sector actually helps in faster funding which in other situations would have been difficult to be funded. Similarly, for the Retail loans, we find that funding is generally disfavored compared to other factors by being funded slower by 5 (ATE $= 4.81$) days. The standard errors for all the 3 models are comparable and so as such the ATE estimates can be compared reliably across the models. These observations suggest that when such economic disparities or similarities exist which can affect lender trust and perceptions of funding a project in a particular sector, biases are bound to arise. Therefore, predictive models which try to model the risk of loan defaults must also incorporate fairness constraints to not allow favoritism towards certain sectors. To this end, we conclude this study by modifying our SSR model to incorporate fairness constraints.

6 Controlling the Disparities from Sectors

To control the disparities arising from the different attributes for different sectors in our regression setting, we adopt the procedure described in [7] and incorporate the constraint in the sampling procedure for the parameter estimates. For each sector s, we divide the dataset as done before into two groups: $D_s^{\uparrow}$ and $D_s^{\downarrow}$ based on s. The specific goal here is to build one regression model for each sector and learn the parameters of that model while minimizing bias associated with predicting the target variables when conditioned over the loan sector attribute. To this end, we use the constraint that ensures that the mean predictions for the two groups $D_s^{\uparrow}$ and $D_s^{\downarrow}$ are equal irrespective of what the target or outcome exhibits.

Adding Regularization: We use the same model based on Bayesian variable selection introduced in Sect. 4.3 with the addition of new regularization terms. We add the sector attribute to the features $\mathbf{X}$, however we now build one single model for each sector with the entire batch of data. We use the balanced means constraint based on the following criteria: $\frac{\sum_{(\mathbf{x}_i,t_i)\in D_s^{\uparrow}} \beta.\mathbf{x}_i}{|D_s^{\uparrow}|} = \frac{\sum_{(\mathbf{x}_i,t_i)\in D_s^{\downarrow}} \beta.\mathbf{x}_i}{|D_s^{\downarrow}|}$, where $D_S^{\uparrow}$ and $D_S^{\downarrow}$ denote control and treatment data. It denotes the constraint that the predictions from our model should be the same for both the treated and the control groups for the loan sector in consideration irrespective of what the target variable differences in the model exhibit. Using the same notations used in Eqs. 2, we make the following adjustment to sample the target variable. Denoting $\frac{\sum_{(\mathbf{x}_i,t_i)\in D_s^{\uparrow}} \beta.\mathbf{x}_i}{|D_s^{\uparrow}|} = \frac{\sum_{(\mathbf{x}_i,t_i)\in D_s^{\downarrow}} \beta.\mathbf{x}_i}{|D_s^{\downarrow}|}$ as $\mathbf{d}$, we add the regularization term as: $y_i \sim N(\beta x_i, \sigma^2) + \lambda \beta \mathbf{d}$, where λ is the hyper-parameter controlling the effect of the regularization term. With this modification, the sampling equations are modified following from Sect. 4.3 and have been added to the Appendix.

Results: Finally, we compare the results of the models with the regularization constraint for the sectors with models discussed prior to this. Additionally, we also compare the results from the model in the absence of external factors and only considering loan attributes available from Kiva data. We adopt a similar validation approach as previous where we perform a 70%–30% train-test split and test on the held-out 30% data. For training the SSR models, we use the same settings as explained in Sect. 5 for the Gibbs sampling procedure. For evaluating the regression models, we use the metric RMSE on the test data as done in the previous section. The regularization hyper-parameter λ, we set it to 0.6 after cross-validating it with several values. The results have been shown in Table 5 - the column LR-LA shows the results for the model with only loan attributes from Kiva. The last column shows results incorporating the regularization term. Additionally, we only test the models with the 4 sectors that showed the highest ATE explained in Sect. 5. We observe that in all these sectors, addition of external factors like the economic attributes and borrower-lender country pair attributes improve the model over the model LR-LA. The model with SSR performs the best in the absence of any regularization for all the sectors having the least RMSE, indicating that variable selection helps improve the predictions. However, when we compare these results with the model SSR (with regularization), we find that the performance drops at the cost of the equality constraints, however what we observe is that the results are still comparable to the simple LR model. We find that for Housing loans, the model with regularization performs comparably worse and this can be attributed to the pre-existing disparities shown by high ATE for these loans as shown in Table 4. Therefore, the equal

Table 5. RMSE results of regression models. Models with LA denote only loan attributes from Kiva are used in the model. The lower values indicate better results.

Sector	LR	LR - LA	SSR	SSR (regularization)
Housing	10.76	11.34	10.61	13.82
Personal use	9.6	10.02	9.46	10.24
Retail	12.06	13.18	11.91	12.73
Arts	9.31	10.25	9.19	9.52

means constraint does result in performance degradation. However, these results suggest that we can still build models by reducing disparities in the resulting predictions while limiting the drop in performance.

7 Related Work and Conclusions

Understanding the effect of loan attributes towards funding speeds have been studied extensively in [16] albeit only with factors from the loans data. The effects of cultural differences have also been studied in [6] where the authors present evidence that lenders prefer culturally similar borrowers in Kiva. However, the extent to which that affects the actual interests towards particular sectors was not presented. Our work here opens an entire body of research into fairness aware recommendation systems [13] that might be necessary when promoting projects so as to lessen the inherent biases arising from existing lenders. Especially when designing portfolio recommendations as a tool for decision support for lenders as done in [24], it is important to adjust the multi-objective optimization problems incorporating constraints as described in this paper. Such conclusions can also be extended to platforms which are designed for lenders to profit from investments such as Lendingclub [18]. In this paper, we first demonstrated how simple economic factors can play a role in deciding the speed of funding for particular loans and how they can be intertwined with the loan sector. We then measured the existing disparities arising from such factors using causal reasoning estimators and proposed a method to control the differences in outcome. One area where our work can be extended is to develop a single model taking all models into account - this is where the Bayesian variable selection method can be extended to incorporate priors that take into account fairness constraints for all sectors and using empirical bayes to drive the priors.

References

1. Alfaro, L., Kalemli-Ozcan, S., Volosovych, V.: Why doesn't capital flow from rich to poor countries? An empirical investigation. Rev. Econ. Stat. **90**(2), 347–368 (2008)
2. Athey, S., Imbens, G., Pham, T., Wager, S.: Estimating average treatment effects: supplementary analyses and remaining challenges. Am. Econ. Rev. **107**(5), 278–81 (2017)

3. Athey, S., Imbens, G.W., Wager, S., et al.: Efficient inference of average treatment effects in high dimensions via approximate residual balancing. Technical report (2016)
4. Banerjee, A., Duflo, E., Glennerster, R., Kinnan, C.: The miracle of microfinance? Evidence from a randomized evaluation. Am. Econ. J.: Appl. Econ. **7**(1), 22–53 (2015)
5. World bank data. World Bank (2013)
6. Burtch, G., Ghose, A., Wattal, S.: Cultural differences and geography as determinants of online prosocial lending. MIS Q. **38**(3), 773–794 (2014)
7. Calders, T., Karim, A., Kamiran, F., Ali, W., Zhang, X.: Controlling attribute effect in linear regression. In: 2013 IEEE 13th International Conference on Data Mining, pp. 71–80. IEEE (2013)
8. Choo, J., Lee, C., Lee, D., Zha, H., Park, H.: Understanding and promoting microfinance activities in kiva.org. In: Proceedings of the 7th ACM International Conference on Web Search and Data Mining, pp. 583–592 (2014)
9. Hill, J.L.: Bayesian nonparametric modeling for causal inference. J. Comput. Graph. Stat. **20**(1), 217–240 (2011)
10. Ishwaran, H., Rao, J.S., et al.: Spike and slab variable selection: frequentist and Bayesian strategies. Ann. Stat. **33**(2), 730–773 (2005)
11. Kang, J.D., Schafer, J.L., et al.: Demystifying double robustness: a comparison of alternative strategies for estimating a population mean from incomplete data. Stat. Sci. **22**(4), 523–539 (2007)
12. Lee, E.L., et al.: Fairness-aware loan recommendation for microfinance services. In: Proceedings of the 2014 International Conference on Social Computing, pp. 1–4 (2014)
13. Li, Y., Ning, Y., Liu, R., Wu, Y., Hui Wang, W.: Fairness of classification using users' social relationships in online peer-to-peer lending. In: Companion Proceedings of the Web Conference 2020, pp. 733–742 (2020)
14. Lin, M., Prabhala, N.R., Viswanathan, S.: Judging borrowers by the company they keep: friendship networks and information asymmetry in online peer-to-peer lending. Manag. Sci. **59**(1), 17–35 (2013)
15. Lunceford, J.K., Davidian, M.: Stratification and weighting via the propensity score in estimation of causal treatment effects: a comparative study. Stat. Med. **23**(19), 2937–2960 (2004)
16. Ly, P., Mason, G.: Individual preferences over NGO projects: evidence from microlending on kiva. Available at SSRN 1652269 (2010)
17. Mayer, T., Zignago, S.: Notes on CEPII's distances measures: the GeoDist database (2011)
18. Nowak, A., Ross, A., Yencha, C.: Small business borrowing and peer-to-peer lending: evidence from lending club. Contemp. Econ. Policy **36**(2), 318–336 (2018)
19. O'Hara, R.B., Sillanpää, M.J., et al.: A review of Bayesian variable selection methods: what, how and which. Bayesian Anal. **4**(1), 85–117 (2009)
20. Pham, T.T., Shen, Y.: A deep causal inference approach to measuring the effects of forming group loans in online non-profit microfinance platform. arXiv preprint arXiv:1706.02795 (2017)
21. Rakesh, V., Lee, W.C., Reddy, C.K.: Probabilistic group recommendation model for crowdfunding domains. In: Proceedings of the Ninth ACM International Conference on Web Search and Data Mining, pp. 257–266 (2016)
22. Robins, J.M.: Robust estimation in sequentially ignorable missing data and causal inference models. In: Proceedings of the American Statistical Association, Indianapolis, IN, vol. 1999, pp. 6–10 (2000)

23. Singh, P., et al.: Peer-to-peer lending and bias in crowd decision-making. PLoS ONE **13**(3), e0193007 (2018)
24. Zhao, H., Liu, Q., Wang, G., Ge, Y., Chen, E.: Portfolio selections in P2P lending: a multi-objective perspective. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 2075–2084 (2016)