CREDIT SCORES THAT PRIORITIZE CUSTOMER WELFARE:THEORY AND EVIDENCE FROM NIGERIA

Anonymous authors
Paper under double-blind review

ABSTRACT

At the core of many consumer lending decisions is a credit score: an algorithmic assessment of a customer's creditworthiness. Traditional credit scores are designed to maximize lender profits, and use machine learning algorithms to predict which customers will repay loans. This paper proposes and tests a different paradigm for consumer lending, in which 'welfare-sensitive' credit scores allow the lender to balance expected profits against the expected welfare impacts of specific loans. Using data from a randomized control trial in Nigeria, we show how machine learning algorithms can be trained to predict the welfare impact of lending to a client, and how those welfare scores can be combined with traditional credit scores to characterize a Pareto-efficient tradeoff between welfare and profits. Our main result suggests that, in the Nigerian context, the lender could achieve an 11% gain in consumer welfare by sacrificing 0.1% of profits.

1 INTRODUCTION

Financial firms are increasingly reliant on data-driven algorithmic tools to make critical decisions, such as deciding which customers are eligible for a loan. At the core of many lending decisions is a credit score: an algorithmic assessment of a customer's creditworthiness. Credit scores are constructed using algorithms which optimize a single objective – typically, the likelihood that a customer will repay the loan, as a proxy for the expected profits of the lender. On the one hand, credit scoring algorithms have enabled poor and historically excluded individuals in the developing world to gain access to loans. However, lending decisions that aim to maximize firm profit might not improve the welfare of the customer. In fact, profit maximizing loans can have detrimental effects (see Skiba & Tobacman (2019), for example).

This paper proposes, and tests a method for constructing a "welfare sensitive" credit score, which balances the expected *welfare impact* of a loan alongside the expected profit of that loan, attempting to mitigate the potential adverse effects of single objective maximization. Following a framework outlined in Rolf et al. (2020), our approach explicitly balances multiple objectives and outcomes, and allows us to explore the trade-offs between consumer welfare and firm profitability. Our method uses the notion of Pareto optimality, which offers an intuitive way to characterize optimal decision rules when balancing multiple objectives. A central challenge to this approach lies in predicting the expected individual impact of a loan, *ex-ante*, or at the time of the lending decision. We explore the use of generalized random forests Athey et al. (2019) to predict the impact of a loan on each customer using data already available to the lender, at the time of the lending decision. The rest of the paper describes our approach to constructing these welfare sensitive credit scores.

2 RELATED WORK

This paper contributes to a growing literature on welfare aware machine learning, building on the contributions of Rolf et al. (2020). Much of this literature evaluates whether algorithms satisfy various "fairness" criteria (cf. Barocas et al., 2019; Dwork et al., 2012). Such approaches have many limitations (Ensign et al., 2018; Mouzannar et al., 2019; Liu et al., 2018), including an inherent inability of algorithmic decision making tools to satisfy multiple definitions of fairness simultaneously (Kleinberg et al., 2017), and tend to focus on between group-differences as opposed to within group

differences (Kasy & Abebe, 2021). In many contexts, the inclusion of these additional criteria can often result in sub-optimal outcomes (Noriega-Campero et al., 2019).

Our approach differs from these papers in that we explicitly attempt to balance different objectives, and is thus relevant to a vast literature on multi-objective optimization (Jin & Sendhoff, 2008; Deb & Kalyanmoy, 2001). It is also related to recent work by Aswani & Olfat (2019) who develop a hierarchical optimization approach to tackle similar questions.

Our analysis relies on constructing welfare scores, or estimates of the predicted individual treatment effect. While we focus on generalized random forests Athey et al. (2019), our method can be extended to any algorithm that allows for the prediction of individual treatment effects,including the x-learner approach proposed by Künzel et al. (2019)

Finally, even though our paper examines the trade-offs between profit and welfare, it is closely related to recent work that aims to identify the most important populations for the welfare maximizing allocations of scarce resources. Haushofer et al. (2022) examine the trade-offs involved between targeting the poorest individuals versus those who are likely to realize the greatest impacts, using data from a cash transfer program in Kenya. Björkegren et al. (2022b) focus on the "dual" of this problem – they develop methods to uncover a social planner's preferences which are consistent with observed policies in the context of an anti-poverty program in Mexico.

3 SETTING AND DATA

We focus on an experiment in Nigeria (previously described in Björkegren et al. (2022a)) where digital loans were made available by a financial service provider (FSP) to a random subset of new loan customers, who would normally have been denied access. Such loan products are very popular in Nigeria, and similar to those commonly offered across Sub-Saharan Africa. The experiment focused on a sample of 1618 customers, and found that customers who would have ordinarily been denied access to loans see large increases in subjective wellbeing outcomes. We obtain the survey data and administrative data from this experiment. We briefly these describe data below.

3.1 DATA

Welfare In this analysis, we focus on a standardized index of subjective wellbeing as our primary measure of welfare, which aggregates information on depression, and life satisfaction. The exact construction of the variable is described in Appendix A.2.

Profit While we do not have any information on fixed costs incurred by the FSP (per loan), we have detailed information on the principal loan amounts, and repayment history. We thus define profit (per customer) as total amount repaid minus the total amount borrowed.

4 METHODS

4.1 NOTATION AND SETUP

We are interested in a particular case of the general class of problems considered in Rolf et al. (2020). We consider a setting in which an FSP has two simultaneous objectives: maximize some measure of profit and maximize some measure of customer welfare. The FSP makes decisions about customers, who are specified by a credit score $x_i \in \mathbb{R}^1$. Loan policies are functions that output a randomized decision $\pi(x_i) \in [0, 1]$, corresponding to the probability that an individual with credit score x_i is approved for a loan. To each customer we associate a value p_i representing the expected profit to be garnered from loan approval and a value w_i encoding the change in welfare. The profit and welfare objectives are expectations over the joint distribution $\mathcal{D}$ of (w_i, p_i, x_i):

$$\mathcal{U}_W(\pi) = \mathbb{E}[w_i \cdot \pi(x_i)] \quad \text{and} \quad \mathcal{U}_P(\pi) = \mathbb{E}[p_i \cdot \pi(x_i)]. \tag{1}$$

Given these two objectives it is not always possible to define a unique optimal loan policy π. We are instead interested in *Pareto-optimal* loan policies, in the sense that they are not strictly dominated by any other alternative policy.

Rolf et al. (2020) provide results to characterize these policies assuming perfect and imperfect knowledge. In an idealized setting where the welfare and profit contributions w_i and p_i can be determined from the credit score x_i via exact score functions $f_W(x_i) = w_i$ and $f_P(x_i) = p_i$, Pareto-optimal loan policies $\pi^*_\alpha(x)$ are given by

$$\pi^*_\alpha(x) = \mathbb{I}((1-\alpha)f_P(x_i) + \alpha f_W(x_i) \geq 0), \quad \alpha \in [0,1] \tag{2}$$

Even though these are all Pareto-optimal loan policies, the policies π^*_α induce different trade-offs between the objectives. The parameter α determines this trade-off, tracing the *Pareto frontier*:

$$\mathcal{P}_{exact} = \{(\mathcal{U}_P(\pi^*_\alpha), \mathcal{U}_W(\pi^*_\alpha)) : \alpha \in [0,1]\} \tag{3}$$

However, in real life contexts, it is highly unlikely that the FSP knows the score functions f_P and f_W. Instead, the FSP will usually resort to estimating score functions $\hat{f}_P$ and $\hat{f}_W$ from data in the hope that these models can provide good predictions on future examples. A natural question arising at this point is whether the *plug-in* version of π^*_α, that we denote by π^{plug}_α and define below, is optimal over the class of policies that act on predicted scores.

$$\pi^{plug}_\alpha(x_i) = \mathbb{I}((1-\alpha)\hat{f}_P(x_i) + \alpha\hat{f}_W(x_i) \geq 0), \quad \alpha \in [0,1] \tag{4}$$

Rolf et al. (2020) show that π^{plug}_α are optimal as long as the predicted score functions are well calibrated, and that even when plug-in policies are not optimal, the sub-optimality of the resulting classifier in terms of the utility function $\mathcal{U}_\alpha = (1-\alpha)\mathcal{U}_p(\pi) + \alpha\mathcal{W}(\pi)$ is bounded by the α-weighted sum of l_1 errors in the profit and welfare scores.

4.2 Generalized random forests

Generalized random forests is a machine learning method for nonparametrical statistical estimation proposed by Athey et al. (2019). We rely on this method to estimate a profit and a welfare scores $\hat{f_P}(x_i)$ and $\hat{f_W}(x_i)$. We rely on generalized random forests for three reasons. First, generalized random forests can be used to fit any quantity of interest identified as the solution to a set of local moment equations. Specifically, given data $(x_i, d_i) \in \mathbb{R} \times \{0,1\}$ where d_i represents loan approval, we seek forest-based estimates of $f(x)$, defined by a local estimating equation of the form

$$\mathbb{E}[\psi_{f(x),c(x)}(d_i)|x_i = x] = 0 \text{ for all } x \in \mathbb{R} \tag{5}$$

where $\psi(\cdot)$ is some scoring function and $c(x)$ is an optional nuisance parameter. In the following subsection we describe the moment conditions involved in the estimation of each score. Second, the generalized random forest estimator is consistent, asymptotically well-behaved and allows for construction of valid confidence intervals. Finally, generalized random forests are computationally efficient and a software implementation is publicly available.

4.3 Moment conditions

Let p_i customer profit measured following the description in section 2.3. In maximizing the profit objective the FSP seeks a good estimate of the conditional mean function (CMF) of p_i conditional on credit score x_i. The local estimation condition we employ is

$$\mathbb{E}[p_i - \hat{f_P}(x)|x_i = x] = 0 \text{ for all } x \in \mathbb{R}. \tag{6}$$

Let y_i customer welfare as measured following the description in section 2.3. In maximizing the welfare objective the FSP seeks a good estimate of the conditional average treatment effect (CATE) of loans on customer welfare d_i conditional on credit score x_i. The local estimation condition we employ is

$$\mathbb{E}[y_i - \hat{f_W}(x) * d_i - c(x)|x_i = x] = 0 \text{ for all } x \in \mathbb{R}. \tag{7}$$

4.4 Estimation

We use the full sample of 1,618 customers to train the generalized random forests using the R language package grf (Athey et al., 2019). We report out-of-bag predictions in the following section. Both forests are grown to 1,000 trees using honest splitting and otherwise default parameter values.

5 RESULTS AND DISCUSSION

Figure 1 shows the estimated profit and welfare score functions. Dark dots are the point estimates, while the faded dots are the 95% confidence intervals. In the upper figure, we observe that the predicted profit score is increasing in credit score. Customers in the right end of the distribution tend to have significant and positive scores, while customers in the left end tend to have negative and significant scores. The estimates of the conditional average treatment effects on welfare in the lower figure are also reassuring: the significant and positive welfare scores are concentrated in the left tail of the distribution. This is largely aligned with the results in Björkegren et al. (2022a), who report finding an estimate of the average treatment effect for customers below the eligibility threshold of similar sign, magnitude and standard error.

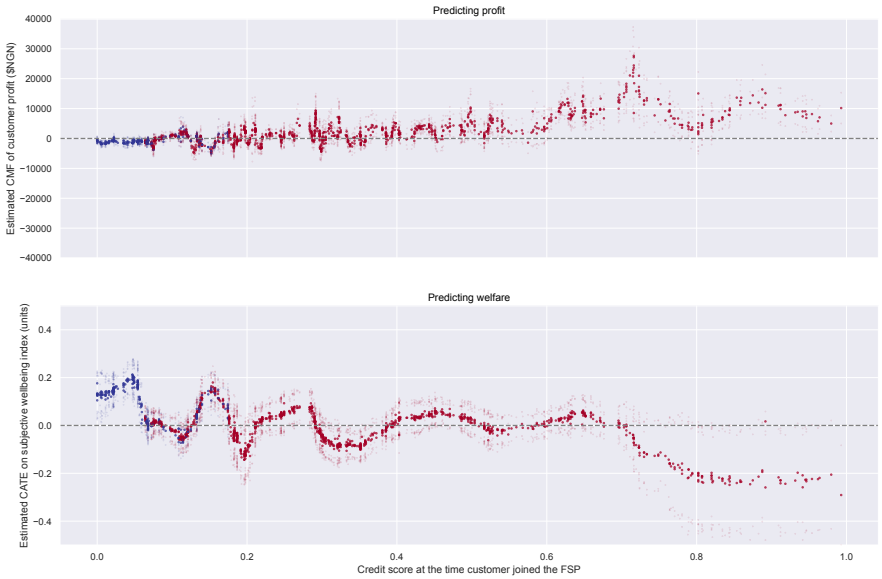

Figure 1: Generalized Random Forest Predictions

Using the estimated profit and welfare score functions, we estimate the empirical Pareto frontier using the optimal policy described in equation 4, in Figure 2 (bottom-left figure). Notably, at the profit maximizing end of the frontier, average welfare is negative – suggesting that the prevailing regime, which purely aims to maximize profit does not generate any welfare gains, on average.

However, large improvements in welfare can be achieved relatively easily. The bottom-right sub-figure describes a scenario where the trade-off rate α is 0.1 (see equation 2 for the definition of the trade-off parameter). In this scenario, customers to right of the green region are approved, while those in the green region are rejected. The policy expands access to a number of customers in the 2nd quadrant who might benefit immensely from gaining access to this loan, even though lending to these customers is unlikely to be profitable. Even though this policy correctly rejects customers in 3rd quadrant who are both unlikely to be profitable, and unlikely to realize gains in welfare, it continues to afford access to customers in the 4th quadrant who might be profitable but might not realize any gains in welfare. This suggests that even the most profit-minded lenders can substantially improve customer welfare by sacrificing relative small amounts of profit, and represents a possible regime for the FSP in Nigeria. In this scenario, a 0.1% loss in profits generates a 11% gain in average consumer welfare, while a 1% loss in profits could generate roughly 61% gain in average consumer welfare.

The top-right sub-figure describes a policy that weights both profit and welfare equally, meaning $\alpha = 0.5$. In this scenario, we observe that a large number of customers who would normally have

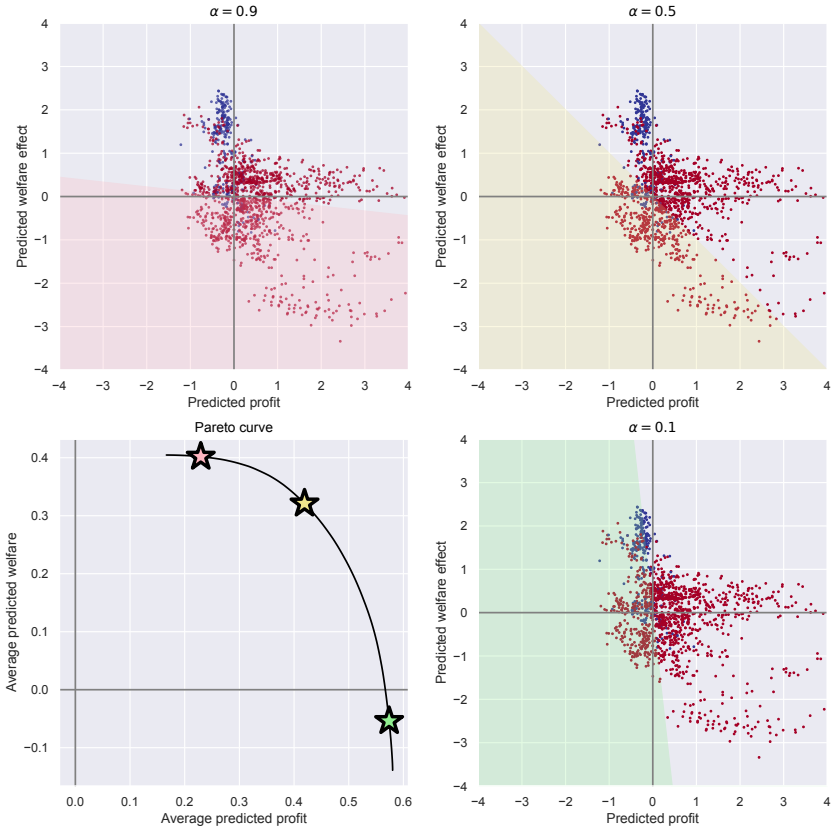

Figure 2: Decision Boundaries and Pareto Frontiers

been denied access to credit (in the 2nd quadrant) are now approved, while a number of customers who might not realize any gains in welfare are rejected (in the 4th quadrant). Such a policy proves to be costly for the lender, but generates large gains in welfare. At this trade-off rate ($\alpha = 0.5$) profits decrease by 27%, relative to the scenario where $\alpha = 0.1$. At the same time, consumer welfare increase by almost 300%. Finally the top-left sub-figure describes a policy that gives a much larger weight to welfare ($\alpha = 0.9$) and thus severely penalizes lender's profit.

There are a number of caveats to these results. Our data is drawn from a single setting, and measures welfare over a relatively short time (3 months). As a result, we are unable to say much about how longer exposure to digital credit might affect customer welfare. We also note that we use very parsimonious models to estimate profit and welfare scores, which results in some amount of prediction error. As a result, the frontiers we estimate might be Pareto sub-optimal.

6 Conclusion

We have shown that it is possible to develop welfare sensitive credit scores, which are easy to implement and interpret. Financial institutions could switch their internal credit scoring algorithms to balance multiple objectives simultaneously, for instance to prioritize loans to clients for whom the loan is predicted to have a positive impact on welfare. Or, a regulator could request disclosure of these welfare measures for auditing or ongoing monitoring.

REFERENCES

Anil Aswani and Matt Olfat. Optimization hierarchy for fair statistical decision problems. 2019. doi: 10.48550/ARXIV.1910.08520. URL https://arxiv.org/abs/1910.08520.

Susan Athey, Julie Tibshirani, and Stefan Wager. Generalized random forests. *The Annals of Statistics*, 47(2):1148 – 1178, 2019. doi: 10.1214/18-AOS1709. URL https://doi.org/10.1214/18-AOS1709.

Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and Machine Learning: Limitations and Opportunities*. fairmlbook.org, 2019. http://www.fairmlbook.org.

Daniel Björkegren, Joshua Blumenstock, Omowunmi Folajimi-Senjobi, Jacqueline Mauro, and Suraj R. Nair. Instant loans can lift subjective well-being: A randomized evaluation of digital credit in nigeria. Working Paper, 2022a. URL https://arxiv.org/abs/2202.13540.

Daniel Björkegren, Joshua E. Blumenstock, and Samsun Knight. (Machine) Learning What Policies Value, June 2022b. URL http://arxiv.org/abs/2206.00727. arXiv:2206.00727 [cs, econ, q-fin].

Kalyanmoy Deb and Deb Kalyanmoy. *Multi-Objective Optimization Using Evolutionary Algorithms*. John Wiley amp; Sons, Inc., USA, 2001. ISBN 047187339X.

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, pp. 214–226, New York, NY, USA, 2012. Association for Computing Machinery. ISBN 9781450311151. doi: 10.1145/2090236.2090255. URL https://doi.org/10.1145/2090236.2090255.

Danielle Ensign, Sorelle A. Friedler, Scott Neville, Carlos Scheidegger, and Suresh Venkatasubramanian. Runaway feedback loops in predictive policing. In Sorelle A. Friedler and Christo Wilson (eds.), *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pp. 160–171. PMLR, 23–24 Feb 2018. URL https://proceedings.mlr.press/v81/ensign18a.html.

Johannes Haushofer, Paul Niehaus, Carlos Paramo, Edward Miguel, and Michael W Walker. Targeting impact versus deprivation. Working Paper 30138, National Bureau of Economic Research, June 2022. URL http://www.nber.org/papers/w30138.

Yaochu Jin and Bernhard Sendhoff. Pareto-based multiobjective machine learning: An overview and case studies. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 38(3):397–415, 2008. doi: 10.1109/TSMCC.2008.919172.

Maximilian Kasy and Rediet Abebe. Fairness, equality, and power in algorithmic decision-making. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, FAccT '21, pp. 576–586, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450383097. doi: 10.1145/3442188.3445919. URL https://doi.org/10.1145/3442188.3445919.

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent Trade-Offs in the Fair Determination of Risk Scores. In Christos H. Papadimitriou (ed.), *8th Innovations in Theoretical Computer Science Conference (ITCS 2017)*, volume 67 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pp. 43:1–43:23, Dagstuhl, Germany, 2017. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-95977-029-3. doi: 10.4230/LIPIcs.ITCS.2017.43. URL http://drops.dagstuhl.de/opus/volltexte/2017/8156.

Sören R Künzel, Jasjeet S Sekhon, Peter J Bickel, and Bin Yu. Metalearners for estimating heterogeneous treatment effects using machine learning. *Proceedings of the national academy of sciences*, 116(10):4156–4165, 2019.

Lydia T. Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. Delayed impact of fair machine learning. In Jennifer Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 3150–3158. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/liu18c.html.

Hussein Mouzannar, Mesrob I. Ohannessian, and Nathan Srebro. From fair decision making to social equality. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, pp. 359–368, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450361255. doi: 10.1145/3287560.3287599. URL https://doi.org/10.1145/3287560.3287599.

Alejandro Noriega-Campero, Michiel A. Bakker, Bernardo Garcia-Bulle, and Alex 'Sandy' Pentland. Active fairness in algorithmic decision making. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '19, pp. 77–83, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450363242. doi: 10.1145/3306618.3314277. URL https://doi.org/10.1145/3306618.3314277.

Esther Rolf, Max Simchowitz, Sarah Dean, Lydia T. Liu, Daniel Bjorkegren, Moritz Hardt, and Joshua Blumenstock. Balancing competing objectives with noisy data: Score-based classifiers for welfare-aware machine learning. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 8158–8168. PMLR, 13–18 Jul 2020. URL https://proceedings.mlr.press/v119/rolf20a.html.

Paige Marta Skiba and Jeremy Tobacman. Do payday loans cause bankruptcy? *The Journal of Law and Economics*, 62(3):485–519, august 2019. doi: https://www.journals.uchicago.edu/doi/full/10.1086/706201.

A APPENDIX

A.1 DESCRIPTION OF EXPERIMENTAL DESIGN IN BJÖRKEGREN ET AL. (2022A)

In this section, we briefly summarize the experimental setting in Nigeria.

Björkegren et al. (2022a) launch an experiment in collaboration with the FSP in Nigeria, which included 8% (randomly selected) of all new customers between August 2019 to February 2020. Customers were cross-randomized across two different treatment arms:

- Auto-approval: Half of all customers (4% of all new customers) were automatically approved for credit, regardless of credit score. The other half ('standard approval' group) were approved only if their credit score exceeded a threshold set by the FSP.
- Loan Value: All customers who were approved received a randomly assigned maximum initial loan offer, selected from NGN 1000, 2000, 5000, 10,000, or 13,000 (between about $2.75 and $35.75). Customers who repaid their initial loan on time would subsequently be eligible for future loans according to the FSP's standard loan ladder.

Data on welfare outcomes was collected roughly three months after customers signed up.

A.2 VARIABLE DEFINITIONS

Subjective well-being index

A Index of subjective well-being (standard deviations): The subjective well-being index is a weighted average of two variables: the respondents' z-score on the PHQ-9 questionnaire, and the z-score of their response to a life satisfaction question, similar to those in the World Values Survey. Note that the respondent's PHQ-9 score can range from 0-27; for ease of visual presentation, we divide the total PHQ-9 score by 27, so that the value ranges from 0 to 1. A lower PHQ-9 score indicates lower levels of depression. The z-scores are constructed by subtracting the mean of the standard approval group and dividing by the standard deviation of the standard approval group. The life satisfaction question we use is: All things considered, how satisfied are you with your life as a whole these days? (Very happy/ quite happy/ not very happy/ not at all happy)

Table 1: Summary Statistics - I

		(1) Mean		(2) Weighted Mean	
PANEL A: DEMOGRAPHICS					
Age		29.936	(8.532)	29.307	(8.384)
Male		0.758	(0.429)	0.760	(0.427)
Location:	Lagos	0.333	(0.471)	0.335	(0.472)
Education:	Primary	0.007	(0.082)	0.007	(0.082)
	Secondary	0.349	(0.477)	0.348	(0.476)
	HND	0.093	(0.290)	0.094	(0.292)
	OND	0.149	(0.356)	0.149	(0.356)
	University	0.357	(0.479)	0.357	(0.479)
Head of household		0.447	(0.497)	0.448	(0.497)
Household size		5.303	(3.199)	5.246	(3.173)
Ethnicity:	Yoruba	0.502	(0.500)	0.500	(0.500)
	Igbo	0.179	(0.383)	0.179	(0.383)
	Hausa	0.043	(0.202)	0.043	(0.204)
PANEL B: EMPLOYMENT/ MISC.					
Primary phone user		0.991	(0.093)	0.992	(0.089)
Uses a bank account		0.997	(0.056)	0.997	(0.054)
Employment:	Self-employed	0.409	(0.492)	0.401	(0.490)
	Salaried (Full-time)	0.269	(0.443)	0.271	(0.445)
	Salaried (Part-time)	0.121	(0.326)	0.125	(0.331)
	Unemployed	0.201	(0.401)	0.202	(0.402)
Days worked last week		3.861	(2.418)	3.871	(2.426)
Runs a business		0.551	(0.498)	0.543	(0.498)
Aspires to open business		0.806	(0.396)	0.790	(0.395)