Credible Interval Estimates of the Size and Legal Composition of the US Foreign-Born Population

Andrés Felipe Mira[1]

Christopher Bollinger[2]

[1]U.S. Census Bureau

[2]University of Kentucky

Motivation and Contribution

Motivation:

- There has been a sharp increase in nonresponse to the citizenship question over the years
- This rise in nonresponse degrades survey quality and estimates from those surveys
- Current point estimates assume non-respondents have the same legal status distribution as respondents
- Missing at Random (MAR) assumption is a strong assumption in this context

Figure 1. Citizenship Question in 2019 ACS
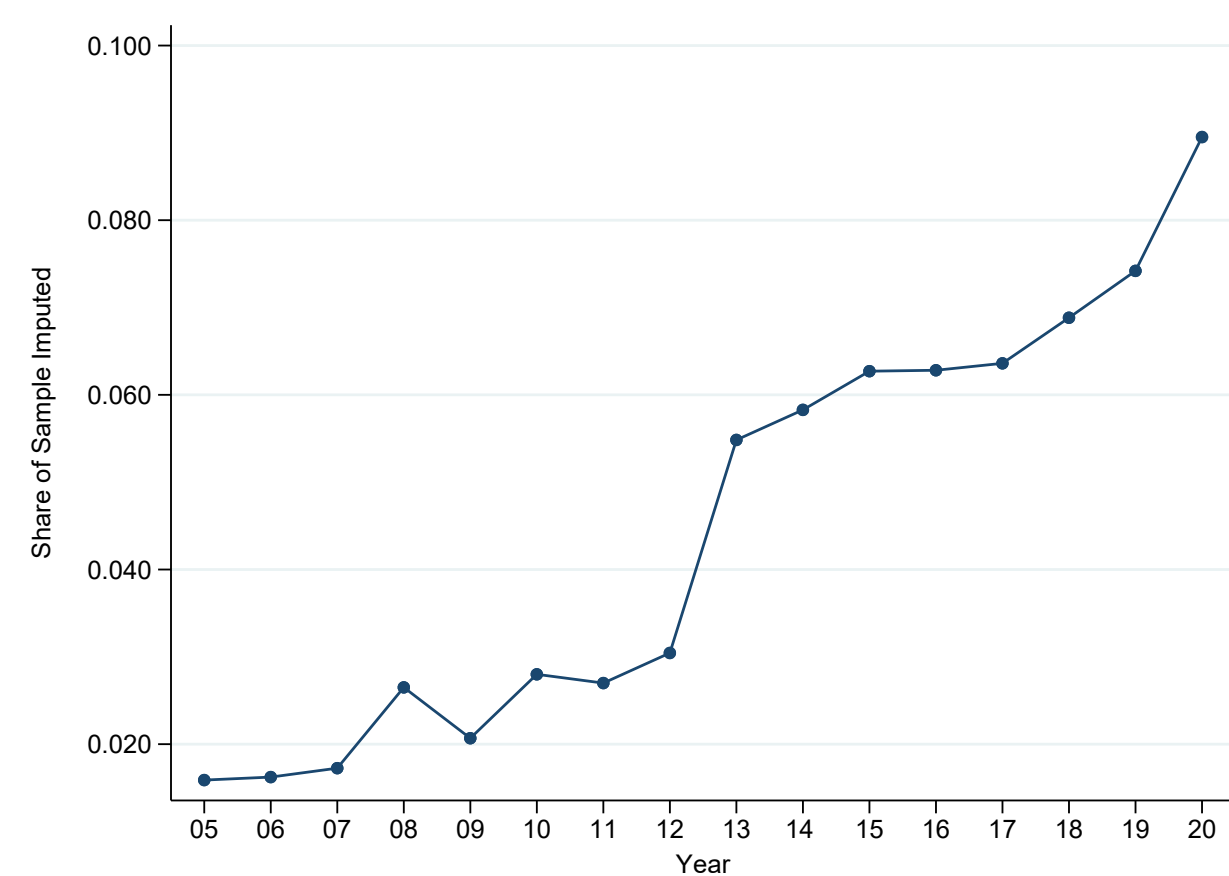
Research Question:

What is the size and legal composition of the US Foreign-Born Population taking into account non-sampling error?

Contribution:

! First to measure total uncertainty in population estimates caused by non-sampling error from item nonresponse.

Data

American Community Survey (ACS) 1-Year:

- Main survey used to produce population estimates of the size and legal composition of the US foreign-born population
- Nationally representative samples of 1% of households.
- Asks individuals whether they are native-born, naturalized citizens, or non-citizens
- Use Public Files from IPUMS

Figure 2. Citizenship Non-response rates in the ACS
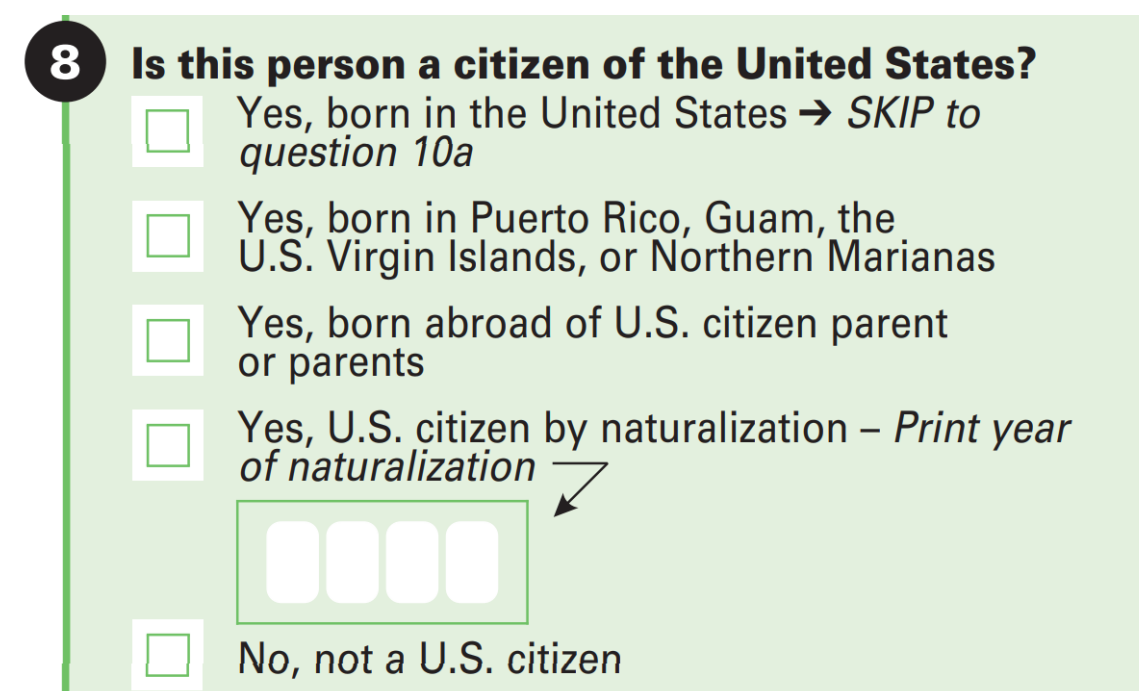
Identifying Undocumented Immigrants in the ACS

! Residual Method proposed by Borjas and Cassidy (2019)

! Assign all immigrant women as having authorized status if:

- that person arrived before 1980;
- that person is a citizen;
- that person receives Social Security benefits, SSI, Medicaid, Medicare, or Military Insurance;
- that person is a veteran, or currently in the Armed Forces;
- that person works in the government sector;
- that person was born in Cuba;
- that person's occupation requires some form of licensing (such as physicians, registered nurses, and lawyers);
- that person's spouse or parent is a legal immigrant or citizen.

! All individuals not identified as a legal immigrant (the residual) are classified as unauthorized.

Missing at Random Assumption

Key underlying assumption: Nonresponse is conditionally random

- The distribution of legal status among respondents is the *same* as that of respondents
- Untested: We do not know the true distribution of non-responding population

Procedure: 'Hot-Deck' Imputation

- A non-respondent is assigned citizenship status based on status of respondent
- Conditional on observables: age, race, and ethnicity, *location*
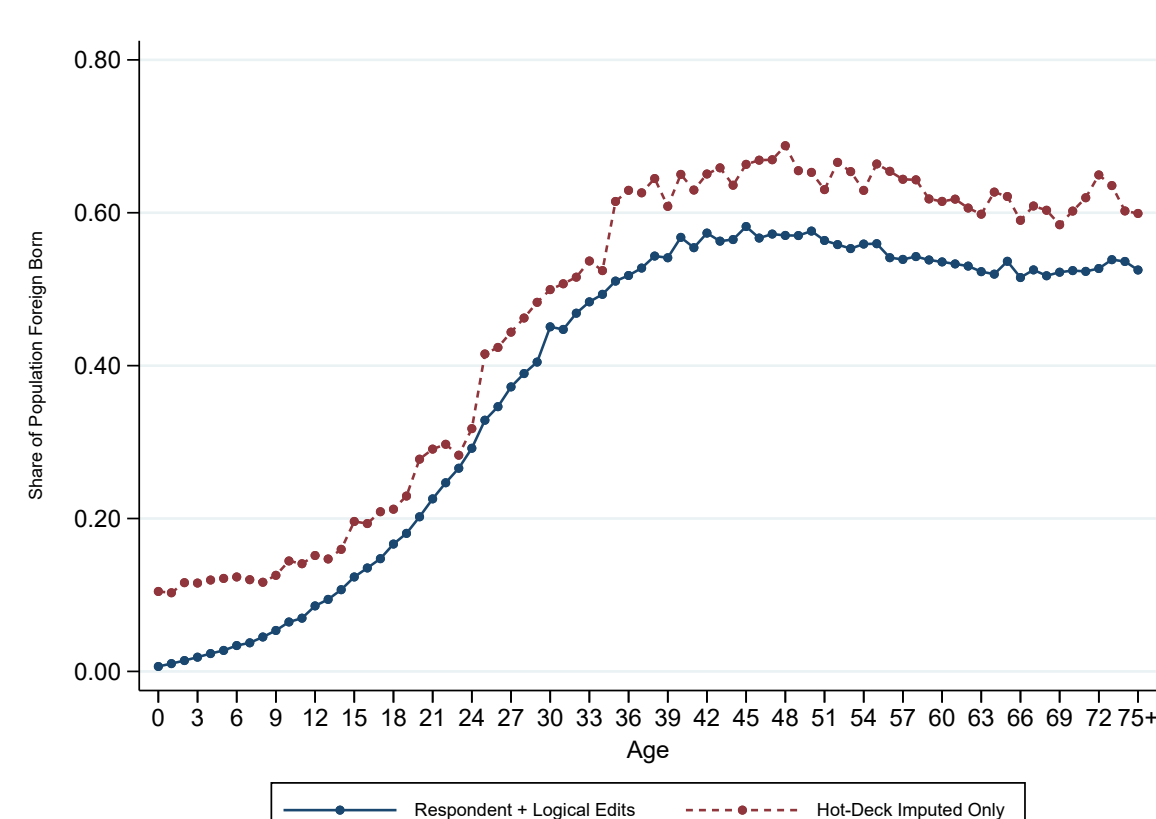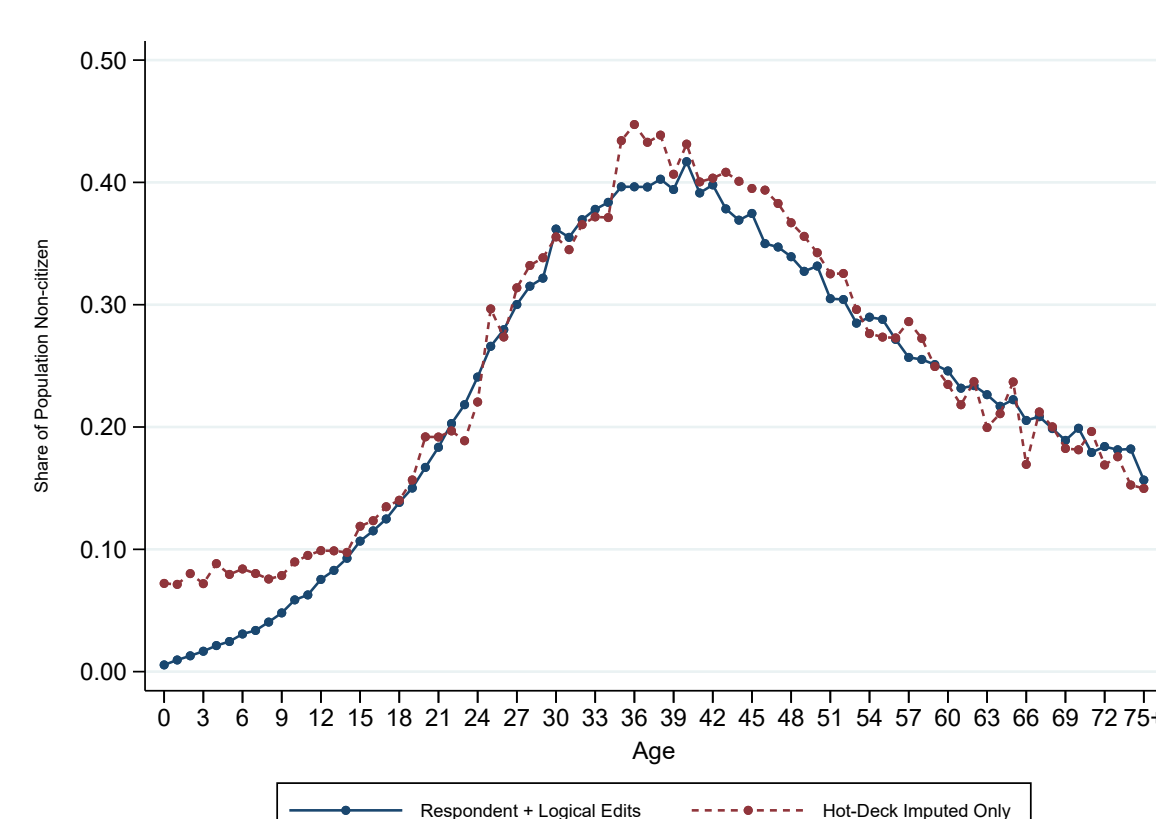

Figure 3. Foreign-Born Hispanics

Figure 4. Non-citizen Hispanics

Manski Bounds

The share of the US population that are foreign-born can be defined as:

$$P(F) = P(F|R = 1) \cdot P(R = 1) + P(F|R = 0) \cdot P(R = 0) \tag{1}$$

where
F = 1 (F = 0): the population unit is foreign-born (native-born)
R = 1 (R = 0): if a population unit did (did not) report citizenship status

- Empirical evidence identifies $P(R)$ and $P(F|R = 1)$
- No empirical information on $P(F|R = 0)$
- Without assuming the exact distribution of foreign-born status among non-respondents, $0 \leq P(F|R = 0) \leq 1$

This yields the following sharp bounds:

$$P(F|R = 1) \cdot P(R = 1) < P(F) < P(F|R = 1) \cdot P(R = 1) + P(R = 0) \tag{2}$$

- To estimate the lower bound: suppose F = 0 for each sample member with missing data in the citizenship question
- To estimate the upper bound: suppose F = 1
- Generates two extreme imputations of each case of missing data

! The point estimate of the share of the foreign-born when using the Census imputations lies between the upper and lower bounds
! Bounds provide maximum degree of uncertainty from item nonresponse
! Estimates within Manski bounds are valid estimates

Interval Estimates of the US Immigrant Population
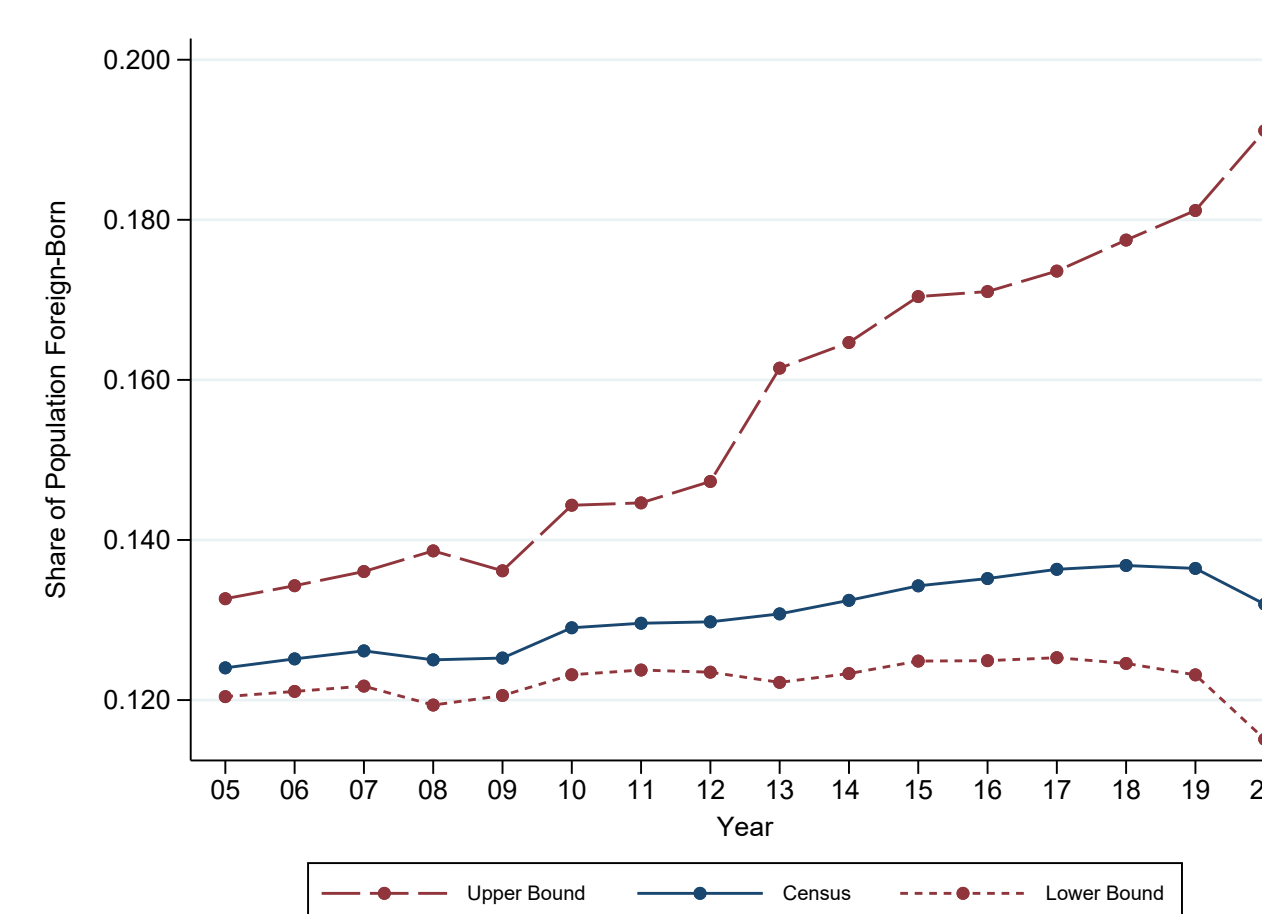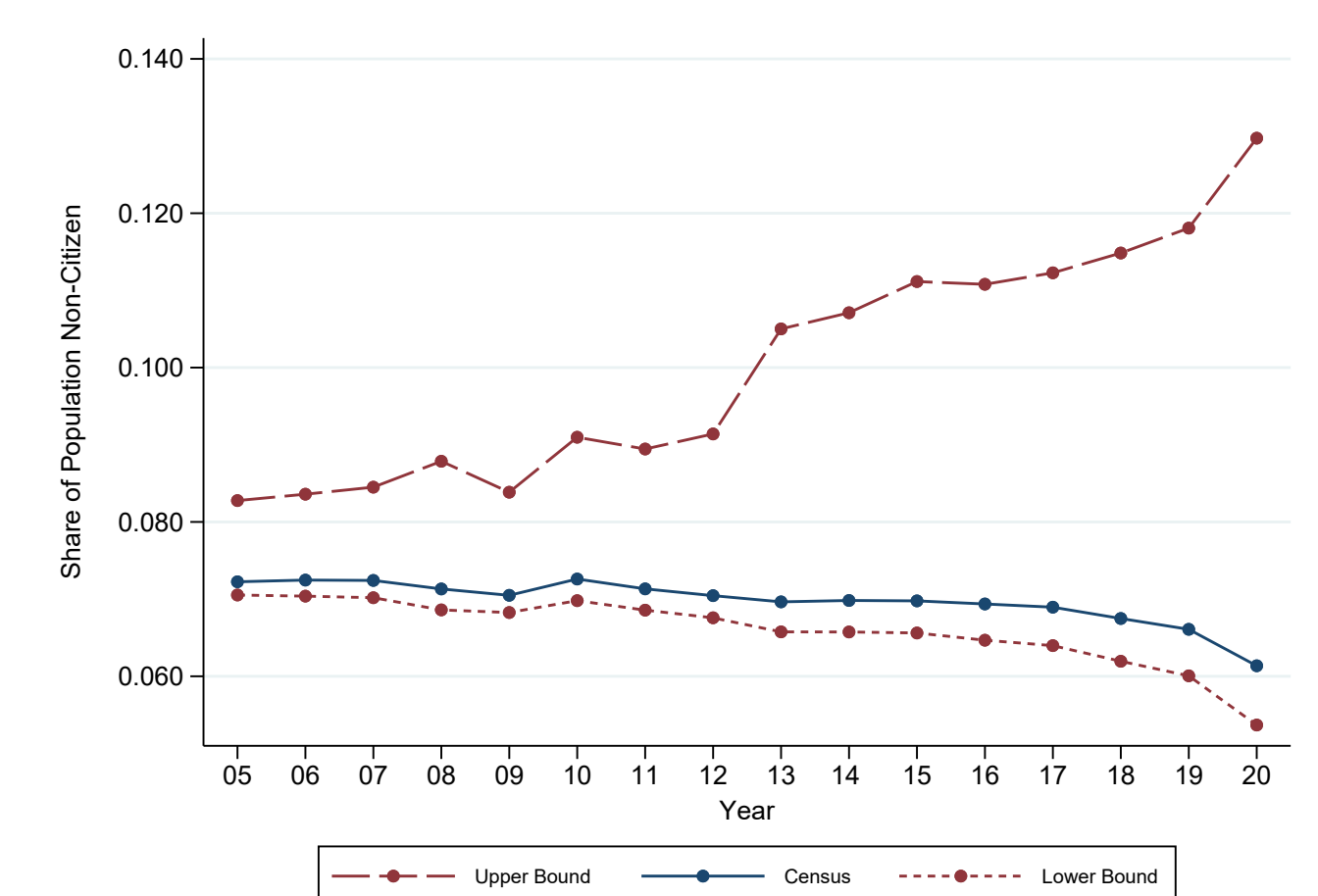

Figure 5. Foreign-Born

Figure 6. Non-Citizen

Size of Foreign-Born Population (2019)

- Bounds: 40.4 and 59.4 Million
- vs. 44.8M point estimate

Size of Non-citizen Population (2019)

- Bounds: 19.7 and 38.7 Million
- vs. 21.7M point estimate

Interval Estimates of the US Undocumented Population

Manski Bounds for Residual Residual Method

Lower Bound:

- All imputed values are assigned so that they **satisfy** the conditions
- Non-respondents assigned as native born
- Residual method is then performed

Upper Bound:

- All imputed values are assigned so that they **fail to satisfy** the conditions
- Non-respondents assigned as Non-citizens
- Residual method is then performed

Figure 7. All Questions in Residual Method
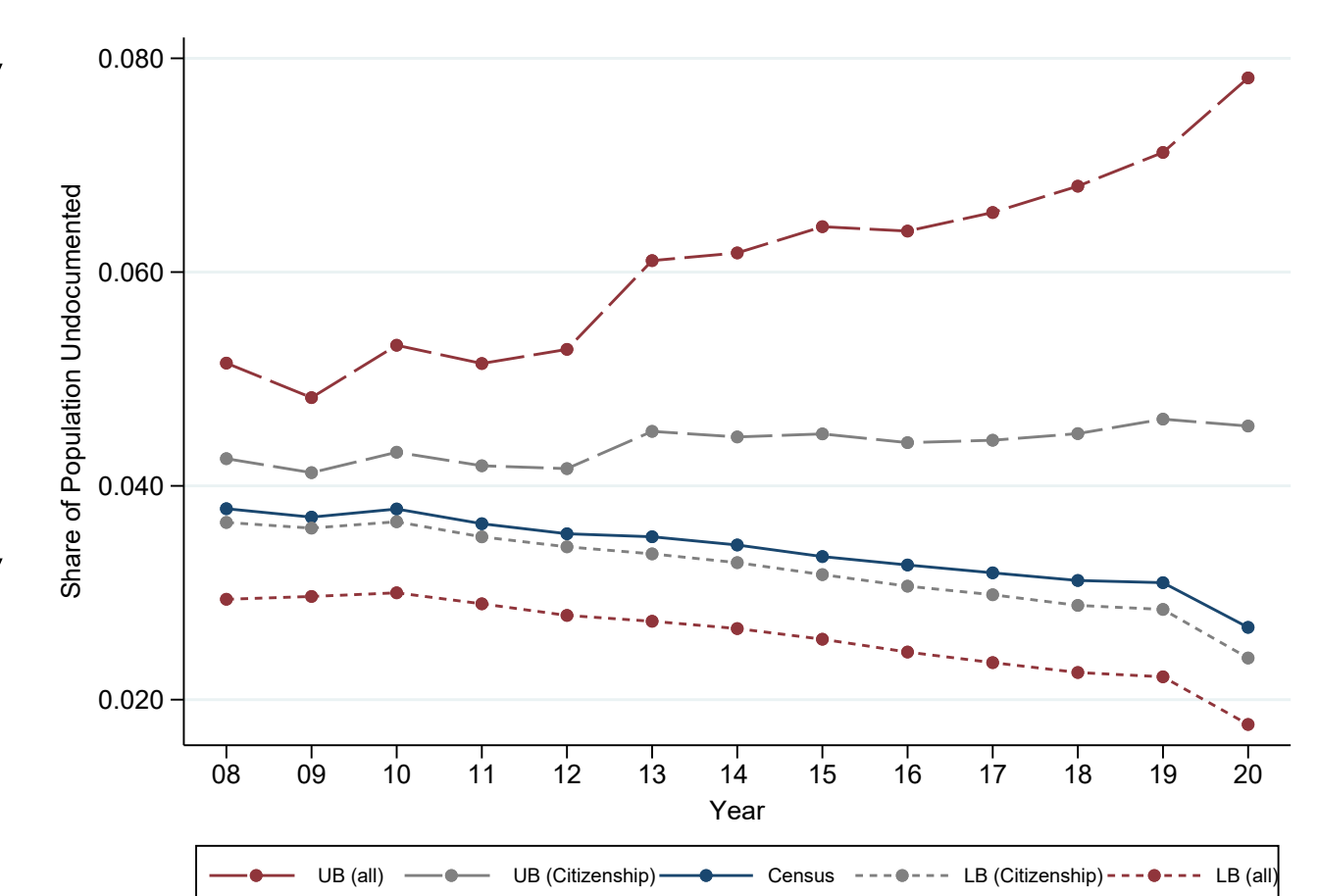
Size of Undocumented Population (2019)

- Bounds: 7.3 and 23.3 Million (vs. 10.15M point estimate)

Conclusion

! Assumptions on the distribution of legal status among non-respondents have to be made to create point estimates
! MAR assumption typically used is a strong assumption in the context of legal status and assumes away all nonrandom error from nonresponse.
! Considerable uncertainty when taking into account nonsampling error
! Without strong assumptions, we can say little about the size and legal composition of the US Foreign-Born population using survey data alone.

Disclaimer: Any opinions and conclusions expressed herein are those of the authors and do not represent the views of the U.S. Census Bureau.

https://www.andresmira.net   2023 ASSA Annual Meeting - New Orleans, LA   andres.f.mira@census.gov