Estimating the Characteristics of Unauthorized Immigrants using US Census Data: Combined Sample Multiple Imputation

James D. Bachmeier, Temple University
Randy Capps, Migration Policy Institute
Jennifer Van Hook, Penn State University

TEMPLE UNIVERSITY®

Key policy questions requiring accurate data on the unauthorized population can be divided into two broad types:

1. What is the size of the population?

    • Stock and flow; rates of growth or decline

1. What are its characteristics?

    • Family and household structure, poverty level, duration of US residence, English language proficiency, etc.

• Large Census surveys (ACS, CPS, NHIS), lack a measure of non-citizens' legal status

• LAFANS, NAWS, MMP, SIPP, are all limited in various ways

• Innovative imputation methods have grown out of necessity

The Residual Method

- Department of Homeland Security (Baker and Rytina 2012), Pew Hispanic Center (Passel 2016), Center for Migration Studies (Warren and Warren 2013)

LPR = Σ New Arrivals (DHS) – (Deaths + Returns)

Unauthorized = (Noncitizens – Legal Non-Immigrants) – LPR

- Useful for estimating the size of the unauthorized population and changes over time.

- Estimation of characteristics, however, limited to information provided on "green card" (Form I-485), and published by DHS: age, sex, state (at application), country of origin, year of (last) arrival

Methods to Estimate More Detailed Characteristics

1. Logical Imputation

   Use of existing information in the CPS or ACS in order identify likely *legal* migrants, leaving a pool of non-citizens with a high likelihood of being unauthorized

1. Survey-Based Probabalistic Imputation

   Probabalistic assignment of legal status in the CPS or ACS based on prediction models estimated in a smaller survey that directly measures immigrants' legal status

Logical Imputation

1. Identify likely legally resident non-citizens in the ACS or CPS using existing variables or combination of variables available in the survey.

    - E.g., immigrants arriving after age 17 who are between, say, age 18 and 25 and are enrolled at the college undergraduate level can be assumed to be legally residing in the U.S. on student visas.

    - Other indicators of probable legal status include welfare recipients, public program participants, government workers, veterans, licensed occupations (e.g., lawyers), and others

1. Assign those remaining in the "unknown" pool as either legal or using some pre-determined criteria.

<u>Logical Imputation</u>

Limitations

- Many LPRs and unauthorized residents can often "look" very similar. How can we be sure we are assigning the remaining pool with accuracy?

- Implicitly assumes no measurement error in the variables used to identify probable legal migrants

  - In an extreme example, of the 3,002 unauthorized immigrants observed in the 2008 Survey of Income and Program Participation, 326 inaccurately reported Medicaid coverage. Logical imputation codes such individuals as LPRs.

Cross-Survey Multiple Imputation (CSMI)

Objective:

1. Pool two surveys: a smaller survey data set that includes direct measures of legal status with a larger data set that does not (e.g., the ACS or CPS)

2. Treat legal status as missing in the larger data set and use multiple imputation techniques to impute it based on a model estimated in the smaller sample

DONOR DATA	TARGET DATA
Unauthorized Status	**------**
Common variables	**Common variables**
Other variables (e.g., labor force, welfare, health)	**Other variables (e.g., labor force, welfare, health)**

Application of the CSMI Method to Impute Legal Status

- Donor data: The Survey of Income and Program Participation (SIPP)

  - Only nationally representative survey with a relatively direct measure of non-citizens' green card status; included in the 1996, 2001, 2004, 2008, and 2014 SIPP panels.

- Target data: American Community Survey (ACS)

  - Satisfies "same universe" condition (as long as the foreign-born population doesn't drastically change in intervening years)

  - Includes many of the same variables in order to build a large prediction equation that satisfies the joint-observation condition

Uses of SIPP-based Imputed ACS Data

- Ultimately, CSMI reproduces a set of multivariate associations between legal status and a set of predictors, observed in the SIPP, on the ACS (or other target sample). So what, exactly, is it good for?

- The main advantage is the increase in sample size, allowing the examination of subpopulations – e.g., DACA / DAPA – with much greater precision (FB: 2008 SIPP, N=˜10,000; 2008 ACS, N=˜320,000)

- The existence of multiple SIPP panels – since 1996 – makes it possible to examine changes in characteristics of the unauthorized population over time, but...

- The change *must* be observed in the SIPP (e.g., one cannot "pick up" the effects of the ACA on health coverage by comparing the 2013 and 2015 ACS if both are imputed using the 2008 SIPP).

Limitations

- CSMI imputations of legal status are strictly limited by the nature of the data.

- While SIPP collects detailed arrival statuses, it only codes LPR and "other" arrivals in public-use data, making it difficult to distinguish between unauthorized and legal non-immigrants (requires logical imputation)

- Moreover, while the 1996, 2001, and 2004 panels provide country of birth detail, the 2008 and 2014 provide only broad regions of birth

- And, while the 1996-2008 panels identify adjustees (those who arrive without a green card but eventually attain one), the 2014 SIPP only asks about arrival status, and excludes the adjustment question. *Can this question be added in on-going waves?*

Uses of SIPP-based Imputed ACS Data (continued)

www.migrationpolicy.org/data/unauthorized-immigrant-population/state/CA

All MPI

MPI
MIGRATION POLICY INSTITUTE

RESEARCH & INITIATIVES | PUBLICATIONS | EVENTS | NEWS | ABOUT MPI

Home » Data Hub » Unauthorized Immigrant Population

Profile of the Unauthorized Population: California

Demographics	Estimate	% of Total
Unauthorized Population	3,019,000	100%
Top Countries of Birth		
Mexico	2,127,000	70%
Guatemala	200,000	7%
El Salvador	114,000	4%
Philippines	82,000	3%
China	74,000	2%
Regions of Birth		

- ▼ DEMOGRAPHICS
- ▼ FAMILY
- ▼ EDUCATION AND LANGUAGE
- ▼ WORKFORCE
- ▼ ECONOMICS
- ▼ DEFERRED ACTION

TEMPLE UNIVERSITY®

<u>Uses of SIPP-based Imputed ACS Data</u> (continued)

- ***Profiles of the unauthorized population and sub-populations*** (Batalova, Hooker and Capps 2013, 2014; Bryant and Bachmeier 2015; Capps, Fix and Zong 2016; Capps et al. 2013; Rosenblum and Ruiz Soto 2015)

- ***(Proposed) Policy impacts*** (Capps and Rosenblum 2014; Capps et al. 2016; Hipsman, Gómez-Aquiñaga and Capps 2016; Mathay and McHugh 2015; McHugh 2014)

- ***Effects of local enforcement efforts*** (Rugh and Hall 2016)

- ***Unauthorized wage penalty*** (Bean, Brown and Bachmeier 2015)

- ***Legal status and the geography of post-IRCA migration flows*** (Bachmeier and Spence N.D.)

- ***Legal status and immigrant integration*** (Bachmeier, Altman, and Van Hook 2015)

Conclusions and On-going Data Issues

- Existing data on the unauthorized population is limited but, building on the pioneering work of Heer, Passel, Warren, etc., innovative methods for estimating the characteristics of the population have emerged over the past decade.

- Can measurement issues in the SIPP be remedied?

- In the absence of better data, continued methodological transparency and collaboration will lead to improvements in estimation.

- Ideally, we need better data, such as a national survey of immigrants.

For More Information

Randy Capps
Director of Research, U.S. Programs,
MPI rcapps@migrationpolicy.org

For additional information and to receive updates:
www.migrationpolicy.org
For interactive data tools on U.S. and International migration data, visit:
http://www.migrationpolicy.org/programs/data-hub