HEC Paris — M1 — Spring 2022

Advanced Finance

[#3] Finance and Data (Part II: Financing and Insurance)

Johan Hombert, Daniel Schmidt

Road map

Introduction

- Prediction is key in finance
  - Corporate valuation (PE, M&A): predict cash flow
  - Credit: predict default
  - Insurance: predict damages
  - Asset management, trading: cf. previous lecture

$\Rightarrow$ Importance of data. New data/AI $\Rightarrow$ New business applications

FinTech in credit markets

- Credit score providers: use data on people and businesses to calculate credit scores sold to banks

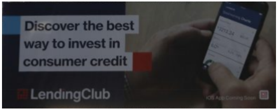
experian. "Social medial insight program that extracts data from Yelp, Facebook, Twitter, and Four Square is offered for use by private lenders and traditional banks. Credit scores for over 1 billion people & businesses, including 235 individuals US consumers and over 25 million US businesses."

ZEST AI Lenddo™ CredoLab kreditech

Also done internally by banks

FinTech in credit markets

- Online lending
  - Crowdlending: match borrowers and lenders, terms of financing set by lenders
  - Marketplace (peer-to-peer) lenders: score borrowers, set the interest rate, match borrowers with lenders
  - Full stack lenders: score borrowers, set the interest rate, make the loans (fintechs, large banks)



LendingClub Closing Down Their Platform for Retail Investors

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

Kabbage "A business could qualify for a higher line of credit if their social media interactions show (...) customers have nice things to say about them."

Big tech in credit markets

- Tech giants are well positioned in credit markets: they have access to consumers + have lots of data on them

THE WALL STREET JOURNAL.

TECH

Goldman Sachs, Apple Team Up on New Credit Card

Card would carry the Apple Pay brand and could launch early next year

Apple buys UK fintech start-up Credit Kudos

Purchase suggests US tech giant will launch greater push into credit services

Tim Bradshaw and **Siddharth Venkataramakrishnan** in London MARCH 23 2022

InsurTech

- Creation of insurance products, scoring with AI

Data machine: the insurers using AI to reshape the industry

Groups are building detailed customer profiles to inform pricing and try to influence behaviour

AI allows insurers such as Ping An to produce highly individualised profiles of customer risk that evolve in real time © FT montage; Alamy, Dreamstime

Ian Smith in London AUGUST 16 2021 50

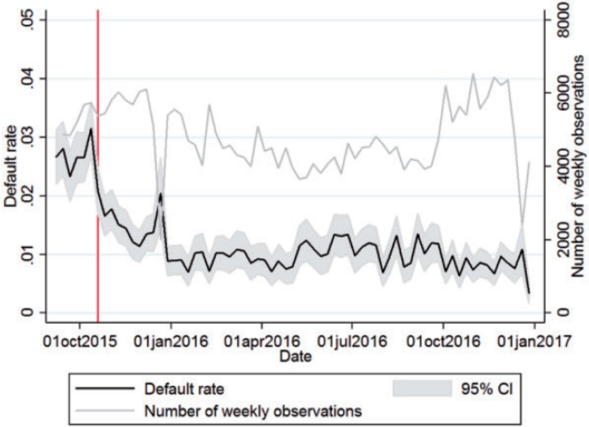
Case study

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

- Scoring with digital footprints at an E-commerce company
  - Goods sent first, paid for later → need to assess buyer's creditworthiness
  - Credit score based on credit history, sociodemographics, past transactions, etc.
  - After Oct 2015: also collected digital footprints (OS, email, etc.)
  - Does this improve prediction of default?

Default regressions (scorable customers)

Variables	(1) Credit bureau bureau score Coef.	z-stat	(2) Digital footprint Coef.	z-stat	(3) Credit bureau score & digital footprint Coef.	z-stat
Credit bureau score	−0.17***	(−7.89)			−0.15***	(−6.67)
Device type & operating system[a]						
Desktop/Windows			Baseline		Baseline	
Desktop/Macintosh			−0.07	(−0.53)	−0.13	(−1.03)
Tablet/Android			0.29***	(3.19)	0.29***	(3.06)
Tablet/iOS			0.08	(1.05)	0.08	(0.97)
Mobile/Android			1.05***	(17.25)	0.95***	(15.34)
Mobile/iOS			0.72***	(9.07)	0.57***	(6.73)
E-mail Host [a]						
Gmx (partly paid)			Baseline		Baseline	
Web (partly paid)			0.00	(0.00)	−0.02	(−0.22)
T-Online (affluent customers)			−0.40***	(−3.90)	−0.35***	(−3.35)
Gmail (free)			0.34***	(3.81)	0.29***	(3.09)
Yahoo (free, older service)			0.75***	(9.19)	0.72***	(8.98)
Hotmail (free, older service)			0.35***	(3.70)	0.28***	(2.72)
Channel						
Paid			Baseline		Baseline	
Affiliate			−0.49***	(−5.35)	−0.54***	(−5.58)
Direct			−0.27***	(−4.25)	−0.28***	(−4.44)
Organic			−0.15*	(−1.79)	−0.15*	(−1.74)
Other			−0.47***	(−4.50)	−0.48***	(−4.36)
Checkout time						
Evening (6 p.m.-midnight)			Baseline		Baseline	
Morning (6 a.m.-noon)			0.28***	(4.50)	0.28***	(4.60)
Afternoon (noon-6 p.m.)			0.08	(1.42)	0.08	(1.47)
Night (midnight-6 a.m.)			0.79***	(7.73)	0.75***	(7.09)
Do-not-track setting			−0.02	(−0.25)	−0.07	(−0.91)
Name in e-mail			−0.28***	(−5.67)	−0.29***	(−5.70)
Number in e-mail			0.26***	(4.50)	0.23***	(3.91)
Is lowercase			0.76***	(13.10)	0.74***	(13.20)
E-mail error			1.66***	(20.00)	1.67***	(20.36)
Constant	12.42***	(5.76)	−4.92***	(−62.87)	9.97***	(4.48)
Control for *Age*, *Gender*, *Item category*, *Loan amount*, and month and region fixed effects	No		No		No	
Observations	254,819		254,819		254,819	
Pseudo R^2	**.0244**		**.0524**		**.0717**	
AUC	**0.683**		**0.696**		**0.736**	
(SE)	**(0.006)**		**(0.006)**		**(0.005)**	
Difference to AUC=50%	**0.183*****		**0.196*****		**0.236*****	
Difference AUC to (1)			**0.013***		**0.053*****	


Default rate
.05
.04
.03
.02
.01
0
Number of weekly observations
8000
6000
4000
2000
0
01oct2015
01jan2016
01apr2016
01jul2016
01oct2016
01jan2017
Date
Default rate
95% CI
Number of weekly observations

New data and AI in finance

- Many opportunities but also pitfalls to avoid for fintech entrepreneurs and for society

1. Financial inclusion
2. Winner's curve
3. Lucas critique
4. Privacy
5. Hirshleifer effect
6. Discrimination

Road map

Introduction

Financial inclusion

Winner's curse

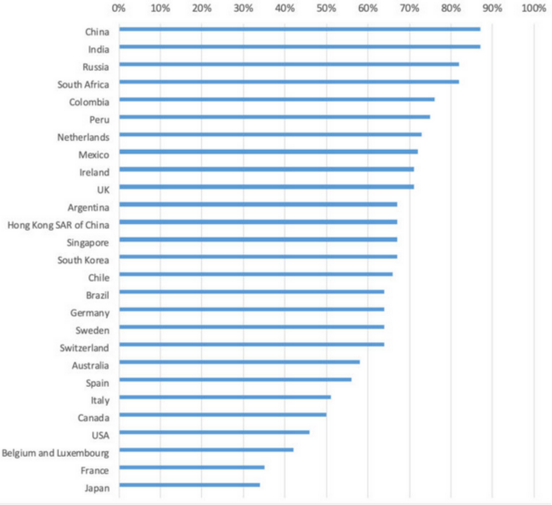
Lucas critique

Hirshleifer effect

Discrimination

Fintech and financial inclusion

Fintech adopters in percentage of digitally active population (source: EY)



Digital banking + Add to myFT

Banks use fintech to make up for lost time on financial inclusion

Institutions are investing in a bid to reach 'unbanked' groups in Africa and the Middle East

Laura Noonan APRIL 24 2019

Fintech + Add to myFT

How developing nations use tech to reach the 'underbanked'

Lenders in Africa and the Middle East circumvent weak digital infrastructure to make progress

Sarah Murray APRIL 24 2019

Fintech and financial inclusion

- Can fintech lenders improve access to credit?

- Fintech lenders' competitive edge over traditional banks:

  A. Lower fixed cost of underwriting loans

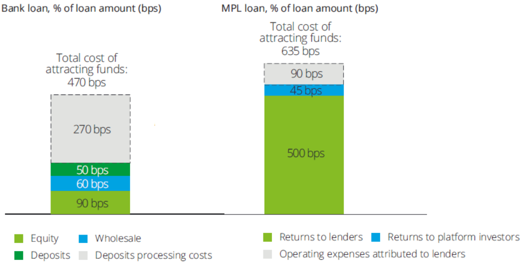
  B. More (or different) information on borrowers

A. Cost structure

A. Fintech lender has:

- lower fixed cost of underwriting a loan
  - automatized process
  - no physical premises, no loan officer
- higher marginal cost of underwriting a loan

Figure 8. Costs of funding an unsecured personal loan: banks and MPLs



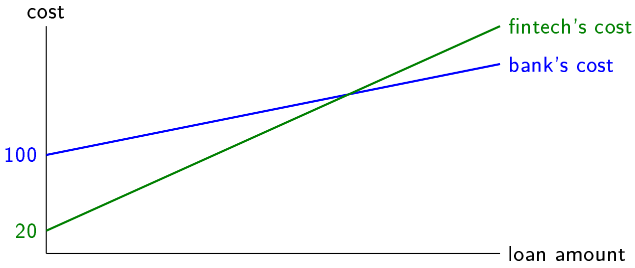
Source: Deloitte analysis

*MPL = Market-Place Lender

A. Cost structure

Bank: underwriting a X € loan costs a fixed 100€ plus 4% of X

Fintech: underwriting a X € loan costs a fixed 20€ plus 6% of X

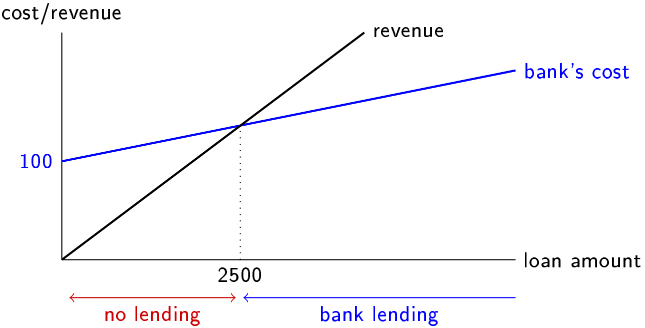


A. Cost structure

- Suppose different borrowers want different loan amounts. All are willing to pay a 8% interest rate

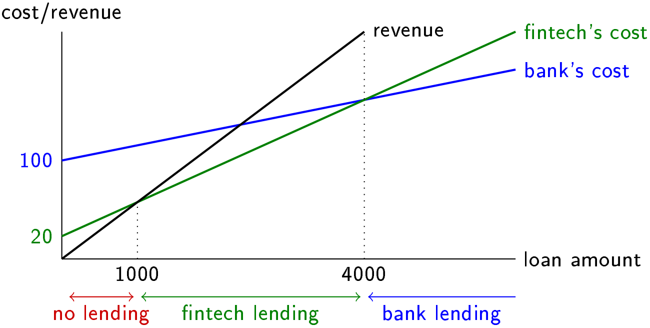
- Before fintech entry: the bank lends if

  $0.08X > 0.04X + 100 \Leftrightarrow$ loan size $X > 2500$ €



A. Cost structure

- After fintech entry: What happens?



1. Small borrowers gain access to credit from fintech lender
2. Fintech lender gains market shares on the intermediate segment
3. Traditional lenders still dominate the large loan segment

B. Data

B. Fintech lenders have alternative data on borrowers

- For ex. digital footprints

- Example
  - Loan amount 100 with interest rate R
  - If borrower repays: $100 \times (1 + R)$ to the lender
  - If borrower defaults: 0 to the lender
  - Lender's funding cost + fixed cost of loan underwriting: 0

Please take a piece of paper and calculate the interest rate at which the lender breaks even if the probability of default is 10%

9.9% 10.0% or 11.1% ?

B. Data

B. Fintech lenders have alternative data on borrowers

- For ex. digital footprints

- Example
  - Loan amount 100 with interest rate R
  - If borrower repays: $100 \times (1 + R)$ to the lender
  - If borrower defaults: 0 to the lender
  - Lender's funding cost + fixed cost of loan underwriting: 0

Please take a piece of paper and calculate the interest rate at which the lender breaks even if the probability of default is 10%

$$-100 + 0.90 \times 100 \times (1 + R) \geq 0 \quad \Rightarrow \quad R \geq \frac{0.10}{0.90} = 11.1\%$$

B. New data

- Consider small entrepreneurs with no credit history or financial statements
  - Ex.: new businesses, informal businesses
  - Entrepreneurs are willing to pay a 8% interest rate
  - Traditional lenders' best estimate of Probability of Default (PD) is 20%

Do traditional lenders lend to these entrepreneurs?

Yes No ?

B. New data

- Consider small entrepreneurs with no credit history or financial statements
  - Ex.: new businesses, informal businesses
  - Entrepreneurs are willing to pay a 8% interest rate
  - Traditional lenders' best estimate of Probability of Default (PD) is 20%

Do traditional lenders lend to these entrepreneurs?

No: lender breaks even only if $R \geq \frac{0.20}{0.80} = 25\%$ but the entrepreneur takes the loan only if $R \leq 8\%$

What is the maximum PD above which traditional lenders stop lending?

7.4%   8.4%   ?

B. New data

- Consider small entrepreneurs with no credit history or financial statements
  - Ex.: new businesses, informal businesses
  - Entrepreneurs are willing to pay a 8% interest rate
  - Traditional lenders' best estimate of Probability of Default (PD) is 20%

Do traditional lenders lend to these entrepreneurs?

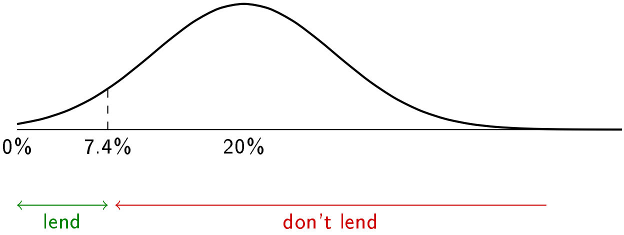
No: lender breaks even only if $R \geq \frac{0.20}{0.80} = 25\%$ but the entrepreneur takes the loan only if $R \leq 8\%$

What is the maximum PD above which traditional lenders stop lending?

Lending can happen only if $\frac{PD}{1-PD} \leq 8\%$ i.e. $PD \leq 7.4\%$

B. New data

- Fintech lenders have access to entrepreneurs' digital footprints
  - They can better identify talented entrepreneurs and estimate a more precise PD for each entrepreneur
  - Distribution of fintech's best estimate of PD



B. New data

- Implications of more data

    1. Talented entrepreneurs with no credit history gain financing

    2. Business opportunity for fintech: can charge 8% to entrepreneurs with low PD and earn profits

B. New data

- Can entrepreneurs lose financing when lenders have more data?

- Example: Entrepreneurs with good credit history

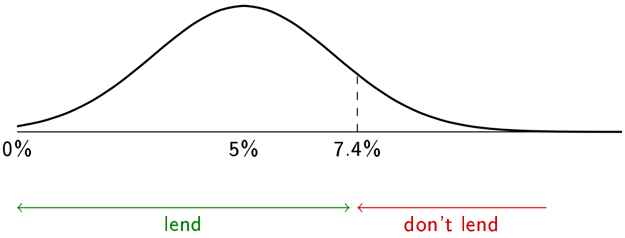
  - Traditional lenders' best estimate is $PD = 5\%$

  - Entrepreneurs are willing to pay up to 8%

$\Rightarrow$ All these entrepreneurs are financed

B. New data

- Fintech lenders have access to entrepreneurs' digital footprints
  - More precise estimate of PD of each entrepreneur
  - Distribution of estimated PD



B. New data

- Implications of more data (cont'd)

  3. Some entrepreneurs lose financing: those with high (previously undetected) default risk

  4. Allocation of credit is better

     – Note the parallel with the previous lecture: better information leads to better allocation of capital

- Problem 1 of Problem Set

- Potential risk: discrimination (more on this later)

Road map

Many potential applications of AI

- Credit: consumer credit, business lending. . .

- Equity: private equity, venture capital. . .

Special Report **Artificial intelligence** + Add to myFT

Artificial intelligence is guiding venture capital to start-ups

Firms are using algorithms to uncover potential investments

- . . . but there's a curse to avoid!

VC game

- You manage a venture capital (VC) fund
  - A startup seeks financing
  - Potential deal: you receive a given share of the company's equity in exchange for cash payment P
- Valuation
  - Value of the company's shares when you cash out, V, is uncertain
  - Company risk is idiosyncratic and risk-free rate is zero
  - A priori estimate of present value = $E[V]$
- Information
  - You analyze the company to form a more precise estimate of V
  - Your best estimate is: check your private signal

VC game

- Information (cont'd)
  - Your best estimate is unbiased: I generated it as the true V (which I know but you don't know) + a random noise with mean 0 and s.d. 20
- Competition
  - You are in competition with other VC funds (other students in the classroom) who also have their own best estimate of V
  - Each VC makes an offer to acquire the company's shares. The highest offer wins the deal
- Profits
  - If you win the deal at price P and the true company value is V, your profit (or loss if negative) is $V - P$
  - If you don't win the deal, your profit is zero
- Let's play! Submit your bid at the link in your mailbox

Winner's curse

- Estimates are unbiased
  - Some are above, some are below, the average is V
  - Naturally, the highest estimate is above V
- If everyone bids their best estimate, the winning price is the highest among all best estimates $\rightarrow$ it is above V $\rightarrow$ the winner overpays
- This is the **winner's curse**

Winner's curse in...

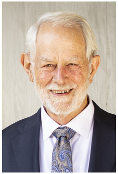
- Lending/financing
  - Borrowers take your loan offer when other lenders don't want to lend to them
- Trading
  - See previous lecture
- Insurance
  - Customers buy insurance from you when other insurers estimate the risk is high and charge high premiums
  - ...or when customers know their risk is higher than you think

Winner's curse in...

- Auctions

Nobel prize 2020

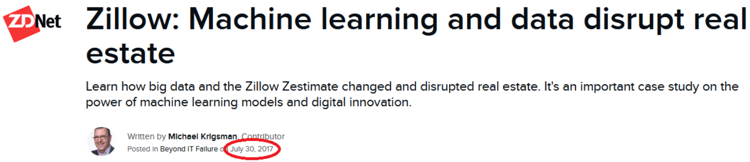
Paul R. Milgrom



Robert B. Wilson

Winner's curse in. . .

- Real estate



ZDNet

Zillow: Machine learning and data disrupt real estate

Learn how big data and the Zillow Zestimate changed and disrupted real estate. It's an important case study on the power of machine learning models and digital innovation.

Written by **Michael Krigsman**, Contributor
Posted in Beyond IT Failure on July 30, 2017

Interview with Zillow's Chief Analytics Officer Stan Humphries in 2017

ZD: How accurate is the Zestimate?

S.H.: Our models are trained such that half of the Earth will be positive and half will be negative; meaning that on any given day, half of [all] homes are going to transact above the Zestimate value and half are going to transact below.

Winner's curse in. . .

- Real estate

ZDNet

Zillow: Machine learning and data disrupt real estate

Learn how big data and the Zillow Zestimate changed and disrupted real estate. It's an important case study on the power of machine learning models and digital innovation.

Written by **Michael Krigsman**, Contributor
Posted in Beyond IT Failure on July 30, 2017

Interview with Zillow's Chief Analytics Officer Stan Humphries in 2017

ZD: How accurate is the Zestimate?

S.H.: Our models are trained such that half of the Earth will be positive and half will be negative; meaning that on any given day, half of [all] homes are going to transact above the Zestimate value and half are going to transact below.

CHRIS STOKEL-WALKER BUSINESS NOV 11, 2021 :00 AM

WIRED

Why Zillow Couldn't Make Algorithmic House Pricing Work

Winner's curse

- Winner's curse is similar to adverse selection (cf. previous lecture)
- Ask yourself which info you DON'T have and others may have
- How to avoid the winner's curse?

1. Be more conservative than your best estimate (margin of safety, "bid shading")
2. Even better: run experiments
   - Experiment pricing on small random sample + backtest
   - Done routinely by tech firms; can be used in consumer finance markets

Road map

Lucas critique

- When fintechs score customers using new data and AI

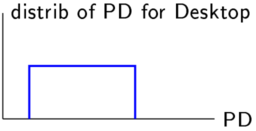
  ...customers' behavior may strategically respond to scoring models

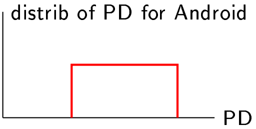
- Example
  - Fintech finds out that Android mobile users have higher probability of default $\rightarrow$ charge them higher interest rate
  - Android mobile owners, if they find out, will connect from their desktop!

- How does it affect the predictive power of the scoring model? How does it affect borrowers?

Lucas critique

- Suppose the fintech observes whether borrowers connect from Android or Desktop and did not use this information in the past





- The fintech starts charging lower rates for Desktop users and higher rates for Android users. Some Android users find out and switch to Desktop

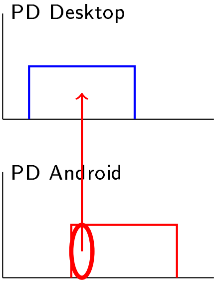
Q. How does this affect the average PD of Desktop users?

Increases; Does not change; Decreases; May increase or decrease

Lucas critique

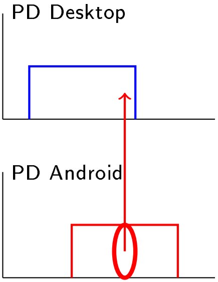
- It depends on who are the switchers!

Case 1: Safe Android users switch



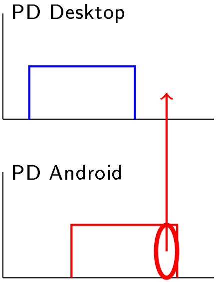
PD of Desktop users =
PD of Android users ↑

Case 2: Intermediate Android users switch



PD of Desktop users ↑
PD of Android users =

Case 3: Risky Android users switch

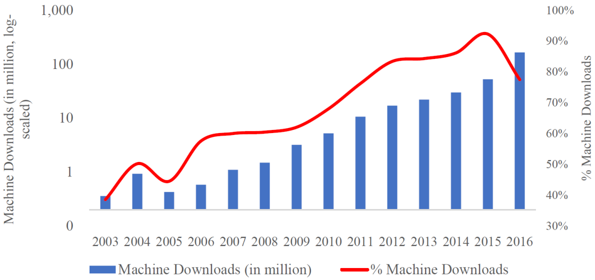


PD of Desktop users ↑
PD of Android users ↓

Case study: Corporate reporting

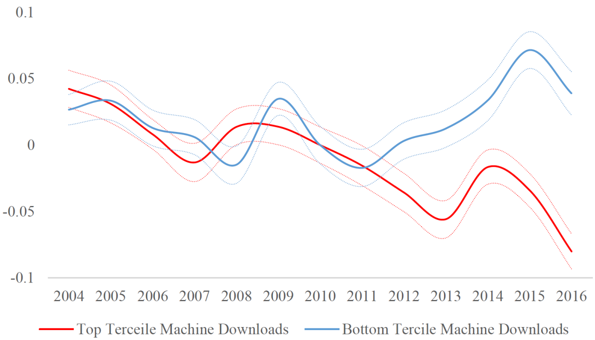
- "How to Talk When a Machine is Listening: Corporate Disclosure in the Age of AI," Cao, Jiang, Yang and Zhang, 2020 [pdf]

Machine downloads of companies' filings with the SEC



Case study: Corporate reporting

- Use of negative words as measured by Loughran-McDonald dictionary widely used by quant funds → decreases after post-2010 rise of algos for companies scrutinized by algos (in red)



Lucas critique: Implications

1. Beware scoring on behavior that can be strategically modified

   ⇒ Ask yourself if data is exogenous or the outcome of a strategic choice

2. A good predictor when not used, can become a poor predictor once used

   - Goodhart's law

   ⇒ Check if the predictive power changes after data is used

Lucas critique: Implications

3. More data can make everyone (fintech AND consumers) worse off

   - Android mobile owners must connect from a desktop or buy a new mobile $\rightarrow$ they're worse off

     AND the predictive power of Android has disappeared $\rightarrow$ fintech is worse off

   - You'd like to buy a health tracking smartwatch but you don't because you're worried your data are used to price your health insurance

     $\Rightarrow$ Both you and the smartwatch producer would be better off if your data were not used

Lucas critique: Implications

4. Commitment NOT TO use data can sometimes create value
   - Both for companies and for consumers
   - Privacy can be a source of economic value
   - Credibility of commitment is key: once personal data exists, it is tempting for companies to use it
   - More on privacy: https://johanhombert.github.io/blog/20210418-privacy-paradox

History of the Lucas critique

- 1960s keynesian economics: central bank can dampen recessions by increasing money supply

- Lucas: ok in unanticipated, but once people anticipate inflation monetary policy becomes ineffective ⇒ 1970s stagflation (and 2022?)

Robert E. Lucas Jr.
(Nobel prize 1995)

Road map

Information in insurance

- Suppose we discover how to predict perfectly who will get sick (but this foreknowledge does not help to prevent or cure diseases)

  Insurers use this information to price health insurance

Q1. Will it make people better or less-well insured?

Q2. Will it make insurers more or less profitable?

Information in insurance

- What will happen?
  - No insurers accept to insure people predicted to be sick
  - People predicted to be healthy don't need insurance

$\Rightarrow$ The health insurance market breaks down

  - People are worse off: they can't get insurance
  - Insurers are worse off: they can't sell insurance

- This is the Hirshleifer effect: Information can destroy insurance

How to overcome the Hirshleifer effect?

- Suppose an insurer announces it will not use the information (and suppose it is credible)

Q3. Does it overcome the Hirshleifer effect and allow the insurer to sell insurance?

How to overcome the Hirshleifer effect?

- Suppose an insurer announces it will not use the information (and suppose it is credible)

Q3. Does it overcome the Hirshleifer effect and allow the insurer to sell insurance?

- No, because of adverse selection
  - People predicted to be healthy are offered cheap insurance from other insurers (or they don't even buy insurance)
  - The insurer only gets people who will be sick, so it cannot insure them
- Problem 2 of Problem Set

How to overcome the Hirshleifer effect?

- Solution 1: Ensure no insurer uses the information
  - Industry ethical standard (?)
  - Regulation (e.g., insurers are prohibited from using genetic information)

How to overcome the Hirshleifer effect?

- Solution 2: Insure before information is revealed
  - Long-term insurance / Premium guaranteed over long period

Road map

Direct discrimination

- Direct discrimination: treatment based on a "protected characteristic" such as race, sex, ethnic or 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: treatment 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 treatment is based on variables correlated with protected characteristics
    - Ex.: interest rate based on borrower's job occupation may end up being different across people with different social or ethnic origins
- May be legal or illegal (e.g. legal for "business necessity" in US law)

Discriminatory algorithms?

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.

MOTHERBOARD
TECH BY VICE

Court Rules Deliveroo Used 'Discriminatory' Algorithm

An Italian court determined that companies can be held liable even if an algorithm unintentionally discriminates against a protected group.

Discriminatory algorithms?

- Algos are (a priori) not subject to human prejudices but...

1. Biased data: algos are fed with human-world data, which may be contaminated by discrimination

   - Unclear how it biases algos

     Ex. Suppose men and women make equally good entrepreneurs but biased venture capitalists set a higher bar for women[1]

     ⇒ In databases of VC-backed startups, the average female-founded company will be better than the average male-founded companies

     ⇒ Algos fed with these data will "learn" that women are better entrepreneurs than men

---

[1] See "Gender Stereotypes and Entrepreneur Financing," Camille Hébert, 2020 [pdf]

Discriminatory algorithms?

2. Triangulation: algos may “triangulate” protected characteristics from other data (without intent to do so) and use them

- Solution
  - Algorithm interpretability: understand how algos make decisions
  - Ex.: feed algo with simulated data and analyze outcomes

Case study: Fintech lenders 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

Case study: Fintech lenders in the US mortgage market

- Good news
  - Less discrimination by algos
  - Discrimination by traditional lender has decreased over time, perhaps as a result of competition from fintech
- Bad news
  - Algos discriminate (although less so than humans)
  - How come?!
  - Likely explanation: Algos "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