Measuring Fairness in the US mortgage market[a]

Hadi Elzayn[1] Simon Freyaldenhoven[2] Minchul Shin[2]

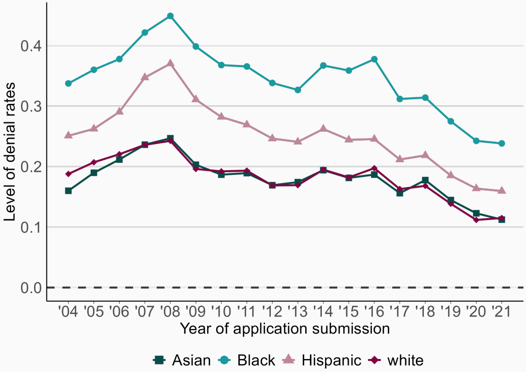
[1]Stanford University

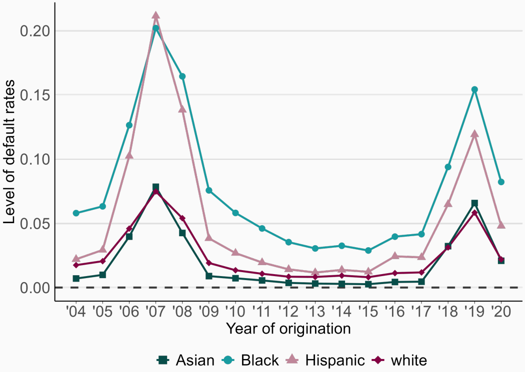
[2]Federal Reserve Bank of Philadelphia

FEDERAL RESERVE BANK
PHILADELPHIA

[a]The views expressed in this presentation are solely those of the authors and do not necessarily reflect the views of the Federal Reserve Bank of Philadelphia or the Federal Reserve System.

Denial and Default rates over time





Fairness in the US mortgage market

Does the outcome of this market look fair?

- Need a definition of fairness.
- Different definitions proposed in the academic literature; e.g. *Translation tutorial: 21 fairness definitions and their politics* [Narayanan, 2018]
- Different definitions used in public debate/news media News headlines
- Different definitions alluded to in the law/regulatory guidance.

Fairness in the US mortgage market

Does the outcome of this market look fair?

- Need a definition of fairness.
- Different definitions proposed in the academic literature; e.g. *Translation tutorial: 21 fairness definitions and their politics* [Narayanan, 2018]
- Different definitions used in public debate/news media News headlines
- Different definitions alluded to in the law/regulatory guidance.

Does it matter what definition of Fairness we use?

Why the US Mortgage market?

1. Mortgage balances are the largest source of debt for most Americans.
2. Mortgage market plays a prominent role in the persistence of wealth gaps across generations (Charles and Hurst [2003], Kuhn et al. [2020]).
3. Mortgage underwriting has seen a shift towards algorithmic underwriting over the past two decades.
4. Protected attributes of the applicant (e.g. race or gender) are directly observable.

Measuring Fairness in consumer finance and the US mortgage market

We're in good company at the Philadelphia Fed!

- Giacoletti, Heimer, and Yu [2022]
- Bhutta and Hizmo [2021]; Bhutta, Hizmo, and Ringo [2022]
- Fuster, Plosser, Schnabl, and Vickery [2019]
- Conklin, Gerardi, and Lambie-Hanson [2022]
- An, Cordell, Geng, and Lee [2022]
- Meursault, Moulton, Santucci, and Schor [2022]
- Carlin and Divringi [2018]

Also of note: Federal Reserve Bulletin (now CFPB). Most papers focus on one particular measure.

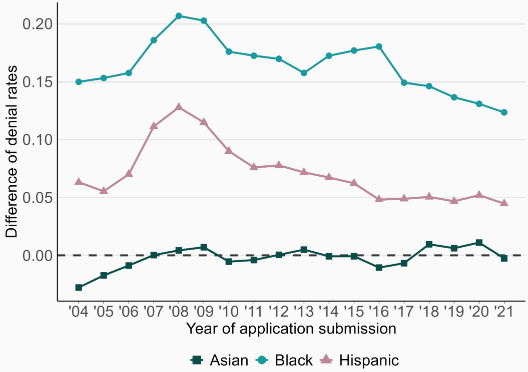
Defining Fairness

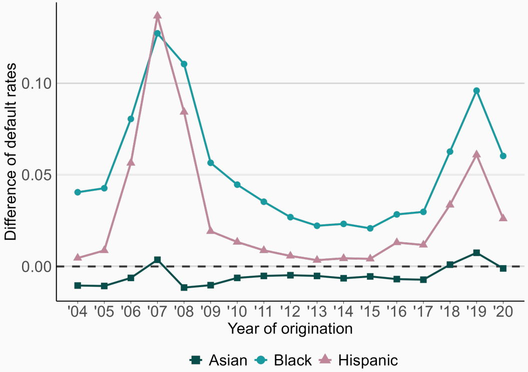
We consider 5 classes of Fairness definitions:

1. Statistical Parity
2. Predictive Parity
3. Marginal Outcome test
4. Equalized Odds
5. Conditional Statistical Parity

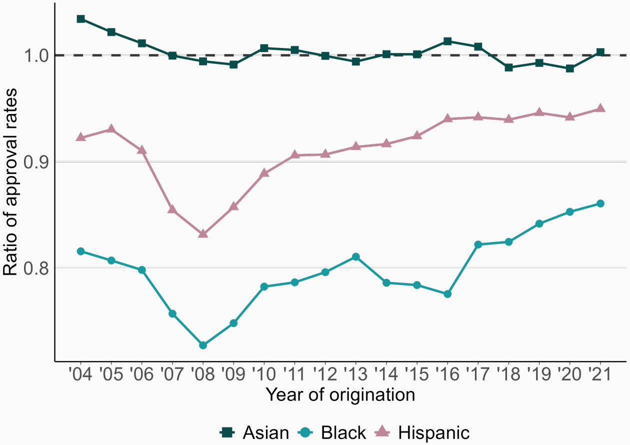
Adverse impact (29 CFR § 1607.4):
“A selection rate for any race [...] which is less than four-fifths [..] will generally be regarded [..] as evidence of adverse impact.”

Statistical Parity and Predictive Parity





The four-fifths rule



Defining Fairness

We consider 5 classes of Fairness definitions:

1. Statistical Parity
2. Predictive Parity
3. Marginal Outcome test
4. Equalized Odds
5. Conditional Statistical Parity

Equal Credit Opportunity Act (Regulation B), 12 CFR § 1002.4(a):
“Equal credit standards”

Marginal outcome test

- Economics has traditionally used the outcomes of "marginal candidates" to assess whether credit standards are equal
- Idea: If the same risk threshold is used for Black and white applicants, people at the threshold should default at the same rate

Interagency Fair Lending Examination Procedures

"The examiner-in-charge should, during the following steps, judgmentally select from the initial sample only those denied and approved applications which constitute marginal transactions."

Marginal outcome test

- Issue: Obtaining “marginal applicants” is often difficult.

Interagency Fair Lending Examination Procedures

“The examiner-in-charge should, during the following steps, judgmentally select from the initial sample only those denied and approved applications which constitute marginal transactions.”

- We develop a novel way to construct marginal applicants by identifying candidates who apply for multiple mortgages.
- An applicant is marginal if she has one approval and one denial.

Identifying Crossapplicants

We use Machine Learning to transform cHMDA into an applicant level dataset.

1. Split the data into partitions, characterized by the distinct outcomes of 9 categorical variables.
   Example: census tract of the property

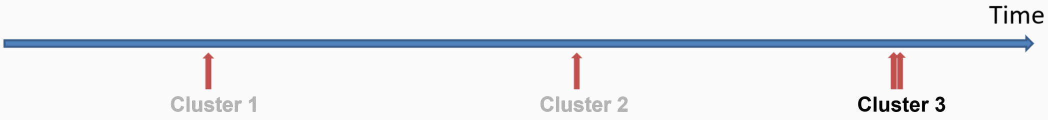
2. Apply a state-of-the-art agglomerative hierarchical clustering algorithm to further break down these partitions into clusters based on 5 additional continuous attributes.
   Example: application date

Identifying Crossapplicants

We use Machine Learning to transform cHMDA into an applicant level dataset.

1. Split the data into partitions, characterized by the distinct outcomes of 9 categorical variables.
   Example: census tract of the property

2. Apply a state-of-the-art agglomerative hierarchical clustering algorithm to further break down these partitions into clusters based on 5 additional continuous attributes.
   Example: application date



Identifying Crossapplicants

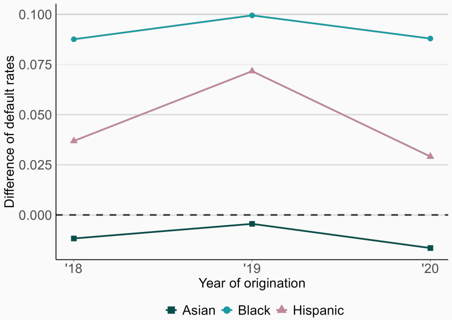
- All applications in a given cluster are “near-identical”.
- Rate at which clusters contain multiple origination informative about whether clusters correspond to single individuals.
  - Bad algorithm - all clusters are pairs of applications from two applicants: Most clusters with two approvals have two originations
  - Perfect algorithm - all clusters are pairs of applications from one applicant: No clusters with two approvals have two originations
- We estimate that 92% of clusters represent single individuals.

Marginal Candidates

- Can subset to those clusters that contain both one approval, and one denial.
- Individuals that submitted two near-identical applications:
  - One loan officer approved the application
  - Another loan officer denied the application
- We consider such an applicant marginal by "revealed preference".

Marginal Candidates

- If same risk threshold is used across lenders, default rate should be the same across groups
- Marginal Black applicants default more frequently
- Not consistent with higher lending standards for Black applicants
- Fairness violation is “negative” for this definition



Defining Fairness - feasibly

We consider 6 broad classes of Fairness definitions:

1. Statistical Parity
2. Predictive Parity
3. Marginal Outcome test
4. Equalized Odds
5. Conditional Statistical Parity

Equal Credit Opportunity Act (Regulation B), 12 CFR § 1002.1(b): “availability of credit to all creditworthy applicants without regard to race [...]”

Equalized Odds - Equality of Opportunity

- Consider crossapplicants with an originated loan that did not default.
- How likely were they denied at least once?
- Intuitively captures notion of “unfairly denied”

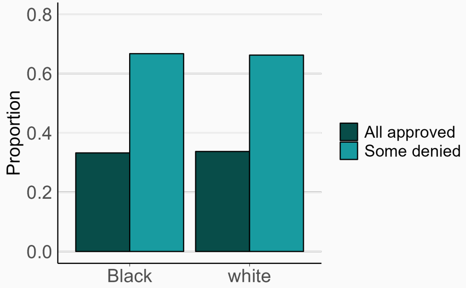
Does the rate of “unfair denials” vary by race?

Equalized Odds - Equality of Goodwill

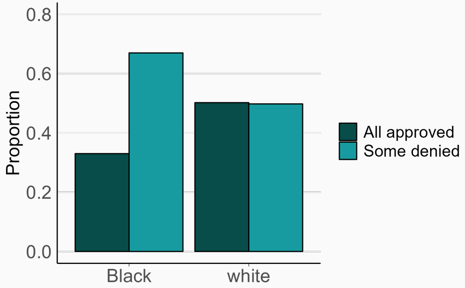
- Consider crossapplicants with an originated loan that did default.
- How likely were they approved on all their applications?
- Intuitively captures notion of “unfairly approved”

Does the rate of “unfair approvals” vary by race?

Equalized Odds (based on crossapplicants)



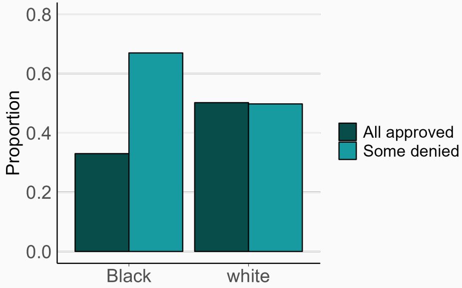
Defaulters



Non-defaulters

Equality of Opportunity (based on crossapplicants)

- Black applicants:
  67% chance at least one denial for non-defaulters
- white applicants:
  50% chance at least one denial for non-defaulters
- Intuitively, captures the rate of being “unfairly denied”



Non-defaulters

Defining Fairness - feasibly

We consider 5 classes of fairness definitions:

1. Statistical Parity
2. Predictive Parity
3. Marginal Outcome test
4. Equalized Odds
5. Conditional Statistical Parity

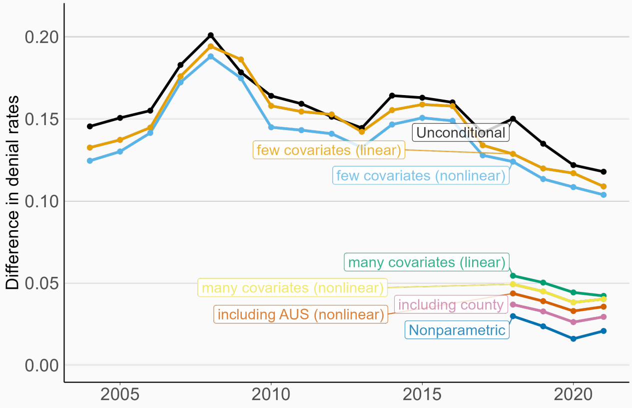
Disparate treatment/impact: “disparate impact tests should only include controls for attributes that are plausibly business justified.”[Ayres, 2010]

Conditional Statistical Parity

Conditional Statistical Parity measures the difference in denial rates across groups, conditional on varying information sets (e.g. a set of covariates X).

- Choice of X is extremely important
- Potential for
  1. missing covariates
  2. included variable bias

Conditional Statistical Parity



Specifications differ in

1. set of covariates
2. functional form

Interactive dashboard

- All results are based on two works-in-progress (joint works with Minchul and Hadi Elzayn)
- Interactive Appendix available, allows the user to explore the different measures we discussed today (and more!) across time and space
- Hoping to expand this to other protected attributes, and make this available through the Philadelphia Fed to Researchers and the general public.

Taking stock

Does it matter what definition of Fairness we use?

- Yes, different definitions lead to different results.
- No “right" definition, but will be context-dependent
- Even then, perhaps a more comprehensive view considers multiple definitions.

Defining Fairness - Theory

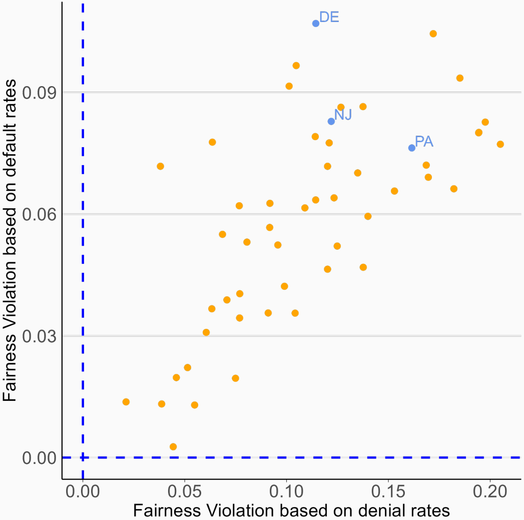
- Theoretical results exist that prove how seemingly reasonable definition are at odds with each other.
- *Inherent Trade-Offs in the Fair Determination of Risk Scores* [Kleinberg, Mullainathan, and Raghavan, 2017] considers three widely used definitions of fairness.
  (related to: Marginal Outcome test, Equality of Opportunity, Equality of Goodwill)

Kleinberg et al. [2017]:

"We prove that except in highly constrained special cases, there is no method that can satisfy these three conditions simultaneously."

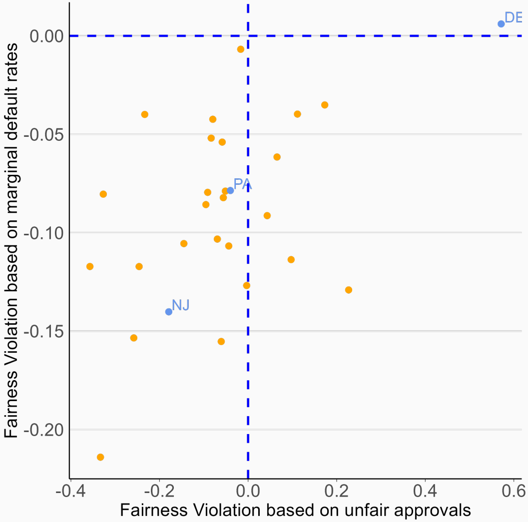
Stylized Facts

- Broad measures point to systemic inequality between demographic groups
- Not necessarily reflective of discrimination in the mortgage sector



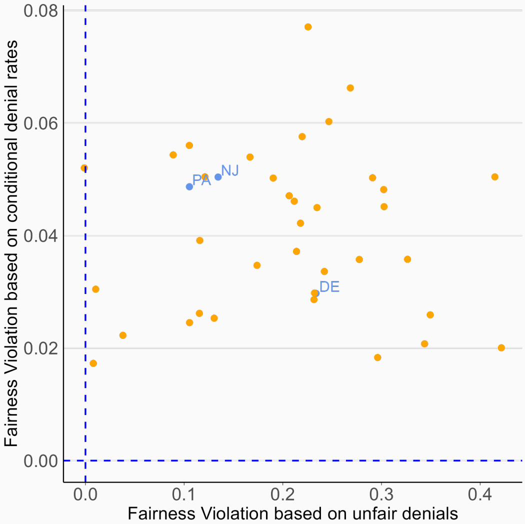
Stylized Facts

- Isolating the mortgage decision yields a more nuanced picture
- On the one hand:
  1. Marginal candidates more likely to default in minority group
  2. “Defaulters” denied at same rate across demographic groups

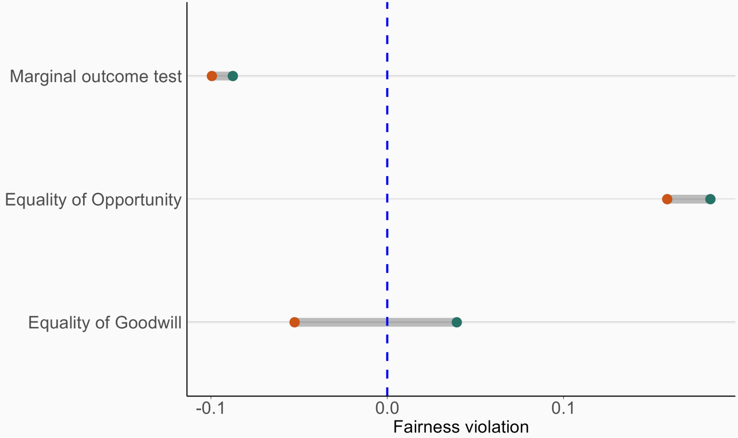


Stylized Facts

- Isolating the mortgage decision yields a more nuanced picture
- On the other hand:
    1. Minority applicants more likely to be denied conditional on rich set of covariates
    2. Minority “Non-defaulters” denied at substantially higher rates



Fairness definition matters



Range for each definition based on yearly measures for 2018-2020

Conclusion

- Definition of fairness matters, both theoretically and empirically.
- Any one measure (or study) may not paint the full picture

Conclusion

- Definition of fairness matters, both theoretically and empirically.
- Any one measure (or study) may not paint the full picture

Does the outcome of the US mortgage market look fair?

- Broad measures of fairness point to systematic inequality
- More narrow measures are more ambiguous

Thank you

References

Xudong An, Larry Cordell, Liang Geng, and Keyoung Lee. Inequality in the time of covid-19: Evidence from mortgage delinquency and forbearance. *Available at SSRN 3789349*, 2022.

Ian Ayres. Testing for discrimination and the problem of “included variable bias”. *Yale Law School Mimeo*, 2010.

Neil Bhutta and Aurel Hizmo. Do minorities pay more for mortgages? *The Review of Financial Studies*, 34(2):763–789, 2021.

Neil Bhutta, Aurel Hizmo, and Daniel Ringo. How much does racial bias affect mortgage lending? evidence from human and algorithmic credit decisions. Technical report, FEDS Working Paper, 2022.

Alex Carlin and Eileen Divringi. Home improvement lending in the third federal reserve district: Patterns by income, race, and gender. *Cascade Focus*, 2018.

Kerwin Kofi Charles and Erik Hurst. The correlation of wealth across generations. *Journal of Political Economy*, 111(6):1155–1182, 2003.

James Conklin, Kristopher Gerardi, and Lauren Lambie-Hanson. Can everyone tap into the housing piggy bank? racial disparities in access to home equity. 2022.

Andreas Fuster, Matthew Plosser, Philipp Schnabl, and James Vickery. The role of technology in mortgage lending. *The Review of Financial Studies*, 32(5):1854–1899, 2019.

Marco Giacoletti, Rawley Heimer, and Edison G Yu. Using high-frequency evaluations to estimate discrimination: Evidence from mortgage loan officers. In *Working Paper*, 2022.

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. In Christos H. Papadimitriou, editor, *8th Innovations in Theoretical Computer Science Conference*, volume 67, pages 43:1–43:23. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik, 2017.

Moritz Kuhn, Moritz Schularick, and Ulrike I Steins. Income and wealth inequality in america, 1949–2016. *Journal of Political Economy*, 128(9): 3469–3519, 2020.

Vitaly Meursault, Daniel Moulton, Larry Santucci, and Nathan Schor. The time is now: Advancing fairness in lending through machine learning. *FRB of Philadelphia Working Paper No. 22-39*, 2022.

Arvind Narayanan. Translation tutorial: 21 fairness definitions and their politics. In *Proc. conf. fairness accountability transp., New York, USA*, volume 1170, page 3, 2018.

Fairness in the US mortgage market

AP

U.S. WORLD POLITICS VIDEO SPOTLIGHT ENTERTAINMENT SPORTS BUSINESS SCIENCE FACT CHECK CLIMATE ...

Lolita the orca dies | Latest on Hurricane Hilary | American Airlines flight crackdown | In Maui, concerns of 'climate gentrification'

The secret bias hidden in mortgage-approval algorithms

Racial Bias in Mortgage Biz? New Data Says No, Researchers Find

By Edward J. Pinto | Tobias Peter | Randall Bloomquist

October 01, 2021

CNBC

Search quotes, news & videos

MARKETS BUSINESS INVESTING TECH POLITICS CNBC TV INVESTING CLUB PRO MAKE

PERSONAL FINANCE

Mortgage denial rate for Black borrowers is twice that of overall population, report finds

PUBLISHED SAT, AUG 27 2022•8:30 AM EDT

Why Default Rates Cannot Shed Light on Mortgage Discrimination

Race and Economy

How mortgage algorithms perpetuate racial disparity in home lending

David Brancaccio and Rose Conlon | Aug 25, 2021

HOUSING

Home lending remains unequal

A new study shows that the homeownership gap between Blacks and whites was wider in 2020 than at any point in the 20th century.

back