

Using Machine Learning to Estimate the Effect of Undocumented Status on Education-Occupation Mismatch for College Graduates

Dr. Veronica Sovero
Mario Arce Acosta

University of California, Riverside



Contents

- 1 Introduction
- 2 Data and Methods
- 3 Results
- 4 Conclusion

Introduction

Motivation

- Many states offer a combination of in-state tuition and financial aid to undocumented immigrants. ¹
- There are roughly 1.71 million undocumented immigrants with a college education residing in the United States. ²
- For college graduates, how does undocumented status affect:
 - Occupation-education mismatch?
 - Wages?

¹<https://www.higheredimmigrationportal.org/states/>

²<https://cmsny.org/educated-immigrants-millet-080122/>

Literature Review

**Undocumented
Status
Imputation**

**Van Hook,
Jennifer et al.,
(2015)**

**Ruhnke, Simon
A., Wilson,
Fernando A.,
and Stimpson,
Jim P. (2022)**

**Education-Occupation
mismatch**

Ortega, Francesc and Hsin, Amy (2018)

**Li, Xiaoguang and Lu,
Yao (2023)**

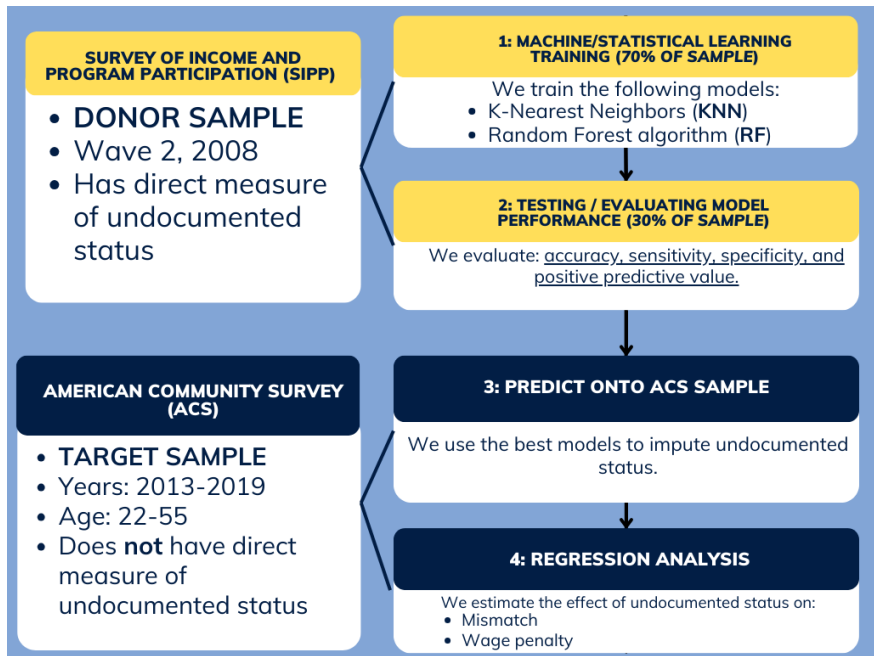
**Wage
penalties**

**Borjas, George
J. and Cassidy,
Hugh (2019)**

Our contribution

- Utilize machine learning methods for undocumented status imputation in the American Community Survey (ACS)
- Estimate the impact of undocumented status on education-occupation mismatch and wage penalties for college graduates (smaller, hidden population)
- Examine the role of federal and state-level policy on labor market outcomes for undocumented immigrants

Data and Methods



Training Sample

We apply *logical imputation* to generate the SIPP training sample, classifying a noncitizen as possibly undocumented if they **do not** meet any of the following conditions:

- Veteran status
- Medicare receipt
- Social Security income receipt
- Arrived before January 1st, 1982

Undocumented status is observed in this training sample.

K-Nearest Neighbors classifier

KNN

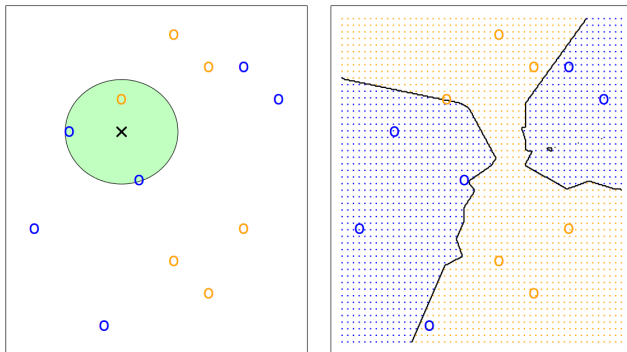
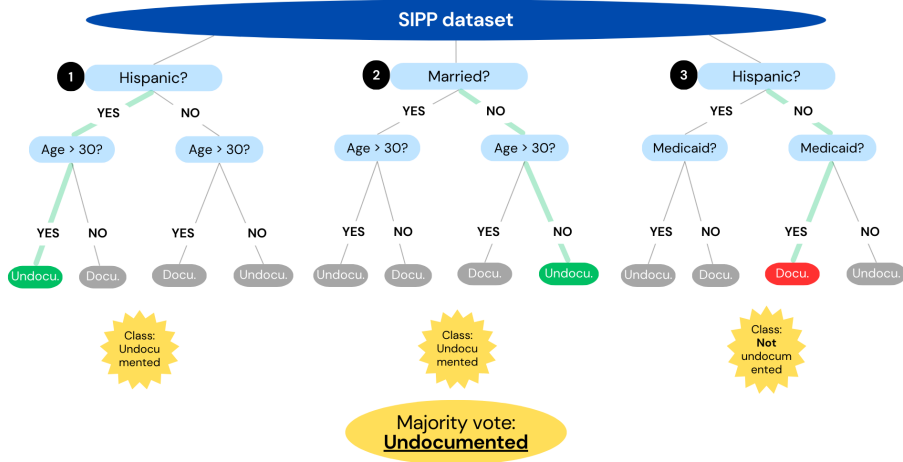


Figure: Source: *Intro to Statistical Learning with R* at <https://www.statlearning.com/>

Random Forest algorithm

RF



Model performance metrics

	Logical edits	Logistic	KNN	RF
Sensitivity	1.0000	0.6401	0.6424	0.6629
Specificity	0.9547	0.6614	0.6028	0.7190
Positive predictive value	0.2857	0.4518	0.4135	0.5070
Accuracy	0.9555	0.6549	0.6148	0.7019

Table: Imputation Methods Model Metrics

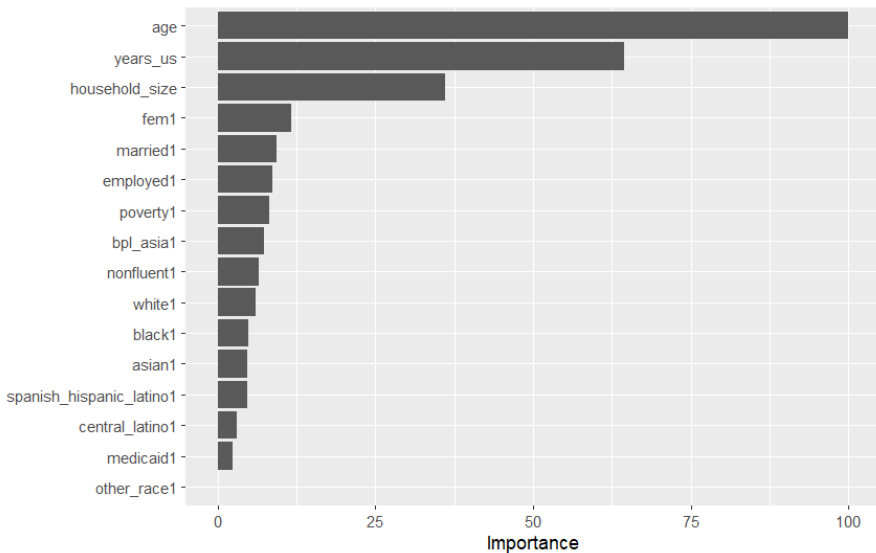
Sensitivity: $\frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$

Specificity: $\frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$

Positive-predictive value: $\frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$

Accuracy: $\frac{\text{True Positives} + \text{True Negatives}}{\text{Total}}$

Feature importance (RF model)



Defining Vertical and Horizontal mismatch

People are considered mismatched within their occupation using the following definitions:³

Vertically mismatched:

Educational attainment \neq Most common educational attainment

Horizontally mismatched:


Field of study \neq Two most common fields of study

Horizontally undermatched:

Median wage of occupation $<$ Median wage of field of study (of those horizontally matched)

Horizontally overmatched:

Median wage of occupation $>$ Median wage of field of study (of those horizontally matched)

³U.S. born workers are the reference group for modal occupations and fields of study. 

Econometric model

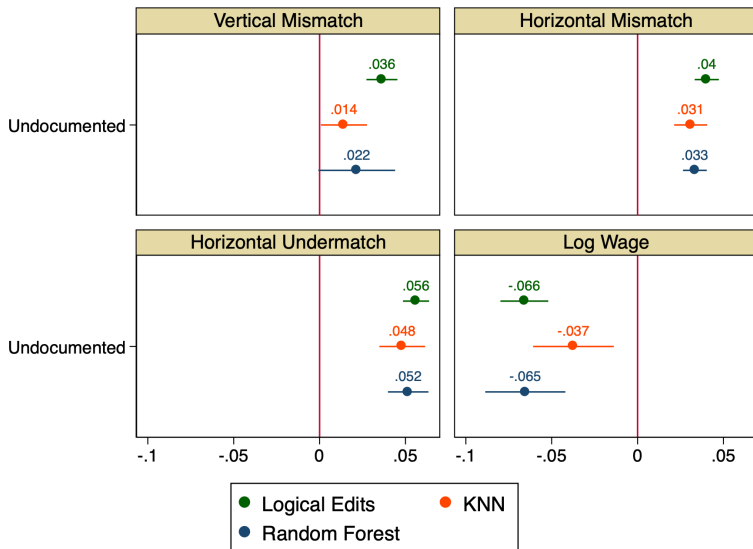
$$(1) \textit{Vertical Mismatch}_i = \beta X_i + \beta_U \textit{Hundermatch}_i + \beta_O \textit{Hovermatch}_i + \beta_u \textit{Undocu}_i + \varepsilon_i$$

$$(2) \textit{Horizontal Mismatch}_i = \beta X_i + \beta_V \textit{Vmismatch}_i + \beta_u \textit{Undocu}_i + \varepsilon_i$$

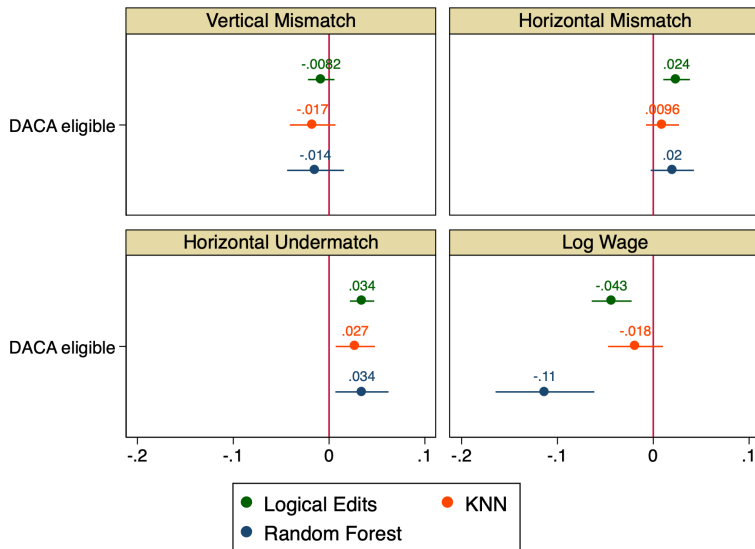
$$(3) \log wage_i = \beta X_i + \beta_V \textit{Vmismatch}_i + \beta_U \textit{Hundermatch}_i + \beta_O \textit{Hovermatch}_i + \beta_u \textit{Undocu}_i + \varepsilon_i$$

Results

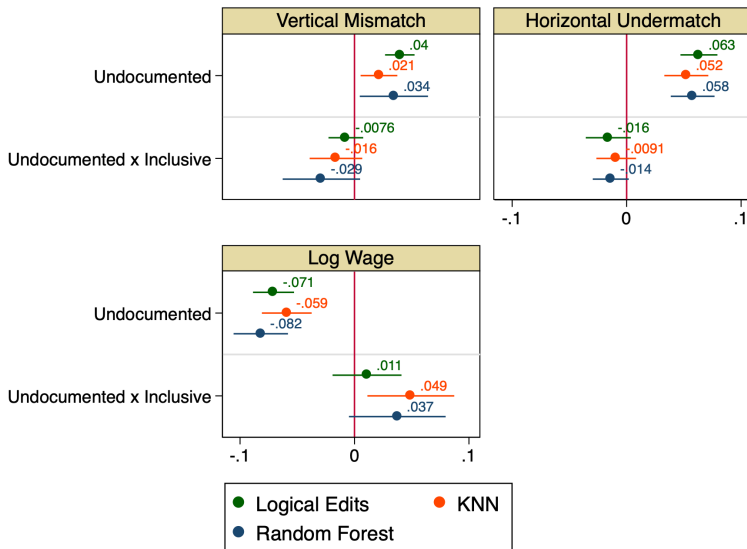
Main Results



DACA



Immigration Policy Climate (Samari Goleen, 2021)



Summary

- RF algorithm has the best positive predictive value compared to previous imputation methods
- Undocumented college graduates experience education-occupation mismatch and wage penalties
- Suggestive evidence that inclusive policy climates help reduce mismatch rates and wage penalties

Conclusion

Contact information

Dr. Veronica Sovero: ysovero@ucr.edu

Mario Arce Acosta: mario.arce1615@gmail.com



[HTTPS://GITHUB.COM/NEWTRINO0/UNDOCU_MISMATCH
_WAGE_RESEARCH_2024](https://github.com/newtrino0/undocu_mismatch_wage_research_2024)

Acknowledgments

Thank you to the University of California, Riverside and the Academic Preparation, Recruitment & Outreach department for making this trip possible!