Optimal Credit Scores Under Adverse Selection

Optimal Credit Scores Under Adverse Selection

NICOLE IMMORLICA, Microsoft Research
ANDRE M. SZTUTMAN, Massachusetts Institute of Technology
ROBERT M. TOWNSEND, Massachusetts Institute of Technology

CCS Concepts: • **Applied computing** → **Economics**.

Additional Key Words and Phrases: information design, applied machine learning, credit scores, rural credit markets

ACM Reference Format:
Nicole Immorlica, Andre M. Sztutman, and Robert M. Townsend. 2022. Optimal Credit Scores Under Adverse Selection. In *Proceedings of the 23rd ACM Conference on Economics and Computation (EC '22), July 11–15, 2022, Boulder, CO, USA.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3490486.3538331

Data is becoming increasingly available and more easily processed. New data and methods are useful across many economic sectors and applications, including credit markets. There is a large population of potential borrowers who have short credit histories and thus are unable to receive credit [1]. The new methods would allow banks to identify the creditworthy among these potential borrowers, giving credit to those who perhaps need it the most [2].

Because of this increased capacity to identify creditworthy individuals, one may hope that the inefficiencies coming from information asymmetries would progressively disappear. However, a key reason that makes information asymmetries generate inefficiencies in these thin credit market segments is adverse selection: as the price of the loans decreases (or interest rates increase), the pool of borrowers can get progressively worse. Those who would be more likely to repay are only willing to borrow at higher prices. The credit market unravels, resulting in too few or no transactions happening. As long as there is some information asymmetry, some heterogeneity in expected repayment rates that lenders cannot observe, there can still be adverse selection problems.

Data owners, such as data-intensive firms and platforms, may hope that by making their data available to financial providers they will improve credit access. However, this hope lacks a theoretical justification. The inefficiencies arising from adverse selection do not necessarily get better with more information, and indeed, may as well get worse. As shown in Levin [3], more information does not necessarily increase the number of transactions and the realized gains from trade. More information can prevent implicit cross-subsidization between different types, making the previously subsidized types leave the market. Hence, as more data and improved technologies for processing data arrive, there remain key issues concerning how much data to share.

To answer the question of how much data to share, we build on the literature on information design and formulate the optimal disclosure problem of a partially informed intermediary with commitment, maximizing the probability of successful transactions weighted by the size of gains from trade. This formulation allows us to answer the question of which variables in a dataset should be shared with financial providers – for example, whether geographic information should

EC '22, July 11–15, 2022, Boulder, CO, USA

ACM ISBN 978-1-4503-9150-4/22/07.
https://doi.org/10.1145/3490486.3538331

be shared and at which level of granularity, or whether only an index that combines different pieces of information should be shared. We construct an optimal disclosure system and derive new conditions for the optimality of a disclosure system in terms of local sufficient statistics. The optimal policy should satisfy three simple properties: i) generically messages should combine at most two signals; ii) there should be an increasing relationship between the price elasticities of the value of the loans to investors and the prices of these loans; and iii) when different signals are combined into a single message, there should be a decreasing relationship between these elasticities and the prices these loans would have if the signals were unbundled.

We apply our results to the rural credit markets in Thailand. This is a particularly fitting setup for at least four different reasons. First, these credit markets are thin and there is not much risk sharing across villages, so the potential welfare gains are large. Second, there is evidence of intensive risk sharing within villages, which makes us think that they, through a platform acting on their behalf, are able to organize and commit to an optimal disclosure policy. Third, a unique feature of this setup benefits us from an identification perspective. There is a main lender, the Bank of Agriculture and Agricultural Cooperatives, a government-owned bank, holding a significant fraction of the market for agricultural loans. This bank uses a rigid set of rules to set interest rates. We explore variation in these rules as a source of identification for slopes of supply and average value curves. These slopes are key ingredients in the computation of the optimal credit scores and appear as sufficient statistics in the necessary conditions we derive for the optimality of disclosure systems. Fourth, we benefit from rich data from Townsend Thai Project, including detailed information on consumption, income and its different sources, crops, livestock, loans, and interest rates. This allows the construction of detailed balance sheets, income and cash flow statements for each household, as well as their credit histories. Assuming that the platform has access to the detailed information in this dataset, while investors do not, we show what pieces of information should be made available to investors and how, effectively constructing "optimal credit scores."

We find that the optimal disclosure policy substantially improves the gains from trade relative to a simple full disclosure policy, with the size of gains being of the order of 0.45% the size of a typical loan per household per month. Moreover, we find that the optimal policy puts higher weight than full disclosure credit scores on variables seemingly related to the solvency of farmers relative to variables that are informative about their current liquidity. Our findings can be instrumental in improving credit access in places where it is most needed by making better use of data.

The complete version of this paper can be found at https://ssrn.com/abstract=4113505.

REFERENCES

[1] Jesse Bricker, Lisa J Dettling, Alice Henriques, Joanne W Hsu, Lindsay Jacobs, Kevin B Moore, Sarah Pack, John Sabelhaus, Jeffrey Thompson, and Richard A Windle. 2017. Changes in US family finances from 2013 to 2016: Evidence from the Survey of Consumer Finances. *Fed. Res. Bull.* 103 (2017), 1.

[2] Julapa Jagtiani and Catharine Lemieux. 2019. The roles of alternative data and machine learning in fintech lending: evidence from the LendingClub consumer platform. *Financial Management* 48, 4 (2019), 1009–1029.

[3] Jonathan Levin. 2001. Information and the Market for Lemons. *RAND Journal of Economics* (2001), 657–666.