American Community Survey
Design and Statistical Methodology

Alfredo Navarro, U.S. Census Bureau


Rio de Janeiro, Brazil
November 10, 2011

U S C E N S U S B U R E A U
Overview of the Session

• Data Collection
• Sample Design
• Weighting and Estimation
  – Major Components
  – Multi-year Estimates
• ACS Research Weighting and Estimation
Data Collection
Data Collection

• Methodology based on best practices from decennial census and demographic surveys

• Monthly samples using overlapping multi-mode data collection methods
  – Mail
  – Telephone
  – Personal Visit
## Data Collection

<table>
<thead>
<tr>
<th>Month of Data Collection</th>
<th>2005</th>
<th>2006</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>November</td>
<td>December</td>
</tr>
<tr>
<td>ACS Sample Panel</td>
<td>Mail</td>
<td>Phone</td>
</tr>
<tr>
<td>November, 2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>December, 2005</td>
<td>Mail</td>
<td>Phone</td>
</tr>
<tr>
<td>January, 2006</td>
<td>Mail</td>
<td>Phone</td>
</tr>
<tr>
<td>February, 2006</td>
<td>Mail</td>
<td>Phone</td>
</tr>
<tr>
<td>March, 2006</td>
<td></td>
<td></td>
</tr>
<tr>
<td>April, 2006</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Sample Design

• Survey designed to include
  – U.S. Stateside and Puerto Rico
  – Population in both housing units and group quarters (group quarters started in 2006)

• Survey designed to produce annually updated single-year and multi-year estimates
Sample Design Frame

• Sample cases selected from an updated Master Address File (MAF)

• MAF updated through the use of…
  – Postal Service updates in most areas
  – Special field updating in more rural areas and presence of non-city style addresses
Sample Design

- Un-clustered one-stage systematic sample of housing units selected as initial sample each month

- Sub-sample of nonrespondents selected after mail and phone attempts for personal visit follow-up
## ACS Initial Sample Design

<table>
<thead>
<tr>
<th>Governmental Unit Size: Estimate of Occupied Housing Units</th>
<th>ACS 1-year Sampling Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 - 200</td>
<td>10.0%</td>
</tr>
<tr>
<td>201 - 800</td>
<td>~7.0%</td>
</tr>
<tr>
<td>801 - 1200</td>
<td>~3.5%</td>
</tr>
<tr>
<td>Census Tract Size</td>
<td></td>
</tr>
<tr>
<td>2000 or less</td>
<td>~2.4%</td>
</tr>
<tr>
<td>Over 2000</td>
<td>~1.7%</td>
</tr>
</tbody>
</table>
ACS Sample Design

Rate Definitions 2005 to 2010
Sampling rates function of base rate (BR)
One fixed rate stratum

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Block MOS Criteria</th>
<th>Sampling Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0 &lt; GUMOS ≤ 200</td>
<td>10% (fixed)</td>
</tr>
<tr>
<td>2</td>
<td>200 &lt; GUMOS ≤ 800</td>
<td>3×BR</td>
</tr>
<tr>
<td>3</td>
<td>800 &lt; GUMOS ≤ 1,200</td>
<td>1.5×BR</td>
</tr>
<tr>
<td>1</td>
<td>TRACTMOS ≤ 2,000</td>
<td>BR</td>
</tr>
<tr>
<td>4</td>
<td>2,000 &lt; TRACTMOS</td>
<td>0.735×BR</td>
</tr>
</tbody>
</table>
Reallocation of the HU Address Sample - Improvement

Increase the number of sampling strata
  – Smaller stratum intervals allows smoother transitions between rates

Increase sampling rates for blocks in the very smallest governmental units
  – Increase reliability of the estimates
Reallocation of the HU Address Sample – 2011 Stratification

New Stratification (small GUs)
- increased number of fixed rate strata
- increased the rates

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Block MOS Criteria</th>
<th>Sampling Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$0 &lt; \text{GUMOS} \leq 200$</td>
<td>15% (fixed)</td>
</tr>
<tr>
<td>2</td>
<td>$200 &lt; \text{GUMOS} \leq 400$</td>
<td>10% (fixed)</td>
</tr>
<tr>
<td>3</td>
<td>$400 &lt; \text{GUMOS} \leq 800$</td>
<td>7% (fixed)</td>
</tr>
<tr>
<td>4</td>
<td>$800 &lt; \text{GUMOS} \leq 1,200$</td>
<td>$2.8 \times \text{BR} \sim 5%$</td>
</tr>
</tbody>
</table>
# Reallocation of the HU Address Sample – 2011 Stratification

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Block MOS Criteria</th>
<th>Sampling Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>0 &lt; TRACTMOS ≤ 400</td>
<td>3.5×BR</td>
</tr>
<tr>
<td>6</td>
<td>0 &lt; TRACTMOS ≤ 400</td>
<td>0.92×3.5×BR</td>
</tr>
<tr>
<td>7</td>
<td>400 &lt; TRACTMOS ≤ 1,000</td>
<td>2.8×BR</td>
</tr>
<tr>
<td>8</td>
<td>400 &lt; TRACTMOS ≤ 1,000</td>
<td>0.92×2.8×BR</td>
</tr>
<tr>
<td>9</td>
<td>1,000 &lt; TRACTMOS ≤ 2,000</td>
<td>1.7×BR</td>
</tr>
<tr>
<td>10</td>
<td>1,000 &lt; TRACTMOS ≤ 2,000</td>
<td>0.92×1.7×BR</td>
</tr>
<tr>
<td>11</td>
<td>2,000 &lt; TRACTMOS ≤ 4,000</td>
<td>BR</td>
</tr>
<tr>
<td>12</td>
<td>2,000 &lt; TRACTMOS ≤ 4,000</td>
<td>0.92×BR</td>
</tr>
<tr>
<td>13</td>
<td>4,000 &lt; TRACTMOS ≤ 6,000</td>
<td>0.6×BR</td>
</tr>
<tr>
<td>14</td>
<td>4,000 &lt; TRACTMOS ≤ 6,000</td>
<td>0.92×0.6×BR</td>
</tr>
<tr>
<td>15</td>
<td>6,000 &lt; TRACTMOS</td>
<td>0.35×BR</td>
</tr>
<tr>
<td>16</td>
<td>6,000 &lt; TRACTMOS</td>
<td>0.92×0.35×BR</td>
</tr>
</tbody>
</table>
### Sub-sampling Rates

#### Nonresponse Follow-up

<table>
<thead>
<tr>
<th>Address and Tract Characteristics</th>
<th>Sub-sampling Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unmailable Addresses</td>
<td>2 – in - 3</td>
</tr>
<tr>
<td>Mailable addresses with the lowest mail/CATI rates</td>
<td>1 – in – 2</td>
</tr>
<tr>
<td>Mailable addresses in tracts with average mail/CATI rates</td>
<td>2 – in – 5</td>
</tr>
<tr>
<td>Other mailable addresses</td>
<td>1 – in – 3</td>
</tr>
</tbody>
</table>
Weighting and Estimation
Annual Weighting Process
3 Major Components

• Initial weights to reflect the probability of selection

• Adjust weights of interviewed households to account for noninterviews

• Adjust weights to independent housing unit and population estimates (controls)
Initial Weight
Probabilities of Selection

• Initial probability of selection is assigned as a function of the sample design

• Nonresponse follow-up (Personal Visit – CAPI) sample design
Nonresponse Adjustment

- The weight of the nonrespondents is transferred to the respondents

- Nonresponse adjustment is carried out at the census tract level for groups of households with characteristics correlated with nonresponse:
  - Census tract
  - Type of building (single vs. multi-unit)
  - Month of data collection
Ratio Adjustments to Housing Unit and Population Controls

- Post-censal estimates are produced by updating the previous census results using various administrative records data.
- In a multi-stage process, housing unit and population adjustment ratios are applied to the weights.
- Applied at the county (or group of counties) level by sub-county areas and race/ethnicity and age/sex groups.
Ratio Adjustments to Controls - Why?

- Reduce variability of the estimates
- Reduce bias
  - Undercoverage of housing units
  - Undercoverage of people within housing units
Multi-Year Estimates

- Combining or “pooling”
- Population controls
- Tabulation geography
- Inflation adjustments
Both Single- and Multi-Year Estimates are Period Estimates

• 2005 single-year estimates are based on Jan 2005 – Dec 2005 interviews (12 months)

• 2005-2007 three-year estimates are based on Jan 2005 – Dec 2007 interviews (36 months)

• 2005-2009 five-year estimates are based on Jan 2005 – Dec 2009 interviews (60 months)
Multi-Year Estimates

Pooling Advantages

- Improved accuracy of estimates – taking advantage of increased number of sample cases
- More up-to-date controls
- Flexibility of developing weighting procedures
- Production of multi-year data products mirror the 1-year data products
Multi-Year Estimates
Population Controls

• Simple average of the set of population controls for the years comprising the multi-year estimate

• For example, for the 2005-2009 five-year estimates, sum the controls released in 2010 for 2005, 2006,…,2009 divided by 5

• Use the most recently released estimates for each year
Multi-Year Estimates

Tabulation Geography

- Boundary changes can occur through annexations during the multi-year period

- Plan is to tabulate using the geography of the most recent year in the multi-year estimate

- For 2005-2009 estimates, tabulate using all interviews for the period of 2005-2009 that were conducted in blocks that define the area in 2009
Multi-Year Estimates
Inflation Adjustments

• The Consumer Price Index is used to compute inflation factors

• Dollar valued data items are inflation adjusted to the most recent year of the period

• For example, for the 2005-2009 estimates, appropriate inflation factors are applied to reported income values for 2005, 2006, …, 2008 to adjust to 2009 constant dollars
Multi-Year Estimates
Medians

• Medians are produced using combined data records from all years

• A 3-year median household income estimate is determined by combining the household records from the 3 years into one data set and determining the median from this combined distribution
ACS Research
Weighting and Estimation
ACS Research
Weighting and Estimation

• Multi-year estimates

• Family equalization

• Use of alternative population estimates as ACS controls

• Small area estimates
ACS Research
Multi-Year Estimates

• Production of a series of multi-year estimates for the ACS test sites (1999 – 2005)

• The objective of this research was to assess issues of stability and comparability

• A secondary objective was to gain insights to develop and improve user materials
ACS Research

Family Equalization

Assess impact of modifying weighting procedures to ensure consistency of the following estimates:

– Households and householders

– Householders and spouses (and unmarried partners) in married (unmarried) couple (partners) households

– Subfamily husbands and wives in married-couple subfamilies
ACS Research
Use of Alternative Population Controls

- The quality of national and state level population estimates is generally accepted

- The benefits of using of these controls by several current surveys is well documented and understood

- We developed and implemented research to assess the feasibility and benefits to control the ACS data to sub-county estimates of total population.
ACS Research
Small Area Estimates

- Research on 1999-2001 and 1999 – 2005 MYE ACS data indicated tract-level standard errors were higher than anticipated
- Lack of tract-level controls identified as likely primary cause
ACS Research
Small Area Estimates

• Research into variance reduction is ongoing
  – Model-assisted weighting adjustment also incorporating administrative records
  – Model-based estimates for specific applications
Contact Information

U.S. DEPARTMENT OF COMMERCE
U.S. Census Bureau
Washington, DC 20233
Alfredo Navarro
Room: 4K071
Phone: 301-763-3600
Email: Alfredo.Navarro@census.gov
Section 203 of the Voting Rights Act – Modeling Approach

ACS Small Area Application
Notation

Let,

$\text{LMG} = \text{Language minority groups}, \ i = 1, \ldots, 68$

$\text{CIT} = \text{Citizens}$

$\text{LEP} = \text{Limited English Proficiency}$

$\text{ILL} = \text{Illiteracy}$

$\text{CVAP} = \text{Citizen Voting Age Population}$

$J = \text{Juridictions (States, Counties, MCDs, AI areas)}$
Language Minority Groups

• Hispanic Origin Population (1)

• Asians – Detailed Groups (16)
  Chinese, Japanese, Filipino, Thai, etc.

• American Indians and Alaska Natives (51)
Jurisdiction Level Qualification

A jurisdiction qualifies if:

\[
\text{LEP} = \frac{\text{LMG-CVAP-LEP/CAVP}}{\text{CAVP}} > 5\%
\]

OR

\[
\text{LMG-CVAP-LEP} > 10,000
\]

and

\[
\text{LMG-IIIL} > \text{National ILL rate}
\]
Model Development

We then form a model in the following mode:
\[ P(\text{CIT},\text{LEP},\text{ILL}) = P(\text{CIT}) \times P(\text{LEP},\text{CIT}) \times P(\text{ILL},\text{CIT} \text{ and } \text{LEP}) \]
Small Area Modelling

BASIC MODEL STRUCTURE:

\[ N(\text{CIT}) \sim \text{Bin} \left( N(\text{VAP}), p(\text{CIT}) \right) \]
\[ N(\text{CIT,LEP}) \sim \text{Bin} \left( N(\text{CIT}), p(\text{LEP,CIT}) \right) \]
\[ N(\text{CIT,LEP,ILL}) \sim \text{Bin} \left( N(\text{LEP}), p(\text{ILL,CIT and LEP}) \right) \]
\[ P \sim \text{Beta}(a, b) \]
Small Area Modelling

• Further refinement was done with geographic stratification

• Design based estimates were regressed on selected survey variables, predictions were computed

• Stratification was developed with the cumsum sqrt of the frequencies – Dalenious (1957)
Estimation

- Beta distribution parameters are estimated using MLE
- The empirical Bayes prediction of the true predictions is defined as the point estimate
- Variance estimates are fully Bayesian – 300K draws are generated from the posterior distribution with vague priors