Researching Administrative Records Usage in the Nonresponse Followup Operation of the 2020 Census

Andrew D. Keller and Scott M. Konicki, U.S. Census Bureau

Presented at the Modern Modeling Methods Conference
University of Connecticut
May 23 - May 26, 2016

OUR CHALLENGE

- 2010 Census – 50 million Nonresponse Folllowup (NRFU) Operation addresses to visit
- Approximately 15 million vacant units, 5 million non-existent units in 2010 – can we resolve without visiting the address?
- Not able to obtain all person characteristics for NRFU in-person interviews of occupied units
- Cost $2 billion

OUR GOALS

- Reduce cost of the NRFU operation while maintaining data quality
- Instead of in-person contact attempts, use administrative records (AR) to resolve cases
- Identify vacant and occupied units using AR - remove these cases from the field workload

OUR SOLUTION

1. **Universe Creation:** Build person- and housing unit- level datasets - See Fig. 1
2. **Model Development:** Build models to predict 2010 NRFU responses using AR data
3. **Model Application:** Apply results to current vintage AR data to identify AR Vacant and AR Occupied units

FIGURE 1
(Hypothetical Data Example)

Person-Level Dataset

Unit ID	Person ID	On IRS 1040	On IRS 1099	On Medicare	Age	Sex
100	1001	X	X		32	F
100	1002	X			5	M
200	2001	X		X	68	M

Housing Unit-Level Dataset

Unit ID	Any IRS1040	Any IRS 1099	Any Medicare	HH Count	HH Composition
100	X	X		2	1 Adult, 1 Child
200	X		X	2	1 Adult, 0 Child

AR VACANT – HOW DO WE DETERMINE VACANT UNITS?

1. Universe Creation

- Data: 2010 Census NRFU addresses
- NRFU addresses combined with AR information from those addresses

2. Model Development

- Multinomial logistic regression model
- Dependent variable: 2010 Census status (1: occupied, 2: vacant, 3: delete)

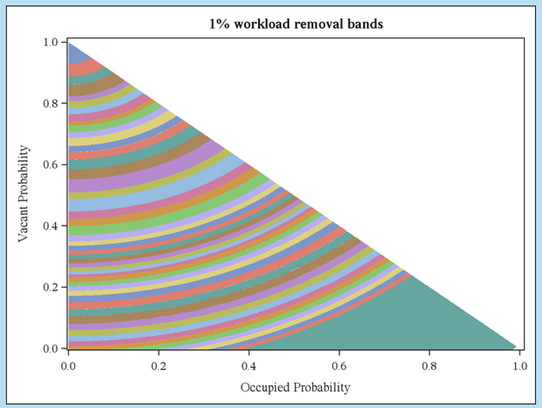
3. Model Application - Which units to identify as AR Vacant?

- Use Euclidean distance function to select cases with high vacant probability (near 1) and low occupied probability (near 0)

$$d_{vac} = \sqrt{(1-\widehat{P_{vac}})^2 + (0-\widehat{P_{occ}})^2}$$

- Given a specified threshold for the distance, all cases below that threshold are identified as AR Vacant
- The distance function can be visualized as successive bands of cases emanating from the point (0,1) - See Fig. 2
- Each successive band represents an additional 1% of the NRFU workload.

FIGURE 2



AR OCCUPIED – HOW DO WE DETERMINE OCCUPIED UNITS?

1. Universe Creation

Person-Place model

- Data: Person-place combinations across AR sources for 2010 Census NRFU addresses

HH Composition model

- Data: 2010 Census NRFU addresses
- NRFU address combined with AR information from those address

2. Model Development

Person-Place: Binomial logistic regression model

- Dependent variable: 2010 Census status (1: AR person-place pair is observed in 2010 Census, 0: AR person-place pair is not observed in 2010 Census)

HH Composition: Multinomial logistic regression model

- Dependent variable: 2010 Census HH composition (0: not occupied, 1: 1 adult, 0 children, 2: 1 adult, 1+ children, 3: 2 adult, 0 children, 4: 2 adult, 1+ children, 5: 3 adult, 0 children, 6: 3 adult, 1+ children, 7: Other)

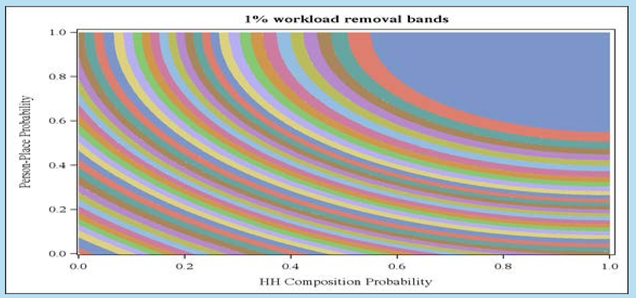
3. Model Application - Which units to identify as AR Occupied?

- Use Euclidean distance function to select cases with high person-place probability (near 1) and high HH composition probability (near 1)

$$d_{OCC} = \sqrt{(1-\widehat{P_{p\text{-}p}})^2 + (1-\widehat{P_{hhc}})^2}$$

- Given a specified threshold for the distance, all cases below that threshold are identified as AR Occupied
- The distance function can be visualized as successive bands of cases emanating from the point (1,1) - See Fig. 3
- Each successive band represents an additional 1% of the NRFU workload.

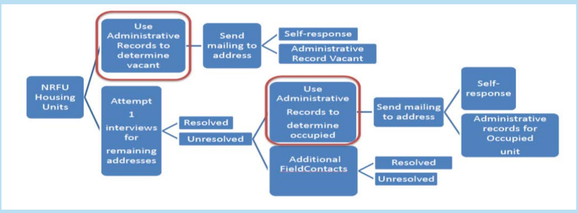
FIGURE 3



IMPLEMENTATION

- **Completed in 2016 Census Test currently ongoing**
- Remove AR Vacant units from NRFU workload at the beginning of NRFU operation - See Fig. 4
- Attempt one contact for all remaining NRFU addresses
- Remove AR Occupied units from NRFU workload if unresolved after first contact
- Send an additional postcard mailing to all AR Vacant and AR Occupied addresses

FIGURE 4



CONCLUSIONS

- Reduced workload via AR will **decrease** cost
- AR vacant or AR occupied misclassification could **decrease** data quality
- Within-household person coverage of AR may **increase** or **decrease** data quality
- Characteristic availability for AR will **increase** and **decrease** data quality

United States Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. Census Bureau
census.gov