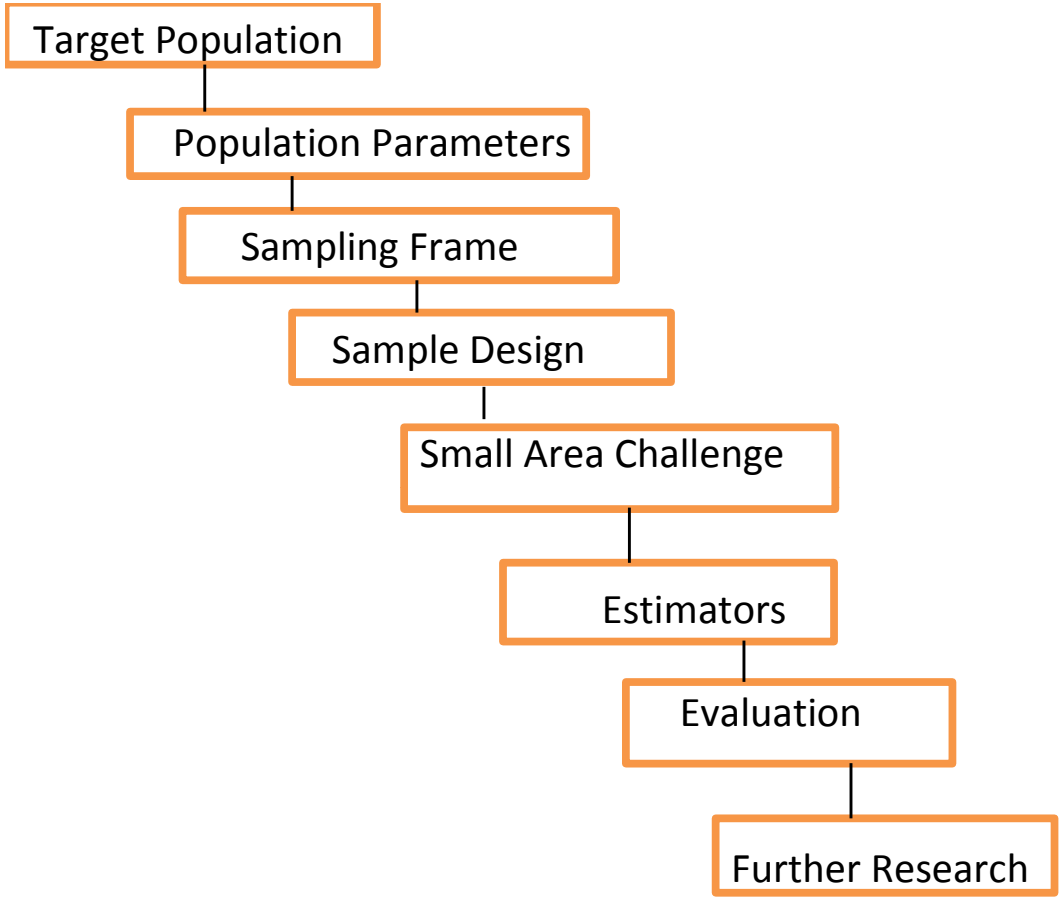
Small Area Estimation Applications in the US Census Bureau

Yang Cheng
Bac Tran
Partha Lahiri
Carma Hogue

U.S. Census Bureau

United States Census Bureau
Economic Statistics

Outline


Target Population
Population Parameters
Sampling Frame
Sample Design
Small Area Challenge
Estimators
Evaluation
Further Research

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
  - Schools 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

http://www2.census.gov/govs/apes/10locmd.txt
2010 Public Employment and Payroll Data
Local Governments
MARYLAND

SOURCE: 2010 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.

Government Function	Full-time employees	Full-time pay ($)	Part-time employees	Part-time pay ($)	Full-Time Equivalent Employment	Total March Pay ($)
Total	189,620	984,236,113	59,634	89,231,689	214,213	1,073,467,802
Financial Administration	2,285	11,454,282	147	268,486	2,350	11,722,768
Other Government Administration	3,300	16,966,287	844	1,565,802	3,692	18,532,089
Judicial and Legal	3,233	16,149,220	363	681,272	3,438	16,830,492
Police Protection Total	15,983	93,050,897	1,381	1,603,148	16,620	94,654,045
Police Officers Only	12,278	75,342,746	148	249,672	12,362	75,592,418
Other Police Employees	3,705	17,708,151	1,233	1,353,476	4,258	19,061,627
Fire Protection Total	6,772	40,058,581	153	252,374	6,845	40,310,955
Firefighters Only	6,222	37,071,603	43	52,648	6,242	37,124,251
Other Fire Employees	550	2,986,978	110	199,726	603	3,186,704
Corrections	3,559	17,501,794	73	144,404	3,608	17,646,198
Highways	5,267	21,153,791	99	165,216	5,313	21,319,007
Air Transportation	39	150,081	45	46,878	56	196,959
Water Transport and Terminals	3	18,757	8	3,388	5	22,145
Public Welfare	2,579	11,536,891	1,321	2,456,860	3,455	13,993,751
Health	3,934	18,597,016	1,114	2,591,935	4,706	21,188,951

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)

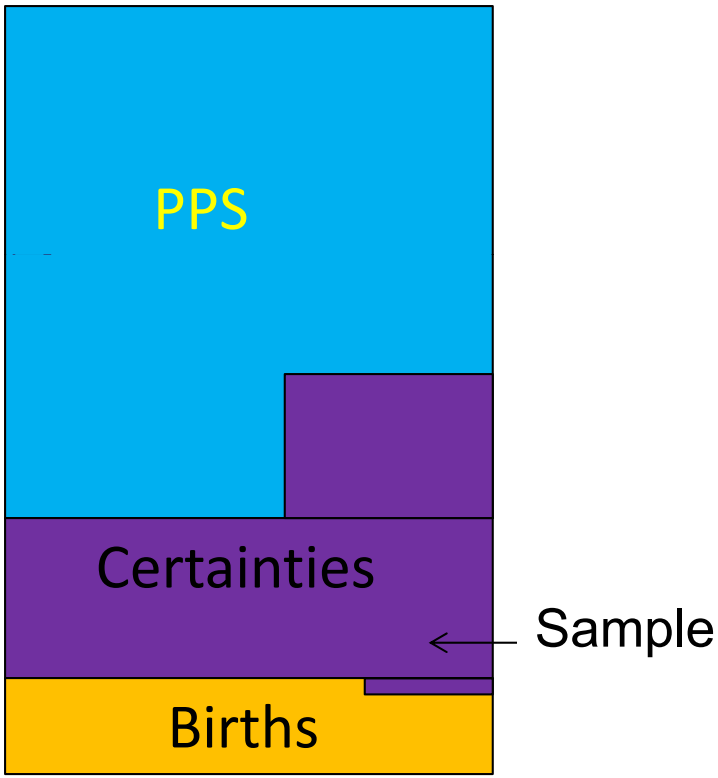
Sample Design

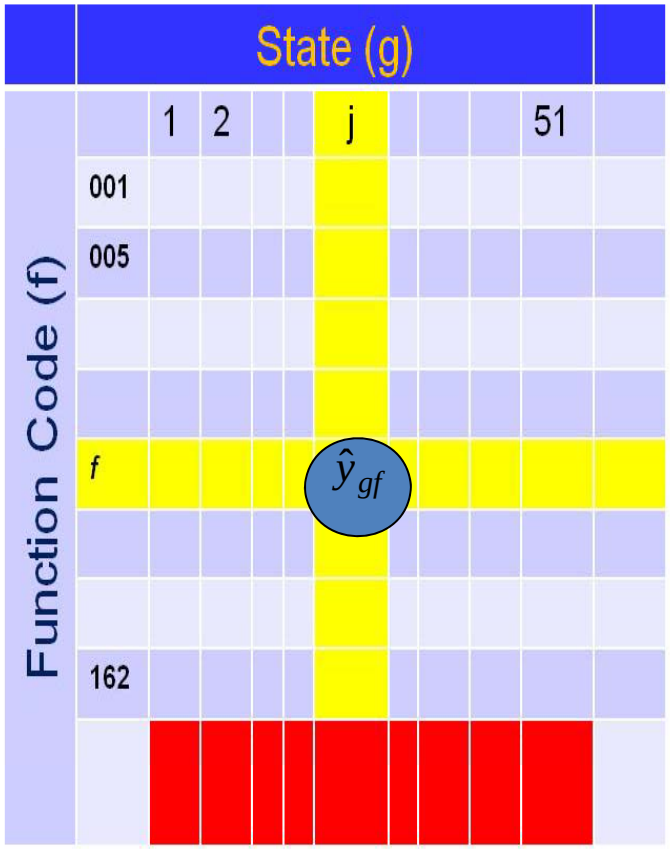
Multistage sample design

- PPS sample
  - Stratified PPS (state x type) based on Total Pay
- Cut-off sampling method in sizable (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





Small Area Challenge

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

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

$$Y_{gf} = \sum_{i \in U_{gf}} Y_{gfi} \; where \; g = state, and \; f = function$$

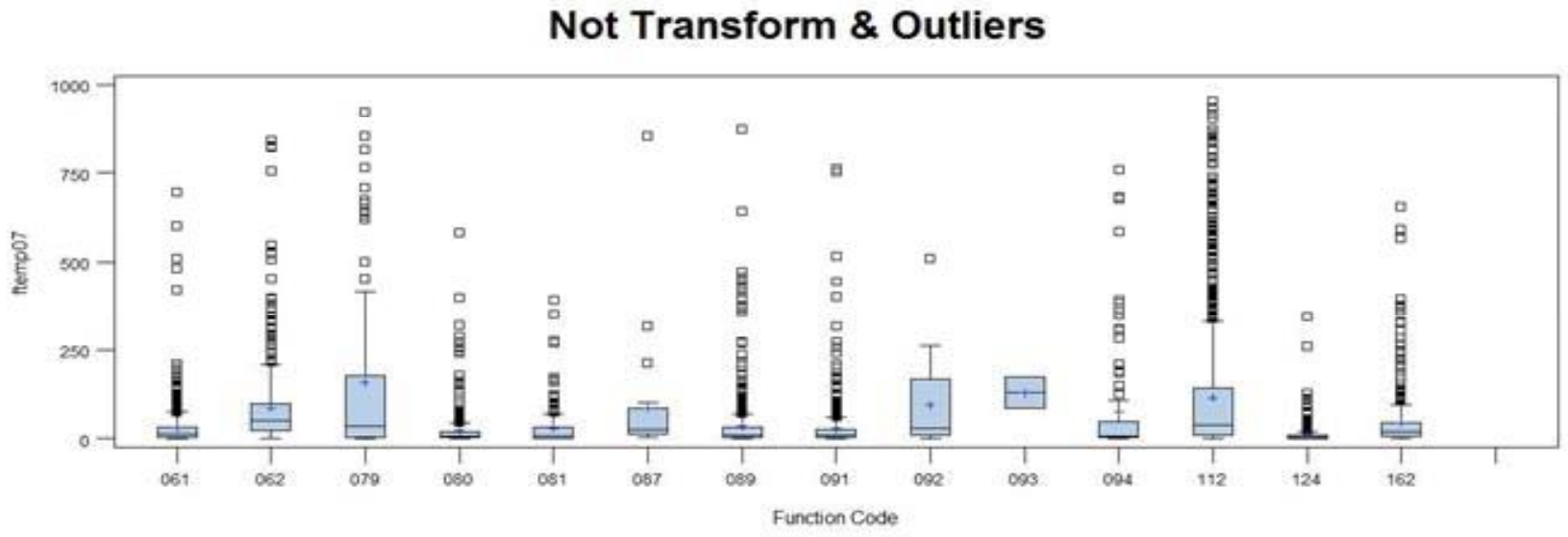
Small Area Challenge (Cont'd)

- Small area: a small geographic area within a larger geographic area or a small demographic group within a larger group

- Most small area estimation methods borrow strength from related or similar small areas using auxiliary data

Other Challenges

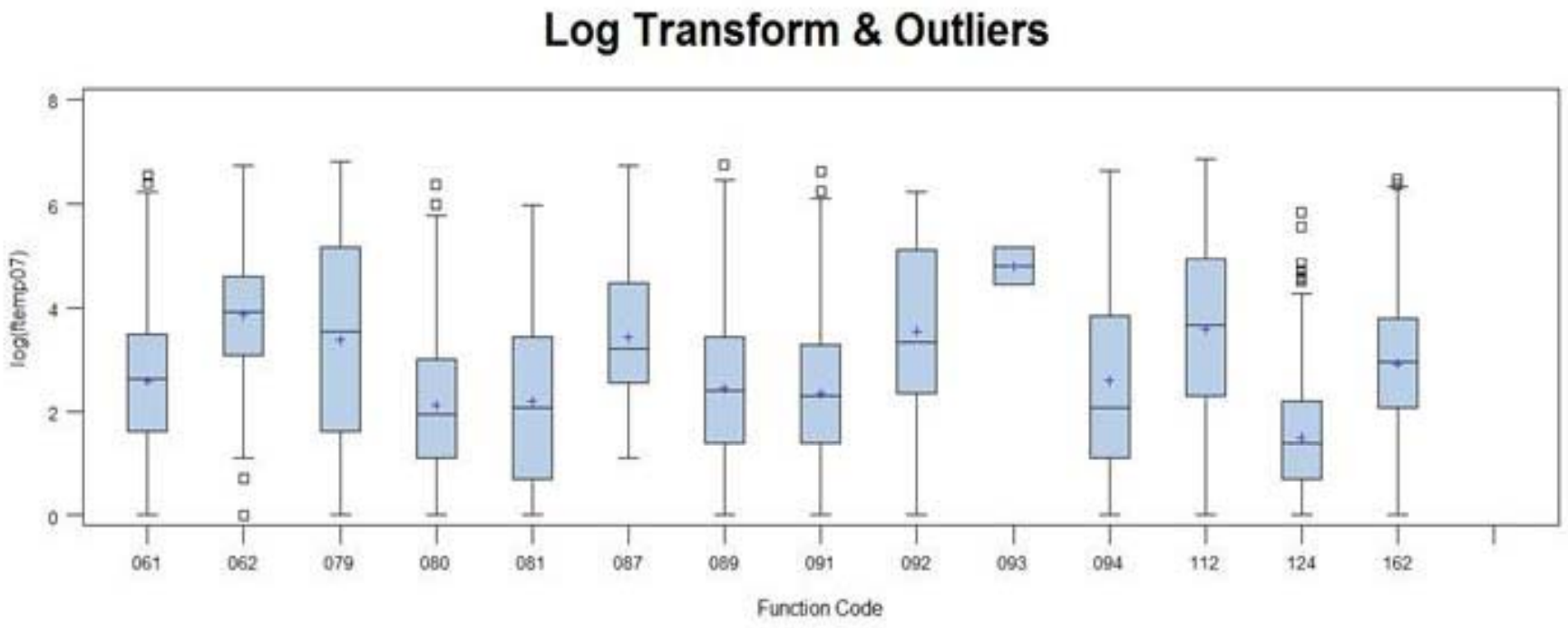
Figure 1: Skew data -Not Transform (California)
(Full-Time Employees, Function)



United States Census Bureau
Economic Statistics

Other Challenges (Cont'd)

Figure 2: Skew data - Log Transform (CA)
(Log(Full-Time Employees), Function)



Estimators- ASPEP

- Direct

  →Horvitz-Thompson: $\hat{y}_{gf}^{HT} = \sum w_{gf} y_{gf}$

- Battese, Harter, Fuller (BHF) Model
- Our Proposed Model

Estimators (Cont'd)
Battese, Harter, Fuller (BHF) Model

$$y_{ij} = \beta_0 + \beta_1 x_i + v_i + \varepsilon_{ij}$$

y_{ij} : the number of full-time employees for the jth governmental unit within the ith small area

x_i : number of full-time employees for the *i*th small area obtained from the previous census

$\beta_0, \text{and } \beta_1$: unknown intercept and slope, respectively; v_i are small area specific random effects

ε_{ij} : errors in individual observations

Estimators (Cont'd)
Our Proposed Model

$$\log(y_{ij}) = \beta_0 + \beta_1 \log(x_i) + v_i + \varepsilon_{ij}$$

where

$$v_i \overset{iid}{\sim} N(0, \tau^2) \quad \text{and} \quad \varepsilon_{ij} \overset{iid}{\sim} N(0, \sigma^2)$$

Evaluation Data

- California 2002 & 2007 Census ASPEP

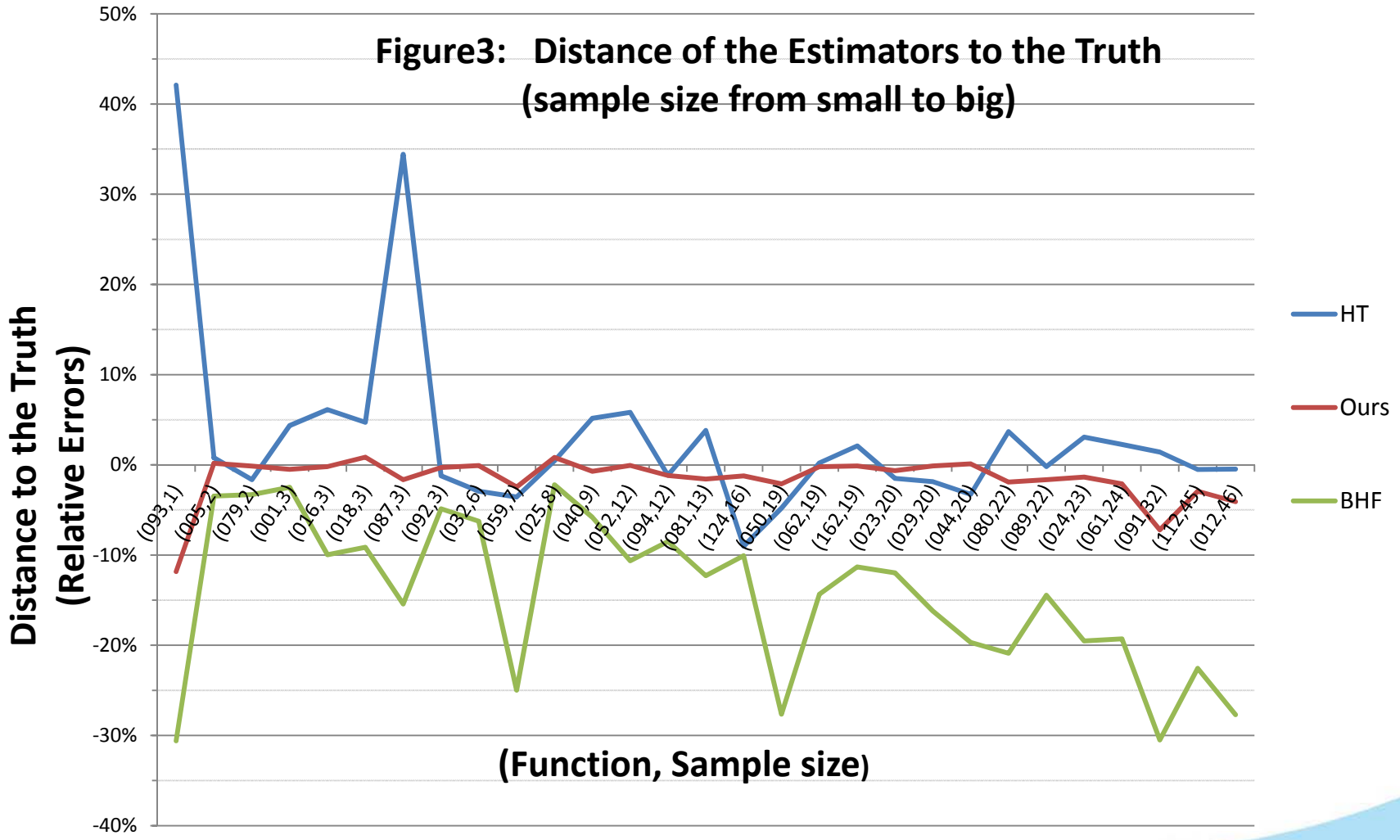
  government units that overlap between the 2002 and 2007 Census of Governments reporting strictly positive numbers of full-time employees.

Evaluation- Results

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

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

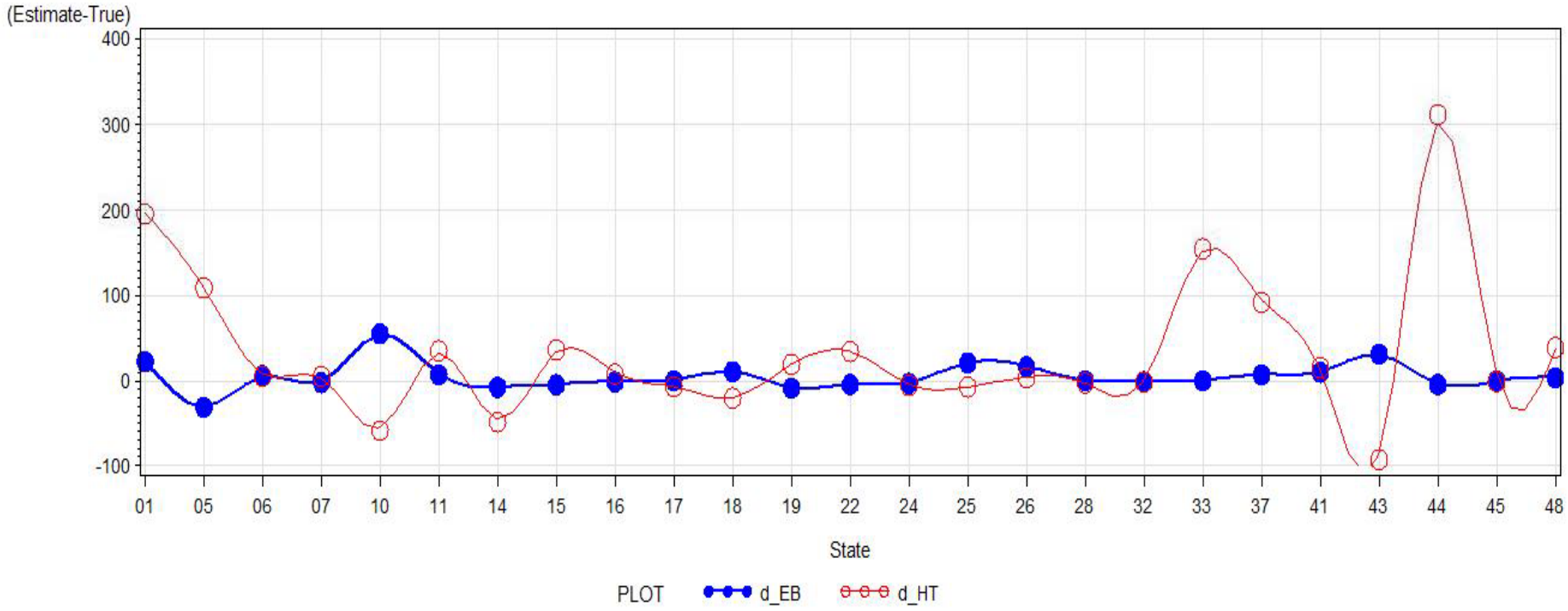
Evaluation (Cont'd)
Visualization of Table 1



United States Census Bureau
Economic Statistics

Evaluation- Results
(For Gas Supply, All States, Average n= 4)

Figure 4: Distances of EB, HT to the Truth

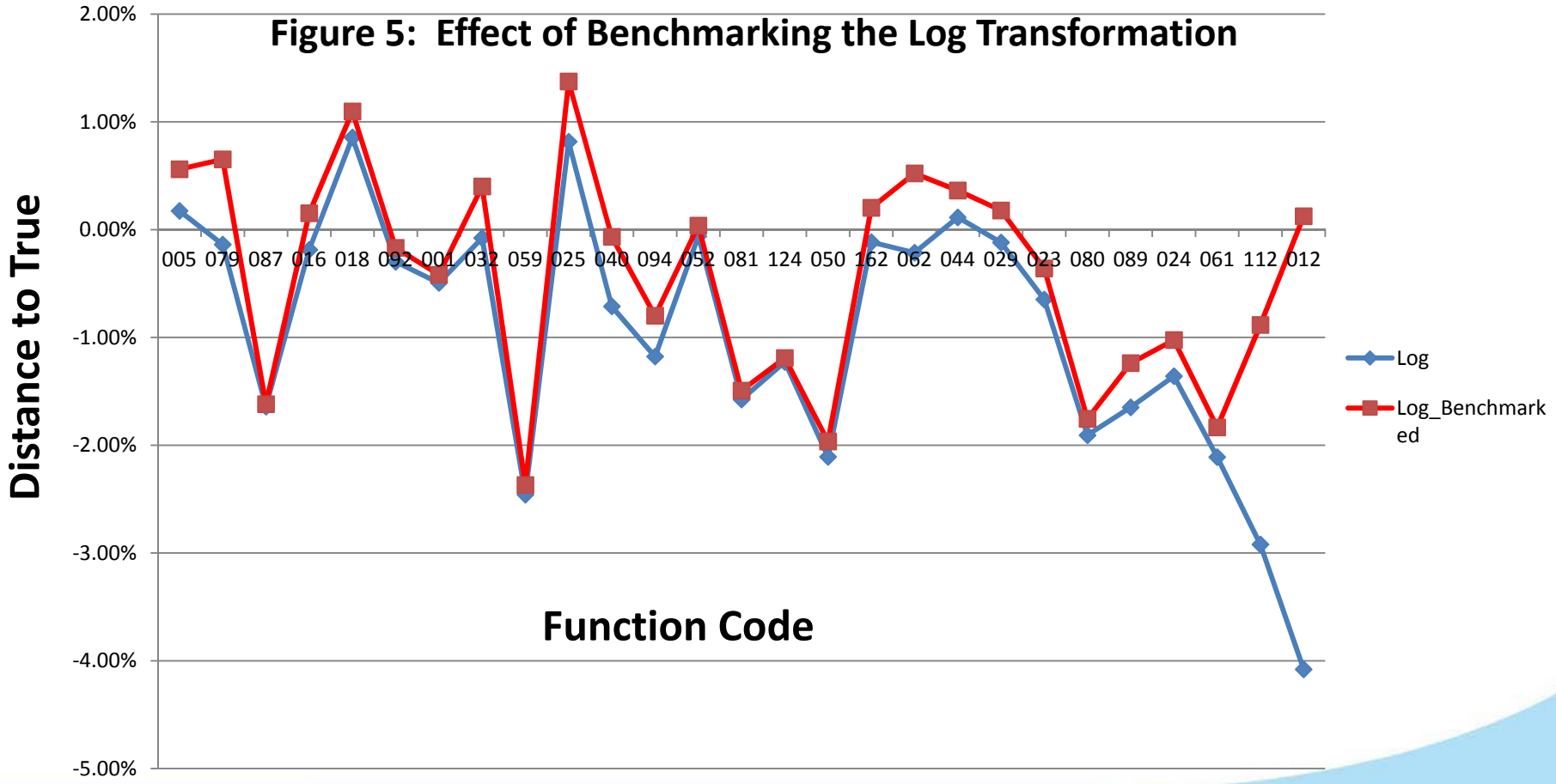


Evaluation (Cont'd)
Overall- Relative Errors

Table 2: Comparison of Overall Relative Errors (CA)

Overall - Absolute Relative Errors			
Σ|(HT-True)/True|	**Σ|(EB-True)/True|**	**Σ|(EB_benchmarked -True)/True|**	**Σ|(BHF-True)/True|**
5.26%	1.67%	1.44%	14.35%
Overall - Relative Errors			
Σ(HT-True)/True	**Σ(EB-True)/True**	**Σ(EB_benchmarked-True)/True**	**Σ(BHF-True)/True**
3.05%	-1.5%	-1%	-14.35%

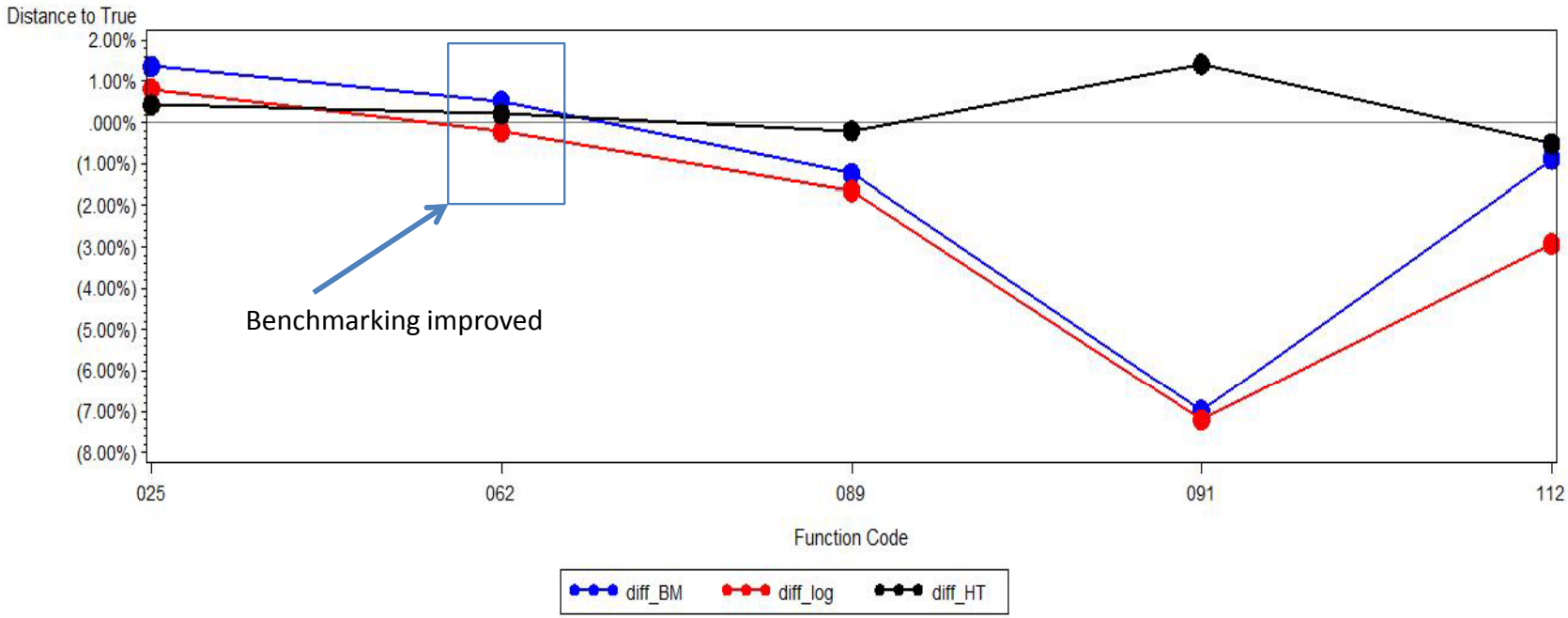
Evaluation (Cont'd)
Raking Log-transformed to HT Base (CA)



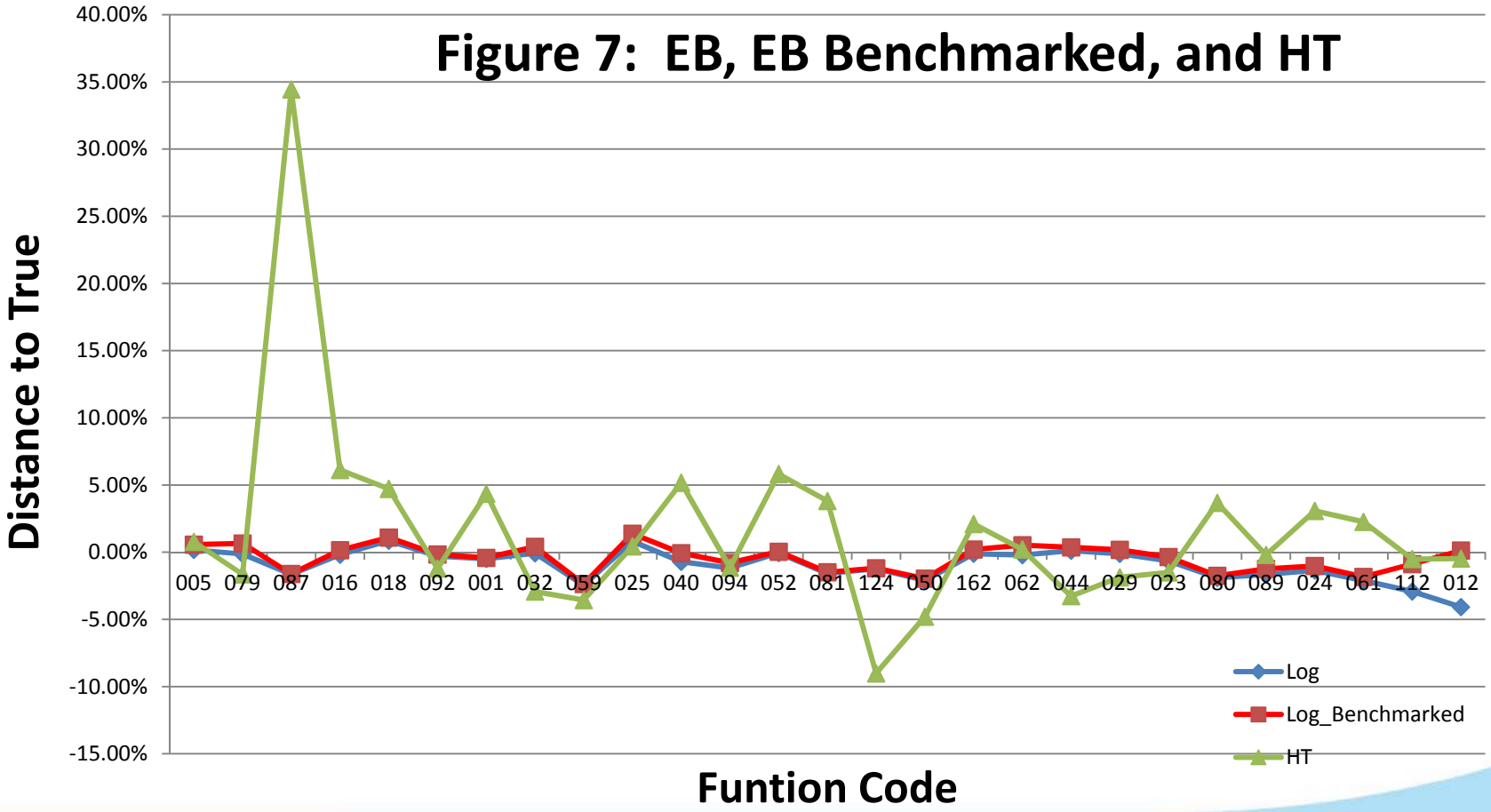
United States Census Bureau
Economic Statistics

Evaluation (Cont'd)

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



Evaluation (Cont'd)
Comparison: EB, EB Benchmarked and HT

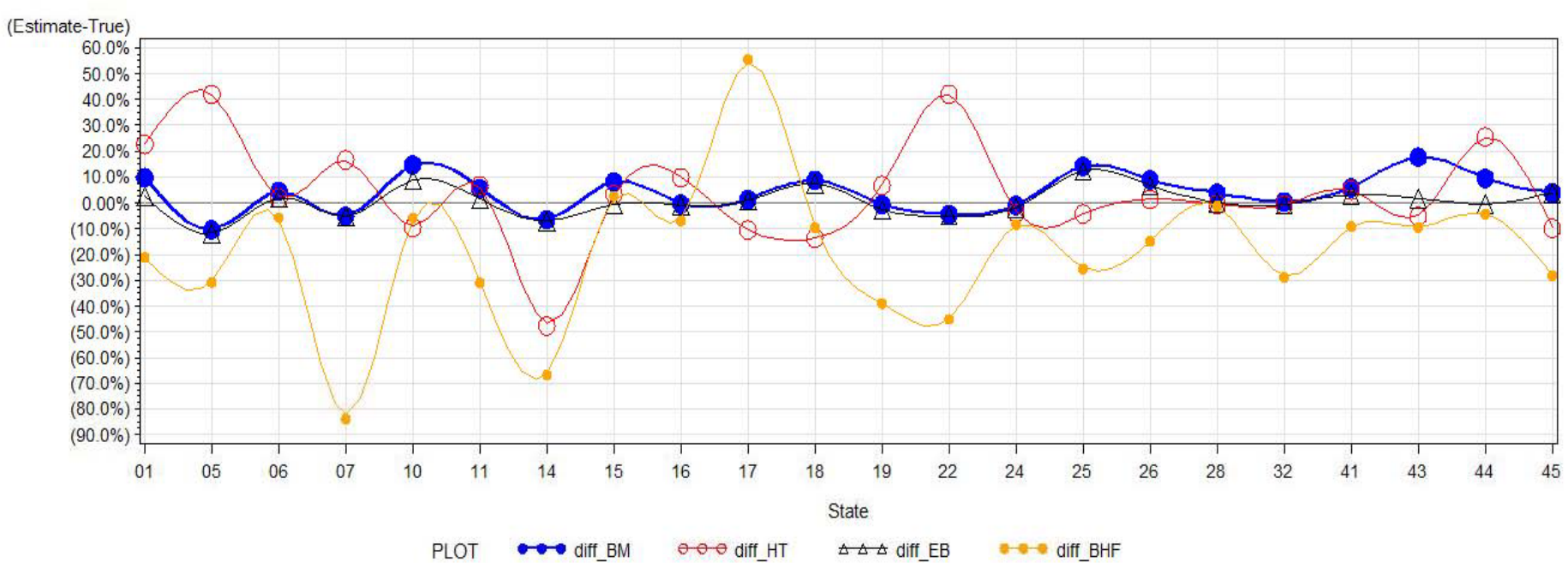


United States Census Bureau
Economic Statistics

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)
Results

- 24 out of 29 function codes (CA), our estimator outperforms the BHF, especially in small area (n <= 8)
- Benchmark Ratio (BR)
  - BR= |∑(estimate-HT)/HT|
  - Indicating how close the estimate is to the HT when considering large areas

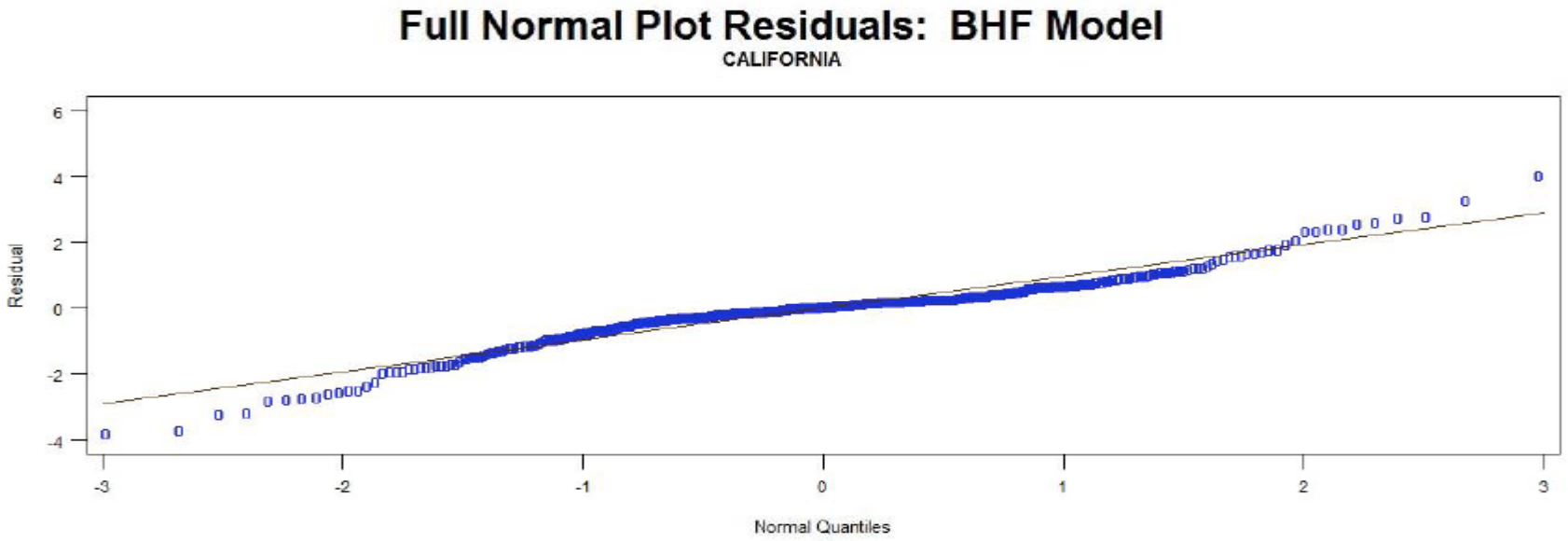
Evaluation (Cont'd)
Results

Comparison of Benchmark Ratios (Nation)

Size	BR for the EB	BR for the BHF	Number of units
< 50	1.5	1.6	1086
≥ 50	1.1	1.5	212

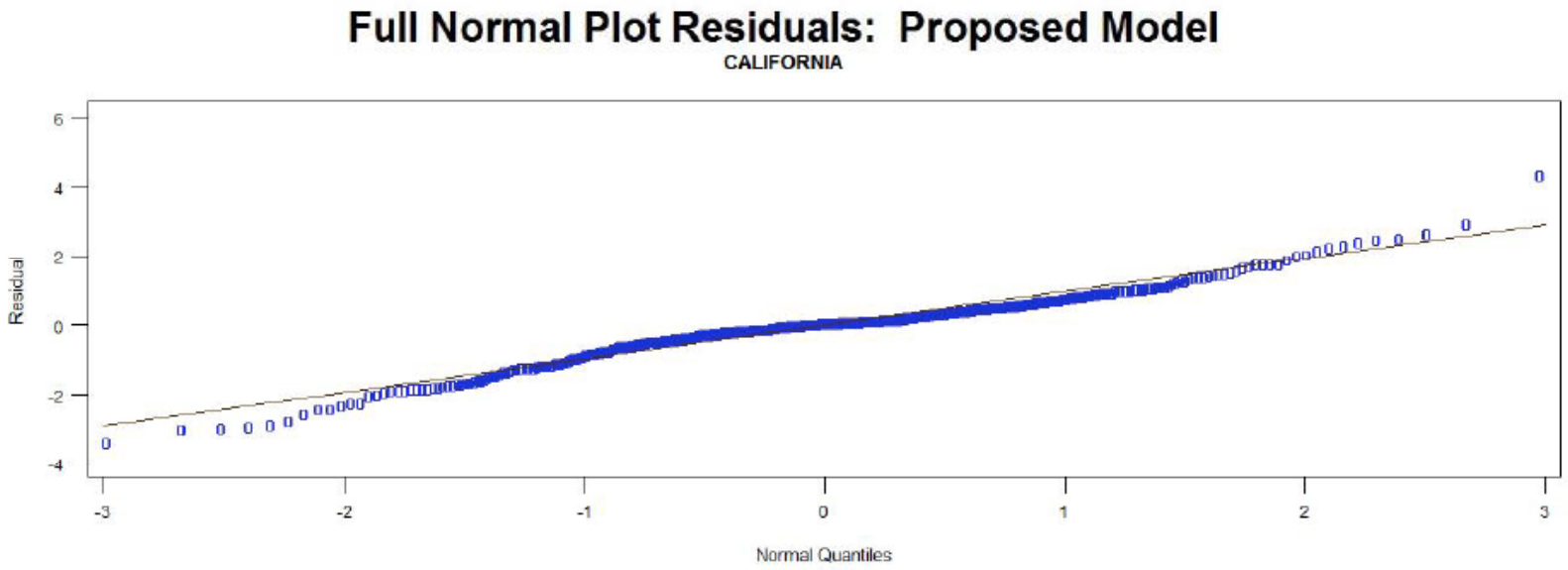
Evaluation (Cont'd)
Results- Diagnostic Analysis

Figure 9: QQ Plot for BHF Model



Evaluation (Cont'd)
Results- Diagnostic Analysis

Figure 10: QQ Plot for Our Model



Contact Information

Partha Lahiri

plahiri@survey.umd.edu

Yang Cheng

Yang.Cheng@census.gov

Bac Tran

Bac.Tran@census.gov

Carma Hogue

Carma.R.Hogue@census.gov