# Discussion by Julie Gershunskaya

# on

# *Small Area Estimation: Its Evolution in Five Decades*

# by Malay Ghosh

**Hansen Lecture**

**October 2019**

# Prospective of a survey practitioner

***<u>Disclaimer</u>*: *Any opinions expressed in this presentation are those of the presenter and do not constitute policy of the Bureau of Labor Statistics***

# The beginning of large-scale survey sampling

Morris Hansen (1987) *"Some History and Reminiscences on Survey Sampling"*

- 1937 Neyman's lectures
- 1937 Enumerative Check Census
- **Total survey design:**
  complex of "*theory and operations*" that includes empirical studies, experimentation, and planning of surveys, following by data collection, and estimation

# Statistical engineering

Around that same time in India, Mahalanobis developed philosophy of "statistical engineering", which is similar in spirit to "total survey design" of the Bureau of Census.

In 1946, Mahalanobis presented his work on large-scale surveys in Bengal to the Royal Statistical Society. He notes a distinction between a "small-scale working in a laboratory" and implementing technologies in large-scale manufacturing. This distinction is reflected in our "…calling large-scale production a matter of chemical engineering rather than of pure chemistry. **In the same way large-scale sample surveys may be appropriately called *statistical engineering*.**"

# Small Area Estimation as "engineering framework" for present day official statistics

Large-scale surveys are designed for production of reliable sample-based estimates of parameters at target population levels.

But there is an ever increasing demand to produce estimates in "unplanned domains". The demand is such that production for small areas now becomes a large-scale enterprise.

What's the plan for "unplanned domains"?

- SAE is a fusion of theory & the art of modeling, plus an operational plan that accounts for specifics of a survey.

Most SAE theory & operations focus on the *estimation stage*, that is, given survey has been already designed and data collected

- What about some planning at the design stage?

A rare exception: Dorfman (2018) "Towards a REEP for SAE": oversampling in a sample of "planned" small domains

# Example: Current Employment Statistics (CES) survey

- Tight production timeline (estimates published 3-4 weeks after data collection)
- Large number of small areas (~ 6000 SA monthly, cross-classified by detailed industry and geography)
- Dynamic and heterogeneous nature of the population of businesses:
  - Influential observations ("representative outliers") often appear in the sample and affect estimates
  - Employment in groups of domains may grow (decline) faster (care when "borrowing strength" across sets of domains)
- Robustness to outliers is an essential requirement for CES small area models

# Model selection and evaluation in CES

Availability of 'gold standard' – data from Quarterly Census of Employment and Wages (QCEW) administrative source – defines the CES SAE strategy:

- research and planning stage (pre-production):

  test candidate models $M_1, \ldots, M_L$ against QCEW (based on MAD, say) on a number of historical employment series over several years. Select best model $M_l$ for use in production

- post-production stage:

  evaluation based on "external data": test against QCEW that become available on a lagged basis

# Building on Fay-Herriot model

$M_0$: classical FH *area-level* model

$$\hat{y}_i \mid \theta_i \overset{ind}{\sim} N(\theta_i, D_i) \qquad \textit{sampling model}$$

$$\theta_i \overset{ind}{\sim} N(\mathbf{x}_i^T \mathbf{b}, A) \qquad \textit{linking model}$$

Model $M_0$ is not robust in two respects:

1) the *sampling model* is not robust to noisy outlying direct estimates ("fixed and known" plugged-in $D_i$ may not reflect outliers in sample data)

2) the *linking model* normality assumption may fail (e.g., groups of domains may form clusters)

Thus, we expand $M_0$ to more robust models $M_1$ and $M_2$

# More flexible model alternatives to $M_0$

**1)** $M_1$: replace "fixed and known" assumption on $D_i$ with joint modeling of $\hat{y}_i$ and $\hat{D}_i$ :

$$\hat{y}_i \mid \theta_i \overset{ind}{\sim} N(\theta_i, D_i); \qquad \theta_i \overset{ind}{\sim} N(\mathbf{x}_i^T \mathbf{b}, A)$$

$$\hat{D}_i \mid D_i \overset{ind}{\sim} G\left(\frac{an_i}{2}, \frac{an_i}{2} D_i^{-1}\right); \qquad D_i \overset{ind}{\sim} IG\left(2, \gamma_0 e^{z_i^T \gamma}\right)$$

Thus we add protection from outlying noisy direct estimates.
(Maiti et al. 2014, Sugasawa et al. 2017; Gershunskaya & Savitsky 2018)

---

**2)** $M_2$: relax the normality assumption of random effects:

use finite mixture of normals (Gershunskaya and Savitsky 2018)

$$\hat{y}_i \mid \theta_i \overset{ind}{\sim} N(\theta_i, D_i); \qquad \theta_i \overset{ind}{\sim} \sum_{k=0}^{K} \pi_k N\left(\mu_k + \mathbf{x}_i^T \mathbf{b}, A\right)$$

$$\hat{D}_i \mid D_i \overset{ind}{\sim} G\left(\frac{an_i}{2}, \frac{an_i}{2} D_i^{-1}\right); \qquad D_i \overset{ind}{\sim} \sum_{k=0}^{K} \pi_k IG\left(2, \gamma_{0k} e^{z_i^T \gamma}\right)$$

(similar in spirit to global-local shrinkage priors of Tang et al. 2018 , also Student-t distribution used by Ghosh et al. 2018)

# Ongoing diagnostics

Thus, we tested models on 10 years of historical data and found that, overall, model $M_2$ performed the best.

The question remains: what if although $M_2$ works in general, in some months during the production, the model will not fit for some domains?

Need to equip analysts with:

- *tests* and *graphical tools* to diagnose possible problems
- *plan, instructions* on how to proceed if problems arise:

  e.g., significant unexpected event (hurricanes, etc.), other changes that become known to subject matter specialists?

# Example of routine domain screening

Suppose model $\mathrm{M}_2$ is selected for production:

- Fit $\mathrm{M}_2$ and compute *posterior predictive distribution* $p(\tilde{y}_i \mid \hat{y}_i)$

- Flag those direct domain estimates $\hat{y}_i$ that have low probability of following $p(\tilde{y}_i \mid \hat{y}_i)$.

- Send the list of “suspect” domains to analysts for checking.

Analysts review the list and determine if extreme $\hat{y}_i$ is due to:

(1) Deficiency of the sample or

(2) Failure of modeling assumptions (analysts may have additional out-of-sample information to support their judgment)

# Summary

- Demand for estimates in "unplanned" domains instigated development of the SAE field

- Theoretical advances in last 5 decades in SAE (and revolution in computers and software) $\Rightarrow$ SAE becomes a *large-scale* adventure

- Need more tools (and philosophy for quality control in production environment) for "statistical engineers": to help ensure objective, reliable and impartial estimates in small domains.

# Contact Information

**Julie Gershunskaya**
**Mathematical Statistician**
**SMS/OEUS**

202-691-6360
**Gershunskaya.Julie@bls.gov**