# Graphical Display of Uncertainty and Related Tools for Inference in Small Domain Estimation

**John L. Eltinge, Bureau of Labor Statistics**
**and Jaime Brugueras, Bureau of Labor Statistics and University of Illinois at Chicago**
**J. L. Eltinge, Office of Survey Methods Research, PSB 1950, BLS, 2 Massachusetts Avenue NE, Washington, DC 20212**
**Eltinge.John@bls.gov**

**Key Words:** Confidence sets; Current Employment Statistics (CES); False discovery rate; Small area estimation; Simultaneous inference; Triple-goal estimation.

**Abstract:** Small domain estimation methods use regression, hierarchical modeling and related tools to combine sample survey data and auxiliary information to produce estimators for a relatively large number of subpopulations. Some practical applications of these methods are complicated by two factors. First, anecdotal evidence indicates that in some cases stakeholders attempt to use published small domain estimates to carry out exploratory analyses and related (informal) simultaneous inference. Second, many stakeholders have relatively limited previous training in statistics. To address these issues, this poster presents some relatively simple graphical methods for display of small domain estimates and related measures of uncertainty. For cases involving univariate estimates, the triple-goal approach of Shen and Louis (1998) provides a useful framework for development of three types of graphical displays. For cases involving multivariate comparisons, projections of confidence sets and related simultaneous inference methods lead to some alternative graphical approaches.

Overview:



## I. Small Domain Estimation in the Current Employment Statistics (CES) Program

### A. Current Employment Statistics Program

1. Monthly establishment survey

2. Sample design: Stratification by state, industry, size class

3. Intended to control variances at national and state levels

4. Strong interest by states, other stakeholders in estimates at the Metropolitan Statistical Area (MSA) level

### B. Users of Small Domain Estimates

1. Varying priorities for specific estimands:

- Level (sub-population totals)
- One-month change (relative or absolute)
- Twelve-month change
- "Trends" (criteria often not defined a priori)

2. Some stakeholders appear to use data in a highly multivariate, exploratory form

3. Users appear to have asymmetric utility functions (overestimation worse than underestimation)

4. Important special feature: Approximately 5-8 months after a reference month t , we obtain from administrative records the nominal true employment level (pop total)

$x_{iat}$

for industry i, geographical area a, and month t

C. Small Domain Estimator

$\hat{x}_{iat}$ = variance-weighted ave of:

$\hat{y}_{1iat}$ = Direct sample-based estimator
(unbiased but inefficient)

$\hat{y}_{2iat}$ = Predicted value from a time series model used with historical data x(i, a, t)

$\hat{y}_{4iat}$ = Based on CES state-level ests of industry growth rates

II. Quantile-Quantile Plots to Identify Lack of Fit

A. Two Classes of Standardized Residuals

1. Contemporaneous: Can produce after data collection for month t

$$t_{C,iat} = \frac{\hat{x}_{iat} - \hat{y}_{1iat}}{se_{C,iat}}$$

where

$$se_{C,iat} = \{\hat{V}(\hat{x}_{iat} -$$

2. Retrospective: Can produce when we receive administrative records for month t

$$t_{R,iat} = \frac{\hat{x}_{iat} - x_{iat}}{se_{x,iat}}$$

where

$$se_{x,iat} = \{\hat{V}(\hat{x}_{iat} - x_{iat})\}^{1/2}$$

B. Individual Quantile-Quantile Plots:

1. Compare order statistics of

$$t_{C,iat}$$

$$t_{C,iat}$$

with quantiles of t dist
on 6 degrees of freedom

2. Look for (cf. Daniel, 1959):

- Extremes in tails
- Shape of central part of plot
- Slope = 1? (else var issues)

3. Full set of estimates

4. Trimmed set of estimates (finer structure after outliers)

C. Quantile-Quantile Plot Matrix

1. Analogous to scatterplot matrix, related displays

2. Complement individual Q-Q plots

3. Look for consistent or distinctive patterns across industries, areas

4. Contemporaneous and retrospective plots

III. Graphical Display of Statistical Significance and Practical Significance of Lack of Fit

A. Three Indicators of Lack of Fit

$$sigx = \frac{\hat{x}_{iat} - x_{iat}}{x_{iat}}$$

$$sigy = \frac{\hat{x}_{iat} - \hat{y}_{iat}}{s_{C,iat}}$$

$$sigz = \frac{\hat{x}_{iat} - x_{iat}}{s_{R,iat}}$$

B. Explore:

1. Identify outliers in three dimensions

Information conveyed beyond single-variable or bivariate plots?

2. Results after removal of outlying domains

IV. Additional Approaches

A. Triple-Goal Approach (Shen and Louis, 1998)

Possible Goals in Hierarchical Modeling (and Small Domain):

1. Inference for individual small area quantities ("coordinate-specific estimates"): Traditional emphasis in small area estimation

2. Ranks among small domains

3. Distribution function, histograms, quantiles of true $\chi_{iat}$

B. Evaluate and Control False Discovery Rate (FDR) (Benjamini and Hochberg, 1995; Black, 2004)

1. Formal hypothesis testing:

FDR = Expected value of ratio

(Number of incorrectly rejected null hypotheses)/ (Total number of rejected null hypotheses)

Cf. Apparent exploratory use of small domain use by many stakeholders

2. Graphical display of uncertainty measures, FDR to address this?

**Acknowledgements:**

The authors thank Julie Gershunskaya for providing the data used in these analyses; and Pat Getz, Larry Huff, Kirk Muller, Ken Shipp, and the members of the Current Employment Statistics Policy Council for many insights into the small domain estimation needs of CES data users.

**References:**

Benamini, Y. and Y. Hochberg (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society, Series B* **57**, 289-300.

Black, M.A. (2004). A Note on the Adaptive Control of False Discovery Rates. *Journal of the Royal Statistical Society, Series B* **66**, 297-304.

Daniel, C. (1959). Use of Half-Normal Plots in Interpreting Factorial Two Level Experiments
*Technometrics*, 1, 311-341.

Rao, J.N.K. (2003). *Small Area Estimation.* New York: Wiley.

Shen, W. and T.A. Louis (1998). Triple-Goal Estimation in Two-Stage Hierarchical Models. *Journal of the Royal Statistical Society, Series B* **60**, 455-471.