



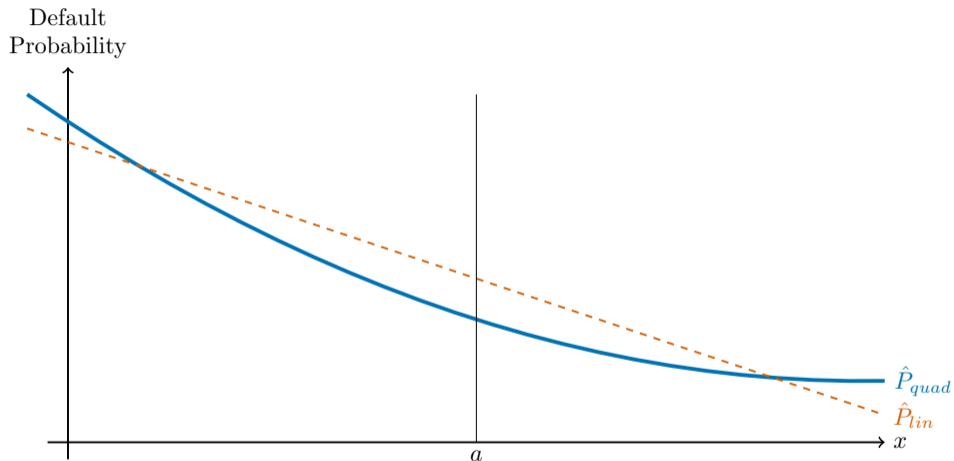
Predictably Unequal? The Effect of Machine Learning on Credit Markets

**A. Fuster, P. Goldsmith-Pinkham, T. Ramadorai,
and A. Walther**
SNB, FRBNY / Yale SOM, Imperial College

Disclaimer

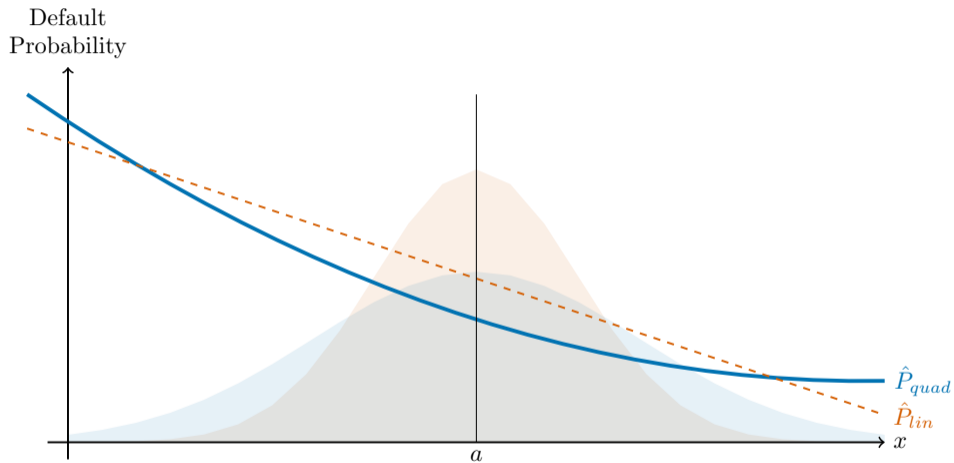
The views expressed do not necessarily reflect the position of the Federal Reserve Bank of New York, the Federal Reserve System, or the Swiss National Bank.

Winners and Losers



Convex quadratic: "extreme" x lose, others gain

Winners and Losers



Two groups: "blue" borrowers lose due to high variance

Sources of Unequal Effects

- Previous example could arise from

$$y = P(x) + \varepsilon,$$

where P is nonlinear and g does not matter for y .

⇒ **Winners/losers** arise from additional **flexibility** of new technology.

Effects across g depend on functional form of new technology, and the differences in distribution of characteristics

Sources of Unequal Effects

- Previous example could arise from

$$y = P(x) + \varepsilon,$$

where P is nonlinear and g does not matter for y .

⇒ Winners/losers arise from additional **flexibility** of new technology.

Effects across g depend on functional form of new technology, and the differences in distribution of characteristics

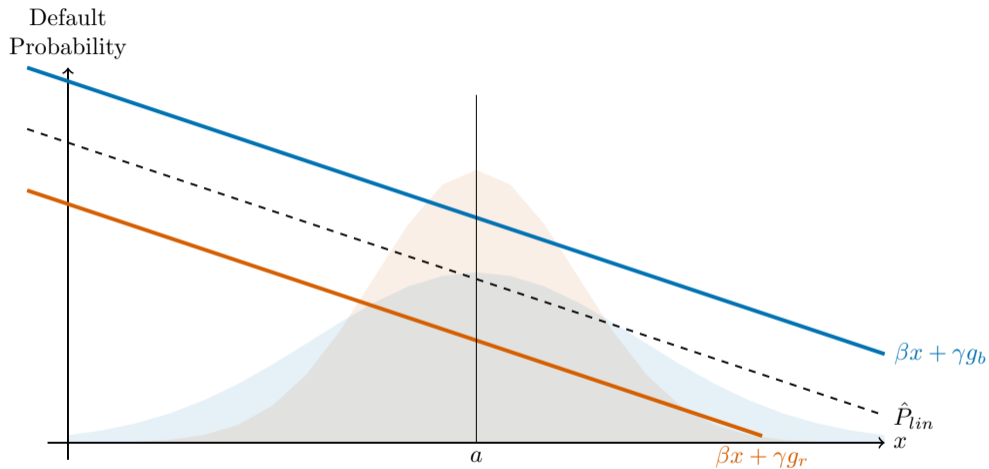
- Alternative:

$$y = \beta \cdot x + \gamma \cdot g + \varepsilon,$$

i.e. true relationship is linear, but g predictive of default.

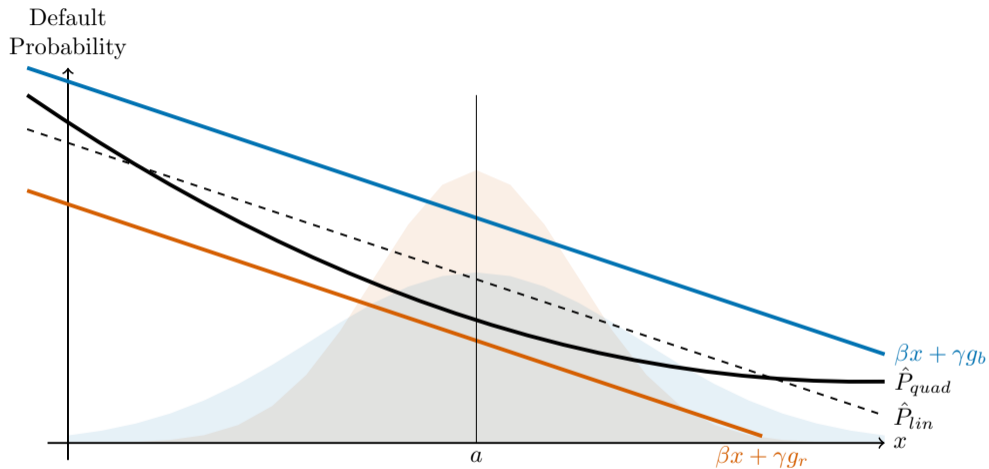
⇒ Effects of new technology arise due to “**triangulating**” g

Triangulation



- No linear correlation between x and $g \rightarrow$ linear model simply recovers average

Triangulation



- Blue borrowers more likely to have extreme $x \rightarrow$ nonlinear model penalizes.

US Mortgage Data

HMDA

- Application date, applicant income, loan type, size, purpose,
- **race, ethnicity, gender**

McDash (Black Knight)

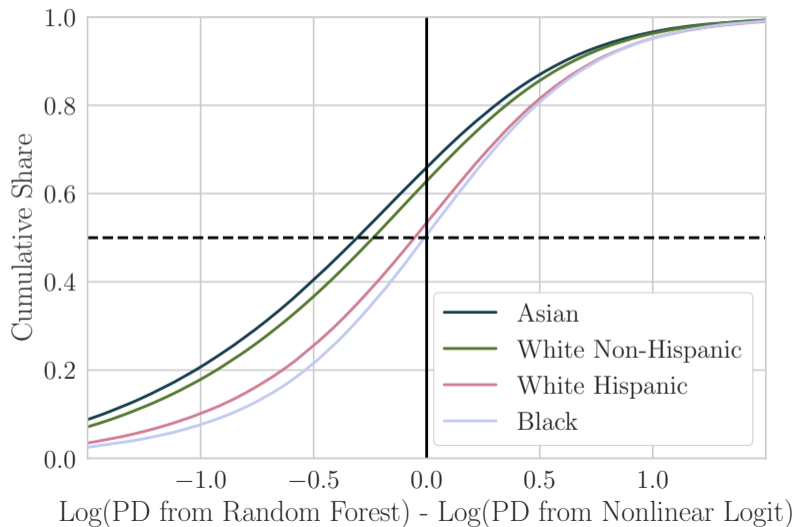
- Underwriting, contract and performance: e.g. FICO, LTV, interest rate, **default status**



Linked Dataset

- 9.4m mortgage loans from 2009-2013
- Portfolio *and* GSE loans, < \$1m
- **Default:** 90+ days delinquent within 3 years of origination

Unequal Effects of New Technology: Population



Flexibility versus Triangulation

Decomposition of model improvements:

1. Add **race** as an explanatory variable to Logit
2. Allow use of ML **technology** to the model with race
(i.e. "add" nonlinear functions / interactions of x as explanatory variables)

Flexibility versus Triangulation

Decomposition of model improvements:

1. Add **race** as an explanatory variable to Logit
2. Allow use of ML **technology** to the model with race
(i.e. "add" nonlinear functions / interactions of x as explanatory variables)

	Race	Technology
ROC-AUC	5.88	94.12
Precision	7.90	92.10
R2	2.04	97.96

⇒ Improved performance mostly due to flexibility, not triangulation

NB: Order of decomposition matters; but our qualitative conclusion is robust

Conclusion

Improvements in statistical technology creates

- Greater predictive power and gains for producers
- Increased disparity in outcomes for consumers

Framework for unequal effects: Flexibility and Triangulation

Empirical assessment in the US mortgage market

- Unequal effects along racial lines
- Appear to be driven by flexibility