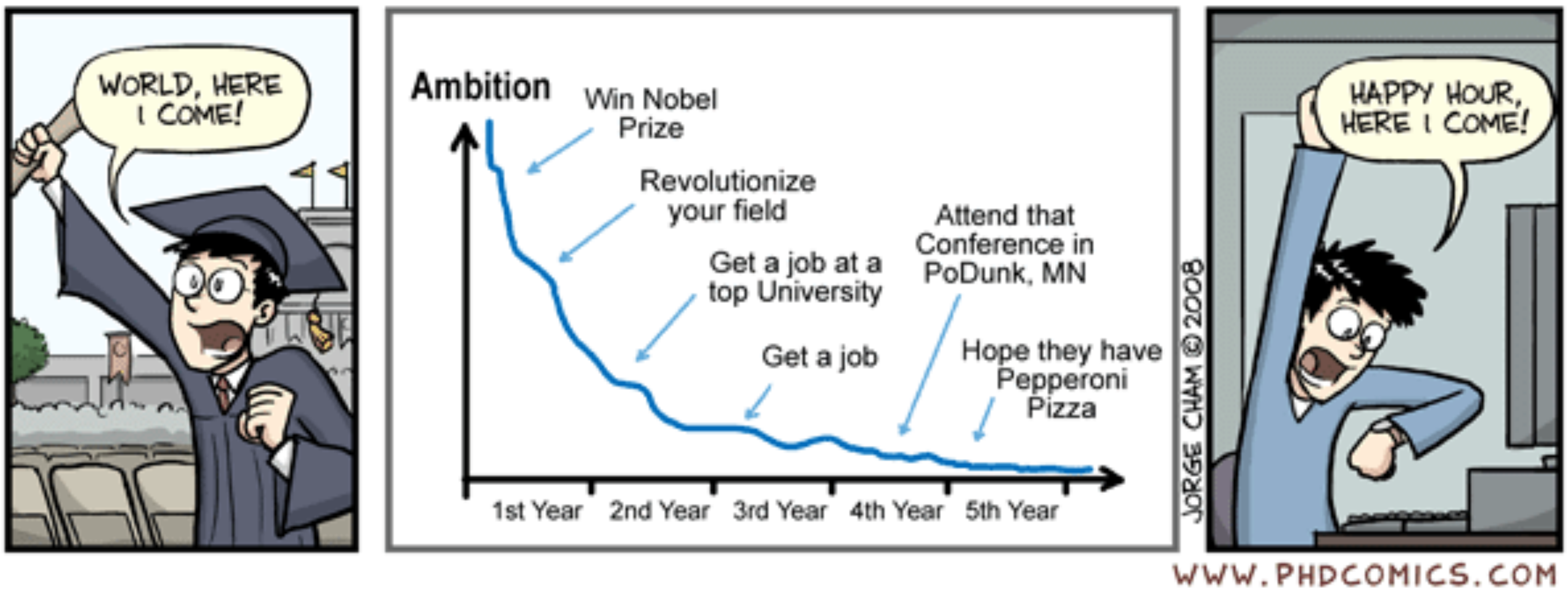
Discussion of Liu (2023)

"Language Frictions in Consumer Credit"

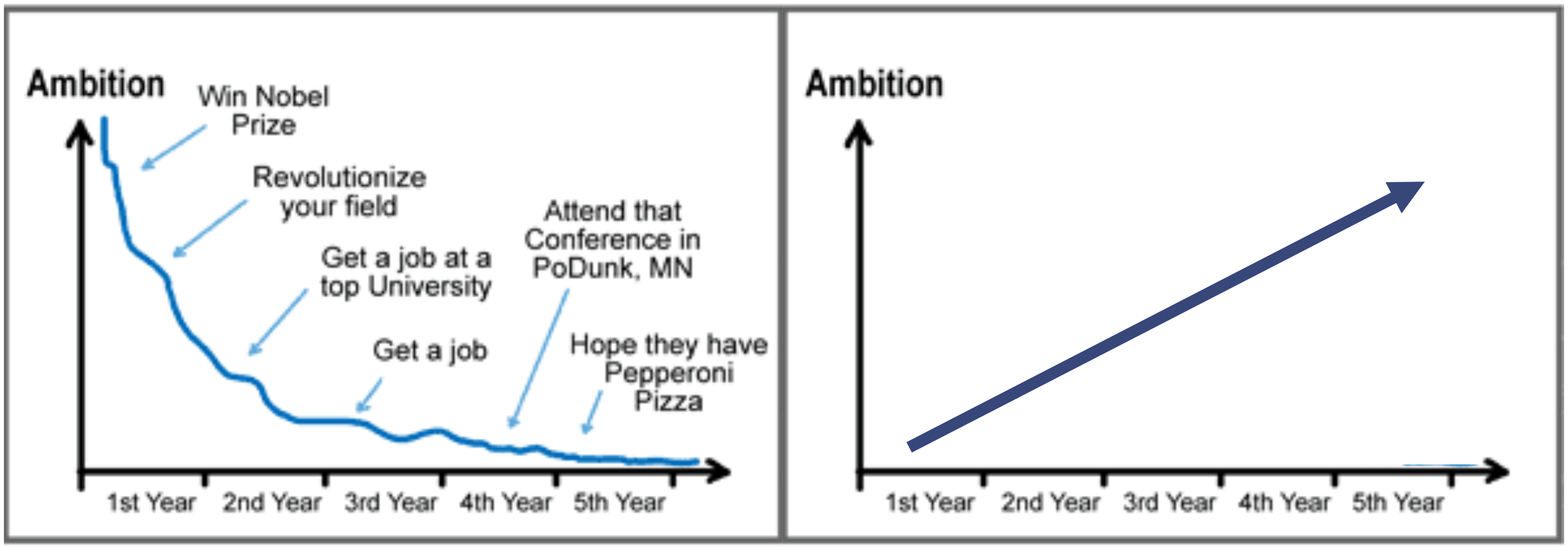
Discussant: Sangmin Simon Oh (Chicago Booth)

MFA 2024

Congratulations on finishing the job market!



Congratulations on finishing the job market!



Recap

Objective

- Study the role of language frictions in accessing mortgage credit

Approach

- National Survey of Mortgage Originations (NSMO) → Individual-level Analysis
- Home Mortgage Disclosure Act (HMDA) → County-level Analysis
- Variation: (i) Policy Shock, (ii) LEP vs. Non-LEP, (iii) Hispanic vs. Non-Hispanic

Result

- Reducing language frictions have a large impact throughout various stages of the process of attaining mortgage credit
  - Dimensions: (i) Application Process, (ii) Approval, (iii) Price

Recap

Objective

- Study the role of language frictions in accessing mortgage credit

Approach

- National Survey of Mortgage Originations (NSMO) → Individual-level Analysis
- Home Mortgage Disclosure Act (HMDA) → County-level Analysis
- Variation: (i) Policy Shock, (ii) LEP vs. Non-LEP, (iii) Hispanic vs. Non-Hispanic

Result

- Reducing language frictions have a large impact throughout various stages of the process of attaining mortgage credit
  - Dimensions: (i) Application Process, (ii) Approval, (iii) Price

Job Market Paper!

In my opinion, this is a great paper to start a new agenda, and I will try to be as constructive as possible in my comments today.

Recap

Objective

- Study the role of language frictions in accessing mortgage credit

Approach

- National Survey of Mortgage Originations (NSMO) → Individual-level Analysis
- Home Mortgage Disclosure Act (HMDA) → County-level Analysis
- Variation: (i) Policy Shock, (ii) LEP vs. Non-LEP, (iii) Hispanic vs. Non-Hispanic

Result

- Reducing language frictions have a large impact throughout various stages of the process of attaining mortgage credit
  - Dimensions: (i) Application Process, (ii) Approval, (iii) Price

Job Market Paper!

In my opinion, this is a great paper to start a new agenda, and I will try to be as constructive as possible in my comments today.

Plan for Discussion

1. Measurement Errors
2. Contemplating Policy Implications
3. Framing

Point 1. Measurement Errors

Main Independent Variable: LEP_i

$$y_{it} = \alpha + \beta_0 LEP_i + \beta_1 Hispanic_i + \beta_2 LEP_i \times Hispanic_i + \beta_3 LEP_i \times Post_t$$
$$+ \beta_4 Hispanic_i \times Post_t + \beta_5 LEP_i \times Hispanic_i \times Post_t + \gamma X_{it} + \delta_t + \epsilon_{it}. \quad (2)$$

- **Goal**: Measure the English proficiency of a given borrower i

Main Independent Variable: LEP_i

$$y_{it} = \alpha + \beta_0 LEP_i + \beta_1 Hispanic_i + \beta_2 LEP_i \times Hispanic_i + \beta_3 LEP_i \times Post_t + \beta_4 Hispanic_i \times Post_t + \beta_5 LEP_i \times Hispanic_i \times Post_t + \gamma X_{it} + \delta_t + \epsilon_{it}. \quad (2)$$

- **Goal**: Measure the English proficiency of a given borrower i
- **Approach 1**. Measure LEP_i using survey
  - $LEP_i = 1$ if the answer is “Yes” to the following question:

Is speaking my primary language, which is not English, important in choosing the mortgage lender or broker?

Main Independent Variable: LEP_i

$$y_{it} = \alpha + \beta_0 LEP_i + \beta_1 Hispanic_i + \beta_2 LEP_i \times Hispanic_i + \beta_3 LEP_i \times Post_t + \beta_4 Hispanic_i \times Post_t + \beta_5 LEP_i \times Hispanic_i \times Post_t + \gamma X_{it} + \delta_t + \epsilon_{it}. \quad (2)$$

- **Goal**: Measure the English proficiency of a given borrower i
- **Approach 1**. Measure LEP_i using survey
  - $LEP_i = 1$ if the answer is "Yes" to the following question:

    Is speaking my primary language, which is not English, important in choosing the mortgage lender or broker?

- **Approach 2**. Measure LEP_i using ML + borrower characteristics
  - Gender, race, ethnicity, household income + state-year FE
  - Train XGBoost on 2015-19 American Community Survey (ACS)
  - Use the model to predict borrowers' LEP status in the HMDA+ data

What are some concerns with approach 2?

- Author: “Since the machine learning prediction generates measurement error of the key independent variable, it is necessary to gauge the bias of the triple-difference estimation that uses the predicted LEP status.”

What are some concerns with approach 2?

- Author: "Since the machine learning prediction generates measurement error of the key independent variable, it is necessary to gauge the bias of the triple-difference estimation that uses the predicted LEP status."

- Issue 1. Non-Classical Measurement Error

  - Suppose we want to estimate $y = \beta x + \epsilon$ but we only have data on $\tilde{x} = x + u$ with $E[u] = 0$. Also assume $Cov(u, x) = 0, Cov(u, \epsilon) = 0$.

  - Then, we have

$$\hat{\beta} = \frac{Cov(x+u, \beta x+\epsilon)}{Var(x+u)} = \beta \frac{Var(x)}{Var(x)+Var(u)}$$

  and $\hat{\epsilon} = \mathrm{y} - \hat{\beta}\tilde{\mathrm{x}} = \mathrm{y} - \hat{\beta}(\mathrm{x}+\mathrm{u}) = \epsilon + \left(\beta - \hat{\beta}\right)x - \hat{\beta}u$.

  - Now when x is a binary variable, we have $Cov(x, u) < 0$.

  - This is what Section VI.B of the paper addresses.

What are some concerns with approach 2?

- Author: “Since the machine learning prediction generates measurement error of the key independent variable, it is necessary to gauge the bias of the triple-difference estimation that uses the predicted LEP status.”

- Issue 2. ML-Specific Considerations

  - Zhang, Xue, Yu, and Tan (2023):
    - When the regressors are generated using ML/AI and directly plugged in regression models, then we can either have over- or under-estimation, depending on the error structures.

What are some concerns with approach 2?

- Author: “Since the machine learning prediction generates measurement error of the key independent variable, it is necessary to gauge the bias of the triple-difference estimation that uses the predicted LEP status.”

- Issue 2. ML-Specific Considerations

  - Zhang, Xue, Yu, and Tan (2023):
    - When the regressors are generated using ML/AI and directly plugged in regression models, then we can either have over- or under-estimation, depending on the error structures.

  - Wei and Malik (2023):
    - Borrowers with LEP = 1 may be on average different from other borrowers with LEP = 0 on “peripheral” features, which may be useful for predicting English proficiency.
    - Concern if these “peripheral” features affect mortgage outcomes

What are some concerns with approach 2?

- Author: “Since the machine learning prediction generates measurement error of the key independent variable, it is necessary to gauge the bias of the triple-difference estimation that uses the predicted LEP status.”

- Issue 2. ML-Specific Considerations

    - Zhang, Xue, Yu, and Tan (2023):
        - When the regressors are generated using ML/AI and directly plugged in regression models, then we can either have over- or under-estimation, depending on the error structures.

    - Wei and Malik (2023):
        - Borrowers with LEP = 1 may be on average different from other borrowers with LEP = 0 on “peripheral” features, which may be useful for predicting English proficiency.
        - Concern if these “peripheral” features affect mortgage outcomes

    - References: Yang et al. (2018), Fong and Tyler (2021), Qiao and Huang (2021), Wei and Malik (2023), Zhang, Xue, Yu, and Tan (2023)

What are some concerns with approach 2?

- Author: “Since the machine learning prediction generates measurement error of the key independent variable, it is necessary to gauge the bias of the triple-difference estimation that uses the predicted LEP status.”

Machine Learning Predictions as Regression Covariates

Christian Fong[1] and Matthew Tyler[2]

known to be unbiased, consistent, and have correct coverage. In fact, simply plugging in the ML predictions as if they were the real covariates would have caused us to arrive at the complete opposite conclusion. The problem is not addressed by intuitive strategies previous studies have employed, such as bootstrapping and integrating over the uncertainty in the predictions. While statistics and econometrics provide methods for measurement error, such as multiple imputation, we show that these solutions perform poorly when most observations contain missing data, as often happens in ML.[1]

What are some concerns with approach 2?

- Author: “Since the machine learning prediction generates measurement error of the key independent variable, it is necessary to gauge the bias of the triple-difference estimation that uses the predicted LEP status.”

Machine Learning Predictions as Regression Covariates

Christian Fong[1] and Matthew Tyler[2]

known to be unbiased, consistent, and have correct coverage. In fact, simply plugging in the ML predictions as if they were the real covariates would have caused us to arrive at the complete opposite conclusion. The problem is not addressed by intuitive strategies previous studies have employed, such as bootstrapping and integrating over the uncertainty in the predictions. While statistics and econometrics provide methods for measurement error, such as multiple imputation, we show that these solutions perform poorly when most observations contain missing data, as often happens in ML.[1]

- **Suggestion:** I encourage the author to (i) clarify the nature of the measurement error and (ii) draw on the ML + causal inference literature to adjust for sources of error

Point 2. Contemplating Policy Implications

Policy Implications

Paper highlights the merits of reducing language frictions and discusses policy implications:

Merits	Demerits
Streamlined application process	
Increased availability of credit	**?**
Lower borrowing costs	
No deterioration of mortgage risk	

Policy Implications

Paper highlights the merits of reducing language frictions and discusses policy implications:

Merits	Demerits
Streamlined application process	
Increased availability of credit	**?**
Lower borrowing costs	
No deterioration of mortgage risk	

Other Considerations:

- Demand: Do borrowers substitute away from minority / community banks?

  Hurtado and Sakong (2023), "The Effect of Minority Bank Ownership on Minority Credit"

  "Over 70% of mortgages go to borrowers of bank owners' same race"

Policy Implications

Paper highlights the merits of reducing language frictions and discusses policy implications:

Merits	Demerits
Streamlined application process	
Increased availability of credit	?
Lower borrowing costs	
No deterioration of mortgage risk	

Other Considerations:

- Demand: Do borrowers substitute away from minority / community banks?

  Hurtado and Sakong (2023), "The Effect of Minority Bank Ownership on Minority Credit"

  "Over 70% of mortgages go to borrowers of bank owners' same race"

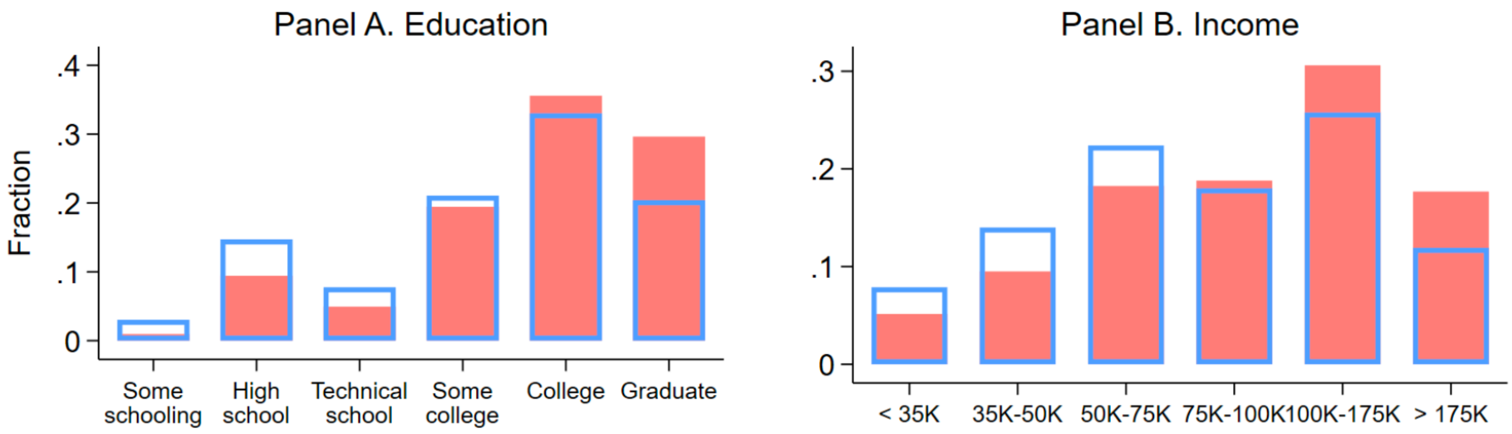
- Supply: Do banks ration credit to other types of borrowers?

  "A policy expert at FHFA said that the mortgage translation disclosure was designed to alleviate lenders' concerns about compliance risks when serving LEP borrowers."

Point 3. Framing

Framing #1: Trace Out Implications for Inequality

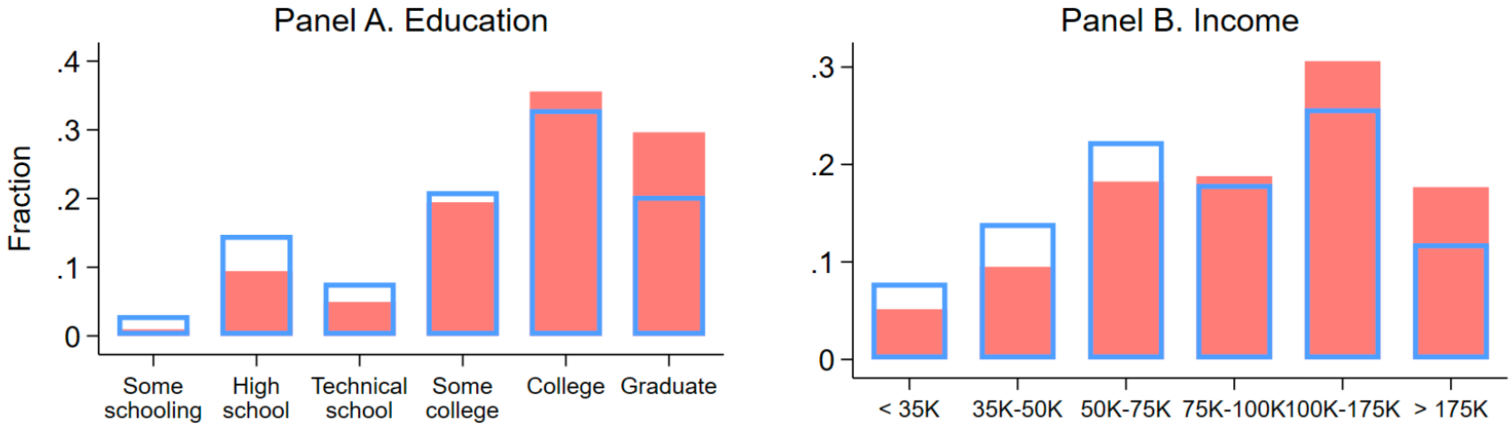
- Language frictions correlate with education attainment and levels of income



- While monetary policy does not target any particular group of people ... we are attentive to disparities in the labor market, rather than just the headline numbers" (Powell, 2019)

Framing #1: Trace Out Implications for Inequality

- Language frictions correlate with education attainment and levels of income



- While monetary policy does not target any particular group of people ... we are attentive to disparities in the labor market, rather than just the headline numbers" (Powell, 2019)

Suggestions

- Borrower-level Heterogeneity by Inequality (in opportunity, wealth, income, …)
- Does the share of LPE status vary across census tracts?
  - Footnote 6: "NSMO reports 3 types of census tract based on income"

Framing #2: Place More Emphasis on Real Outcomes

Paper has very interesting results on real outcomes:

Section IV: “LEP borrowers pay a lower interest rate when they had access to mortgage documents in their primary languages”

- Mechanism: Reducing language frictions effectively reduces search costs

Section V.C: “Reducing language frictions did not introduce extra risks to mortgage market”

- Measure(s): Average credit scores, Unconditional delinquency rates

Section V.D: “Positive policy impact on local lender competition for LEP borrowers”

- Measure(s): (i) # of active lenders and (ii) HHI in different market segments

Framing #2: Place More Emphasis on Real Outcomes

Paper has very interesting results on real outcomes:

Section IV: "LEP borrowers pay a lower interest rate when they had access to mortgage documents in their primary languages"

- Mechanism: Reducing language frictions effectively reduces search costs

Section V.C: "Reducing language frictions did not introduce extra risks to mortgage market"

- Measure(s): Average credit scores, Unconditional delinquency rates

Section V.D: "Positive policy impact on local lender competition for LEP borrowers"

- Measure(s): (i) # of active lenders and (ii) HHI in different market segments

Suggestions

- More evidence on the marginal borrower (extensive vs. intensive margin)
- Decompose alleviation of borrower (demand) vs. lender (supply)-side frictions

Final Thoughts

- Author studies a creative and important question that augments our current understanding of important frictions in the U.S. mortgage market

- **Punchline:** Reducing language frictions have a large impact throughout various stages of the process of attaining mortgage credit

- Key empirical challenge is measurement, which the author addresses in multiple ways

- I think there are ways to make the paper more appealing to a broader audience, for which I have some suggestions.

Final Thoughts

- Author studies a <u>creative</u> and important question that augments our current understanding of important frictions in the U.S. mortgage market

- **Punchline:** Reducing language frictions have a large impact throughout various stages of the process of attaining mortgage credit

- Key empirical challenge is <u>measurement</u>, which the author addresses in multiple ways

- I think there are ways to make the paper more appealing to a broader audience, for which I have some suggestions

- Congratulations again on completing the academic job market!
- Hope to see you soon in future conferences!