A Novel Portfolio Optimization Framework for Online Loan Investments

Ram Gopal Xiao Qiao Moris Strub **Zonghao Yang**

Information Systems Research (2025)

FA590 Statistical Learning in Finance

FinTech and Peer-to-Peer Lending

- Financial technology: Integration of technology in financial services
- P2P lending: One of the most successful examples of fintech
  - Underwriting, application screening, credit risk assessment, etc.
- Online platforms where individuals lend and borrow from one another
- Disintermediation of banks in the lending practice
  - Borrowers: Greater access to credit and lower interest rates
  - Lenders: A novel investment opportunity
- Platforms: Zopa, **LendingClub**, Prosper, and many more

P2P Loans

Unsecured personal loans: Installment loans with fixed terms and without collaterals

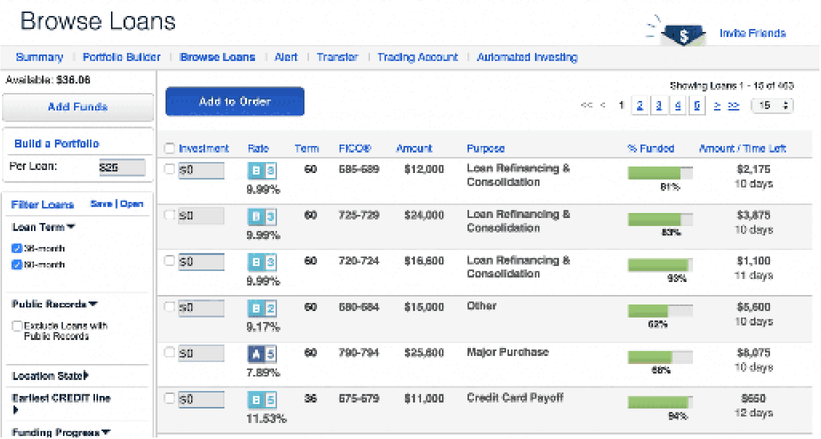
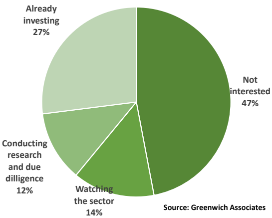


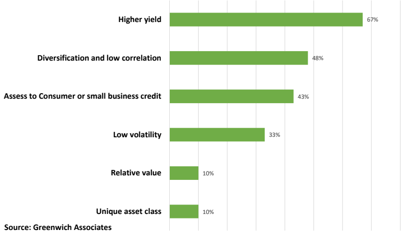
Figure: Investor Interface (LendingClub)

STEVENS INSTITUTE OF TECHNOLOGY

Online Lending

- Expanded investor base from retail to institutional investors
  - European Investment Bank, Aviva, BlackRock, etc.





- Rapid growth: \$112.9 billion in 2021[1]
  - Stock - \$100 trillion; Bond - \$119 trillion[2]
- Challenges to continued growth: Risk of losing money, stringent government regulations, and low awareness of the benefits

[1] Research and Markets, "Peer to Peer (P2P) Lending Market: Global Industry Trends, Share, Size, Growth, Opportunity and Forecast 2022-2027," April 2022

[2] Estimates obtained from the World Bank and SIFMA.

STEVENS INSTITUTE OF TECHNOLOGY

This Paper

- Online loans as a novel asset class
  - Building a portfolio of online loans
  - Online loans vs. stocks, bonds, and real estate
- General Characteristics-based Portfolio Policies (GCPP) framework
  - Nonlinear portfolio policy based on **neural networks**
  - The nonlinear policy leads to an average annual internal rate of return of **13.08%** compared to **6.55%** of an equal-weight (EW) portfolio
- Online loans expand the investor's opportunity set
  - Competitive rates of return with stocks and bonds
  - Limited comovement offers diversification benefits

STEVENS
INSTITUTE OF TECHNOLOGY

Data

- 1,158,476 loans on LendingClub originating between January 2013 and May 2017
- A total of 83 loan and borrower characteristics [Detail]
  - Loan characteristics (20): Loan amount, interest rate, grade, status, total payment, last payment date, etc.
  - Borrower characteristics (63): Delinquent loans, gross income, location, debt-to-income ratio, etc.
- After data preprocessing, there are a total of 144 features
  - One-hot encoding
  - Indicator variables of missing data

Performance Measures

Consider a sequence of cash flows for a loan portfolio: $P_t, t \in \{0, 1, \ldots, T\}$

- Internal rate of return (IRR)

$$0 = NPV = \sum_{t=1}^{T} \frac{P_t}{(1 + IRR)^t} - P_0 \qquad (1)$$

- Return on investment (ROI): Calculated based on cummulative discount payment (CDP)

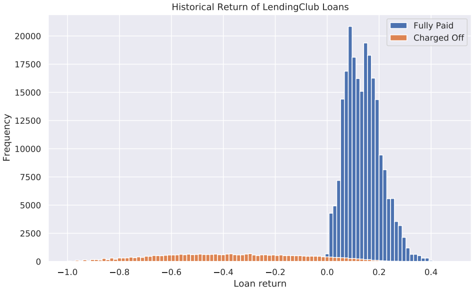
$$ROI = \frac{CDP - P_0}{P_0}, CDP = \frac{P_1}{(1 + \frac{d}{12})^1} + \cdots + \frac{P_T}{(1 + \frac{d}{12})^T} \qquad (2)$$

where d is the discount rate, representing the time value of money

STEVENS INSTITUTE OF TECHNOLOGY

Literature: Credit Risk Assessment

- [Iyer et al., 2016, Emekter et al., 2015]: The charge-off risk of a borrower and the determinants for credit risk
- Binary classification: F1 score $\sim$ 0.30 and AUC score $\sim$ 65%
- [Fu et al., 2021]: EW on predicted fully-paid loans
- Binary prediction alone is insufficient to fully capture credit risk



Literature: A Portfolio of Loans

- Portfolio optimization methods that leverage information in return distribution beyond the binary loan status
- [Guo et al., 2016]: Mean-variance optimization framework

$$
\begin{aligned}
\min_{\omega} \quad & \omega^T \Omega \omega \\
\text{subject to} \quad & \omega^T R = R^* \\
& \sum_{i=1}^{N} \omega_i = 1
\end{aligned}
\tag{3}
$$

- Expected returns (R) and risk (Ω) are difficult to estimate
- [Michaud, 1989]: Unstable solutions given estimation error

STEVENS INSTITUTE OF TECHNOLOGY

Building a Portfolio of Loans

- General characteristics-based portfolio policies (GCPP)
- Key idea: Parameterize portfolio weights as a function of characteristics
- N available loans, a direct mapping from the loan characteristics to the **goodness** of a loan $g(\hat{x}, \theta)$
  - $\hat{x}$ is the vector of standardized characteristics
  - θ is a vector of parameters
- Portfolio weight for each loan i

$$\omega_i = \frac{g(\hat{x}_i; \theta)}{\sum_{j=1}^{N} g(\hat{x}_j; \theta)} \tag{4}$$

STEVENS INSTITUTE OF TECHNOLOGY

Portfolio Optimization

- r_i is the ROI of loan i

$$\sup_{\theta \in \Theta} \mathbb{E}\left[u(r_\omega)\right] = \sup_{\theta \in \Theta} \mathbb{E}\left[u\left(\sum_{i=1}^{N} \omega_i \cdot r_i\right)\right]$$
$$= \sup_{\theta \in \Theta} \mathbb{E}\left[u\left(\sum_{i=1}^{N} \frac{g(\hat{x}_i;\theta)}{\sum_{j=1}^{N} g(\hat{x}_j;\theta)} \cdot r_i\right)\right] \tag{5}$$

- The lender's problem is to choose the value of parameters $\theta^* \in \Theta$ that **maximize the expected utility**
- Constant relative risk aversion (CRRA) utility function:

$$u(r_\omega) = \frac{(1+r_\omega)^{1-\gamma}}{(1-\gamma)}$$

where γ is the risk-aversion coefficient

STEVENS INSTITUTE OF TECHNOLOGY

The Goodness Function

- Mapping between loan characteristics and goodness
- Truncated linear policy [Brandt et al., 2009]

$$g_L(\hat{x}_i;\theta) = \max(0, \bar{\omega} + \frac{1}{N}\theta^\top \hat{x}_i) \tag{6}$$

where $\bar{w} = \frac{1}{N}$ is the benchmark weight, i.e., an equal-weight portfolio

- Nonlinear policy: Neural networks, $g_{NN}(\hat{x}_i;\theta)$
  - Richer interactions and more complex transformations of features
  - Larger portfolio weight space

Nonlinear Policy Specification

- A shallow neural network with a single hidden layer
  - Network size: $\mathbf{144} \rightarrow \mathbf{64} \rightarrow \mathbf{1}$
  - Activation function: Sigmoid function
- Shallow vs. deep learning [Gu et al., 2020]
  - Financial data: Low signal to noise ratio
- Mini-batch optimization with 256 loans
  - Escape from the saddle point in the objective

STEVENS
INSTITUTE OF TECHNOLOGY

Architecture

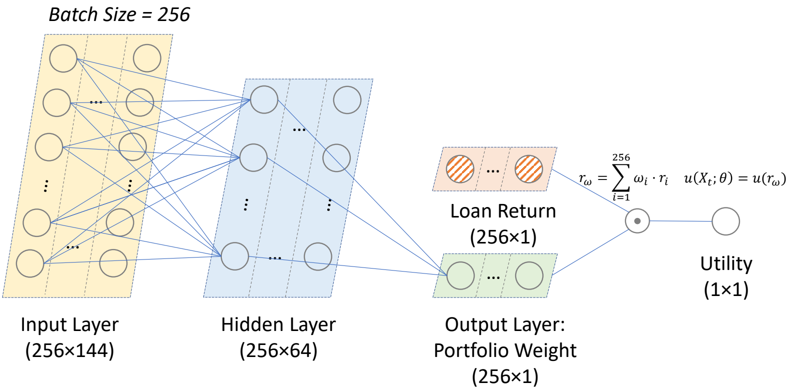
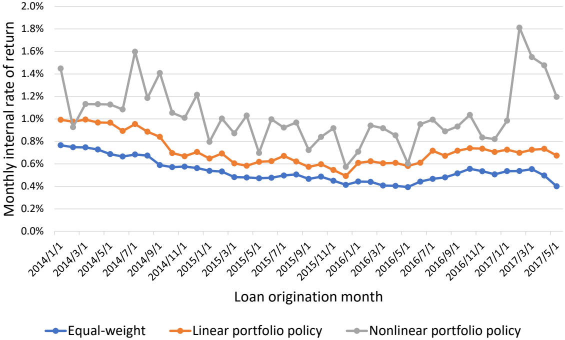


Figure: Neural Network with Batch Training

Performance vs. Equal-Weight

The nonlinear GCPP outperforms various benchmarks

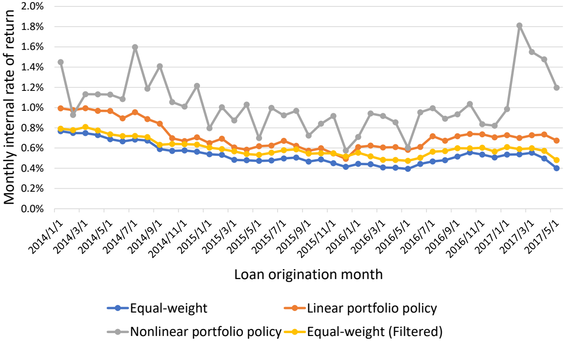
- [DeMiguel et al., 2009]: EW is difficult to beat with portfolio optimization methods



Performance vs. Credit Risk Filtering

The nonlinear GCPP outperforms various benchmarks

- XGBoost to predict charged-off probability, EW portfolio of loans with predicted charged-off rate < 0.15



Performance vs. Mean Variance

The nonlinear GCPP outperforms various benchmarks

- The average IRR of the MV portfolio is close to the nonlinear portfolio
- Out-of-sample portfolio performance is very volatile

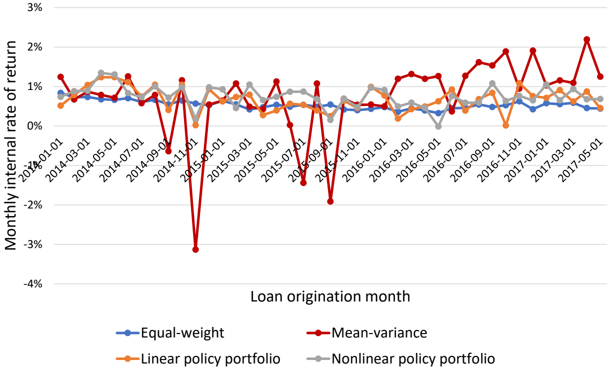


Figure: Mean-Variance Framework

STEVENS INSTITUTE OF TECHNOLOGY

Portfolio Composition

- Mismatch of risk and interest rate
- GCPP advocates for greater allocation to credit-constrained borrowers

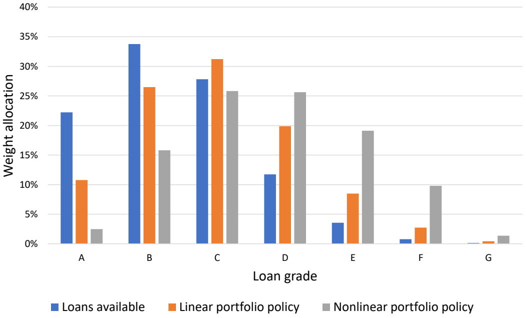


Figure: Average Composition of Loan Portfolios

STEVENS INSTITUTE OF TECHNOLOGY

Bias in Interest Rate Assignment

- African American borrowers tend to receive higher interest rates than their risk levels
- No evidence for bias against female borrowers

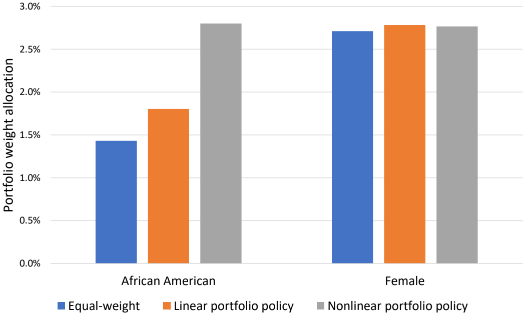


Figure: Portfolio Allocation to Borrower Groups

Investment Opportunity Set

- Comparison with stocks, bonds, and real estate
  - S&P 500 Index, Bloomberg U.S. Aggregate Bond Index, S&P 1-3/3-5/10-20 Year U.S. Treasury Bond Indexes, and MSCI U.S. REIT Index

- Public market equivalent (PME) based on portfolio cash flows

$$PME = \frac{IRR_{\text{loan}}}{IRR_{\text{index}}} \tag{7}$$

STEVENS INSTITUTE OF TECHNOLOGY

Index Portfolio and PME

- Loan portfolio: A cash outflow at the beginning (P_0) and monthly cash inflows from loan repayment ($P_t, t \in \{1, 2, \ldots, T\}$)
- Index portfolio: Given the cash flows of a loan portfolio, the index portfolio is constructed to mirror the cash flows of the loan portfolio, using equivalent investments in the index
- Public market equivalent: $PME = \frac{IRR_{\text{loan}}}{IRR_{\text{index}}}$

Example: S&P 500 Index

Suppose an investor forms a $100,000 portfolio of loans. A hypothetical investment agent will buy or sell the same amount of the S&P 500 as the investor lends or receives loan payments. In the last month, the agent will sell all the shares left or buy the shares she shorted.

Time	0	1	2	...	34	35	36
Loan	-$100,000	P_1	P_2	...	P_{34}	P_{35}	P_{36}
S&P 500	-$100,000	P_1	P_2	...	P_{34}	P_{35}	P'_{36}

Online Loans Versus Stocks, Bonds, and Real Estate

- High rates of return and low correlations with other asset classes

Table: Online Loans Versus Stocks, Bonds, and Real Estate

Loan strategy	PME	Corr.	Pr(win)	PME	Corr.	Pr(win)
	A: S&P 500 Index			D: 3-5 Year T-Bond Index		
Equal-weight	0.93	-0.32	44%	3.38	0.49	100%
Linear	1.49	-0.12	49%	4.64	0.26	100%
Nonlinear	1.93	-0.21	66%	6.00	0.13	100%
	B: Aggregate Bond Index			E: 10-20 Year T-Bond Index		
Equal-weight	2.41	0.44	100%	1.51	0.46	100%
Linear	3.56	0.31	100%	2.14	0.58	100%
Nonlinear	4.84	0.14	100%	2.81	0.24	100%
	C: 1-3 Year T-Bond Index			F: REIT Index		
Equal-weight	8.79	0.44	100%	2.18	0.48	76%
Linear	13.47	0.26	100%	2.49	0.46	88%
Nonlinear	17.99	0.08	100%	2.82	0.18	95%

STEVENS
INSTITUTE OF TECHNOLOGY

Conclusion

- GCPP is a novel portfolio optimization framework
  - The nonlinear GCPP outperforms various benchmarks and alternatives from the literature
  - Applicable to other investment, such as stocks
- Online loans can expand the investor's opportunity set
  - Competitive rates of return compared to stocks, bonds, and real estate
  - Limited comovement offers diversification benefits
- Takeaways
  - Mismatch between interest rate and risk
  - Potential bias against African American borrowers

STEVENS INSTITUTE OF TECHNOLOGY

Appendix

Loan Characteristics [Main]

Feature Name	Description
id	A unique LC assigned ID for the loan listing
title	The loan title provided by the borrower
grade	LC assigned loan grade
installment	The monthly payment owed by the borrower if the loan originates
int_rate	Interest Rate on the loan
issue_d	The month which the loan was funded
loan_amnt	The listed amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this value
funded_amnt	The total amount committed to that loan at that point in time
loan_status	Current status of the loan
purpose	A category provided by the borrower for the loan request.
collection_recovery_fee	Post charge off collection fee
sub_grade	LC assigned loan subgrade
term	The number of payments on the loan. Values are in months and can be either 36 or 60
total_pymnt	Payments received to date for total amount funded
total_rec_int	Interest received to date
total_rec_late_fee	Late fees received to date
total_rec_prncp	Principal received to date
recoveries	post charge off gross recovery
last_pymnt_d	Last month payment was received
verification_status	Indicates if income was verified by LC, not verified, or if the income source was verified

STEVENS INSTITUTE OF TECHNOLOGY

Borrower Characteristics (1) [Main]

Feature Name	Description
emp_title	The job title supplied by the borrower when applying for the loan
emp_length	Employment length in years
home_ownership	The home ownership status provided by the borrower during registration or obtained from the credit report
annual_inc	The self-reported annual income provided by the borrower during registration
zip_code	The first 3 numbers of the zip code provided by the borrower in the loan application
addr_state	The state provided by the borrower in the loan application
dti	A ratio calculated using the borrower's total monthly debt payments on the total debt obligations, excluding mortgage and the requested LC loan, divided by the borrower's self-reported monthly income
delinq_2yrs	The number of 30+ days past-due incidences of delinquency in the borrower's credit file for the past 2 years
earliest_cr_line	The month the borrower's earliest reported credit line was opened
fico_range_low	The lower boundary range the borrower's FICO at loan origination belongs to
fico_range_high	The upper boundary range the borrower's FICO at loan origination belongs to
inq_last_6mths	The number of inquiries in past 6 months (excluding auto and mortgage inquiries)
mths_since_last_delinq	The number of months since the borrower's last delinquency
mths_since_last_record	The number of months since the last public record
open_acc	The number of open credit lines in the borrower's credit file
pub_rec	Number of derogatory public records
revol_bal	Total credit revolving balance
revol_util	Revolving line utilization rate, or the amount of credit the borrower is using relative to all available revolving credit
total_acc	The total number of credit lines currently in the borrower's credit file
collections_12_mths_ex_med	Number of collections in 12 months excluding medical collections
mths_since_last_major_derog	Months since most recent 90-day or worse rating
acc_now_delinq	The number of accounts on which the borrower is now delinquent.

Borrower Characteristics (2) [Main]

Feature Name	Description
tot_coll_amt	Total collection amounts ever owed
tot_cur_bal	Total current balance of all accounts
mths_since_rcnt_il	Months since most recent installment accounts opened
total_rev_hi_lim	Total revolving high credit/credit limit
avg_cur_bal	Average current balance of all accounts
bc_open_to_buy	Total open to buy on revolving bankcards
bc_util	Ratio of total current balance to high credit/credit limit for all bankcard accounts
chargeoff_within_12_mths	Number of charge-offs within 12 months
delinq_amnt	The past-due amount owed for the accounts on which the borrower is now delinquent
mo_sin_old_il_acct	Months since oldest bank installment account opened
mo_sin_old_rev_tl_op	Months since oldest revolving account opened
mo_sin_rcnt_rev_tl_op	Months since most recent revolving account opened
mo_sin_rcnt_tl	Months since most recent account opened
mort_acc	Number of mortgage accounts
mths_since_recent_bc	Months since most recent bankcard account opened
mths_since_recent_bc_dlq	Months since most recent bankcard delinquency
mths_since_recent_inq	Months since most recent inquiry
mths_since_recent_revol_delinq	Months since most recent revolving delinquency
num_accts_ever_120_pd	Number of accounts ever 120 or more days past due
num_actv_bc_tl	Number of currently active bankcard accounts
num_actv_rev_tl	Number of currently active revolving trades
num_bc_sats	Number of satisfactory bankcard accounts

Borrower Characteristics (3) [Main]

Feature Name	Description
num_bc_tl	Number of bankcard accounts
num_il_tl	Number of installment accounts
num_op_rev_tl	Number of open revolving accounts
num_rev_accts	Number of revolving accounts
num_rev_tl_bal_gt_0	Number of revolving trades with balance greater than 0
num_sats	Number of satisfactory accounts
num_tl_120dpd_2m	Number of accounts currently 120 days past due (updated in past 2 months)
num_tl_30dpd	Number of accounts currently 30 days past due (updated in past 2 months)
num_tl_90g_dpd_24m	Number of accounts 90 or more days past due in last 24 months
num_tl_op_past_12m	Number of accounts opened in past 12 months
pct_tl_nvr_dlq	Percent of trades never delinquent
percent_bc_gt_75	Percentage of all bankcard accounts greater than 75% of limit
pub_rec_bankruptcies	Number of public record bankruptcies
tax_liens	Number of tax liens
tot_hi_cred_lim	Total high credit/credit limit
total_bal_ex_mort	Total credit balance excluding mortgage
total_bc_limit	Total bankcard high credit/credit limit
total_il_high_credit_limit	Total installment high credit/credit limit
hardship_flag	Flags whether or not the borrower is on a hardship plan

STEVENS
INSTITUTE OF TECHNOLOGY

Practical Portfolio Constraints

- Investment-amount constraints on individual loans
  - Minimum amount of \$25 on LendingClub (Why \$25?)
  - Requested loan amount
- Binary search method for a constrained nonlinear portfolio
  - Investment amount is truncated
  - The truncated investment amount is a monotonically increasing function [Binary Search]
- Portfolio performance of varying portfolio investment amounts from \$1,000 to \$100 million

STEVENS INSTITUTE OF TECHNOLOGY

Binary Search Method [Main]

- With an initial amount \$M, the truncated portfolio weight is

$$\tilde{\omega}(M) = \begin{cases} 0 & \text{if } \omega_i * M < 25 \\ U_i/M & \text{if } \omega_i * M > U_i \\ w_i & \text{otherwise} \end{cases} \tag{8}$$

  where ω_i is the portfolio weight given by the nonlinear portfolio and U_i is the loan amount

- The investment amount of the truncated portfolio

$$f(M) = \left(\sum_{i=1}^{N} \tilde{\omega}_i(M) \right) \times M \tag{9}$$

  - $f(M)$ is a monotone increasing function of M

- Suppose an investor wants to invest in $\$F$, we propose a binary search method to search for M such that $|f(M) - F| < 25$

Portfolio Performance with Constraints

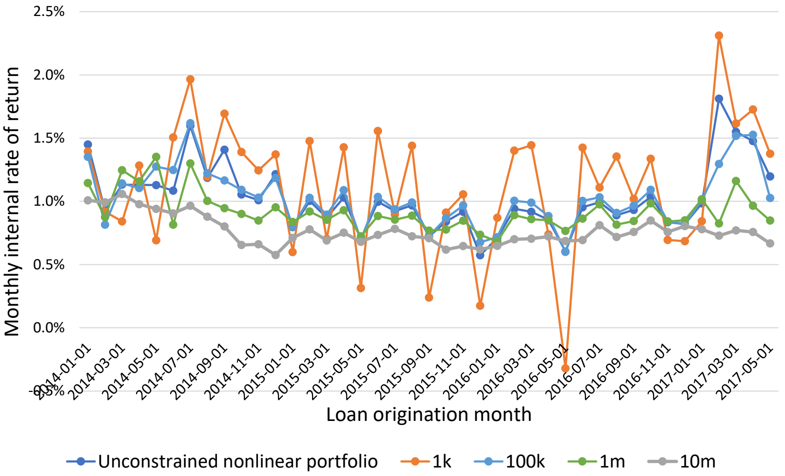
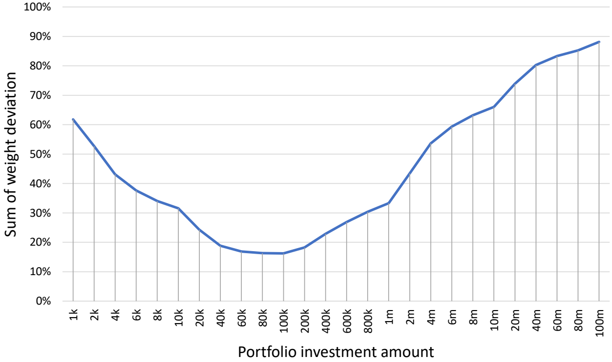


Figure: Investment-Amount Constrained Portfolio Performance

STEVENS INSTITUTE OF TECHNOLOGY

Sum of Weight Deviation

- Sum of Weight Deviation $= \frac{1}{2}\sum_{i=1}^{N} |\omega_i - \omega_{i,b}|$
- The smallest and largest investment amount both have marked weight deviation from the unconstrained portfolio
- \$100,000 portfolio deviates the least



Constrained Nonlinear Portfolio

- The constrained nonlinear portfolios have competitive performance with the benchmark unconstrained nonlinear portfolio
- The nonlinear policy takes larger positions, but not extreme positions
- The nonlinear policy has lower charged-off risk than the market

Table: Summary Statistics for Investment-amount Constrained Portfolios

	Utility	ROI	SD - ROI	IRR	SD - IRR	# funded	$\max \omega_i$	Charged-off rate
Equal weight	-0.9546	4.78%	1.40%	0.53%	0.10%	21384	0.01%	14.78%
Benchmark	-0.8938	12.01%	3.90%	1.03%	0.27%	21384	5.63%	13.13%
1k	-0.8833	13.66%	7.17%	1.12%	0.51%	19	13.86%	11.94%
10k	-0.8867	12.94%	4.46%	1.09%	0.31%	109	7.52%	12.70%
20k	-0.8875	12.82%	4.13%	1.08%	0.29%	174	6.85%	12.69%
40k	-0.8896	12.53%	3.78%	1.06%	0.26%	275	6.01%	12.95%
60k	-0.8912	12.32%	3.57%	1.05%	0.25%	358	5.38%	13.13%
80k	-0.8923	12.17%	3.45%	1.04%	0.24%	427	4.90%	13.27%
100k	-0.8933	12.04%	3.37%	1.03%	0.23%	493	4.59%	13.36%
200k	-0.8977	11.48%	3.04%	0.99%	0.20%	748	3.67%	13.83%
400k	-0.9012	11.03%	2.75%	0.96%	0.19%	1125	3.05%	14.14%
600k	-0.9029	10.81%	2.53%	0.95%	0.17%	1426	2.51%	14.07%
800k	-0.9047	10.58%	2.34%	0.93%	0.16%	1688	2.11%	14.19%
1m	-0.9065	10.35%	2.20%	0.92%	0.15%	1923	1.91%	14.37%
10m	-0.9252	8.10%	1.64%	0.77%	0.12%	7200	0.35%	15.48%
100m	-0.9455	5.78%	1.35%	0.60%	0.10%	20487	0.04%	15.47%

STEVENS INSTITUTE OF TECHNOLOGY

Minimum-Investment Amount: $25?

- Fix the investment amount as $5,000
- Vary the minimum amount per loan from $0 to $200
- Investors can benefit from reduced minimum investment limits on individual loans

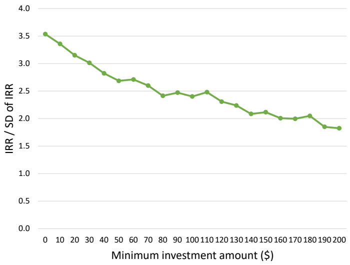


Figure: Mean & SD

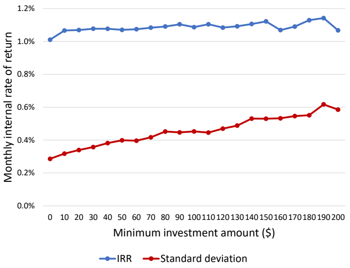


Figure: Mean / SD

STEVENS INSTITUTE OF TECHNOLOGY

Extension on GCPP Framework

- Credit risk assessment model: XGBoost
- F1 score: 0.344; AUC: 0.685
- Market (in)efficiency

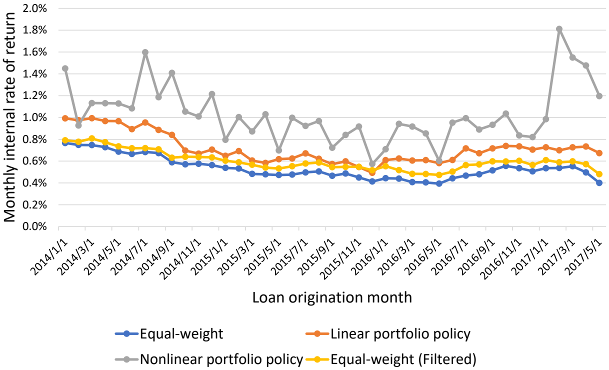


Figure: Credit Risk Filtering Approach

GCPP with Charged-off Prediction

- Embedding the charged-off prediction improves the portfolio performance
- XGBoost can capture information not learned by the shallow neural network

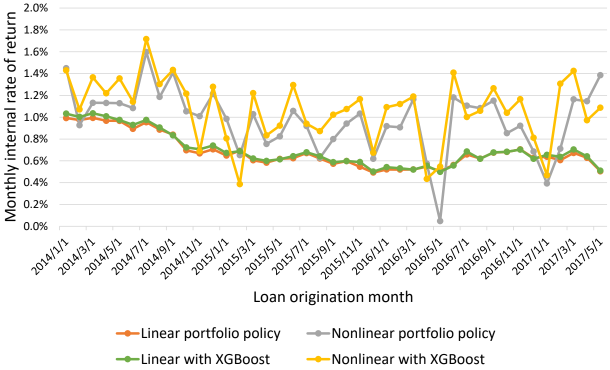


Figure: GCPP with Charged-off Prediction

STEVENS INSTITUTE OF TECHNOLOGY

References I

Brandt, M. W., Santa-Clara, P., and Valkanov, R. (2009).
Parametric portfolio policies: Exploiting characteristics in the cross-section of equity returns.
Review of Financial Studies, 22(9):3411–3447.

DeMiguel, V., Garlappi, L., and Uppal, R. (2009).
Optimal versus naive diversification: How inefficient is the 1/N portfolio strategy?
Review of Financial Studies, 22(5):1915–1953.

Emekter, R., Tu, Y., Jirasakuldech, B., and Lu, M. (2015).
Evaluating credit risk and loan performance in online Peer-to-Peer (P2P) lending.
Applied Economics, 47(1):54–70.

References II

Fu, R., Huang, Y., and Singh, P. V. (2021).
Crowds, lending, machine, and bias.
Information Systems Research, 32(1):72–92.

Gu, S., Kelly, B., and Xiu, D. (2020).
Empirical asset pricing via machine learning.
Review of Financial Studies, 33(5):2223–2273.

Guo, Y., Zhou, W., Luo, C., Liu, C., and Xiong, H. (2016).
Instance-based credit risk assessment for investment decisions in P2P lending.
European Journal of Operational Research, 249(2):417–426.

Iyer, R., Khwaja, A. I., Luttmer, E. F., and Shue, K. (2016).
Screening peers softly: Inferring the quality of small borrowers.
Management Science, 62(6):1554–1577.

References III

Michaud, R. O. (1989).
The Markowitz optimization enigma: Is "optimized" optimal?
Financial Analysts Journal, 45(1):31–42.

STEVENS
INSTITUTE OF TECHNOLOGY