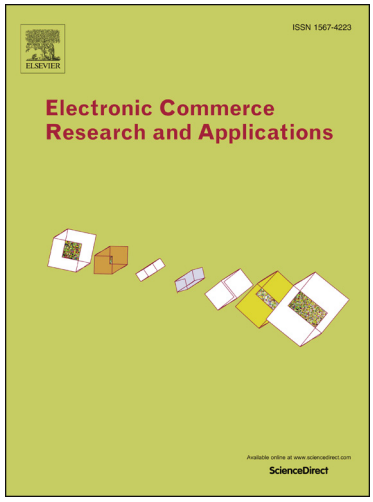

A novel behavioral scoring model for estimating probability of default over time in Peer-to-Peer lending

Zhao Wang, Cuiqing Jiang, Yong Ding, Xiaozhong Lv, Yao Liu

PII: S1567-4223(17)30099-6
DOI: https://doi.org/10.1016/j.elerap.2017.12.006
Reference: ELERAP 749

To appear in: *Electronic Commerce Research and Applications*

Received Date: 12 September 2017
Revised Date: 18 December 2017
Accepted Date: 18 December 2017

Please cite this article as: Z. Wang, C. Jiang, Y. Ding, X. Lv, Y. Liu, A novel behavioral scoring model for estimating probability of default over time in Peer-to-Peer lending, *Electronic Commerce Research and Applications* (2017), doi: https://doi.org/10.1016/j.elerap.2017.12.006

A NOVEL BEHAVIORAL SCORING MODEL FOR ESTIMATING PROBABILITY OF DEFAULT OVER TIME IN PEER-TO-PEER LENDING

Zhao Wang [a], Cuiqing Jiang (corresponding author) [a], Yong Ding [a], Xiaozhong Lv [a], Yao Liu [a]

[a] School of Management, Hefei University of Technology, Hefei 230009, Anhui, P.R. China

Last revised: December 18, 2017

ABSTRACT

Traditional behavioral scoring models applying classification methods that yield a static probability of default may ignore the borrowers' dynamic characteristics because borrower repayment behavior evolves dynamically. In this study, we propose a novel behavioral scoring model based on a mixture survival analysis framework to predict the dynamic probability of default over time in peer-to-peer (P2P) lending. A random forest is utilized to identify whether a borrower will default, and a random survival forest is introduced to model the time to default. The results of an empirical analysis on a Chinese P2P loan dataset show that the proposed *ensemble mixture random forest* (EMRF) has a better performance in terms of predicting the monthly dynamic probability of default, while compared with standard mixture cure model, Cox proportional hazards model and logistic regression. It is also concluded that the proposed EMRF model provides a meaningful output for timely post-loan risk management.

Keywords: Behavioral scoring; dynamic probability, P2P lending; random forest; random survival forest; risk management; survival analysis.
