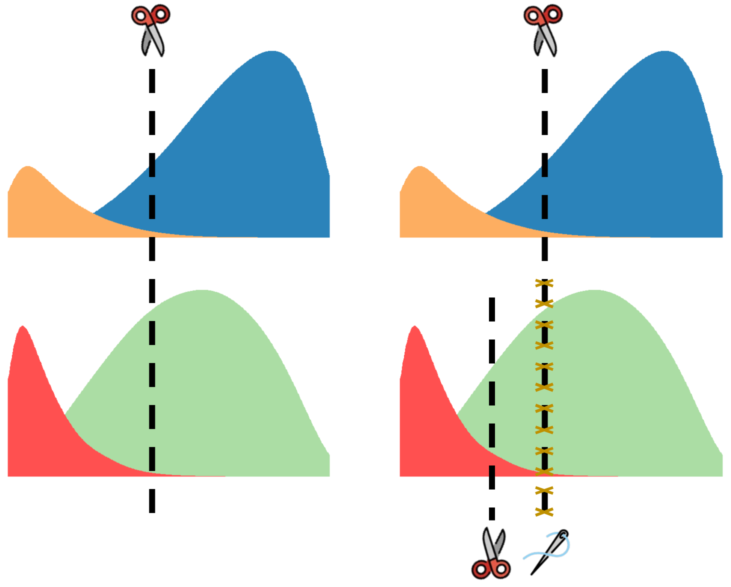
One Threshold Doesn't Fit All:
Tailoring Machine Learning Predictions of Consumer Default for Lower-Income Areas

Vitaly Meursault, Dan Moulton,
Larry Santucci, Nathan Schor
Federal Reserve Bank of Philadelphia



September 30, 2022

Disclaimer: The views expressed here do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System. No statements here should be treated as legal advice.

Machine learning in lending involves a tension

- ML allows lenders to predict default better overall

Machine learning in lending involves a tension

- ML allows lenders to predict default better overall
- Concern: protected groups benefit disproportionately less

Machine learning in lending involves a tension

- ML allows lenders to predict default better overall
- Concern: protected groups benefit disproportionately less

We propose to link model improvement with relaxation of lending thresholds for lower-income areas

Machine learning in lending involves a tension

- ML allows lenders to predict default better overall
- Concern: protected groups benefit disproportionately less

We propose to link model improvement with relaxation of lending thresholds for lower-income areas

- Ensure lower-income areas benefit by equalizing their true positive rates with higher income areas

Machine learning in lending involves a tension

- ML allows lenders to predict default better overall
- Concern: protected groups benefit disproportionately less

We propose to link model improvement with relaxation of lending thresholds for lower-income areas

- Ensure lower-income areas benefit by equalizing their true positive rates with higher income areas
- Still predict default better overall than the baseline model

Machine learning in lending involves a tension

- ML allows lenders to predict default better overall
- Concern: protected groups benefit disproportionately less

We propose to link model improvement with relaxation of lending thresholds for lower-income areas

- Ensure lower-income areas benefit by equalizing their true positive rates with higher income areas
- Still predict default better overall than the baseline model
- This comes at a cost of higher false positive rates in lower income areas

Machine learning in lending involves a tension

- ML allows lenders to predict default better overall
- Concern: protected groups benefit disproportionately less

We propose to link model improvement with relaxation of lending thresholds for lower-income areas

- Ensure lower-income areas benefit by equalizing their true positive rates with higher income areas
- Still predict default better overall than the baseline model
- This comes at a cost of higher false positive rates in lower income areas
- This approach requires explicit use of geography in lending

Machine learning in lending involves a tension

- ML allows lenders to predict default better overall
- Concern: protected groups benefit disproportionately less

We propose to link model improvement with relaxation of lending thresholds for lower-income areas

- Ensure lower-income areas benefit by equalizing their true positive rates with higher income areas
- Still predict default better overall than the baseline model
- This comes at a cost of higher false positive rates in lower income areas
- This approach requires explicit use of geography in lending

We contribute to the literature at the intersection of consumer lending, ML and inequality

- "Predictably unequal" (Fuster et al., 2021)
- "How costly is the noise?" (Blattner and Nelson, 2021 WP)

We contribute to the literature at the intersection of consumer lending, ML and inequality

- “Predictably unequal” (Fuster et al., 2021)
- “How costly is the noise?” (Blattner and Nelson, 2021 WP)
- “The input fallacy” (Gillis, 2022)

We contribute to the literature at the intersection of consumer lending, ML and inequality

- “Predictably unequal” (Fuster et al., 2021)
- “How costly is the noise?” (Blattner and Nelson, 2021 WP)
- “The input fallacy” (Gillis, 2022)
- “Equality of Opportunity in Supervised Learning”

(Hardt et al., 2016)

We contribute to the literature at the intersection of consumer lending, ML and inequality

- “Predictably unequal” (Fuster et al., 2021)
- “How costly is the noise?” (Blattner and Nelson, 2021 WP)
- “The input fallacy” (Gillis, 2022)
- “Equality of Opportunity in Supervised Learning”

(Hardt et al., 2016)

This paper: combine fairness constraints + model improvement, characterize the resulting trade-offs

Paper outline

- Train default prediction models to generate credit scores for population

Paper outline

- Train default prediction models to generate credit scores for population
- Confirm results from recent literature regarding model improvement and inequality

Paper outline

- Train default prediction models to generate credit scores for population
- Confirm results from recent literature regarding model improvement and inequality
- Generate single and group-specific decision thresholds (for higher and lower income areas)

Paper outline

- Train default prediction models to generate credit scores for population
- Confirm results from recent literature regarding model improvement and inequality
- Generate single and group-specific decision thresholds (for higher and lower income areas)
- Generate profit – fairness trade-offs for more and less complex models

Paper outline

- Train default prediction models to generate credit scores for population
- Confirm results from recent literature regarding model improvement and inequality
- Generate single and group-specific decision thresholds (for higher and lower income areas)
- Generate profit – fairness trade-offs for more and less complex models
- Two dimensions: more / less advanced model and stronger / weaker fairness constraints

Meet the characters!

- Model improvement
- LMI and non-LMI areas
- True positive rate (TPR) and false positive rate (FPR)
- ΔTPR
- Profit

- Logistic (ridge) – traditional
- XGBoost – ML

Meet the characters!

- Model improvement
- LMI and non-LMI areas
- True positive rate (TPR) and false positive rate (FPR)
- ΔTPR
- Profit

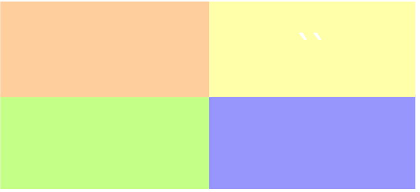
LMI (Lower- and Moderate-Income):

census tracts with median income <80% of the MSA median income

Non-LMI:

census tracts with median income ≥80% of the MSA median income

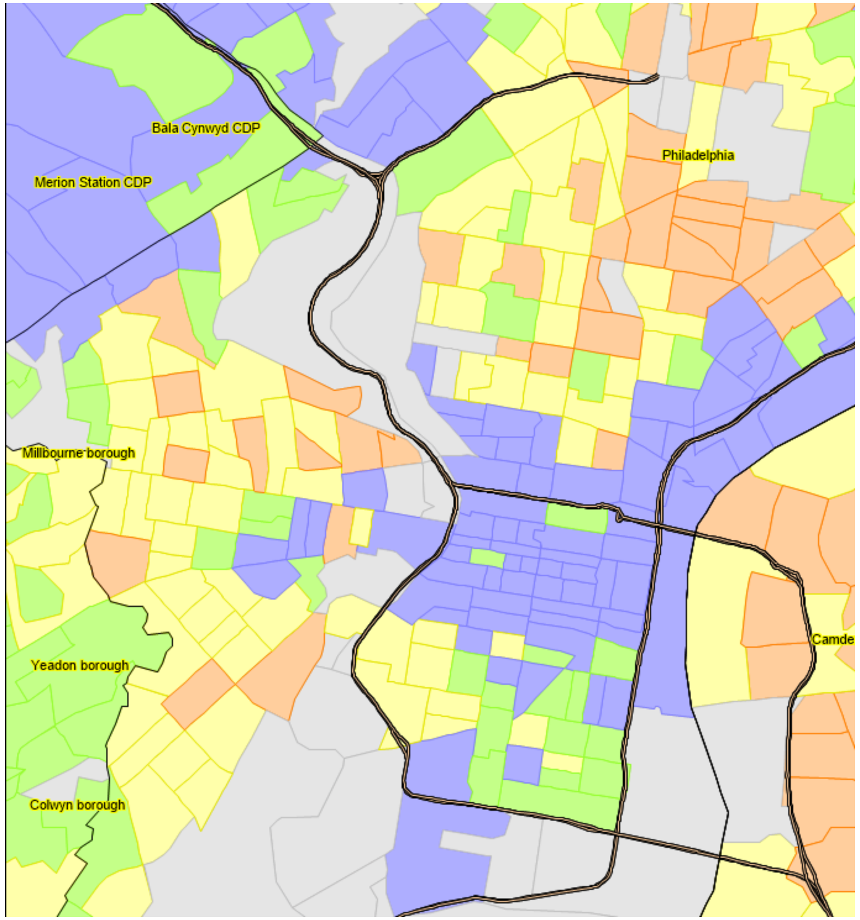
Meet the characters!



LMI (<80% MSA income)

Non-LMI

- Model improvement
- LMI and non-LMI areas
- True positive rate (TPR) and false positive rate (FPR)
- ΔTPR
- Profit
- Single threshold
- Separate thresholds



Created with VDA Web Gis

Meet the characters!

- Model improvement
- LMI and non-LMI areas
- True positive rate (TPR) and false positive rate (FPR)
- ΔTPR
- Profit

- TPR = TP / (TP + FN)

"Out of all people who will pay back, how many were correctly identified by the model?"

- FPR = FP / (FP + TN)

"Out of all people who will default on the loan, how many were incorrectly identified by the model?"

Meet the characters!

- Model improvement
- LMI and non-LMI areas
- True positive rate (TPR) and false positive rate (FPR)
- ΔTPR
- Profit

We focus on

ΔTPR = TPR(non-LMI) – TPR(LMI)

as the fairness metric

LMI (Lower- and Moderate-Income): census tracts with median income <80% of the MSA median income

Meet the characters!

- Model improvement
- LMI and non-LMI areas
- True positive rate (TPR) and false positive rate (FPR)
- ΔTPR
- Profit

Lender cares about FP more than about TP.

We assume lender profits are

TP - λFP, λ = 4 in main specification

Default prediction models (skip today)

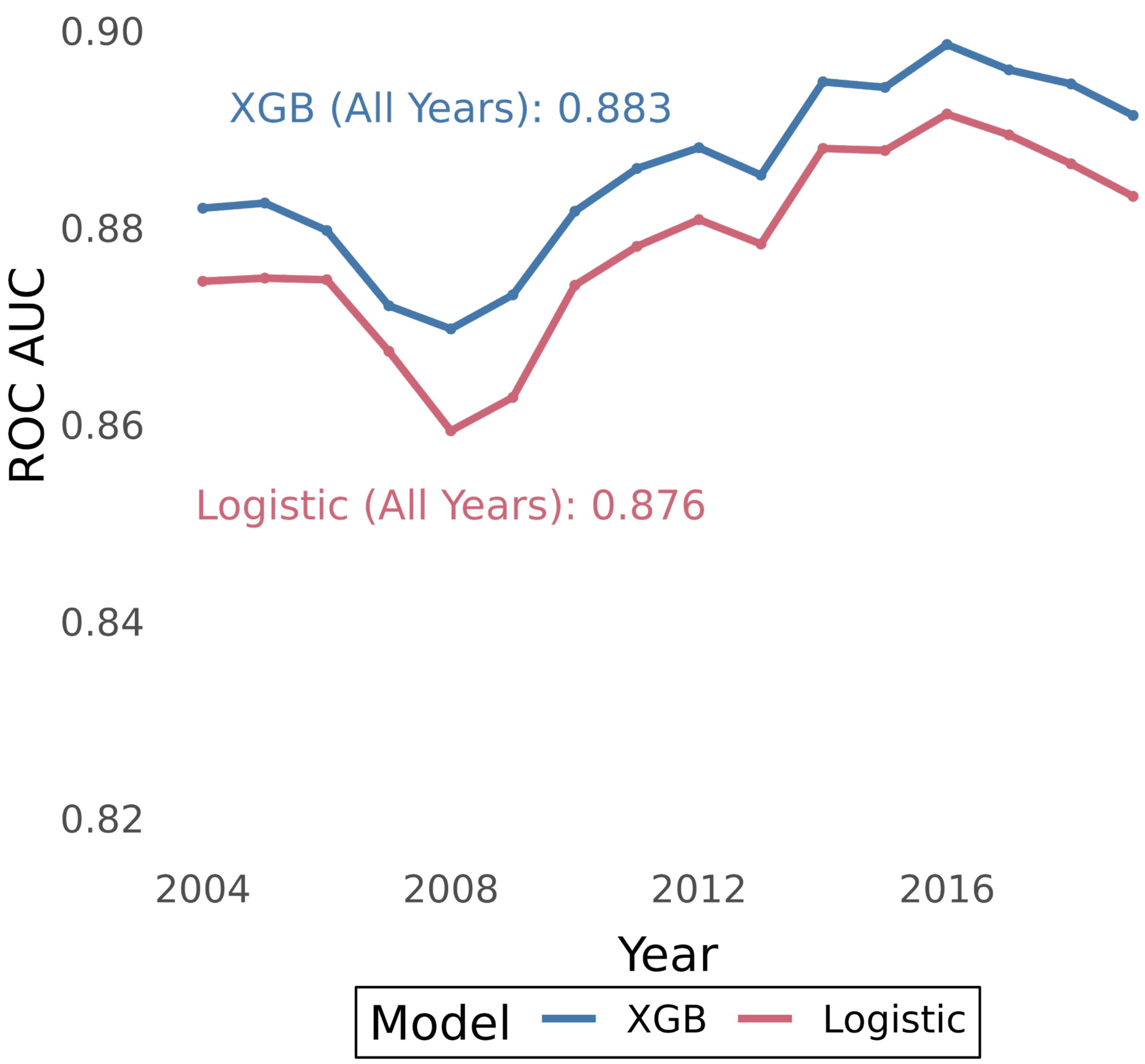
Model improvement and inequality

Single and group-specific thresholds

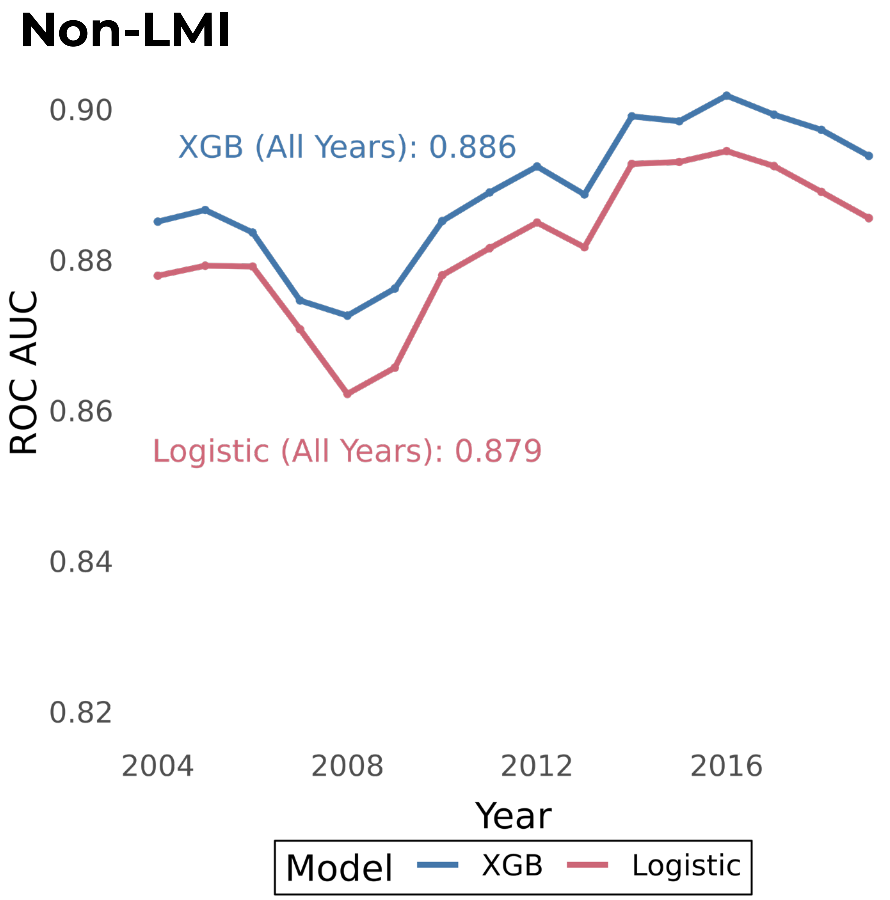
Profit—fairness trade-offs + model change

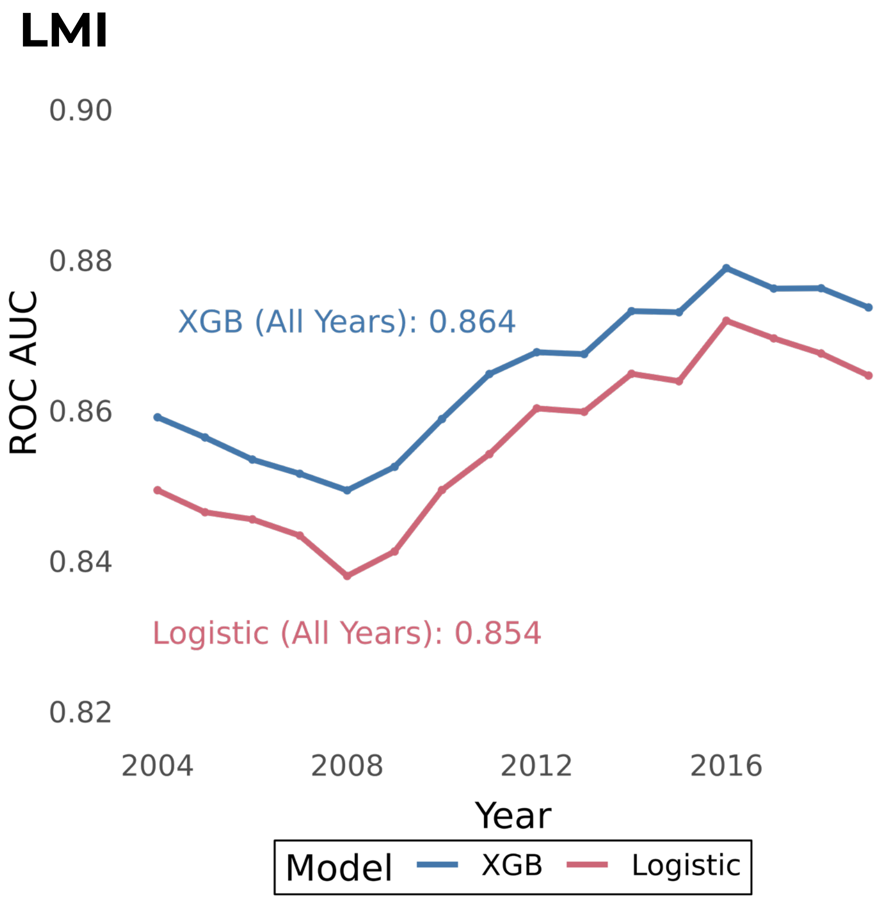
Modeling choices improve overall default prediction

- This difference translates into about 1% increase in profits (under assumptions discussed below)

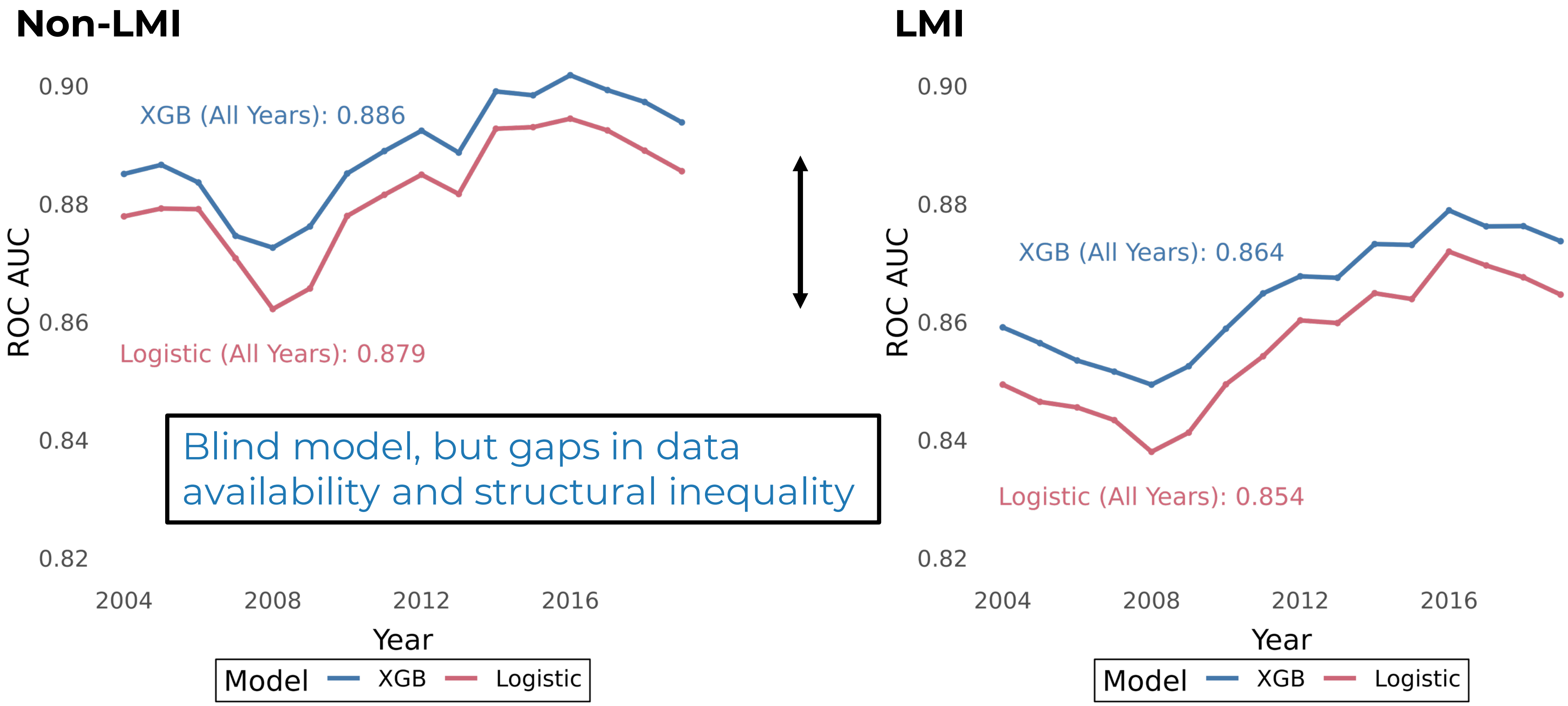


Predictive power isn't same for everyone





Predictive power isn't same for everyone



Blind model, but gaps in data availability and structural inequality

Default prediction models

Model improvement and inequality

Single and group-specific thresholds

Profit—fairness trade-offs + model change

Lending decision is a binary prediction problem

- Consider CS $\in[0,100]$, decreasing in P(default)

Lending decision is a binary prediction problem

- Consider CS $\in[0,100]$, decreasing in P(default)
- Lender picks CS threshold and lends above it

Lending decision is a binary prediction problem

- Consider CS $\in[0,100]$, decreasing in P(default)
- Lender picks CS threshold and lends above it
- The truth is known when repayment or default happens

Lending decision is a binary prediction problem

- Consider CS ∈[0,100], decreasing in P(default)
- Lender picks CS threshold and lends above it
- The truth is known when repayment or default happens

Threshold: 60	Actual Positive	Actual Negative
Predicted Positive	40	10
Predicted Negative	20	30

Positive outcome is **non-default**

Table is based on simulated data

Lending decision is a binary prediction problem

- Consider CS ∈[0,100], decreasing in P(default)
- Lender picks CS threshold and lends above it
- The truth is known when repayment or default happens

Threshold: 60	Actual Positive	Actual Negative
Predicted Positive	40	10
Predicted Negative	20	30

Threshold: 40↓	Actual Positive	Actual Negative
Predicted Positive	50↑	20↑
Predicted Negative	10↓	20↓

Table is based on simulated data

Incentives of lenders and the regulator differ

- Regulator cares about TP and True Positive Rate (TPR):
  TP/(TP + FN)

Threshold: 60	Actual Positive	Actual Negative
Predicted Positive	40	10
Predicted Negative	20	30

Table is based on simulated data

Incentives of lenders and the regulator differ

- Regulator cares about TP and True Positive Rate (TPR):
  TP/(TP + FN)
- Lender cares about profit:
  TP - λFP

Threshold: 60	Actual Positive	Actual Negative
Predicted Positive	40	10
Predicted Negative	20	30

Table is based on simulated data

Incentives of lenders and the regulator differ

- Regulator cares about TP and True Positive Rate (TPR):

  TP/(TP + FN)

- Lender cares about profit:

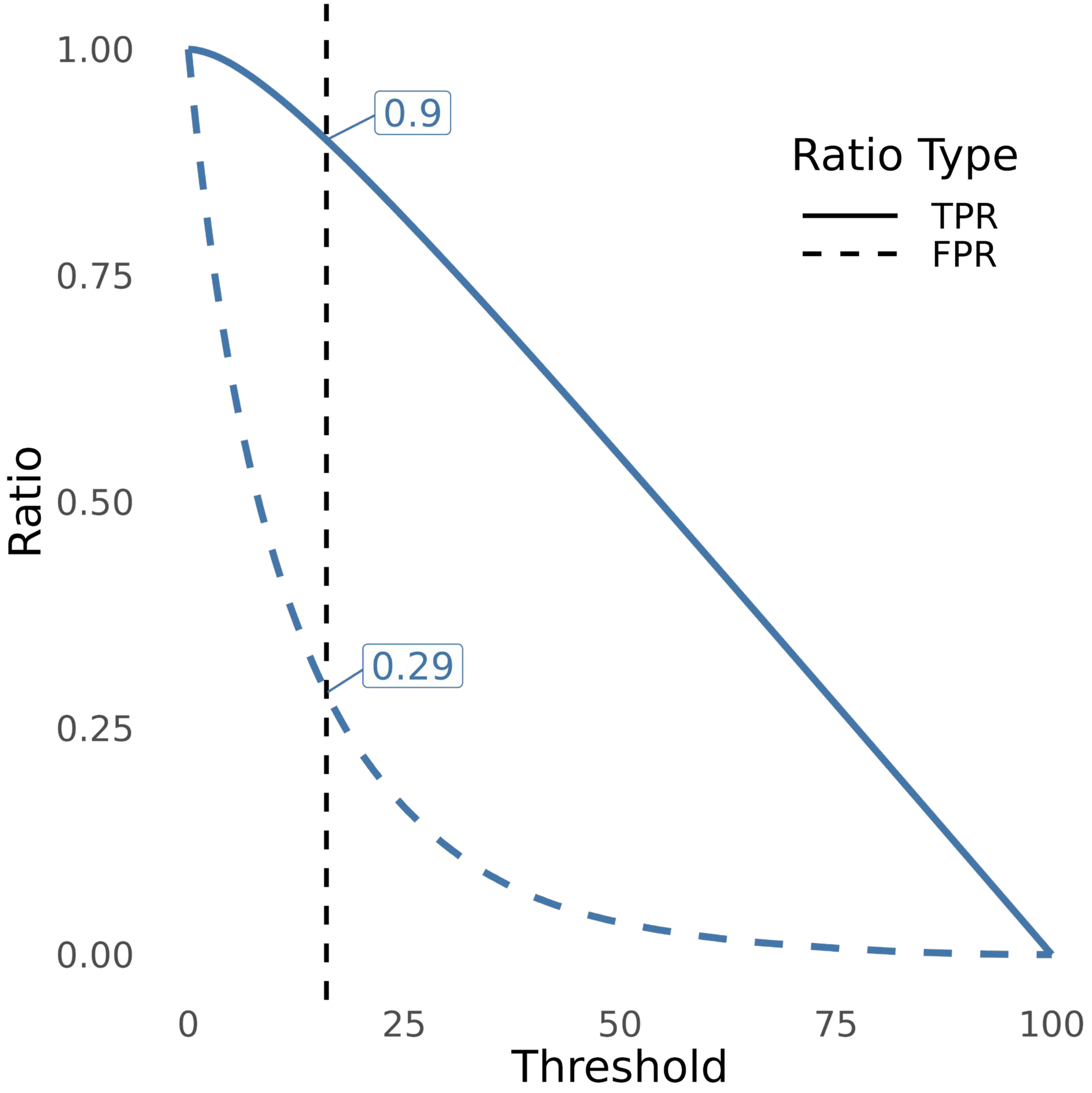
  TP - λFP

- We set λ to 4

Threshold: 60	Actual Positive	Actual Negative
Predicted Positive	40	10
Predicted Negative	20	30

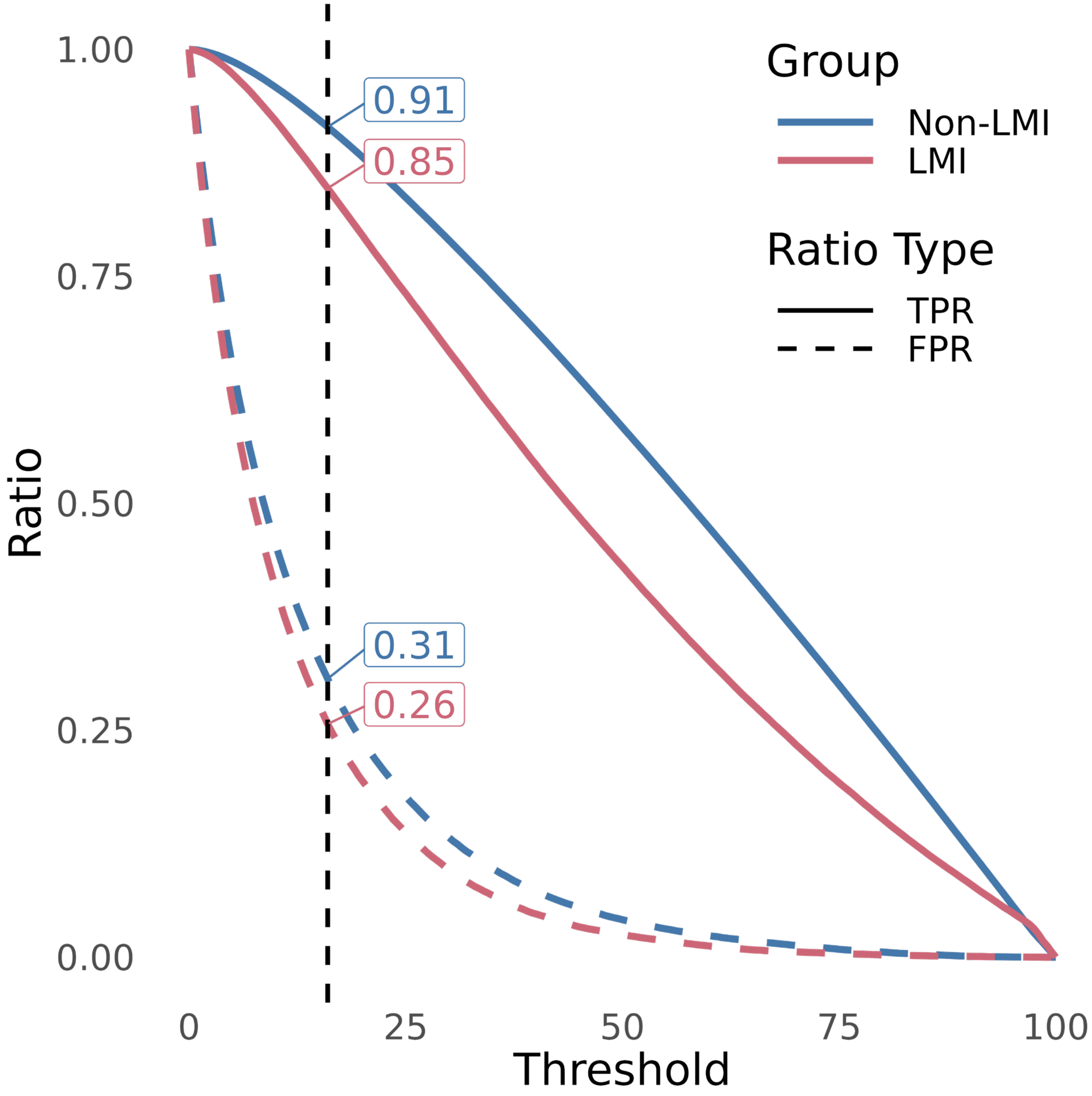
Table is based on simulated data

TPR and FPR overall



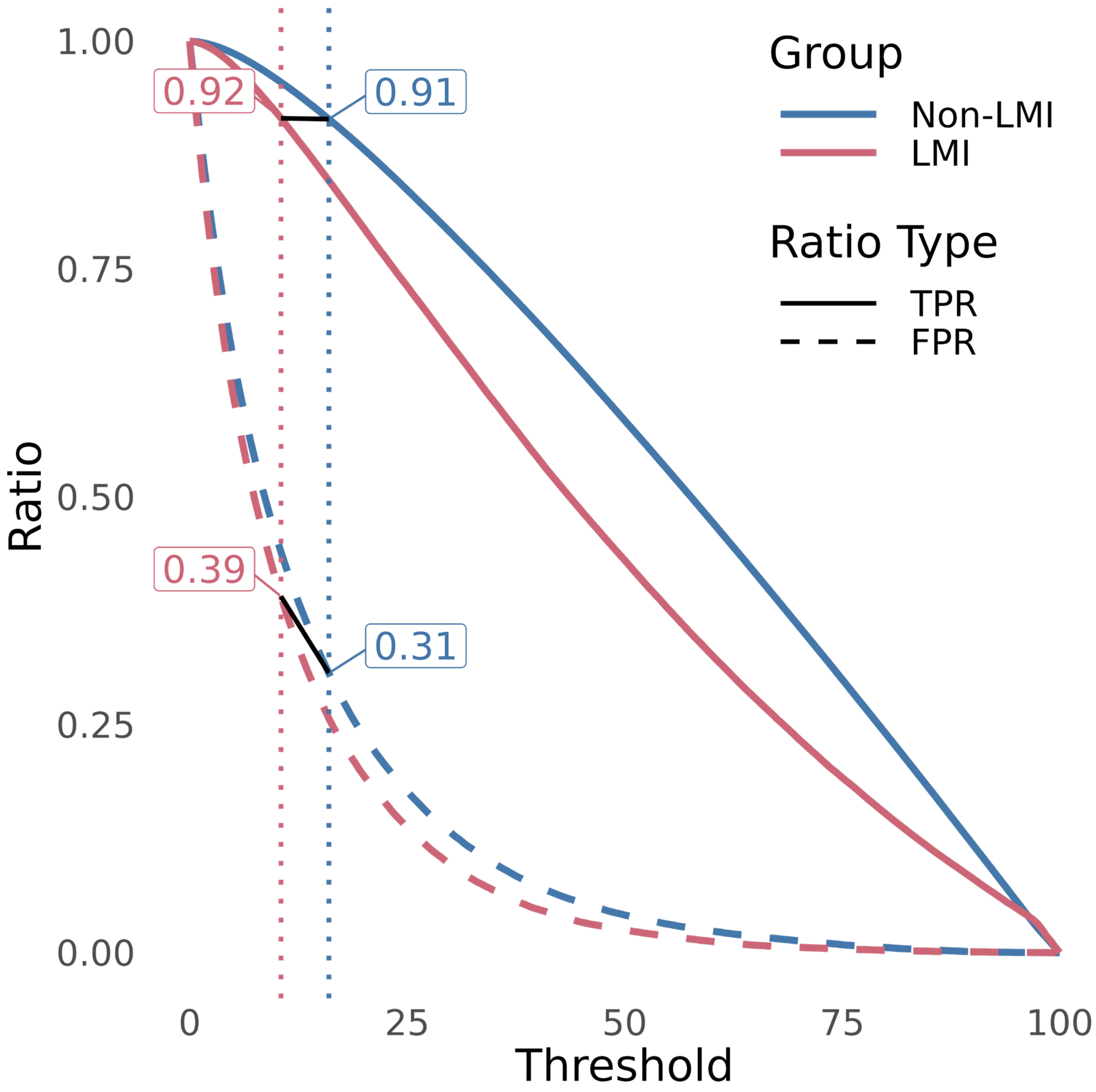
Single threshold leads to TPR disparity

- Regulator cares about equalizing TPR
- Lender cares about profit: TP - λFP
- We set λ to 4



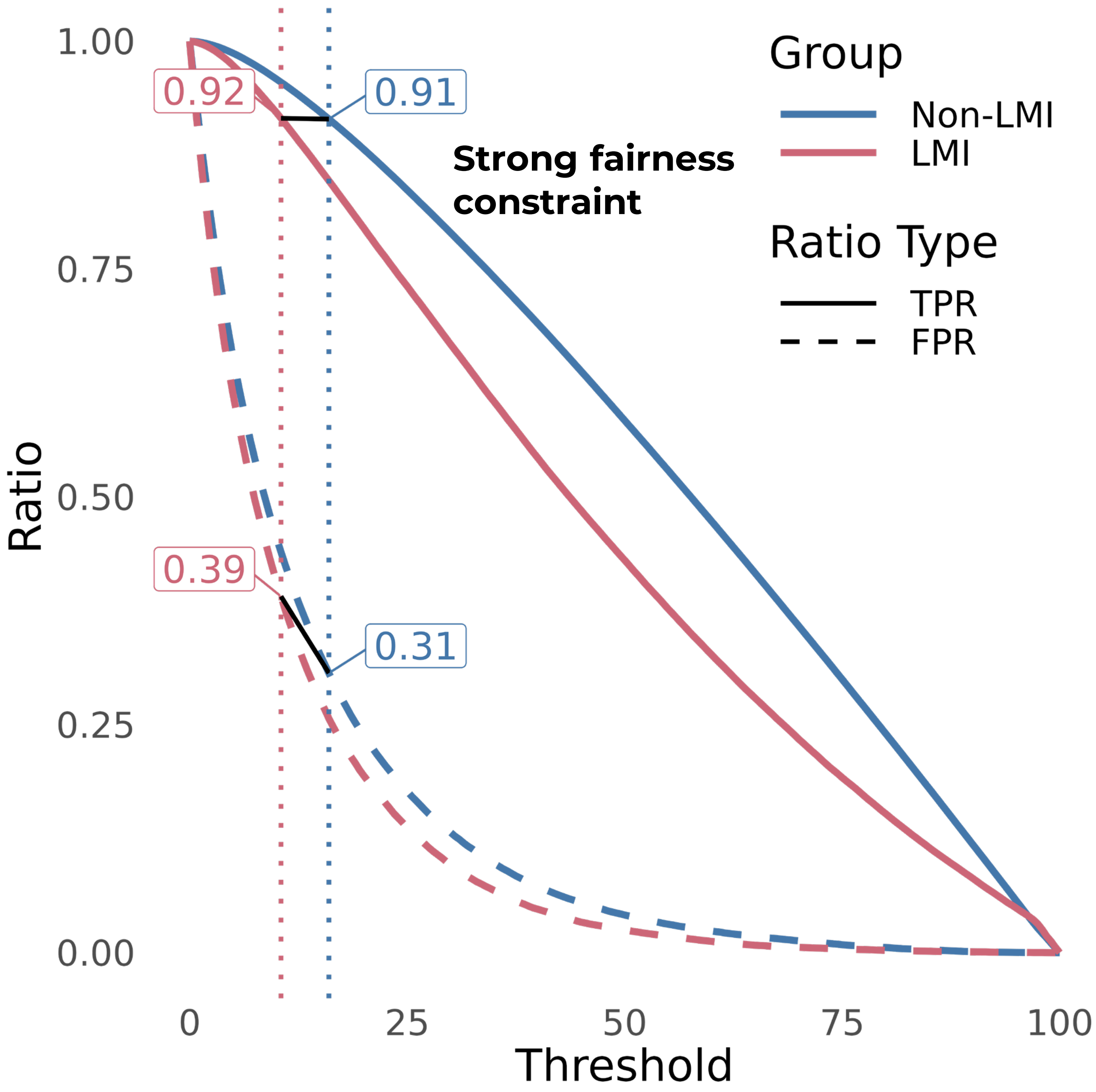
Introducing separate thresholds can reduce ΔTPR

- Regulator cares about equalizing TPR
- Lender cares about profit: TP - λFP
- We set λ to 4



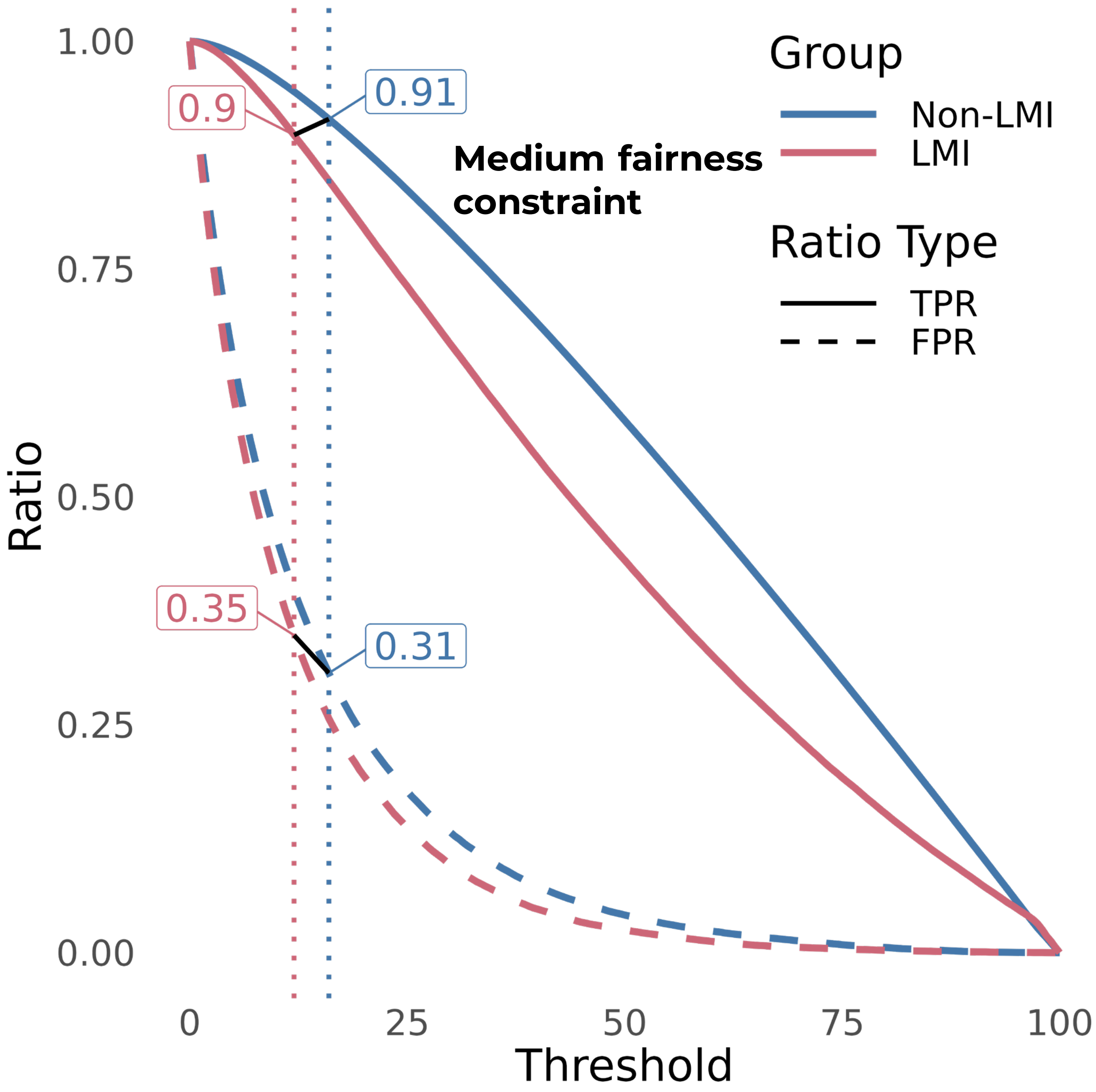
Picking the separate thresholds optimally

- Hardt et al. (2016)
- Consider all pairs of thresholds that equalize TPR
- Out of those pick thresholds that maximize profit (TP - 4×FP)



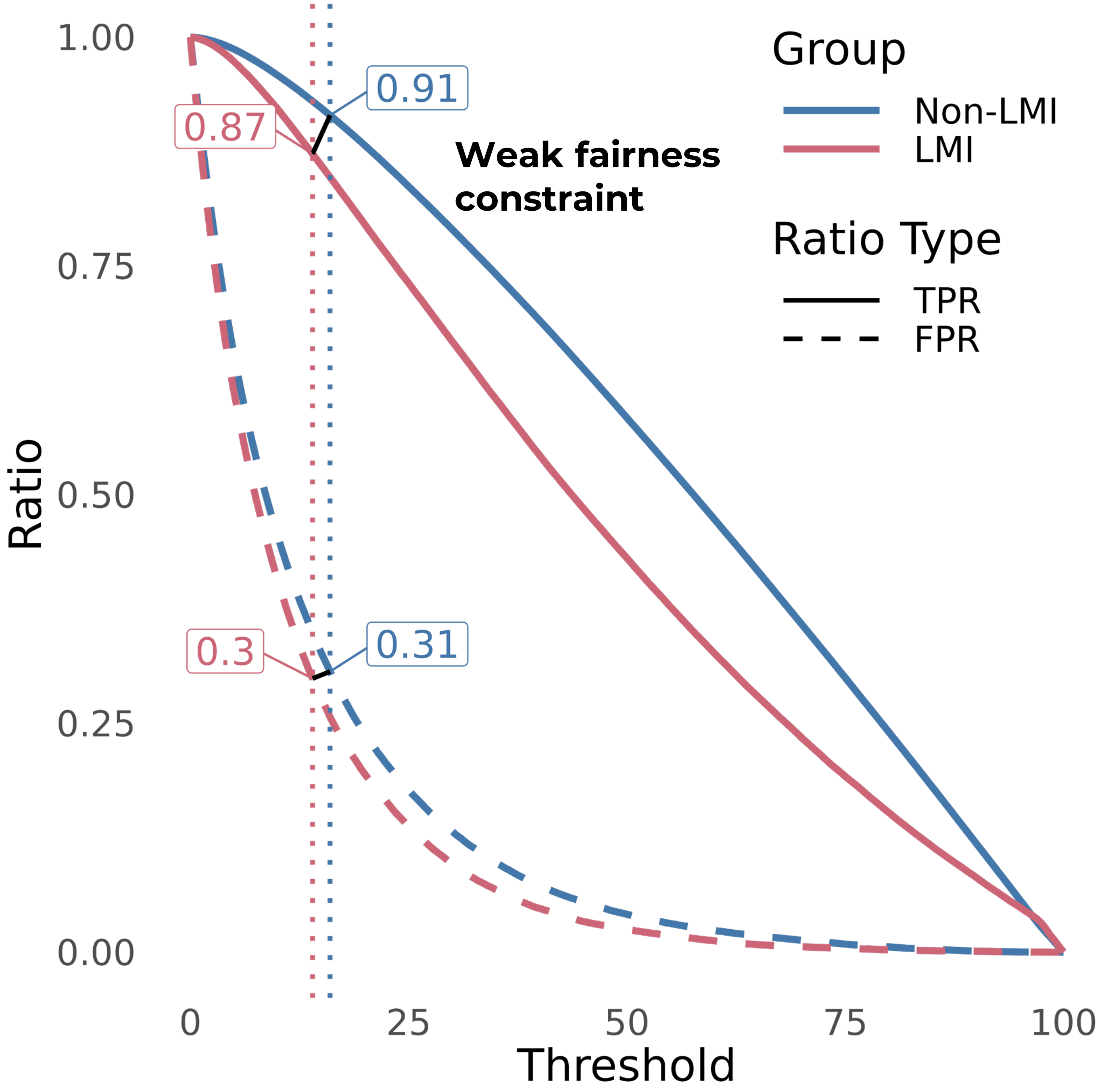
Fairness constraint is easy to relax

- Picking points between the eq. opp. threshold and single threshold relaxes the fairness constraint



Fairness constraint is easy to relax

- Picking points between the eq. opp. threshold and single threshold relaxes the fairness constraint



Default prediction models

Model change and inequality

Single and group-specific thresholds

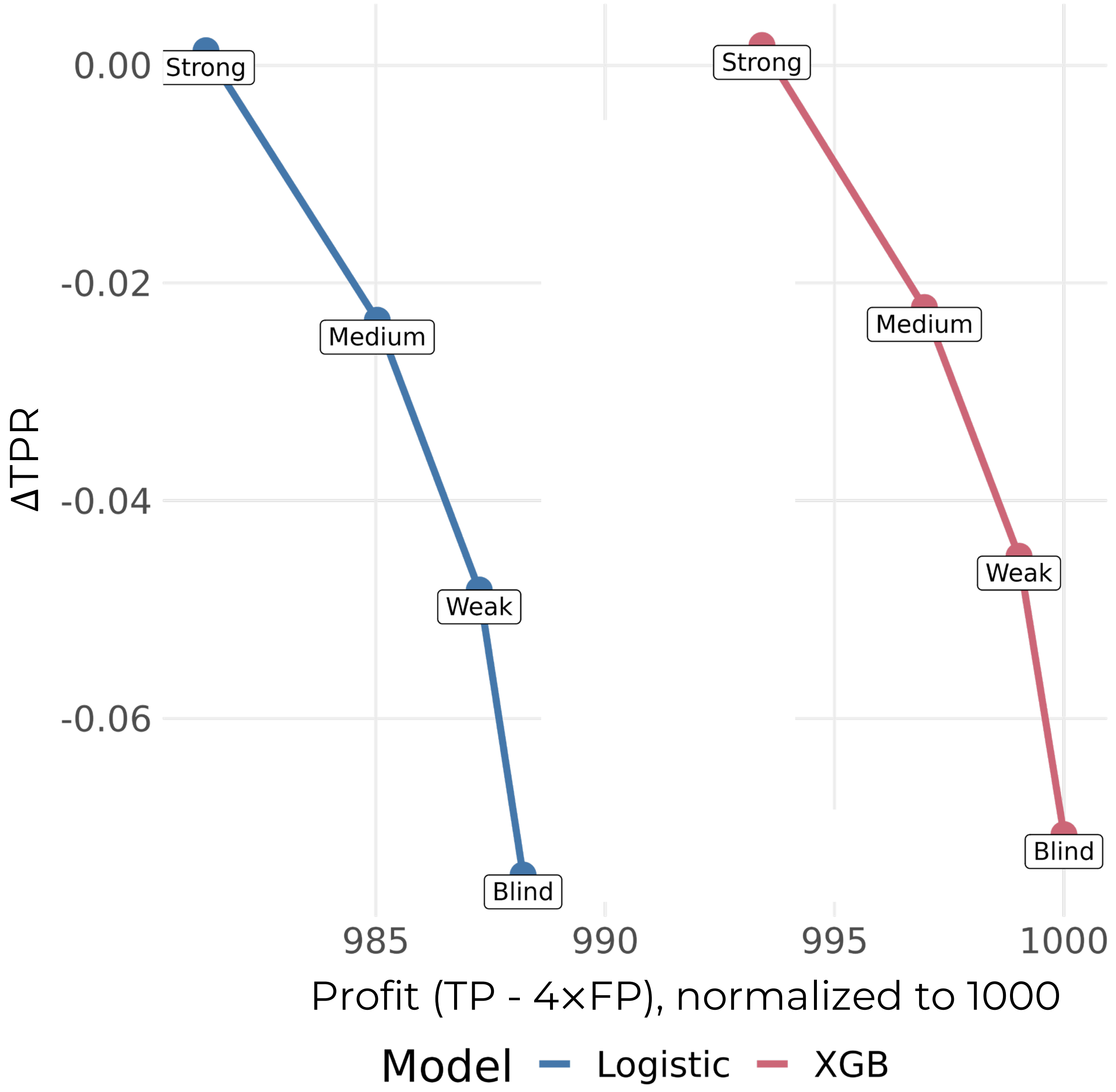
Profit—fairness trade-offs + model change

Fairness—profit tradeoff

- Regulator cares about equalizing TPR
- Lender cares about profit: TP - λFP
- We set λ to 4

ΔTPR = TPR(non-LMI) – TPR(LMI)

Fairness constraints:
Strong: 0 ΔTPR in-sample
Medium: 2/3 between Strong and Blind
Weak: 1/3 between Strong and Blind
Blind: Same threshold for non-LMI and LMI

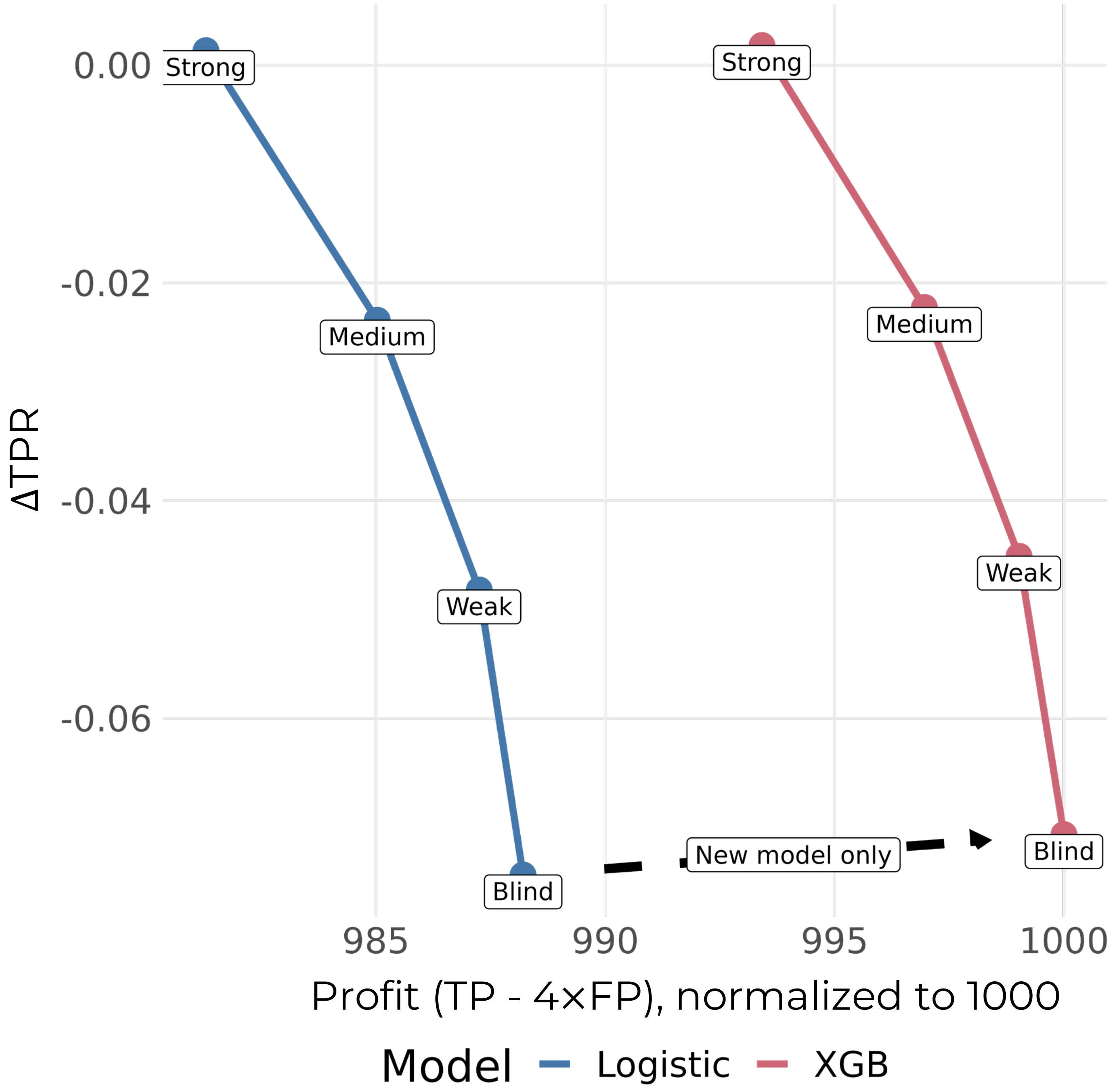


Fairness—profit tradeoff

- Regulator cares about equalizing TPR
- Lender cares about profit: TP - λFP
- We set λ to 4

ΔTPR = TPR(non-LMI) – TPR(LMI)

Fairness constraints:
Strong: 0 ΔTPR in-sample
Medium: 2/3 between Strong and Blind
Weak: 1/3 between Strong and Blind
Blind: Same threshold for non-LMI and LMI

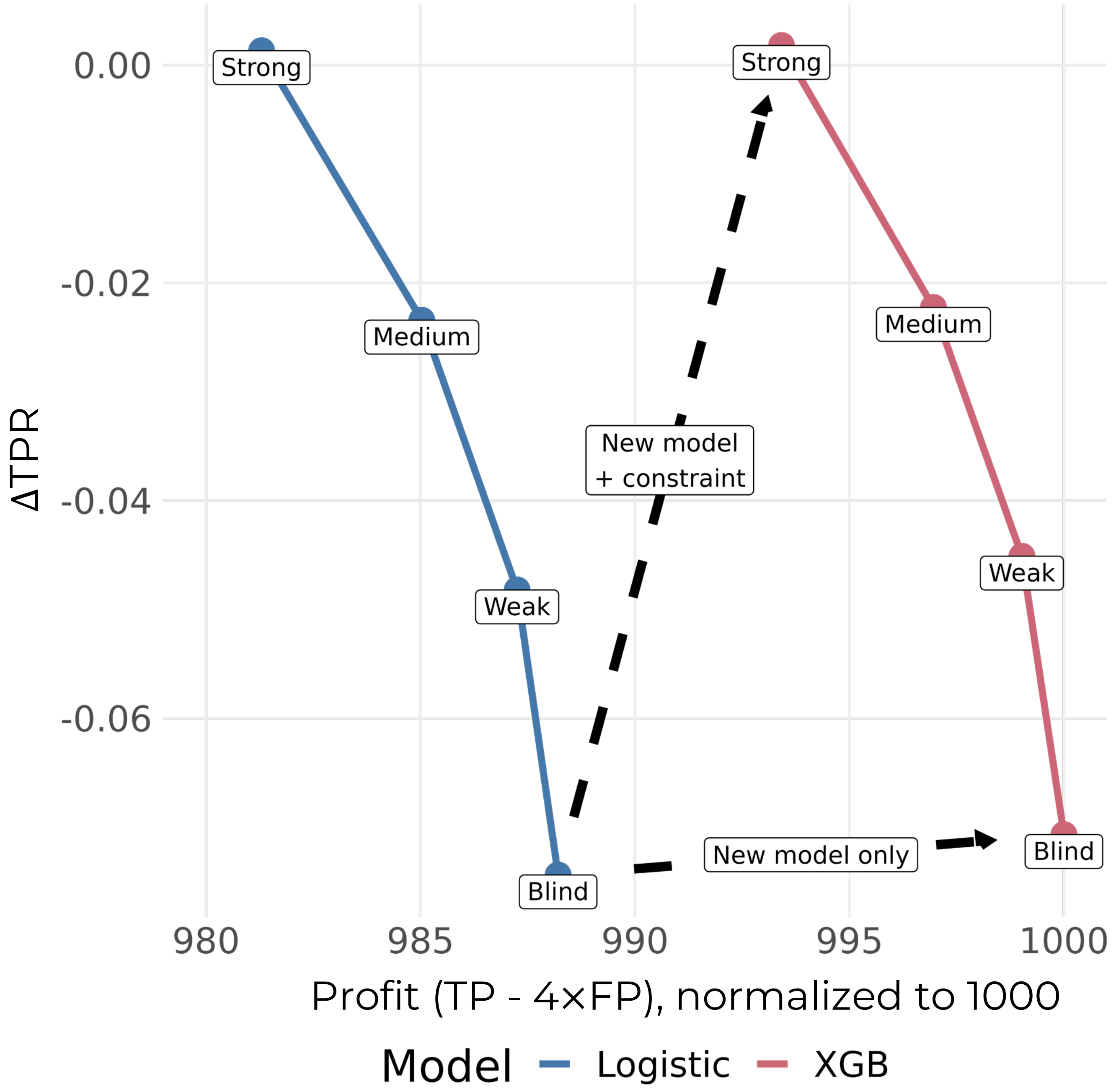


Fairness—profit tradeoff

- Regulator cares about equalizing TPR
- Lender cares about profit: TP - λFP
- We set λ to 4

ΔTPR = TPR(non-LMI) – TPR(LMI)

Fairness constraints:
Strong: 0 ΔTPR in-sample
Medium: 2/3 between Strong and Blind
Weak: 1/3 between Strong and Blind
Blind: Same threshold for non-LMI and LMI

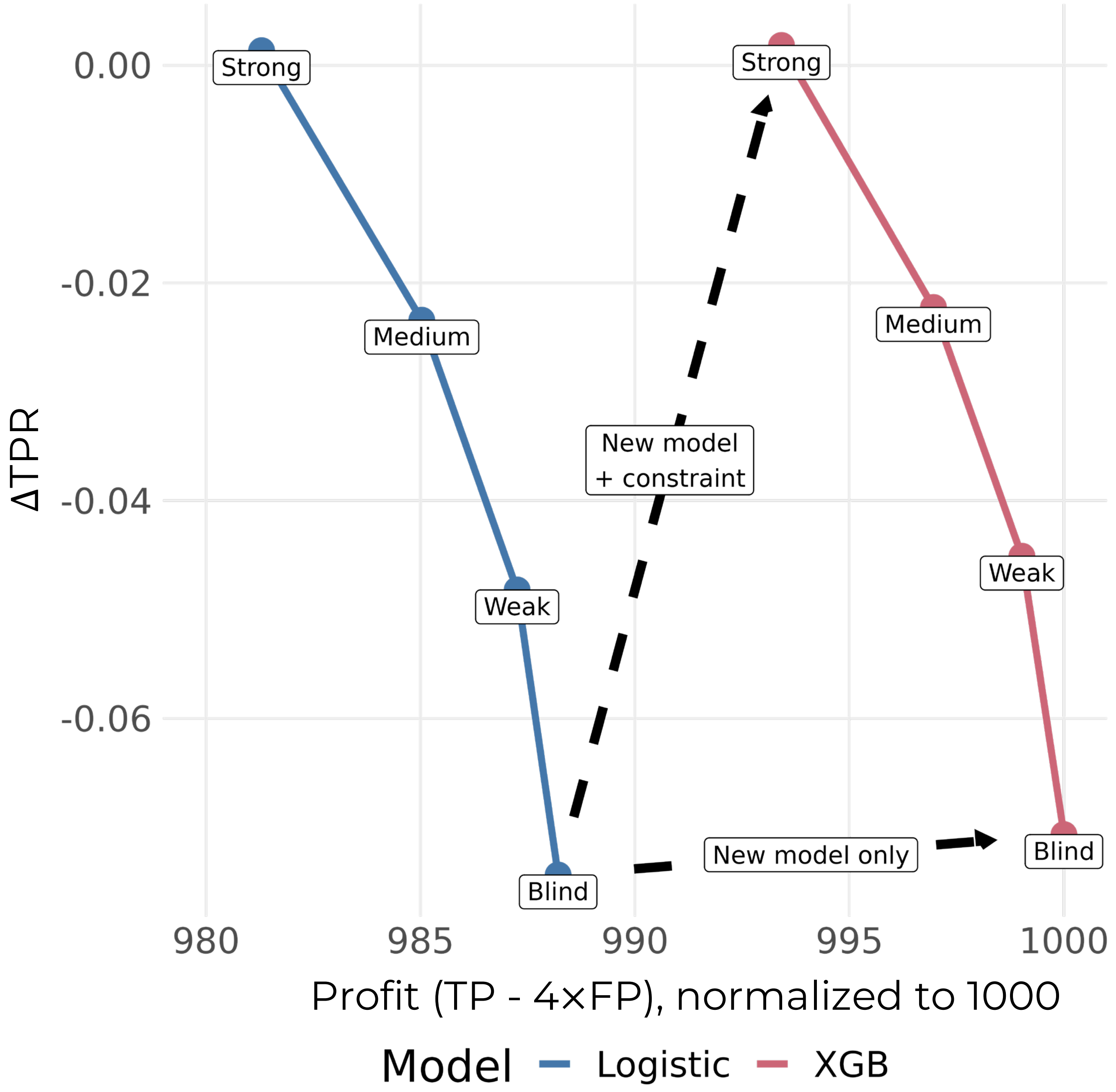


Fairness—profit tradeoff

- If better models are coupled with fairness constraints, profits rise and fairness improves

ΔTPR = TPR(non-LMI) – TPR(LMI)

Fairness constraints:
Strong: 0 ΔTPR in-sample
Medium: 2/3 between Strong and Blind
Weak: 1/3 between Strong and Blind
Blind: Same threshold for non-LMI and LMI



Separate thresholds are not a silver bullet…

Separate thresholds are not a silver bullet...

- There are ethical, legal and practical reasons behind the current framework

Separate thresholds are not a silver bullet…

- There are ethical, legal and practical reasons behind the current framework
- Legislation is naturally concerned with lenders using protected attributes like race and income to discriminate

Separate thresholds are not a silver bullet…

- There are ethical, legal and practical reasons behind the current framework
- Legislation is naturally concerned with lenders using protected attributes like race and income to discriminate
- ML adoption itself is meeting challenges

Separate thresholds are not a silver bullet…

- There are ethical, legal and practical reasons behind the current framework
- Legislation is naturally concerned with lenders using protected attributes like race and income to discriminate
- ML adoption itself is meeting challenges
- Evaluating long-term impact of separate threshold policy is challenging

Separate thresholds are not a silver bullet…

- There are ethical, legal and practical reasons behind the current framework
- Legislation is naturally concerned with lenders using protected attributes like race and income to discriminate
- ML adoption itself is meeting challenges
- Evaluating long-term impact of separate threshold policy is challenging

…but hopefully this talk at least made you think about the unintended consequences of protected attribute blindness

2020’s might be the new 1970’s

1970’s

ECOA constrained the use of technological advances to limit discrimination

My concern with putting it in like this is that I'm worried we may be put in a position where if it can be -- where when we allow statistically sound information to be used in order -- in a point scoring system, I'm not sure we don't run the risk of allowing discriminatory statistically sound data to be used in the point scoring system. I guess I just don't like the point scoring system.

Senator Joseph Biden,
markup session of the Senate Banking Committee,
discussion of ECOA amendments, 09/29/1975

2020's might be the new 1970's

1970's

ECOA constrained the use of technological advances to limit discrimination

My concern with putting it in like this is that I'm worried we may be put in a position where if it can be -- where when we allow statistically sound information to be used in order -- in a point scoring system, I'm not sure we don't run the risk of allowing discriminatory statistically sound data to be used in the point scoring system. I guess I just don't like the point scoring system.

Senator Joseph Biden,
markup session of the Senate Banking Committee,
discussion of ECOA amendments, 09/29/1975

2020's

A new wave of technological advances, renewed interest in disparities In lending

"When consumers and regulators do not know how decisions are made by the algorithms, consumers are unable to participate in a fair and competitive market free from bias."

Director Rohit Chopra
joint DOJ, CFPB, and OCC Press Conference, 10/22/2021

New aspect: more attention to disparities in outcomes

Conclusion

- Better models lead to better default prediction, but gaps in predictive power remain

Conclusion

- Better models lead to better default prediction, but gaps in predictive power remain
- Separate thresholds are a way to reduce disparities in TPR at some cost to profits

Conclusion

- Better models lead to better default prediction, but gaps in predictive power remain
- Separate thresholds are a way to reduce disparities in TPR at some cost to profits
- The costs can be mitigated by linking fairness constraints to model change

Conclusion

- Better models lead to better default prediction, but gaps in predictive power remain
- Separate thresholds are a way to reduce disparities in TPR at some cost to profits
- The costs can be mitigated by linking fairness constraints to model change
- Under the right conditions, explicit use of sensitive attributes can reduce disparities in outcomes

Thank you!