Small Area Estimation Applications in the US Census Bureau Annual Survey of Employment and Payroll Evaluation

Bac Tran
Program Research Branch, Chief Governments Division
U.S. Census Bureau
Outline

- Target Population
- Population Parameters
- Sampling Frame
- Sample Design
- Small Area Challenges
- Estimators
- Evaluation
Target Population

- Individual governments

A government is an organized entity which, in addition to having governmental character, has sufficient discretion in the management of its own affairs to distinguish it as separate from the administrative structure of any other governmental unit.

- Types
  - Counties
  - Municipalities
  - Townships
  - Special Districts
  - School Districts
Parameters of Interest
Annual Survey of Employment and Payroll (ASPEP)

Full-time Employees
Full-time Pay
Part-time Employees
Part-time Pay
Part-time Hours
Parameters of Interest (Cont’d)
ASPEP Publication

Statistics on the number of federal, state, and local government employees and their gross payrolls

2011 Public Employment and Payroll Data
Local Governments
United States Total

SOURCE: 2011 Annual Survey of Public Employment and Payroll. For information on sampling and nonsampling errors and definitions, see http://www.census.gov/govs/apes/how_data_collected.html. Data users who create their own estimates from these tables should cite the U.S. Census Bureau as the source of the original data only.

Detail may not sum to totals because of rounding.

<table>
<thead>
<tr>
<th>Government Function</th>
<th>Full-time employees</th>
<th>Full-time Pay ($)</th>
<th>Part-time employees</th>
<th>Part-time Pay ($)</th>
<th>Full-Time Equivalent Employment ($)</th>
<th>Total March Pay ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>10,781,323</td>
<td>46,773,983,844</td>
<td>3,318,107</td>
<td>3,631,752,914</td>
<td>11,999,059</td>
<td>50,405,736,765</td>
</tr>
<tr>
<td>Financial Administration</td>
<td>210,190</td>
<td>938,862,938</td>
<td>41,461</td>
<td>40,556,655</td>
<td>222,584</td>
<td>979,419,583</td>
</tr>
<tr>
<td>Other Government Administration</td>
<td>191,865</td>
<td>860,103,178</td>
<td>162,573</td>
<td>89,278,684</td>
<td>217,307</td>
<td>949,381,862</td>
</tr>
<tr>
<td>Judicial and Legal</td>
<td>232,038</td>
<td>1,089,735,867</td>
<td>33,339</td>
<td>46,238,354</td>
<td>243,960</td>
<td>1,135,974,221</td>
</tr>
<tr>
<td>Police Protection Total</td>
<td>782,739</td>
<td>4,208,166,047</td>
<td>106,504</td>
<td>100,359,135</td>
<td>819,445</td>
<td>4,308,524,163</td>
</tr>
<tr>
<td>Police Officers Only</td>
<td>609,849</td>
<td>3,507,627,164</td>
<td>38,434</td>
<td>39,386,380</td>
<td>623,259</td>
<td>3,547,013,544</td>
</tr>
<tr>
<td>Other Police Employees</td>
<td>172,890</td>
<td>700,538,883</td>
<td>60,070</td>
<td>60,971,756</td>
<td>196,146</td>
<td>761,510,639</td>
</tr>
<tr>
<td>Fire Protection Total</td>
<td>320,311</td>
<td>1,905,253,370</td>
<td>98,773</td>
<td>45,352,826</td>
<td>330,408</td>
<td>1,950,606,196</td>
</tr>
<tr>
<td>Firefighters Only</td>
<td>295,576</td>
<td>1,781,236,979</td>
<td>88,445</td>
<td>38,626,738</td>
<td>311,562</td>
<td>1,819,863,717</td>
</tr>
<tr>
<td>Other Fire Employees</td>
<td>24,735</td>
<td>124,016,391</td>
<td>8,327</td>
<td>6,726,088</td>
<td>26,846</td>
<td>130,742,475</td>
</tr>
<tr>
<td>Corrections</td>
<td>249,396</td>
<td>1,064,980,702</td>
<td>14,465</td>
<td>20,131,246</td>
<td>255,391</td>
<td>1,085,111,949</td>
</tr>
<tr>
<td>Highways</td>
<td>279,685</td>
<td>1,124,072,016</td>
<td>24,422</td>
<td>24,039,982</td>
<td>290,970</td>
<td>1,148,111,997</td>
</tr>
<tr>
<td>Air Transportation</td>
<td>41,340</td>
<td>205,379,744</td>
<td>3,847</td>
<td>4,943,085</td>
<td>43,065</td>
<td>210,322,839</td>
</tr>
<tr>
<td>Water Transport and Terminals</td>
<td>7,706</td>
<td>46,652,215</td>
<td>1,022</td>
<td>1,204,473</td>
<td>9,057</td>
<td>47,856,688</td>
</tr>
<tr>
<td>Public Welfare</td>
<td>248,993</td>
<td>997,087,389</td>
<td>38,604</td>
<td>54,808,883</td>
<td>269,225</td>
<td>1,051,896,723</td>
</tr>
<tr>
<td>Health</td>
<td>219,377</td>
<td>950,113,069</td>
<td>56,300</td>
<td>85,354,374</td>
<td>244,339</td>
<td>1,035,467,444</td>
</tr>
<tr>
<td>Hospitals</td>
<td>492,601</td>
<td>2,349,529,064</td>
<td>134,898</td>
<td>307,882,497</td>
<td>565,162</td>
<td>2,657,411,562</td>
</tr>
</tbody>
</table>
Parameters of Interest

Statistical Aggregation

- Totals
  - by (state, function)

- Level of government totals
  - Local, state, state and local
  - Nation
Parameters of Interest (Cont’d)
Some Function Codes of ASPEP

001, Airport
002, Space Research & Technology (Federal)
005, Correction
006, National Defense and International Relations (Federal)
012, Elementary and Secondary - Instruction
112, Elementary and Secondary - Other Total
014, Postal Service (Federal)
016, Higher Education - Other
018, Higher Education - Instructional
021, Other Education (State)
022, Social Insurance Administration (State)
023, Financial Administration
024, Firefighters
124, Fire - Other
025, Judicial & Legal
029, Other Government Administration
032, Health

040, Hospitals
044, Streets & Highways
050, Housing & Community Development (Local)
052, Local Libraries
059, Natural Resources
061, Parks & Recreation
062, Police Protection - Officers
162, Police-Other
079, Welfare
080, Sewerage
081, Solid Waste Management
087, Water Transport & Terminals
089, Other & Unallocable
090, Liquor Stores (State)
091, Water Supply
092, Electric Power
093, Gas Supply
094, Transit
Sampling Frame

- Governments Integrated Directory (GID) → Created in 2007

- Unit ID: 14 digits

| State (2) | Type (1) | County (3) | Unit (3) | SUP (3) | SUB (2) |
Sampling Frame (Cont’d)

Example of an unit ID

→ 33 2 031 001 000 00 = New York City

33 2 031 001 301 00 = New York City public school system (dependent on the city government)

33 2 031 001 302 00 = Fashion Institute (dependent post-secondary education agency)

33 2 031 001 303 00 = CUNY, City University of New York (dependent on the city government)

33 2 031 001 303 01 = Manhattan Community College (one campus of CUNY)
Sample Design

Multistage sample design

- **PPS sample**
  - Stratified PPS \((\text{state} \times \text{type})\) based on Total Pay

- **Cut-off sampling method in sizable \((\text{state, type})\) strata**
  - Construct a cut-off point to determine small and large size units (two strata)

- **Modified cut-off sampling** (a stratified PPS sample method)
  - Sub-sampling on small strata
Sample

Sampling Frame

\[ \pi ps \]

Certainties

Births

Sample

<table>
<thead>
<tr>
<th>Function Code (f)</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>j</td>
<td>51</td>
<td></td>
</tr>
<tr>
<td>001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>005</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>162</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

\[ \hat{y}_{gf} \]
Small Area Challenges

- Designed at (state, type) level, estimated at state by function level

- Estimate total employees and total payroll at state by function level

\[ Y_{gf} = \sum_{i \in U_{gf}} Y_{gfi} \text{ where } g = \text{state}, \text{ and } f = \text{function} \]
Other Challenges

Skew data - Not Transform
Other Challenges (Cont’d)

Skew data - Log Transform
Estimators- ASPEP

- Direct
  Horvitz-Thompson: \( \hat{y}_{gf}^{HT} = \sum w_{gfi} y_{gfi} \)

- Composite

- Battese, Harter, Fuller (BHF) Model

- Our Proposed Model
Composite Estimator

\[ \hat{y}_{gf}^{\text{composite}} = \phi_g \hat{y}_{gf}^{HT} + (1 - \phi_g) \hat{y}_{gf}^{\text{synthetic}} \]

where \( g = \) state, \( f = \) function code

\[ \hat{y}_{gf}^{\text{synthetic}} = \hat{K}_{gf} \hat{Y}_g \]
Estimators- ASPEP
Composite Weight (Cont’d)

- Purcell & Kish (1979)

\[ w_{gf}^{\text{Sf}} = 1 - \frac{\sum_{g \in G, f \in F} v(\hat{Y}_{gf}^D)}{\sum_{g \in G, f \in F} (\hat{Y}_{i}^S - \hat{Y}_{i}^D)^2} \]

- Issue: Negative in some
  \( i = \) (state, function code)
  - Fixable (Lahiri & Pramanik, 2010)
Composite Estimators (Cont’d)

\[
\hat{Y}_{gf} = \hat{K}_{gf} \hat{Y}_g
\]

Direct (HT):

\[
\hat{Y}^{HT}_{gf}
\]

Synthetic:

\[
y_{gf}^{syn} = \hat{K}_{gf} \hat{Y}_g
\]

Composite:

\[
y_{gf}^{composite}
\]

\[
\hat{K}_{gf} = \frac{x_{gf}}{\sum_f x_{gf}}
\]

2009 ASPEP regress on 2007 Census (decision-based)
Estimators (Cont’d)
Battese, Harter, Fuller (BHF) Model

\[ y_{ij} = \beta_0 + \beta_1 x_i + v_i + \epsilon_{ij} \]

- \( y_{ij} \): the number of full-time employees for the \( j^{th} \) governmental unit within the \( i^{th} \) small area
- \( x_i \): number of full-time employees for the \( i^{th} \) small area obtained from the previous census
- \( \beta_0 \) and \( \beta_1 \): unknown intercept and slope, respectively; \( v_i \) are small area specific random effects
- \( \epsilon_{ij} \): errors in individual observations
Estimators (Cont’d)
Our Proposed Model

\[
\log(y_{ij}) = \beta_0 + \beta_1 \log(x_i) + \nu_i + \varepsilon_{ij}
\]

where

\[
\nu_i \sim N(0, \tau^2) \quad \text{and} \quad \varepsilon_{ij} \sim N(0, \sigma^2)
\]
Data for Evaluation

Government units that overlap between the 2002 and 2007 Census of Governments reporting strictly positive numbers of full-time employees.
Evaluation

- Performance of log transform EB
  - Results
  - Residuals Diagnostic
  - EB performance in small area
  - Benchmark Ratio (BR)
    - EB $\rightarrow$ HT when n becomes larger

- Smoothening the EB
  - One-way raking state totals to the direct (HT)
  - Two-way raking state by function totals to the HT
Evaluation - Results

- Out of 1,225 (CA, function code) cells
  - 671 cases (clear winner) → our model
  - 324 cases → HT
  - 230 cases → Composite

- No significant difference
  - 160 cases between log-transformed model and the HT
  - 145 cases between the composite and the HT

- HT won in cells where more than 70% of the units were large certainties

- Testing for significance, our model can be used in 831 out of 1,225 cells (≈68%)
Evaluation - Results

Table 1: Percent Relative Error for Differences Estimates of Full Time Employees to the Truth (California)

<table>
<thead>
<tr>
<th>Function</th>
<th>HT</th>
<th>Proposed</th>
<th>BHF</th>
<th>n_pps</th>
<th>n_pps/n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gas Supply</td>
<td>42.1%</td>
<td>(11.8%)</td>
<td>(30.6%)</td>
<td>1</td>
<td>50.0%</td>
</tr>
<tr>
<td>Correction</td>
<td>0.77%</td>
<td>0.17%</td>
<td>(3.46%)</td>
<td>2</td>
<td>5.41%</td>
</tr>
<tr>
<td>Welfare</td>
<td>(1.65%)</td>
<td>(0.14%)</td>
<td>(3.30%)</td>
<td>2</td>
<td>3.45%</td>
</tr>
<tr>
<td>Water Transport &amp; Terminals</td>
<td>34.4%</td>
<td>(1.64%)</td>
<td>(15.4%)</td>
<td>3</td>
<td>27.3%</td>
</tr>
<tr>
<td>Higher Education - Other</td>
<td>6.12%</td>
<td>(0.19%)</td>
<td>(9.97%)</td>
<td>3</td>
<td>5.66%</td>
</tr>
<tr>
<td>Higher Education - Instructional</td>
<td>4.72%</td>
<td>0.86%</td>
<td>(9.14%)</td>
<td>3</td>
<td>5.66%</td>
</tr>
<tr>
<td>Electric Power</td>
<td>(1.22%)</td>
<td>(0.30%)</td>
<td>(4.87%)</td>
<td>3</td>
<td>15.8%</td>
</tr>
<tr>
<td>Airports</td>
<td>4.35%</td>
<td>(0.49%)</td>
<td>(2.49%)</td>
<td>3</td>
<td>6.67%</td>
</tr>
<tr>
<td>Health</td>
<td>(2.93%)</td>
<td>(0.08%)</td>
<td>(6.26%)</td>
<td>6</td>
<td>9.09%</td>
</tr>
<tr>
<td>Natural Resources</td>
<td>(3.56%)</td>
<td>(2.46%)</td>
<td>(25.0%)</td>
<td>7</td>
<td>14.0%</td>
</tr>
<tr>
<td>Judicial &amp; Legal</td>
<td>0.44%</td>
<td>0.82%</td>
<td>(2.21%)</td>
<td>8</td>
<td>7.77%</td>
</tr>
<tr>
<td>Hospitals</td>
<td>5.17%</td>
<td>(0.71%)</td>
<td>(5.81%)</td>
<td>9</td>
<td>23.1%</td>
</tr>
<tr>
<td>Transit</td>
<td>(1.15%)</td>
<td>(1.18%)</td>
<td>(8.49%)</td>
<td>12</td>
<td>21.8%</td>
</tr>
<tr>
<td>Local Libraries</td>
<td>5.82%</td>
<td>(0.06%)</td>
<td>(10.6%)</td>
<td>12</td>
<td>13.3%</td>
</tr>
<tr>
<td>Solid Waste Management</td>
<td>3.81%</td>
<td>(1.58%)</td>
<td>(12.3%)</td>
<td>13</td>
<td>13.1%</td>
</tr>
<tr>
<td>Fire - Other</td>
<td>(9.02%)</td>
<td>(1.23%)</td>
<td>(10.1%)</td>
<td>16</td>
<td>17.0%</td>
</tr>
<tr>
<td>Housing &amp; Community Development (Local)</td>
<td>(4.80%)</td>
<td>(2.11%)</td>
<td>(27.6%)</td>
<td>19</td>
<td>14.5%</td>
</tr>
<tr>
<td>Police-Other</td>
<td>2.10%</td>
<td>(0.12%)</td>
<td>(11.3%)</td>
<td>19</td>
<td>13.8%</td>
</tr>
<tr>
<td>Police Protection - Officers</td>
<td>0.21%</td>
<td>(0.21%)</td>
<td>(14.4%)</td>
<td>19</td>
<td>14.4%</td>
</tr>
<tr>
<td>Streets &amp; Highways</td>
<td>(3.27%)</td>
<td>0.11%</td>
<td>(19.7%)</td>
<td>20</td>
<td>13.3%</td>
</tr>
<tr>
<td>Other Government Administration</td>
<td>(1.87%)</td>
<td>(0.12%)</td>
<td>(16.2%)</td>
<td>20</td>
<td>13.2%</td>
</tr>
<tr>
<td>Financial Administration</td>
<td>(1.50%)</td>
<td>(0.65%)</td>
<td>(12.0%)</td>
<td>20</td>
<td>13.1%</td>
</tr>
<tr>
<td>Sewerage</td>
<td>3.68%</td>
<td>(1.91%)</td>
<td>(20.9%)</td>
<td>22</td>
<td>20.6%</td>
</tr>
<tr>
<td>Other &amp; Unallocable</td>
<td>(0.20%)</td>
<td>(1.65%)</td>
<td>(14.5%)</td>
<td>22</td>
<td>15.4%</td>
</tr>
<tr>
<td>Firefighters</td>
<td>3.08%</td>
<td>(1.36%)</td>
<td>(19.5%)</td>
<td>23</td>
<td>22.1%</td>
</tr>
<tr>
<td>Parks &amp; Recreation</td>
<td>2.26%</td>
<td>(2.11%)</td>
<td>(19.3%)</td>
<td>24</td>
<td>16.2%</td>
</tr>
<tr>
<td>Water Supply</td>
<td>1.42%</td>
<td>(7.20%)</td>
<td>(30.5%)</td>
<td>32</td>
<td>28.3%</td>
</tr>
<tr>
<td>Elementary and Secondary - Other Total</td>
<td>(0.51%)</td>
<td>(2.92%)</td>
<td>(22.6%)</td>
<td>45</td>
<td>19.3%</td>
</tr>
<tr>
<td>Elementary and Secondary - Instruction</td>
<td>(0.48%)</td>
<td>(4.08%)</td>
<td>(27.7%)</td>
<td>46</td>
<td>19.7%</td>
</tr>
</tbody>
</table>
Evaluation (Cont’d)
Results- Diagnostic Analysis

Full Normal Plot Residuals: BHF Model
CALIFORNIA
Evaluation (Cont’d)
Results - Diagnostic Analysis

- QQ Plot for Our Model
Evaluation - Results
(For Gas Supply, All States, Average n= 4)

Figure 4: Distances of EB, HT to the Truth
Evaluation (Cont’d)

Benchmark Ratio (BR)

- $BR = \left| \sum (\text{estimate} - \text{HT})/\text{HT} \right|$
- Indicating how close the estimate is to the HT when considering large areas
Evaluation (Cont’d)
Results

Comparison of Benchmark Ratios (Nation)

<table>
<thead>
<tr>
<th>Size</th>
<th>BR for the EB</th>
<th>BR for the BHF</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 50</td>
<td>1.5</td>
<td>1.6</td>
</tr>
<tr>
<td>≥ 50</td>
<td>1.1</td>
<td>1.5</td>
</tr>
</tbody>
</table>
Figure 3: Distance of the Estimators to the Truth

(Function, Sample size)
From small n to big
Evaluation (Cont’d)
Raking: Log-transformed to HT Base (CA)

Figure 5: Effect of Benchmarking the Log Transformation

Function Code

Distance to True

Log
Log_Benchmarked
Evaluation (Cont’d)
Effect of Raking

Figure 6: The Effect of Benchmarking the Log Transform Where the HT is Better

Benchmarking improved
Evaluation (Cont’d)
Comparison: EB, Raking EB and HT

Figure 7: EB, EB Benchmark, and HT
Evaluation (Cont’d)
Domain Analysis (Gas Supply, AVG n=4)

EB= log(full-time employees), Benchmarked-EB= EB benchmarked to HT (one-way raking to nation total)

Figure 8: EB, Benchmarked-EB, HT, and BHF
Evaluation (Cont’d)
Overall - Relative Errors

Table 2: Comparison of Overall Relative Errors (CA)

<table>
<thead>
<tr>
<th>Overall - Absolute Relative Errors</th>
<th>Overall - Relative Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Sigma</td>
<td>(HT-True)/True</td>
</tr>
<tr>
<td>5.26%</td>
<td>1.67%</td>
</tr>
<tr>
<td>3.05%</td>
<td>-1.5%</td>
</tr>
</tbody>
</table>
Evaluation (Cont’d)
Two-way Raking: (States, Functions)

- Two-way raking:
  - All states to National total
  - All functions to National functions

- 255 underestimated cases goes down to 210 cases.
Acknowledgements

- Thankfully for strong support to this research
  - Carma Hogue (Assistant Division Chief)
  - Lisa Blumerman (Division Chief)

- Technical advice/review
  - Dr. Partha Lahiri
Contact Information

Bac Tran

Bac.Tran@census.gov

Program Research Branch, Chief
Governments Division
U.S. Census Bureau
Thank you for your time!

Questions?