

The relation between microfinancing and corruption by country: An analysis of an open source dataset

Heather Roy and Sue Kase
Computational and Information Sciences Directorate
U.S. Army Research Laboratory
Aberdeen Proving Ground, U.S.
heather.e.roy2.ctr@mail.mil; sue.e.kase.civ@mail.mil

Abstract—Examining the relation between global microlending and corruption may inform how trust and influence propagate through crowds. Building this understanding may help U.S. Army intelligence officers leverage crowds for humanitarian efforts as well as, to detect signs of adversarial influence. A dataset was created combining open source data from Kiva, a non-profit microfinancing institution, and Transparency International, a global coalition against corruption that publishes an annual Corruption Perceptions Index (CPI). The CPI was merged with Kiva microfinancing variables related to Kiva field partners. A preliminary analysis was conducted on a subset of the data in an effort to determine a near real-time microfinancing proxy for the CPI using the Kiva microfinancing data. Results suggest that when controlling for time on Kiva, the average loan size in dollars, delinquency rate, average loan size per GDP, and average time to fund loan all significantly predict CPI.

Keywords—open source dataset, microfinancing, corruption

I. INTRODUCTION

The U.S. Army has invested in crowds to design gear, advance democracy, and provide life-enhancing advances for battle injuries. With the technological advances leading to increasing connectivity and social interaction between individuals online across the world, it is important for the U.S. Army to build an understanding of how online crowdsourcing networks may be leveraged and influenced. In order to build this understanding, we must begin to develop a strong framework for understanding how influence and trust grow in crowd networks.

As part of the Network Science Collaborative Technology Alliance (NS CTA), researchers from the U.S. Army Research Laboratory are collaborating with academic researchers to determine what network and node based trust mechanisms drive crowd engagement and selection. For instance, consider the challenges raised by crowdfinancing, where crowdsourcers must evaluate and identify persons who they believe will responsibly employ their resources. Additionally, government agencies have long used microlending to cool global hotspots by raising the human condition and fostering loyalty towards the U.S. and its allies. Building an understanding of factors contributing to trust and influence in a crowdsourcing network may help U.S. intelligence officers better understand how to

leverage crowds for humanitarian purposes as well as, assist in the detection of adversarial influence. One such factor that may intuitively affect trust and influence in a network is corruption.

This paper presents an initial investigation of how corruption may influence crowdsourcing, and microfinancing in particular. A body of data was built using open source information. Data was collected from two public websites: Kiva and Transparency International. Kiva is a non-profit organization that uses crowdsourcing to connect people through lending to decrease poverty. Kiva works with microfinance institutions on five continents called field partners. These field partners represent the critical on the ground link from lenders to borrowers; responsible for administering loans. Transparency International is a global coalition against corruption. It was founded in 1993 as the world's first international anti-corruption organization. Since 1995, Transparency International has been publishing the Corruption Perceptions Index (CPI) to measure the perceived levels of public sector corruption by surveying analysts, businesspeople, and experts worldwide. The 2014 CPI measures corruption in 175 countries and territories. Transparency International generally defines corruption as “the abuse of entrusted power for private gain” [1]. By combining information from Kiva and Transparency International, we were able to conduct a preliminary investigation of the relation between microfinancing and corruption at the country level of analysis.

II. METHODS

A. Building an Open Source Dataset

Open source data was collected and combined into a database using the Statistical Package for the Social Sciences (SPSS), from two public websites: www.kiva.org/partners and www.transparency.org/cpi2014. Data from Kiva was collected in two stages. In an attempt to keep the data consistent across partners, screen shots were taken of the list of the 285 Kiva field partners from the Kiva website on November 10, 2014. From these screen shots, the following variables were collected pertaining to the Kiva field partners: name of organization, countries approved to post loans from, total number of countries approved to post loans from, months on Kiva, Field Partner Social Performance Badges, total number of

performance badges awarded, partner risk rating, delinquency rate, and default rate.

Once the initial Kiva data entry was complete, the following variables were collected for each field partner from the Kiva website directly: number of borrowers, dollar amount of total loans, refund rate in percentage, amount of refunded loans in dollars, number of refunded loans, percentage of loans to women, average loan size in dollars, average individual loan size in dollars, average group loan size in dollars, average gross domestic product (GDP) per capita based on purchasing power parity (PPP) in local country, average loan size over GDP per capita PPP, average time to fund a loan in days, average dollars raised per day per loan, and average loan term in months. This data was added in January; therefore, a limitation is the data for these variables is more recent (e.g., approximately 1.5 months). During that time period, five field partners were no longer listed on the Kiva website and were removed from the dataset, leaving 280 field partners.

Kiva assigns a risk rating to calculate the risk of institutional default for each field partner. However, Kiva does not assign a risk rating for field partners completing a basic due diligence process or experimental partnership, or for field partners completing a full due diligence process in an industry without a Kiva risk model. Given that Kiva has identified a difference between those field partners, the preliminary analysis was limited to the subset of data with field partner risk ratings. Furthermore, three field partners that were missing both default and delinquency rates were excluded from analysis, resulting in 116 field partners.

Corruption information was collected from the Transparency International website. The 2014 CPI was recorded and merged with the Kiva data by the first country listed for each field partner. To control for confounding variables, analysis was limited to the field partners approved to post loans from no more than one country with a CPI listing.

B. Automatic Linear Modeling

A preliminary analysis was conducted on the data from 116 field partners to develop a near real-time proxy for the CPI using the microfinancing variables. A forward stepwise multiple regression analysis was conducted using the Automatic Linear Modeling (ALM) analysis in SPSS to determine which microfinancing predictors contributed substantially to the model's ability to predict CPI. The following microfinancing variables were included as predictors: amount refunded, average group loan, average loan, average loan size per GDP, average time to fund loan, delinquency rate, number of Kiva borrowers, number of loans to women, number of refunds, refund rate, and total loans. In order to control for time, each of these original variables were divided by the total number of months the field partner had been on Kiva, creating new variables that were input into the analysis as the predictor variables.

Based upon initial observations of the scatter plot showing the predicted and observed CPI values, six outliers (three field partners from Iraq, one from South Sudan, one from Ukraine, and one from the Philippines) were identified. These outliers were removed when the predicted scores appeared

significantly higher than the observed values. The outliers corresponded to six of the fourteen field partners identified as possible outliers through the ALM output.

Countries were coded in the dataset into one of six regions defined by Transparency International. According to this classification, within the defined subset of 110 field partners (116 minus outliers), there were 43 field partners operating in the Americas region, 30 in the Sub-Saharan African region, 18 in Asia Pacific, 14 in Eastern Europe and Central Asia, 5 in the Middle East and North Africa, and 0 in the European Union and Western Europe region.

With the outliers removed, the ALM was rerun with the best subsets option. The ALM method for multiple regression was selected for its best subsets capability. ALM automatically prepares data for the purpose of improving the predictive power; for example, by winsorizing outliers. The information criterion was selected as the basis for entry into or removal from the model. The ALM process excluded three field partners from analysis; results reported below are pertaining to the final subset of 107 field partners (110 minus three removed via ALM). As this was an exploratory analysis, the alpha level for all tests was set at 0.05, so that the regression model would be sensitive enough to detect any possible predictors of merit.

III. RESULTS

A multiple regression was run to determine the best model to predict CPI from 11 microfinancing variables. The assumptions of linearity, independence of errors, homoscedasticity, unusual points and normality of residuals were met. The best model, with the highest adjusted R-square value of 0.535, determined the average loan size in dollars per months on Kiva, average loan size per GDP per months on Kiva, delinquency rate per months on Kiva, and the average time to fund loans in days per months on Kiva statistically predicted CPI, $F(4,102) = 31.504, p < 0.05, R^2 = 0.5833$. All four variables added statistical significance to the prediction, $p < 0.05$. Regression coefficients and standard errors can be found in Table 1.

TABLE I. SUMMARY OF MULTIPLE REGRESSION ANALYSIS

Table 1	Summary of Multiple Regression Analysis		
	B	SE B	β
Intercept	32.848	1.624	
Average Loan Size in Dollars / Months on Kiva	0.635	0.067	0.361
Average Loan Size per GDP / Months on Kiva	-10.544	2.405	0.228
Delinquency Rate / Months on Kiva	23.031	8.235	0.207
Average Time to Fund Loans in Days / Months on Kiva	-26.870	10.433	0.204

^a $R^2 = .5833$, Adjusted $R^2 = 0.535$.

IV. DISCUSSION

The adjusted R-square value suggests that 53.5% of the variance in CPI may be explained by the four predictor variables included in the model. The standardized beta coefficients (Table 1) indicate average loan size per months on Kiva has the strongest effect on CPI, followed by average loan size per GDP per months on Kiva, delinquency rate per months on Kiva, and average time to fund loans in days per months on Kiva.

Additionally, it is important to consider the direction of the relations between the predictors and CPI. The unstandardized coefficients indicate the direction of the relation between the predictors and CPI (Table 1). Results indicate that average loan size in dollars and delinquency rate per months on Kiva are positively related to CPI. When interpreting these relations it is important to remember that a low CPI score indicates a greater corruption level whereas a high CPI score indicates the field partner is operating in a country perceived to be highly clean of corruption. Considering the positive relations, results suggest average loan size in dollars per month increases as CPI increases. This may suggest that countries perceived to be cleaner of corruption, are more likely to distribute larger average loans. Interestingly, delinquency rate per month is also positively related to CPI such that countries cleaner of corruption have higher delinquency rates. It is possible that wealthier countries can afford to distribute higher average loans while being more lenient on delinquency.

The average loan size per GDP and the average time to fund loans per months on Kiva are negatively related to CPI (Table 1). Kiva calculates the average GDP per capita PPP in a local country by dividing the GDP by the total population of the country to approximate the average yearly income of an individual. The average loan size in dollars divided by this estimate provides a proportion of the average loan in U.S. dollars relative to the average income of a country, per months on Kiva. The negative relation between this predictor and CPI suggests that field partners operating in countries perceived as more corrupt distribute loans that represent a larger proportion of the average income in that country.

Kiva defines the average time to fund loans in days as an indirect measure of the popularity of the loans posted by a field partner [2]. The average time to fund loans refers to the number of days it takes for a loan to be fully funded through crowdsourcing. When controlling for the time a field partner has been on Kiva, this variable negatively relates to CPI such that field partners operating in countries perceived as more corrupt, on average take longer to fund loans.

Applying the above observations in a field partner vignette, we informally evaluate the effectiveness of ALM for understanding microfinancing related precursors to corruption. As an example, in 2013 Yemen earned its lowest CPI score (18) in 11 years. A field partner included in our analysis, Al-

Amal Microfinance Bank (AAMB), operates in Yemen's capital city of Sana'a and in nine governorates representing 80% of the population. AAMB has a low delinquency rate of 0.23% compared to the average 5.75% across Kiva field partners, an average loan size of \$311 compared to \$417 across Kiva partners, an average loan size GDP per capita (PPP) of 11.50% compared to 12.26% across Kiva partners, and an average time to fund a loan of 10.02 days compared to 5.99 days across Kiva partners.

For the most part, AAMB's statistics support a low CPI value. The low delinquency rate and small loan amount may result from AAMB operating in a conflict-affected region which can greatly increase the cost of safely delivering financial services to borrowers. A limited or poorly functioning banking system would also make it difficult to access funding locally. The average time for AAMB to fund a loan is nearly twice the Kiva partner average. AAMB conducts business in areas with very poor infrastructure, such as hazardous roadways, increasing the time and cost of finding clients and maintaining branch offices. With 18 of AAMB's branch offices located outside the capital city of Sana'a in rural areas, travel to find and serve clients becomes costly.

In the future analyses, more thought is required regarding the transformations of the variables to control for effects of a given country's financial status across all variables, interaction effects, and additional indices of corruption. Modeling subsets of the data should also be considered; for example, by region in order to build a clearer understanding of the predictive relation between various microfinancing variables and CPI.

V. CONCLUSION

Results of a preliminary ALM analysis suggest the predictive variables average loan size in dollars per months on Kiva, average loan size per GDP per months on Kiva, delinquency rate per months on Kiva, and the average time to fund loans per months on Kiva statistically predicted CPI. These results suggest further exploration is warranted in order to better understand microfinancing data as a proxy indicator of corruption.

ACKNOWLEDGMENT

This research was supported in part by an appointment to the Student Research Participation Program at the U.S. Army Research Laboratory administered by the Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and USARL. The authors thank Dr. Barry Bodt (ARL) for his guidance of the statistical analysis and Dr. Jonathon Bakdash (ARL) for his assistance in formulating the microfinance and corruption problem space.

REFERENCES

- [1] Transparency International. (2015). *FAQS on Corruption*. Retrieved http://www.transparency.org/whoweare/organisation/faqs_on_corruption/9/
- [2] Kiva. (2015). *XacBank*. Retrieved <http://www.kiva.org/partners/116>