HEC Paris — EMIF/EMBA — Spring 2022

Big Data & Finance

Johan Hombert

Data in Finance: Some Applications

Credit scoring

- Use alternative data to score borrowers, sell score to lenders

- Example:

experian

Social Media Insight™

Making business decisions with limited data is a huge risk. When it comes to new or emerging businesses with thin or no credit profiles, lenders must be armed with the right data to confidently and quickly assess a business.

As a breakthrough, alternative data source for risk assessment, Social Media Insight leverages social data to help lenders build a more complete picture of businesses with thin credit files.

For businesses that have thin credit profiles, a strong social media reputation can be a good measure of health. Social Media Insight aggregates social data that is directly sourced and turns it into predictive attributes:

- Number of reviews revealing if the volume trend indicates sales growth or decline.
- Unique business profile information, such as licensing, hours of operations, pricing levels, and more.
- In depth business description that goes beyond standard SIC or NAICs codes, providing potentially critical information on the type of goods or services supplied.

A brief history of P2P lending

PROSPER From crowdlending to marketplace lender

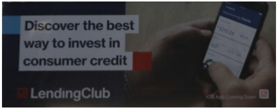
- Crowdlending: Match borrowers and lenders, terms of financing set by lenders
- Marketplace: Score borrowers, set the interest rate, match borrowers with lenders

A brief history of P2P lending

PROSPER From crowdlending to marketplace lender

- Crowdlending: Match borrowers and lenders, terms of financing set by lenders
- Marketplace: Score borrowers, set the interest rate, match borrowers with lenders

LendingClub From marketplace to (shadow) bank



LendingClub Closing Down Their Platform for Retail Investors

Peter Renton · Peer to Peer Lending · Oct. 7, 2020 · 5 min read

- (Shadow) Bank: Score borrowers, set the interest rate, lend using its balance sheet

Mini case: Credit scoring using digital footprints

- "On the Rise of FinTechs: Credit Scoring Using Digital Footprints," 2019, Berg, Burg, Gombovic and Puri, *Review of Financial Studies* [pdf]

- E-commerce company in Germany
  - Buy Now Pay Later: Give credit ⇒ Must assess buyer's creditworthiness
  - Traditional credit score based on credit history and sociodemographics
  - After Oct 2015: also collected digital footprints (OS, email, etc.)
  - Does this improve prediction of default?

Credit scoring using digital footprints

Variable	Value	Default rate (%)
Credit bureau score (by quintile)	All	0.94
	Q1 - lowest	2.12
	Q2	1.02
	Q3	0.68
	Q4	0.47
	Q5 - highest	0.39
Device	All	0.94
	Desktop	0.74
	Tablet	0.91
	Mobile	2.14
	Do-not-track setting	0.88
Operating system	All	0.94
	Windows	0.74
	iOS	1.07
	Android	1.79
	Macintosh	0.69
	Other	1.09
	Do-not-track setting	0.88
E-mail host	All	0.94
	Gmx (partly paid)	0.82
	Web (partly paid)	0.86
	T-Online (affluent customers)	0.51
	Gmail (free)	1.25
	Yahoo (free, older service)	1.96
	Hotmail (free, older service)	1.45
	Other	0.90

Buyers with lower credit score have higher default rate

Buyers who connect from a mobile have high default rate

Buyers with Android mobile have high default rate

Mac users have low default rate

T-Online email users have low default rate

Yahoo email users have high default rate

Credit scoring using digital footprints

Variable	Value	Default rate (%)
Checkout time	All	0.94
	Evening (6 p.m.-midnight)	0.85
	Night (midnight-6 a.m.)	1.97
	Morning (6 a.m.-noon)	1.09
	Afternoon (noon-6 p.m.)	0.89
Do-not-track setting	All	0.94
	No	0.94
	Yes	0.88
Name in e-mail	All	0.94
	No	1.24
	Yes	0.82
Number in e-mail	All	0.94
	No	0.84
	Yes	1.41
Is lowercase	All	0.94
	No	0.84
	Yes	2.14
E-mail error	All	0.94
	No	0.88
	Yes	5.09

Night buyers are more likely to default

A name in the email predicts low default

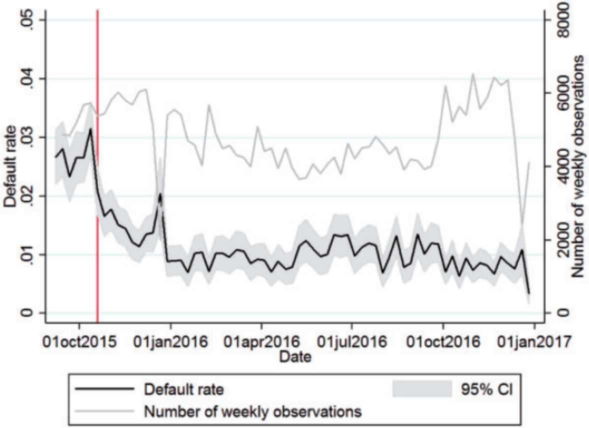
A number in the email predicts high default

An all-lowercase application form predicts high default

A typo in the email predicts high default

Credit scoring using digital footprints

- After starting to score using digital footprints, default rate decreased from 3% to 1%



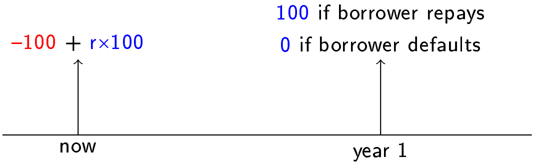
Remarks

- Big data/AI useful in many other markets
  - Credit markets: mortgages, consumer loans, business loans
  - Insurance: predict losses + monitor and influence behavior
  - Venture capital: deal sourcing
  - Fraud detection
  - etc.
- Tech firms are well positioned to enter these markets: access to consumers + lots of data on people and businesses

Fintech Simulation

Credit FinTech

- Your team runs a fintech that makes business loans
- Loan terms
  - Maturity: 1 year
  - Principal amount: **100**, paid at maturity
  - Interest rate: **r** paid upfront
  - Credit risk: borrower defaults with some probability. In case of default, recovery rate for lender is zero
  - Cash flow to lender:



Scoring

- Your company employ data scientists, who determine a credit score for each loan application based on traditional credit scoring information (financials, credit history)

  Traditional score

  A (lowest default risk)
  B
  C
  D
  E
  F (highest default risk)

- You are active in two markets: **Home** and **Foreign**

- In the Home market, your company's data scientists developed an additional score based on online media data

  Online score

  Green (lower default risk)

  Red (higher default risk)

Scoring

- Your information in each market
  - Home market: Traditional score + Online score
  - Foreign market: Traditional score
- Open the file DefaultRates.xlsx

Profit

- To keep the simulation focused on information issues, assume
  - Your operating costs: **0**
  - Your cost of capital: **0%**

NB: In practice, the fintech may be a subsidiary of a large bank that funds mostly through deposits (cost ≃0%) and some equity (ROE ≃8%)

- Profit per loan = NPV = CF at issuance + CF at maturity

NB: With cost of capital >0: $\text{NPV} = \dfrac{\text{CF at issuance}}{1+\text{cost of capital}} + \text{CF at maturity}$

- Go to the worksheet Profit in DefaultRates.xlsx

Competition

- You have 1 competitor (=1 other team) in each market: every potential borrower submits a loan application both to you and to the competing lender
- Your Home market is another team's Foreign market
  - The other team has the same traditional score as yours, but not the online score
- Your Foreign market is another team's Home market
  - The other team has the same traditional score as yours and the online score

Competition

- In your Home market

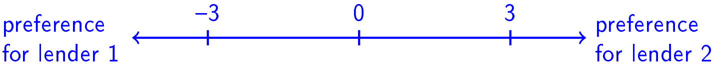
	Traditional score (A,B,C,D,E,F)	Online score (green,red)
Your company	✓	✓
Competitor	✓	

- In your Foreign market

	Traditional score (A,B,C,D,E,F)	Online score (green,red)
Your company	✓	
Competitor	✓	✓

Borrowers

- Loan applicants never accept interest rates above 20%
- Half of loan applicants have a preference tilt for one lender while the other half have a preference tilt for the other lender
- This preference can be more or less intense



- Loan applicants are distributed uniformly on the preference line between −3 and 3
- A consumer with preference for lender 2 with intensity 1.5 takes

  ... lender 2's offer if R(lender 2) < R(lender 1) + 1.5%

  ... lender 1's offer if R(lender 2) > R(lender 1) + 1.5%

Pricing

- In Foreign market: decide on interest rate offered to loan applicants in each traditional score category (A,B,C,D,E,F)
- In Home market: decide on interest rate offered to loan applicants for each combination of credit score and online score (A green, A red, B green, B red, C green, ...)
- Remarks
  - Interest rate between 0% and 20%
  - You are in competition with another team, so your loan offer may not be accepted
  - You can also decide to make no loan offer: leave the cell empty

Round 1

- Input your interest rates in Round1.xlsx in the yellow highlighted cells

- Write a short report (about ½ page) describing your approach to set interest rates in each market

- Send the excel spreadsheet and the report to hombert@hec.fr

Debriefing

- What was your team's strategy in the Home market?
- What was your team's strategy in the Foreign market?
- Do you expect to have higher market share in Home or Foreign?
- Do you expect to earn higher profit in Home or Foreign?

Winner's curse

- Consider loan applications in a given score bucket (D for example)
  - Home lender sorts out higher risk borrowers (red) from lower risk ones (green) → It charges high rate to high risk, low rate to low risk
  - Foreign lender offers the same rate to all because it does not observe the online score
- **Winner's curse:** The foreign lender "wins" when its interest rate is cheaper than the Home lender. This happens with the high risk borrowers, who generate losses to the lender
- Winning when you are less informed than competitor is a curse
- Cf. this morning's session with Jean-Edouard

Winner's curse

- Consider loan applications in a given score bucket (D for example)
  - Home lender sorts out higher risk borrowers (red) from lower risk ones (green) → It charges high rate to high risk, low rate to low risk
  - Foreign lender offers the same rate to all because it does not observe the online score
- **Winner's curse:** The foreign lender "wins" when its interest rate is cheaper than the Home lender. This happens with the high risk borrowers, who generate losses to the lender
- Winning when you are less informed than competitor is a curse
- Cf. this morning's session with Jean-Edouard
- How to avoid the winner's curse?

Round 2

- Price a new batch of loan applications using the experience from the first round

- Same market environment as before

- File Round2.xlsx

Round 2

- Input your interest rates in Round2.xlsx in the yellow highlighted cells

- Write a short report (about ½ page) describing your approach to set interest rates in each market

- Send the excel spreadsheet and the report to hombert@hec.fr

Debriefing

- How did you revise your strategy in each market?
- Do you expect to have higher market share in Home or Foreign?
- Do you expect to earn higher profit in Home or Foreign?

Data and financial inclusion

- Focus on loan applicants with traditional score F
- What would happen if no lender had developed the online score?

Data and financial inclusion

- Focus on loan applicants with traditional score F
- What would happen if no lender had developed the online score?
- More data can improve financial inclusion
  - Only traditional score: F-rated applicants don't have access to credit
  - With additional information: some of them gain access to credit

Data and discrimination

Data and discrimination

Artificial intelligence

FinTech at HEC found to discriminate against minorities

Paris JUNE 02 2022

"My friend and I have the same income and credit history but I belong to a minority group and I'm paying a higher interest rate."

Discrimination

- The law distinguishes between direct discrimination and indirect discrimination

Direct discrimination

- Direct discrimination: Decision based on a "protected characteristic" such as race, sex, ethnic origin, social origin, religion
  - Called "disparate treatment" in US law
- May happen because of prejudice
  - Ex.: Job opening for white men only
- ...or because the protected characteristic is a predictor of risk (a.k.a. statistical discrimination)
  - Ex.: Lower car insurance premium for women because they have fewer accidents
- Illegal in both cases in EU and US

Indirect discrimination

- Indirect discrimination: Decision is not based on protected characteristics but ends up being different for people with a protected characteristic
  - Called "disparate impact" in US law
- Happens when decision is based on information correlated with protected characteristics
  - Ex.: Interest rate based on borrower's job occupation may end up being different for people with different ethnic origins
- May be legal or illegal (legal if for "business necessity" in US law)

Report from your companies' data scientists

- Minority applicants more likely to have red online score
- Algorithm uses no protected characteristic
- Algorithm uses information <u>correlated</u> with protected characteristics

Ex. The online score uses whether the applicant is Android or iPhone user. iPhone users are:

20% more likely to be women

30% more likely to live in a city

60% more likely to be Black (US data)

Report from your companies' data scientists

- Focus on applicants with traditional credit score C

- Open the file Round3.xlsx

Round 3: Non-discriminatory pricing

- You know if loan applicants are minorities
- Set interest rate in Home market for applicants with traditional credit score C
  - Non-minority applicants with green online score
  - Minority applicants with green online score
  - Non-minority applicants with red online score
  - Minority applicants with red online score
- Write a short report describing your approach
- Send the excel spreadsheet and the report to hombert@hec.fr

Debriefing

- What is your team's strategy to avoid discriminatory pricing?
- Did you have to trade off different notions of fairness?
- Other ideas that could not be implemented in the simulation but could be relevant in practice?

Data and discrimination

The New York Times

Apple Card Investigated After Gender Discrimination Complaints

A prominent software developer said on Twitter that the credit card was “sexist” against women applying for credit.

Data and discrimination

Wells Fargo, Upstart criticized after study finds loan disparities

Published Feb. 6, 2020 • Updated Feb. 14, 2020

> The request comes a week after the nonprofit Student Borrower Protection Center found that an Upstart borrower who attended historically black Howard University would pay thousands of dollars more on average for a five-year loan than a borrower with an identical credit profile who studied at New York University.

Data and discrimination

Invisible Primes: Fintech Lending with Alternative Data*

Marco Di Maggio[†], Dimuthu Ratnadiwakara [‡], Don Carmichael [§]

May 28, 2022

Abstract

We exploit anonymized administrative data provided by a major fintech platform to investigate whether using alternative data to assess borrowers' creditworthiness results in broader credit access. Comparing actual outcomes of the fintech platform's model to counterfactual outcomes based on a "traditional model" used for regulatory reporting purposes, we find that the latter would result in a 70% higher probability of being rejected and higher interest rates for those approved. The borrowers most positively affected are the "invisible primes"–borrowers with low credit scores and short credit histories, but also a low propensity to default. We show that funding loans to these borrowers leads to better economic outcomes for the borrowers and higher returns for the fintech platform.

*We thank Upstart Network, Inc for providing access to its data. We also thank seminar and conference participants at the University of Chicago Booth, Yale SOM, MIT (Sloan), USC (Marshall), Colorado Leeds, University of Houston, Florida University Business School, Baruch College, Louisiana State University, UCLA (Anderson), 11th FDIC Consumer Research Symposium, and 2022 GSU-RFS Fintech Conference for helpful comments.

[†]Harvard Business School and NBER: mdimaggio@hbs.edu.

[‡]E. J. Ourso College of Business, Louisiana State University: dratnadiwakara2@lsu.edu.

[§]Upstart: don@upstart.com.

Discrimination in the US mortgage market

- "Consumer Lending Discrimination in the FinTech Era," Bartlett, Morse, Stanton and Wallace, *Journal of Financial Economics*, 2021 [pdf]
- For given borrower characteristics, Latin and African-American mortgage borrowers pay higher interest rates
  - if traditional lender: 8 basis points per year
  - if fintech lender: 5 basis points per year

Discrimination in the US mortgage market

- Half full glass
  - Less discrimination by fintech
  - Discrimination by traditional lender has decreased over time, perhaps as a result of competition from fintech
- Half empty glass
  - Algorithms still discriminate (although less so than humans)
  - Algorithms "learn" that Latin/African-American borrowers are less likely to get a good rate from a traditional lender, so they can be charged higher rates

Thank You! 🎉