Understanding Credit Pricing in Peer-to-Peer Lending

DATASCI 203 | Section 9 | Team 4

Austin Lee, Bianka Paul, River Jiang, Krishna Tummalapalli


The credit dilemma of
Mr. Kweelly

Who is Mr. Kweelly

- A recent graduate who recently got his first job
- Does not have a lot of disposable income
- Has education loans and multiple credit cards debts from too much partying during college days
- Believes in good financial planning
- Wants to consolidate debt
- Lives in a rental apartment
- Struggling but haven’t missed any debt payments in the last 6 months
- Looking for a loan at a low interest rate
- Has friends pursuing MIDS and hears crazy things about DATASCI203

Introduction

Context

- Peer-to-peer lending allows individuals to bypass traditional financial institutions
- Lending Club: One of the largest P2P platforms in the U.S. founded in 2007
- Platform evaluates applications, assigns risk grades, and presents to lenders

Research Question: What financial factors influence peer-to-peer credit pricing?

Why This Matters

- **For borrowers:** Understand which factors impact rates
- **For lenders:** Improve investment decisions
- **For marketplaces:** Design fair and efficient pricing models

Data & Methodology

Data Source

- Lending Club Loan Data (2007-2015)
- 2,260,668 observations with 145 variables per loan
- Training dataset: 1,579,978 loans (70%)
- Validation dataset: 677,130 loans (30%)

Methodology

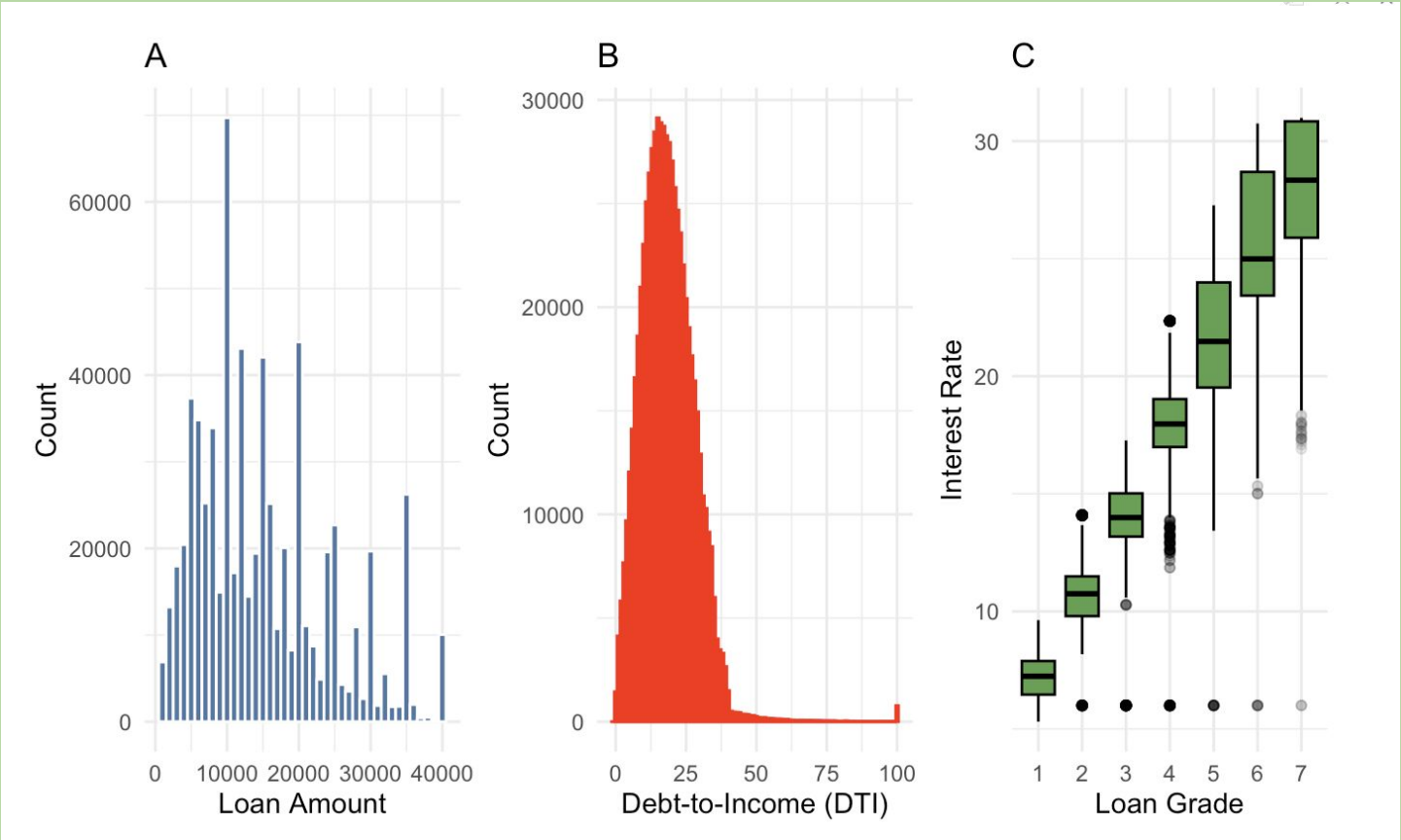
- **Dependent variable:** Interest rate at loan origination
- **Approach:** Ordinary Least Squares (OLS) regression models
    1. Model using Lending Club's credit grade
    2. Models using specific borrower and loan variables
- Goal: Unpack individual contributions of key variables

Data Preparation & Sample Characteristics

Data Wrangling

- Selected 10 relevant variables for credit pricing:
  a. Loan details (**interest rate**, **loan amount**, **repayment period**, **loan purpose,** and **loan grade**)
  b. Borrower financials (**annual income**, **debt-to-income ratio,** and **home ownership status**)
  c. Borrower credit history (**delinquencies in the past two years**, **recent credit inquiries**, and **number of open accounts**)
- Capped extreme values (**annual income** $0-$1M, **DTI**, **delinquencies**, **inquiries**, and **open accounts**)
- Created indicator variables for categorical variables (**home ownership**, and **loan purpose**)

Baseline Characteristics



Our Models : Key Drivers of Interest Rate

Model 1 (Interest Rate ~ Grade)

Adjusted R^2 = 0.91 (Grade alone explains 91% variance)
Intercept = 3.35 (starting rate)
Residual = 1.466 (relatively small)
Grade Coefficient = +3.66 (A to G)
p-value = 2.2e-16 (essentially zero)

Borrower's Grade, assigned by Lending Club, is the strongest predictor of interest rate.

Model 2 (Interest Rate ~ 3 variables)

Adjusted R^2 = 0.16 (explains 16% variance)
Intercept = 11 (not interpretable alone)
Residual = ± 4.1%
Key Predictors:
Term (60m) → 4.09% ▲ from 36m
Debt-to-Income → 0.75% ▲ per +10 DTI
Loan Amount → 0.32% ▼ per +$10K
p-value = 2.2e-16 (essentially zero)

Longer term = higher Rate
More Debt = riskier = more expensive

Model 3 (Interest Rate ~ 10 variables)

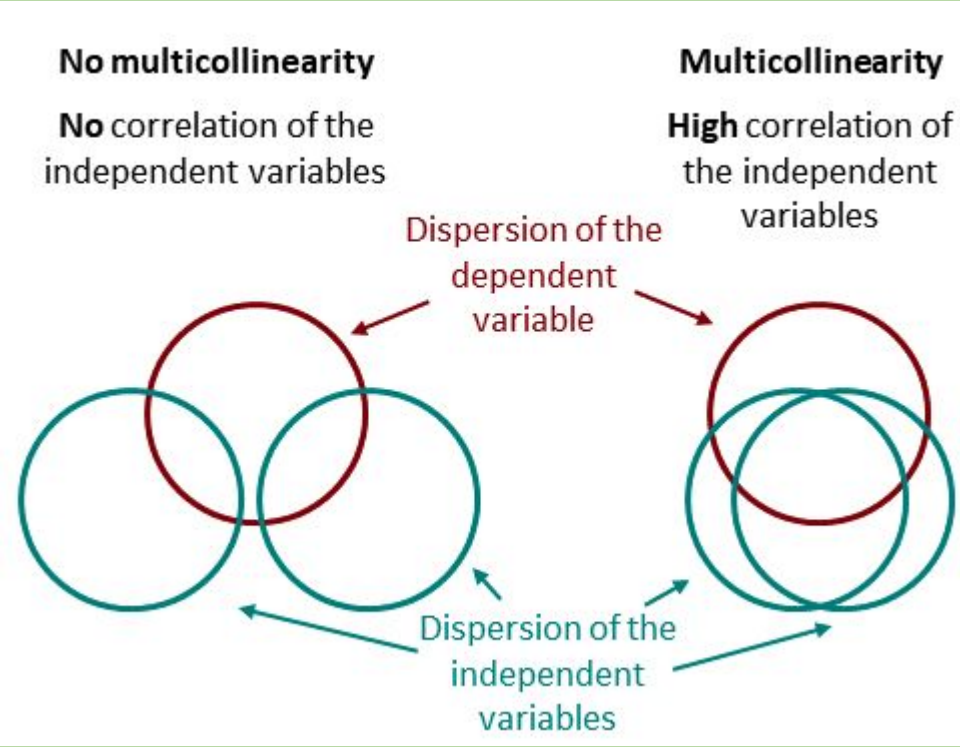
Adjusted R^2 = 0.24 (explains 24% variance)
Intercept = 11.57
Residual = ± 4.2%
Strong Predictors:
Term (60m) → 4.02% ▲
Delinquencies (6m) → 1.14% ▲
Renting vs. Own → 0.61% ▲
Mortgage vs. Own → 0.52% ▼
Loan Purpose → 0.95% ▼
Other Predictors: Loan Amount, Annual Income, Delinquency (2yrs), Open Credit Lines
p-value = 2.2e-16 (essentially zero)
Credit term & loan purpose impact rates. Renters pay more; delinquencies hurt.

Comparing Model 2 and Model 3 (ANOVA Test)

F-Statistic = 9451.6 (strong effect of additional model 3 predictors)
Residual Sum Squares **(RSS)** difference = 1,176,937 less for model 3 indicating better fit
p-value = 2.2e-16 (essentially zero)

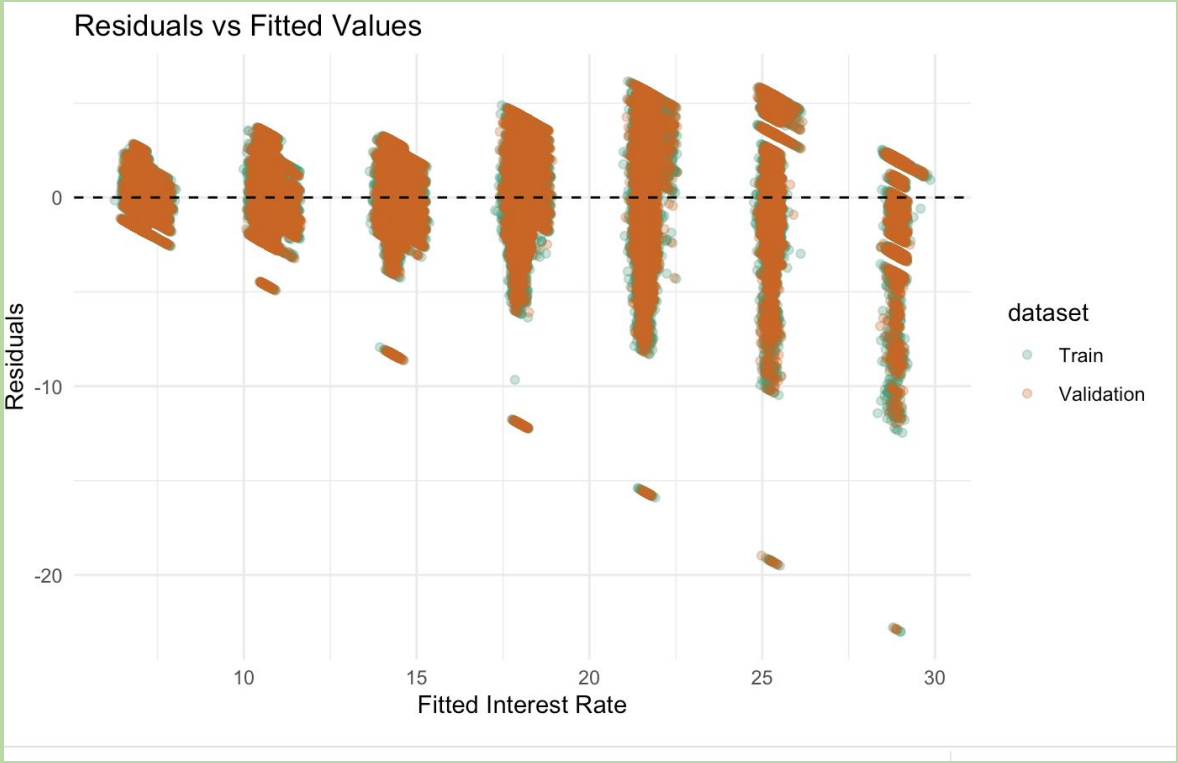
Model Assumptions - IID

Model Assumptions - No Perfect Collinearity



Variable	VIF
Loan Amount	1.53
60-Month Term	1.2
Annual Income	1.42
Debt-to-Income Ratio	1.21
Delinquencies (2 yrs)	1.01
Inquiries (6 mos)	1.03
Open Credit Lines	1.21
Home: Mortgage	2.78
Home: Rent	2.76
Purpose: Debt Consolidation	1.05

Model Assumptions -
Linear Expectation Condition + Practical Implications



Conclusion

Lending Club

- An internally assigned credit grade looks to be their strongest driver of interest rate
- No transparency to borrowers or lenders on what constitutes credit grade
- Peer to peer pricing seems to be more complex to understand than it appears

Our Recommendation for Mr. Kweelly

- Consider a 36-term loan
- Continue to be a renter
- Do not close open credit lines but keep the usage low
- Consider taking your MIDS friends and their professor out for a treat :)

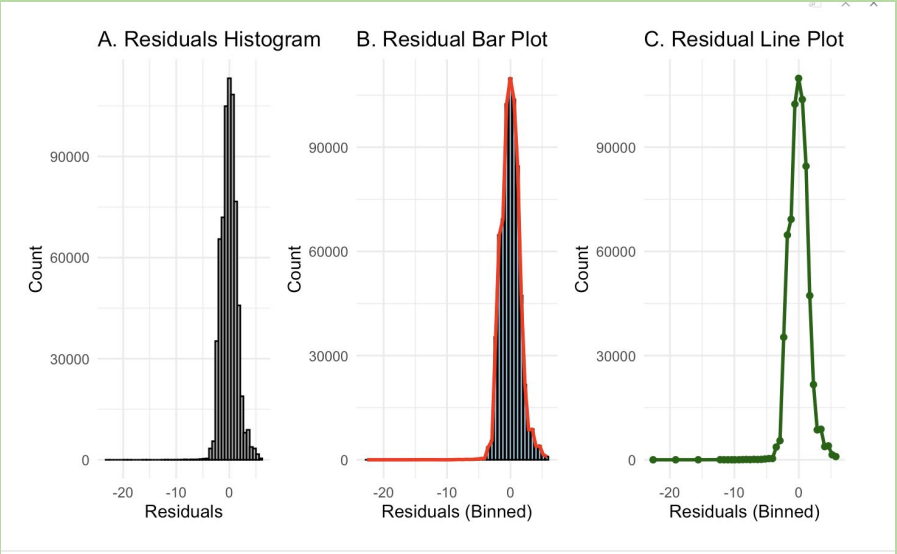
Thank You / Questions

Appendix

Additional Model Specifications and Testing:

- We also tested several univariate regressions (e.g., int_rate ~ loan_amnt, term (60 months) , annual_inc, debt-to-income, delinquencies (2yrs), open_accounts, etc.) to further assess how individual financial features do impact peer to peer interest rate.

  - Each of the predictors demonstrated some level of association related to the interest rate, particularly:
    1. Debt-to-income
    2. Loan term
    3. Delinquencies and recent credit histories



Links:

Reference: Dutt, Aditya, and Vishal Gupta. 2021. *Understanding Loan Pricing on Peer-to-Peer Lending Platforms*. SSRN. https://doi.org/10.2139/ssrn.3785323

Data Source: https://www.kaggle.com/datasets/adarshsng/lending-club-loan-data-csv/data

Model Output

Table 3: Linear Regression Models Predicting Loan Interest Rate

	Dependent variable:		
	Interest Rate (%)		
	Grade Only	3 Predictors	10 Predictors
	(1)	(2)	(3)
Grade (Numeric)	3.658***		
	(0.001)		
Loan Amount		−0.00003***	0.00002***
		(0.00000)	(0.00000)
Annual Income			−0.000008***
			(0.006)
Debt-to-Income		0.075***	0.083***
		(0.001)	(0.00000)
Delinquencies (2 yrs)			0.434***
			(0.001)
Inquiries (6 mo)			1.135***
			(0.006)
Open Accounts			−0.071***
			(0.006)
60-Month Term		4.089***	4.022***
		(0.013)	(0.001)
Home: Mortgage			−0.515***
			(0.017)
Home: Rent			0.614***
			(0.017)
Purpose: Debt Consolidation			−0.951***
			(0.013)
Constant	3.346***	11.000***	11.567***
	(0.004)	(0.014)	(0.023)
Observations	677,130	677,130	677,130
R^2	0.908	0.163	0.238
Adjusted R^2	0.908	0.163	0.238
Residual Std. Error	1.466 (df = 677128)	4.419 (df = 677126)	4.218 (df = 677119)
F Statistic	**6,679,174.000*** (df = 1; 677128)**	**44,011.910*** (df = 3; 677126)**	**21,109.660*** (df = 10; 677119)**
Note:			*p<0.1; **p<0.05; ***p<0.01