

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.

Government Function	Full-time employees	Full-time pay (\$)	Part-time employees	Part-time pay (\$)	Full-Time Equivalent Employment	Total March Pay (\$)
Total	10,781,323	46,773,983,844	3,318,107	3,631,752,914	11,999,059	50,405,736,758
Financial Administration	210,180	938,862,938	41,481	40,556,655	222,584	979,419,593
Other Government Administration	191,865	860,103,178	152,573	89,278,684	217,307	949,381,862
Judicial and Legal	232,038	1,089,735,867	33,339	46,238,354	243,980	1,135,974,221
Police Protection Total	782,739	4,208,166,047	106,504	100,358,135	819,445	4,308,524,183
Police Officers Only	609,849	3,507,627,164	38,434	39,386,380	623,259	3,547,013,544
Other Police Employees	172,890	700,538,883	68,070	60,971,756	196,186	761,510,639
Fire Protection Total	320,311	1,905,253,370	96,773	45,352,826	338,408	1,950,606,196
Firefighters Only	295,576	1,781,236,979	88,445	38,626,738	311,562	1,819,863,717
Other Fire Employees	24,735	124,016,391	8,327	6,726,088	26,846	130,742,479
Corrections	248,396	1,064,980,702	14,465	20,131,246	255,391	1,085,111,949
Highways	279,685	1,124,072,016	24,422	24,039,982	288,970	1,148,111,997
Air Transportation	41,340	205,379,744	3,847	4,943,095	43,085	210,322,839
Water Transport and Terminals	7,706	46,652,215	1,022	1,204,473	8,087	47,856,688
Public Welfare	248,393	997,087,839	38,804	54,808,888	268,225	1,051,896,728
Health	219,377	950,113,069	56,300	85,354,374	244,339	1,035,467,444
Hospitals	492,601	2,349,529,064	134,898	307,882,497	565,162	2,657,411,562

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	040, Hospitals
002, Space Research & Technology (Federal)	044, Streets & Highways
005, Correction	050, Housing & Community Development (Local)
006, National Defense and International Relations (Federal)	052, Local Libraries
012, Elementary and Secondary - Instruction	059, Natural Resources
112, Elementary and Secondary - Other Total	061, Parks & Recreation
014, Postal Service (Federal)	062, Police Protection - Officers
016, Higher Education - Other	162, Police-Other
018, Higher Education - Instructional	079, Welfare
021, Other Education (State)	080, Sewerage
022, Social Insurance Administration (State)	081, Solid Waste Management
023, Financial Administration	087, Water Transport & Terminals
024, Firefighters	089, Other & Unallocable
124, Fire - Other	090, Liquor Stores (State)
025, Judicial & Legal	091, Water Supply
029, Other Government Administration	092, Electric Power
032, Health	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

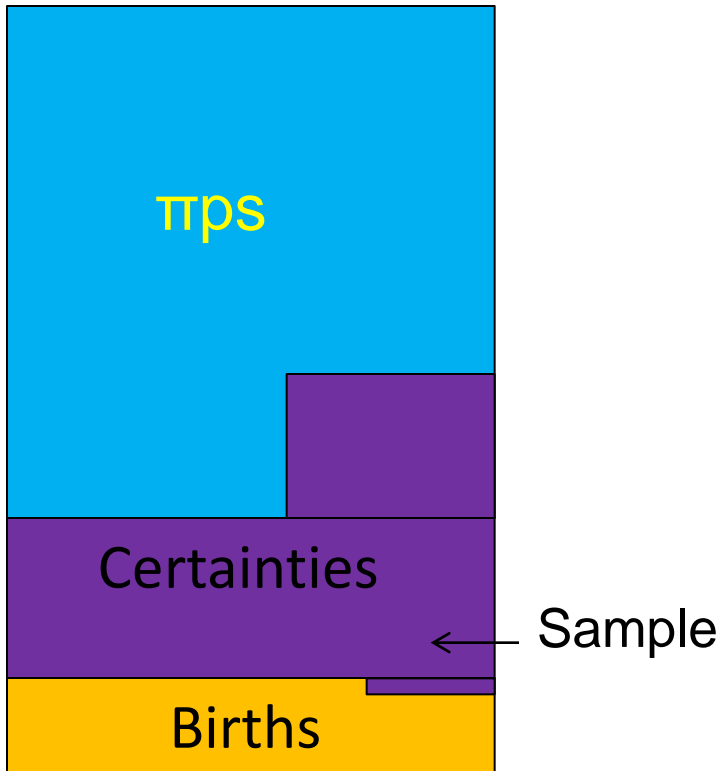
- Construct a cut-off point to determine small and large size units (two strata)

❑ Modified cut-off sampling (a stratified PPS sample method)

- Sub-sampling on small strata

Sample

Sampling Frame



		State (g)						
		1	2	j			51	
Function Code (f)	001							
	005							
	f							
	162							

The intersection of the row labeled **f** and the column labeled **j** is highlighted in yellow and contains a blue circle with the label \hat{y}_{gf} .

Small Area Challenges

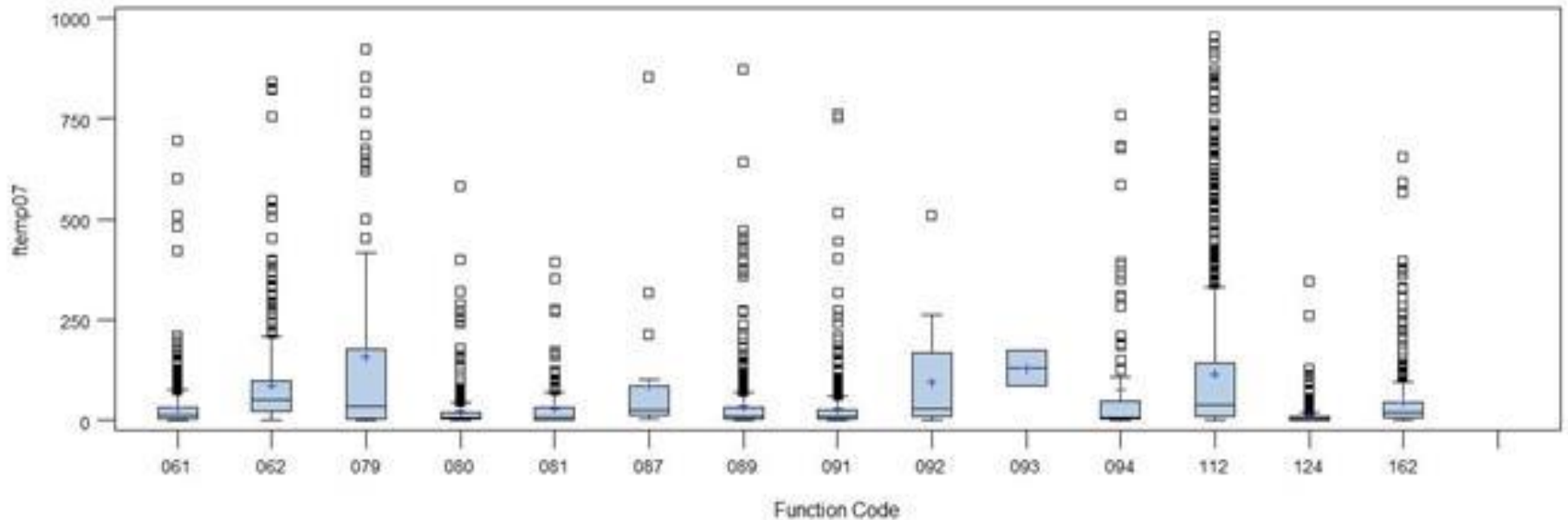
- ❑ Designed at (state, type) level, estimated at state by function level
- ❑ Estimate total employees and total payroll at state by function level

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

Other Challenges

Skew data- Not Transform

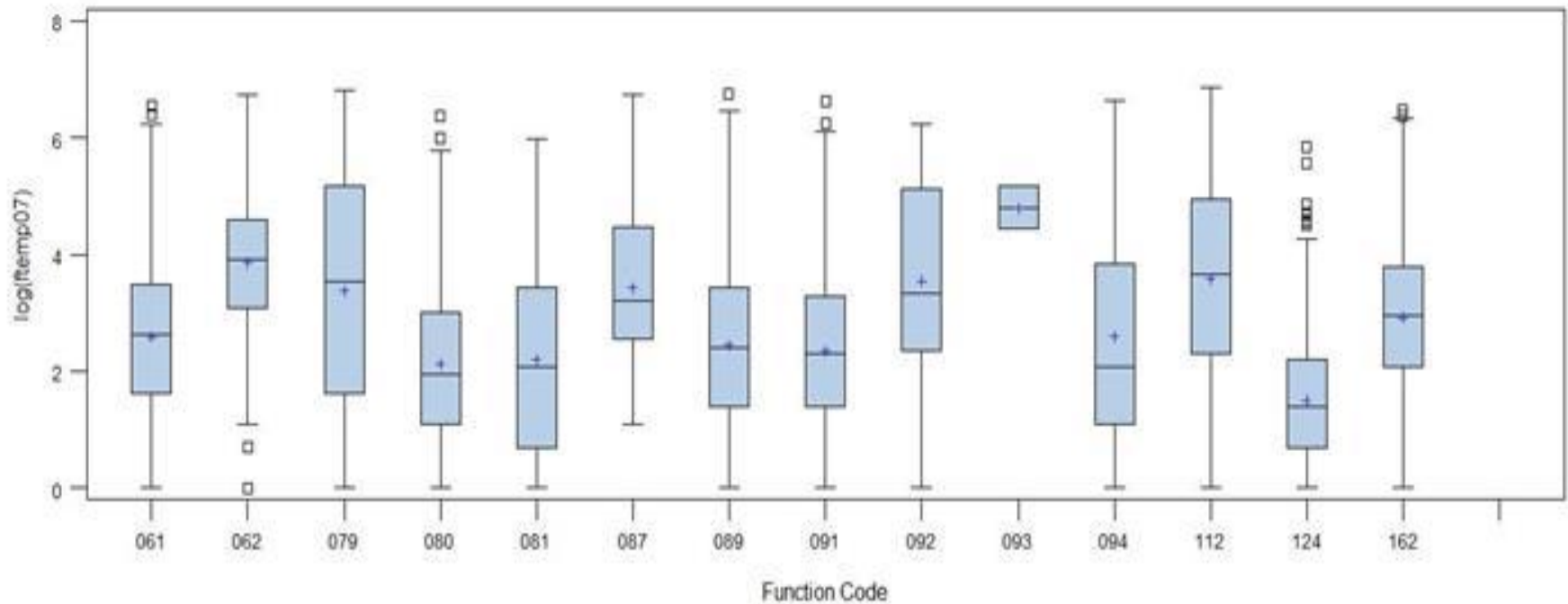
Not Transform & Outliers



Other Challenges (Cont'd)

Skew data- Log Transform

Log Transform & Outliers



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

$$\hat{y}_{gf}^{composite} = \hat{\phi}_g \hat{y}_{gf}^{HT} + (1 - \hat{\phi}_g) \hat{y}_{gf}^{synthetic}$$

where g= state, f= function code

$$\hat{y}_{gf}^{synthetic} = \hat{K}_{gf} \hat{Y}_g$$

Estimators- ASPEP

Composite Weight (Cont'd)

- Purcell & Kish (1979)

$$w^{gf} = 1 - \frac{\sum_{g \in G, f \in F} v(\hat{Y}_{gf}^D)}{\sum_{g \in G, f \in F} (\hat{Y}_i^S - \hat{Y}_i^D)^2}$$

- Issue: → Negative in some
 $i =$ (state, function code)
→ Fixable (Lahiri & Pramanik, 2010)

Composite Estimators (Cont'd)

		State (g)				
		1	2	j	...	51
Function Code (f)	001					
	005					
	f					
	162					

Direct (HT): \hat{y}_{gf}^{HT}
 Synthetic: $y_{gf}^{syn} = \hat{K}_{gf} \hat{Y}_g$
composite
 Composite: y_{gf}

$$\hat{K}_{gf} = \frac{x_{gf}}{\sum_f x_{gf}}$$

2009 ASPEP regress on 2007 Census (decision-based)

Estimators (Cont'd)

Battese, Harter, Fuller (BHF) Model

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

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

β_0 and β_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

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

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

□ Smoothing 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

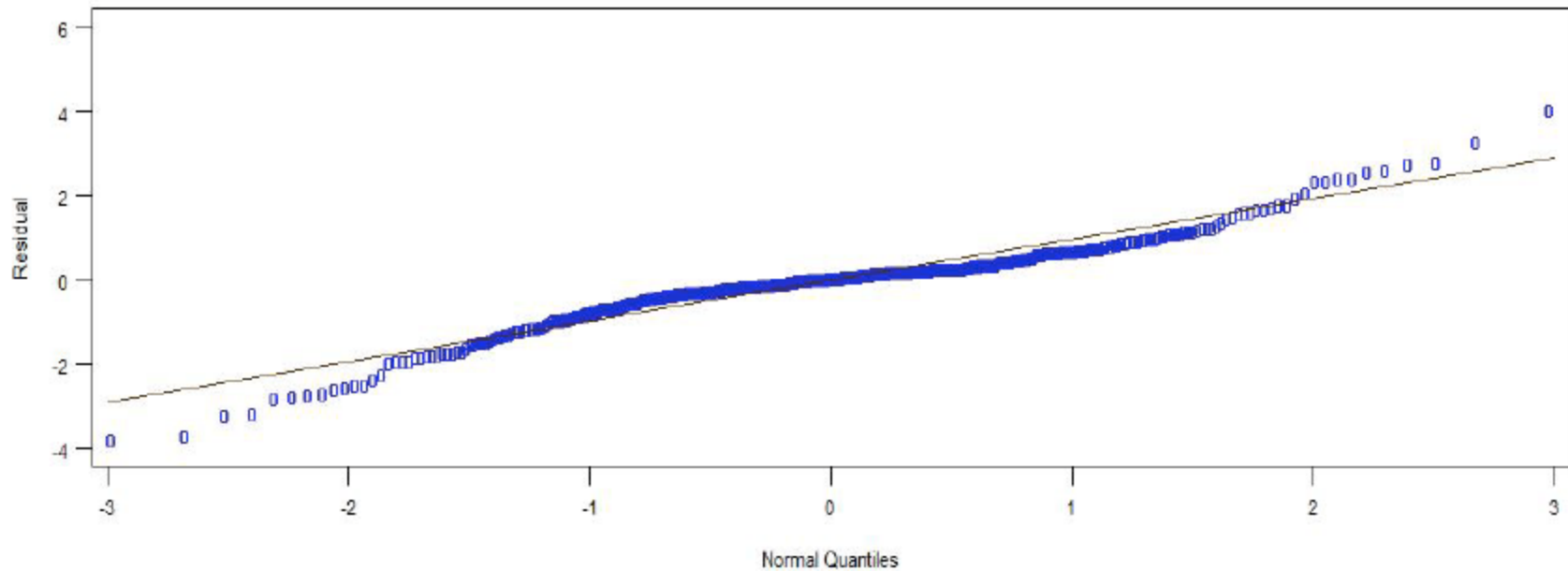
Table 1: Percent Relative Error for Differences Estimates of Full Time Employees to the Truth (California)

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%
Judicial & 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)

Results- Diagnostic Analysis

Full Normal Plot Residuals: BHF Model
CALIFORNIA



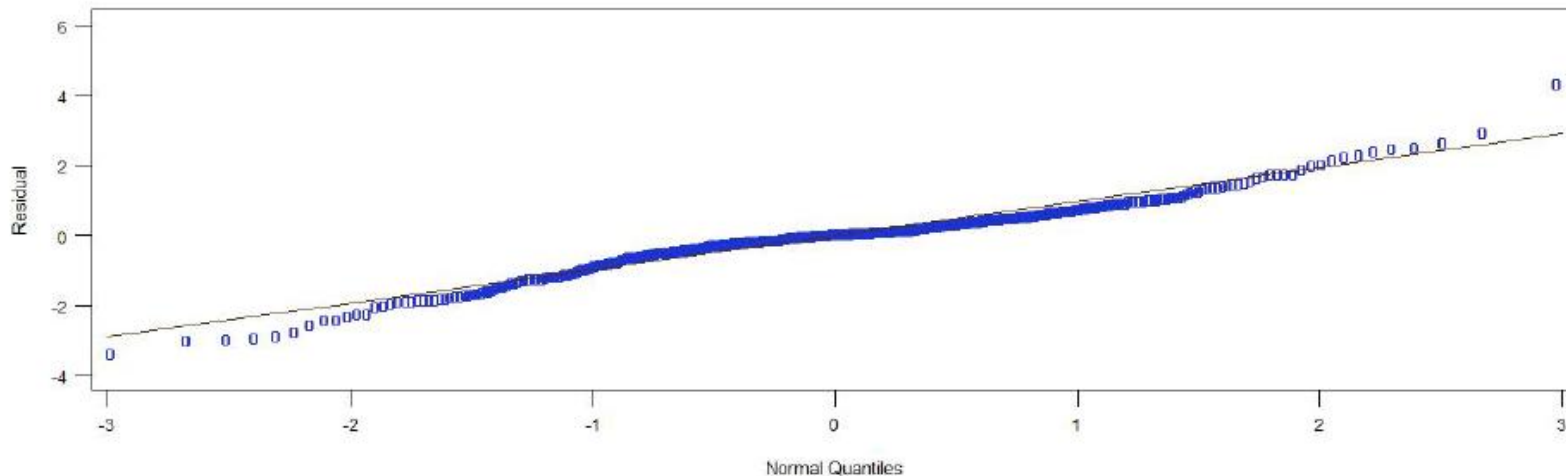
Evaluation (Cont'd)

Results- Diagnostic Analysis

□ QQ Plot for Our Model

Full Normal Plot Residuals: Proposed Model

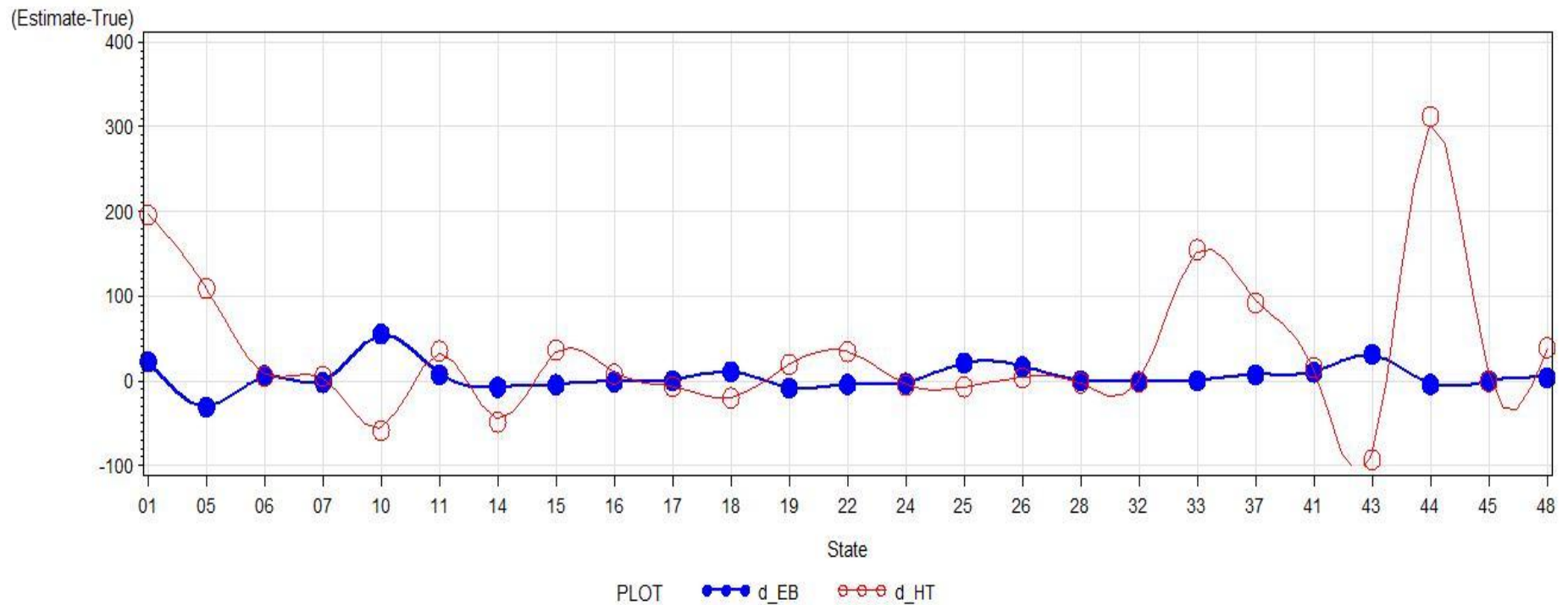
CALIFORNIA



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)

- $BR = |\sum(\text{estimate} - HT) / HT|$
- Indicating how close the estimate is to the HT when considering large areas

Evaluation (Cont'd)

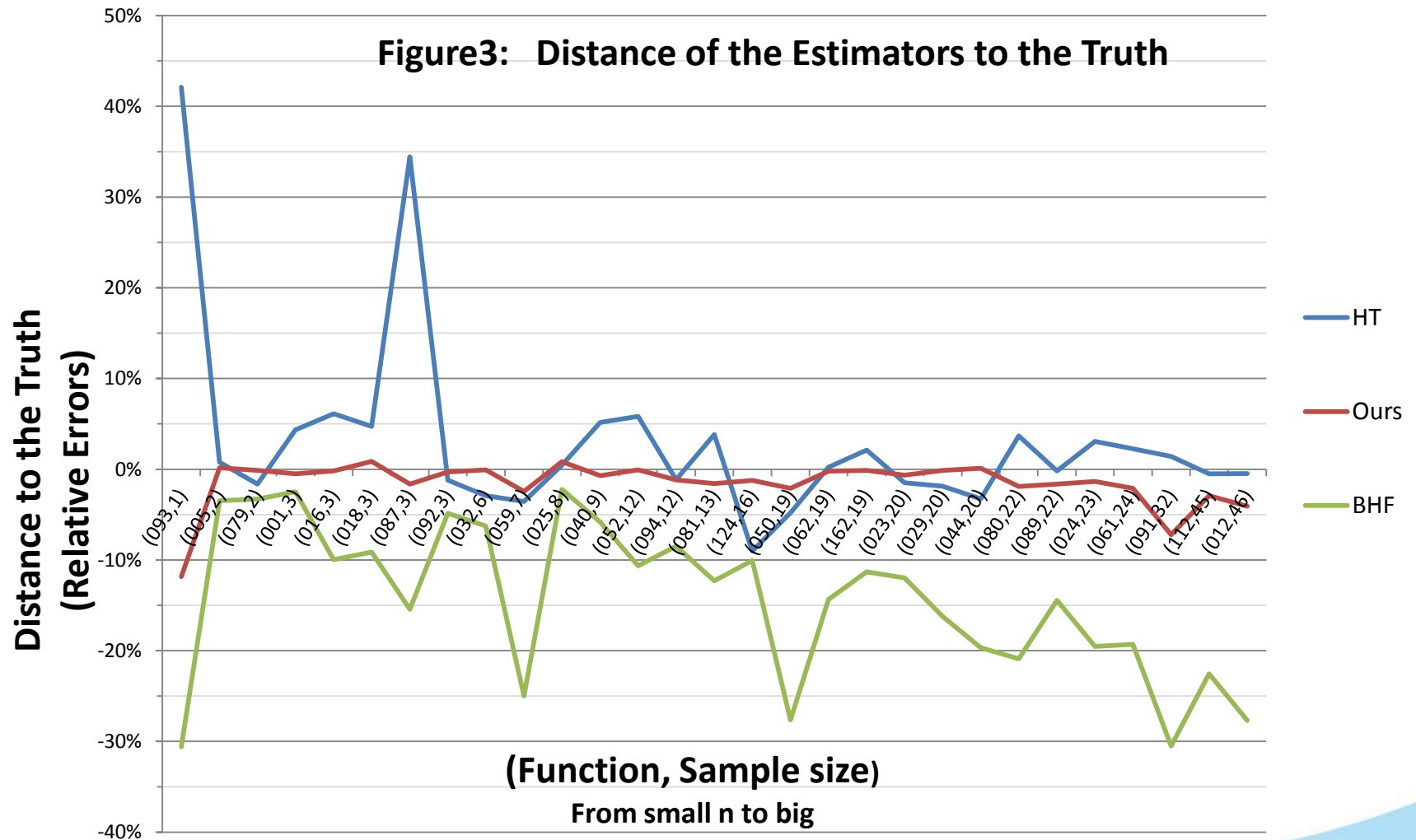
Results

Comparison of Benchmark Ratios (Nation)

Size	BR for the EB	BR for the BHF
< 50	1.5	1.6
≥ 50	1.1	1.5

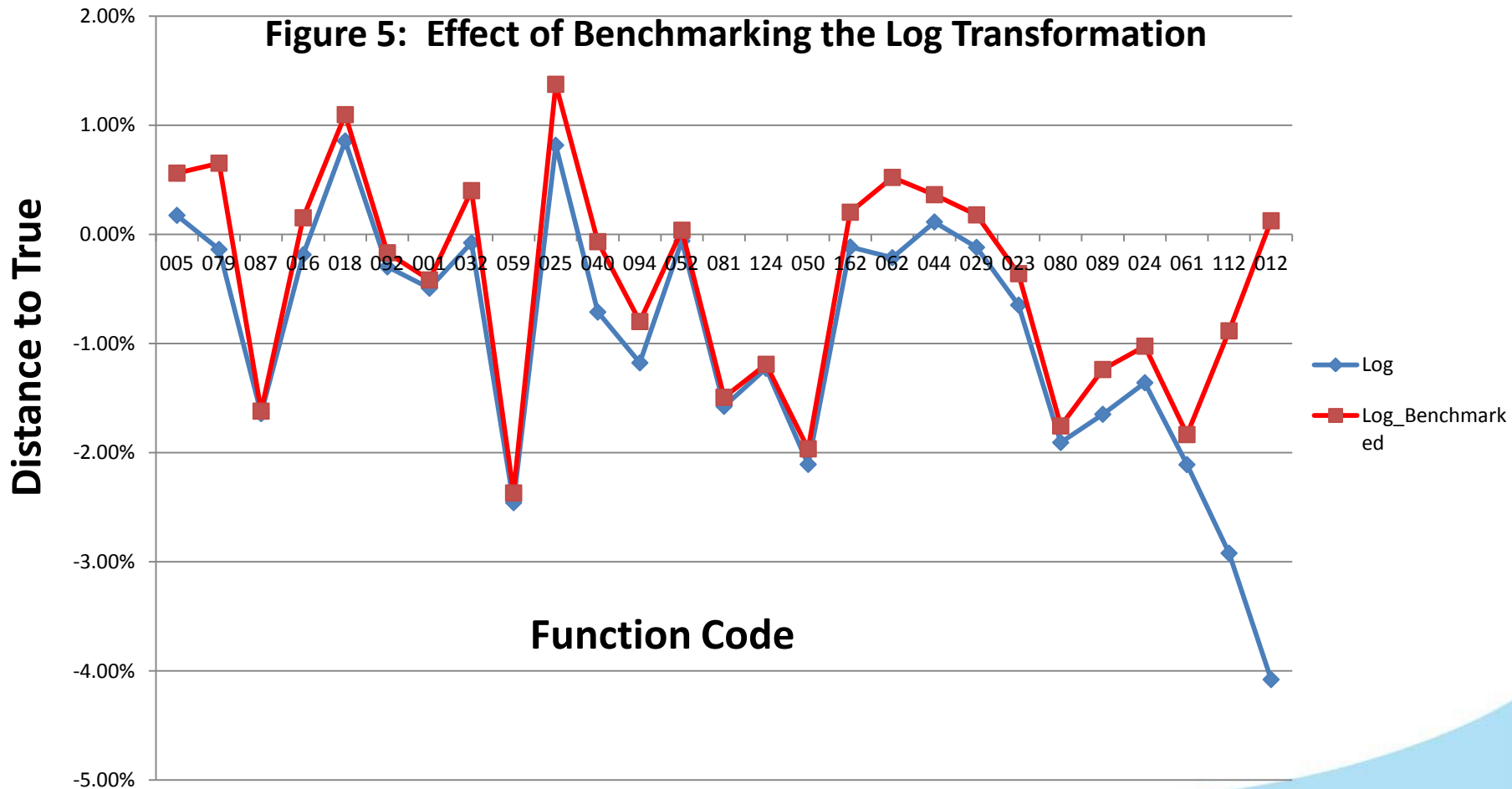
Evaluation (Cont'd)

Visualization of Table 1



Evaluation (Cont'd)

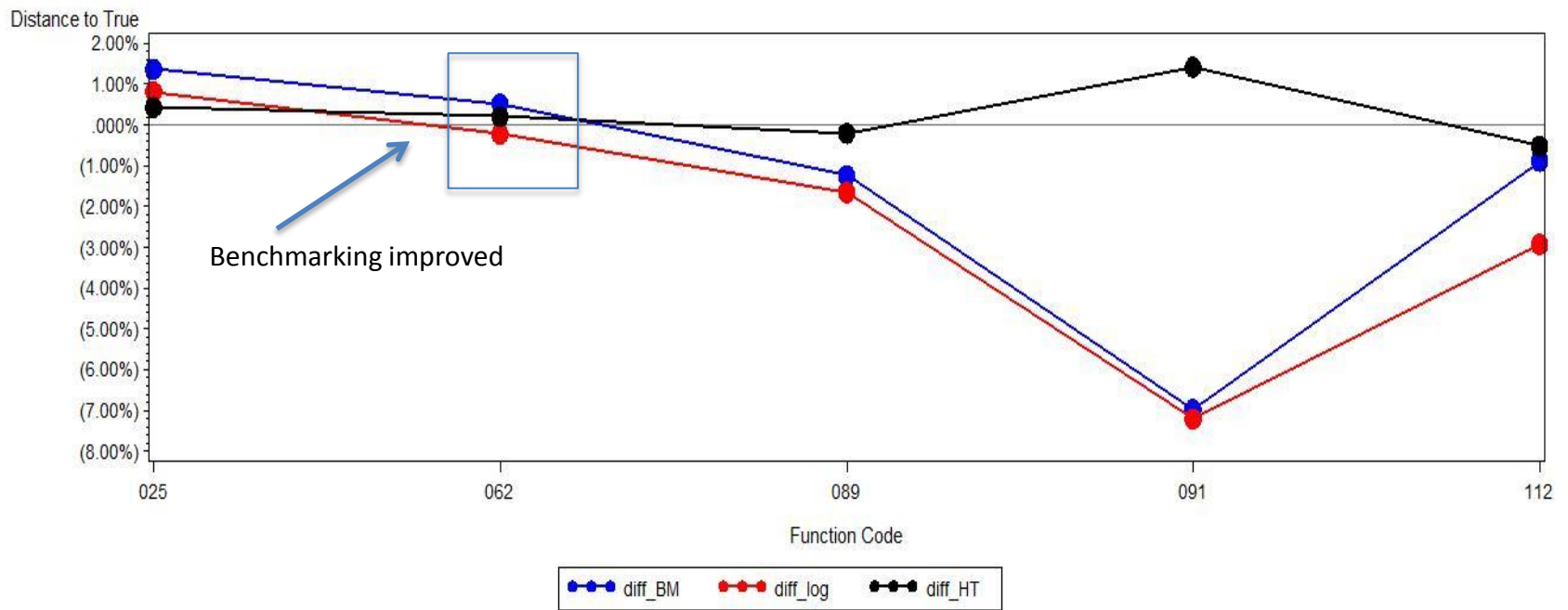
Raking: Log-transformed to HT Base (CA)



Evaluation (Cont'd)

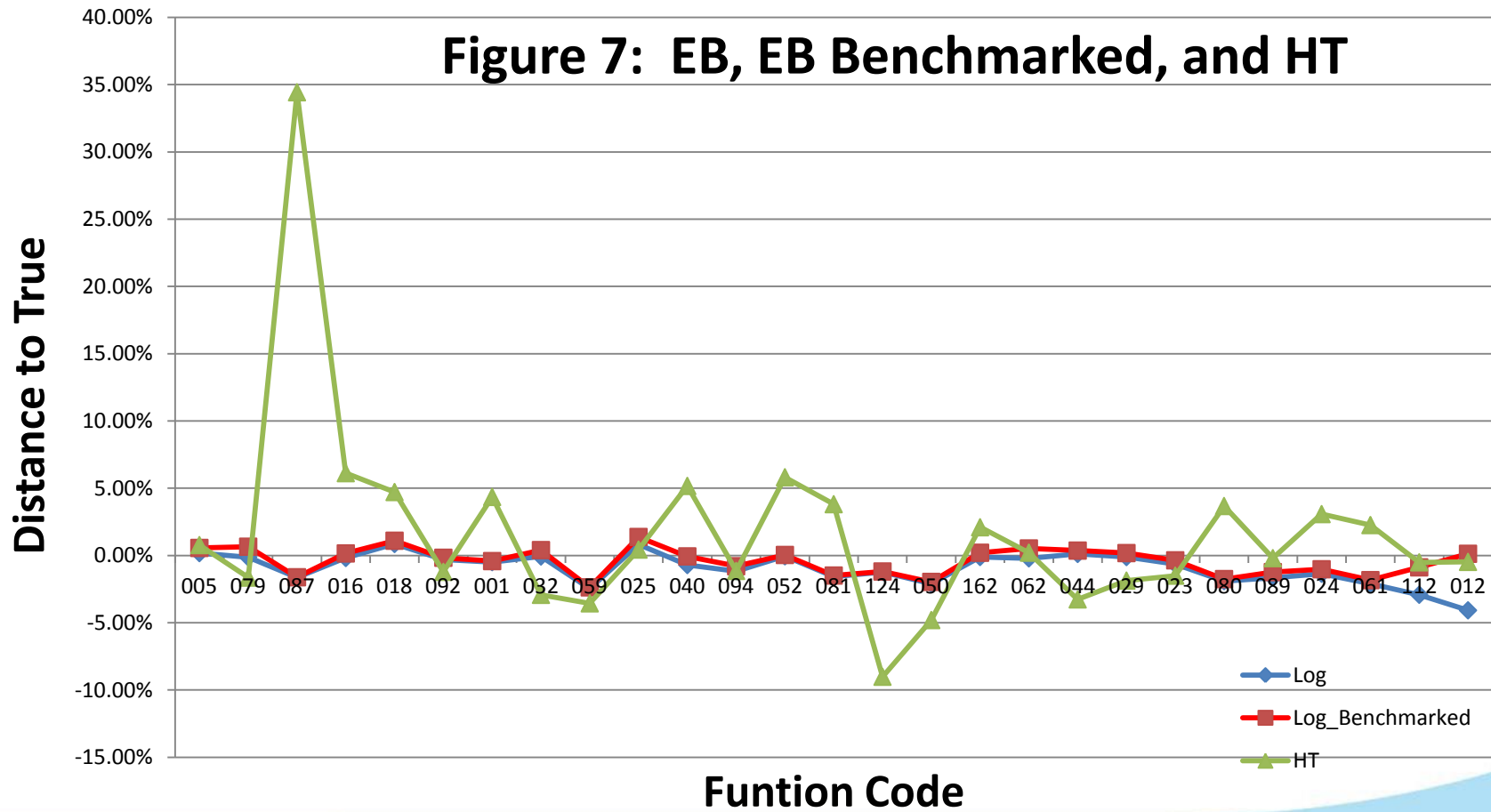
Effect of Raking

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



Evaluation (Cont'd)

Comparison: EB, Raking EB and HT

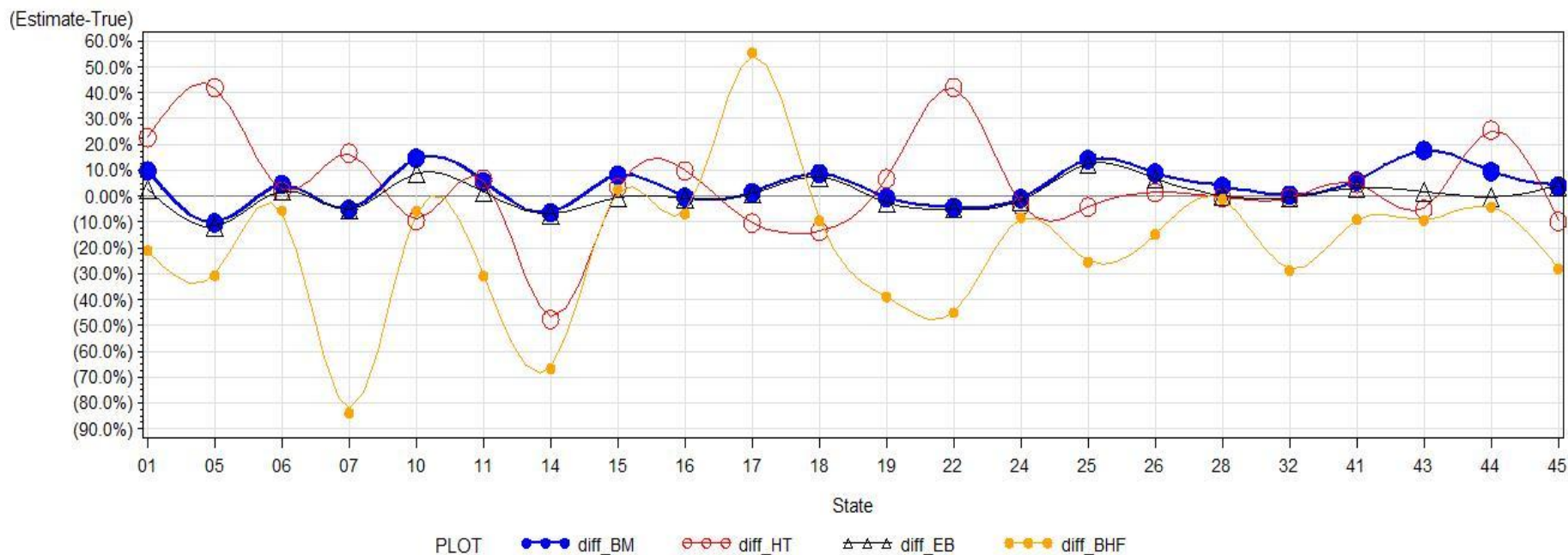


Evaluation (Cont'd)

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)

Overall - Absolute Relative Errors			
$\Sigma (HT-True)/True $	$\Sigma (EB-True)/True $	$\Sigma (EB_benchmarked-True)/True $	$\Sigma (BHF-True)/True $
5.26%	1.67%	1.44%	14.35%
Overall - Relative Errors			
$\Sigma (HT-True)/True$	$\Sigma (EB-True)/True$	$\Sigma (EB_benchmarked-True)/True$	$\Sigma (BHF-True)/True$
3.05%	-1.5%	-1%	-14.35%

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?