Chapter 9: Estimation 2

Jae-Kwang Kim

Iowa State University

Fall, 2014
1. GREG estimator

2. Optimal Estimation

3. Summary
• Generalized regression (GREG) model

\[ E_\zeta (y_i) = x_i' \beta \]

\[ \text{Cov}_\zeta (y_i, y_j) = \begin{cases} 
  c_i \sigma^2 & \text{if } i = j \\
  0 & \text{if } i \neq j 
\end{cases} \]

where \( c_i = c(x_i) \) and \( c(x) \) is a known function of \( x \).
Examples: GREG model

1. Ratio model

\[ E_\zeta(Y_i) = x_i \beta \]
\[ V_\zeta(Y_i) = x_i \sigma^2 \]

2. Regression model

\[ E_\zeta(Y_i) = \beta_0 + x_i \beta_1 \]
\[ V_\zeta(Y_i) = \sigma^2 \]

3. Group mean model (or ANOVA model)

\[ E_\zeta(Y_i) = \mu_g \]
\[ V_\zeta(Y_i) = \sigma^2_g \]

for \( i \in U_g \) and \( U = U_1 \cup U_2 \cup \cdots \cup U_G \)
GREG estimator

- Definition

\[ \hat{Y}_{\text{GREG}} = \sum_{i=1}^{N} \hat{y}_i + \sum_{i \in A} \frac{1}{\pi_i} (y_i - \hat{y}_i) \]

where \( \hat{y}_i = x'_i \hat{\beta} \) with

\[ \hat{\beta} = \left( \sum_{i \in A} \frac{1}{\pi_i c_i} x_i x'_i \right)^{-1} \sum_{i \in A} \frac{1}{\pi_i c_i} x_i y_i. \]

- Alternative representation

\[ \hat{Y}_{\text{GREG}} = \hat{Y}_{\text{HT}} + (X - \hat{X}_{\text{HT}})' \hat{\beta} \]
Examples: GREG estimators

1. Ratio model

\[ \hat{Y}_{GREG} = \hat{Y}_{ratio} = \left( \sum_{i=1}^{N} x_i \right) \frac{\sum_{i \in A} \frac{1}{\pi_i} y_i}{\sum_{i \in A} \frac{1}{\pi_i} x_i} \]

2. Regression model

\[ \hat{Y}_{GREG} = \hat{Y}_{reg} = \sum_{i \in A} \frac{1}{\pi_i} y_i + \left( \sum_{i=1}^{N} x_i - \sum_{i \in A} \frac{1}{\pi_i} x_i \right) \hat{\beta} \]

where

\[ \hat{\beta} = \frac{\sum_{i \in A} \frac{1}{\pi_i} (x_i - \bar{x}_\pi) (y_i - \bar{y}_\pi)}{\sum_{i \in A} \frac{1}{\pi_i} (x_i - \bar{x}_\pi)^2} \]

3. Group mean model (or ANOVA model)

\[ \hat{Y}_{GREG} = \sum_{g=1}^{G} N_g \frac{\hat{Y}_g}{\hat{N}_g} \]
Algebraic Properties

- Linear in $y$

$$\hat{Y}_{\text{GREG}} = \sum_{i \in A} \frac{1}{\pi_i} g_i(A) y_i$$

where

$$g_i(A) = 1 + (X - \hat{X}_{HT})' \left( \sum_{k \in A} \frac{1}{\pi_k c_k} x_kx_k' \right)^{-1} \frac{x_i}{c_i}. $$

Thus, “Final weight = Design weight * g-factor”.
Algebraic Properties

- Calibration property

$$\sum_{i \in A} \pi_i g_i (A) x_i = \sum_{i=1}^{N} x_i.$$  

In fact, the final weights are chosen to minimize

$$\sum_{i \in A} c_i \pi_i \left( w_i - \frac{1}{\pi_i} \right)^2$$ subject to $$\sum_{i \in A} w_i x_i = \sum_{i=1}^{N} x_i.$$  

Kim (ISU) Ch. 9: Estimation 2 Fall, 2014
Algebraic Properties (Cont’d)

- If \( c_i = \lambda' x_i \), then \( \sum_{i \in A} \frac{1}{\pi_i} \hat{y}_i = \sum_{i \in A} \frac{1}{\pi_i} y_i \). Thus,

\[
\hat{Y}_{GREG} = \sum_{i=1}^{N} x'_i \hat{\beta}
\]
Statistical Properties

1. Approximately design unbiased
2. Asymptotic Variance

\[
\text{Var} \left( \hat{Y}_{\text{GREG}} \right) = \text{Var} \left\{ \sum_{i \in A} \pi_i^{-1} (y_i - \mathbf{x}_i' \mathbf{B}) \right\}
\]

where

\[
\mathbf{B} = \left( \sum_{i=1}^{N} \frac{1}{c_i} \mathbf{x}_i \mathbf{x}_i' \right)^{-1} \sum_{i=1}^{N} \frac{1}{c_i} \mathbf{x}_i y_i.
\]

3. Variance estimation: Use \( \hat{E}_i = y_i - \mathbf{x}_i' \hat{\beta} \) instead of \( E_i = y_i - \mathbf{x}_i' \mathbf{B} \) in the HT variance estimator.
Example 1

Group Ratio Model

\[ E_\zeta (Y_i) = \beta_g x_i \]
\[ V_\zeta (Y_i) = \sigma^2_g x_i \]

for \( i \in U_g \) and \( U = U_1 \cup U_2 \cup \cdots \cup U_G \). The \( x_i \) are observed throughout the population. Note that if \( x_i = 1 \) then it reduces to the group mean model.
Example 1 (Cont’d)

- Let $x'_i = (x_{1i}, x_{2i}, \ldots, x_{Gi})$ where

$$x_{gi} = \begin{cases} x_i & \text{if } i \in U_g \\ 0 & \text{otherwise} \end{cases}$$

and $\beta = (\beta_1, \beta_2, \ldots, \beta_G)'$. Then, $E_\zeta (Y_i) = x'_i \beta$ and

$$B = \left( \sum_{i \in U} \frac{x_i x'_i}{\sigma_i^2} \right)^{-1} \sum_{i \in U} \frac{x_i y_i}{\sigma_i^2} = \left( \frac{\sum_{i \in U_1} \frac{y_i}{x_i}}{\sum_{i \in U_1} x_i}, \ldots, \frac{\sum_{i \in U_G} \frac{y_i}{x_i}}{\sum_{i \in U_G} x_i} \right)'$$

and

$$\hat{B} = \left( \sum_{i \in A} \frac{x_i x'_i}{\pi_i \sigma_i^2} \right)^{-1} \sum_{i \in A} \frac{x_i y_i}{\pi_i \sigma_i^2} = \left( \frac{\sum_{i \in A_1} \frac{y_i}{\pi_i}}{\sum_{i \in A_1} \frac{x_i}{\pi_i}}, \ldots, \frac{\sum_{i \in A_G} \frac{y_i}{\pi_i}}{\sum_{i \in A_G} \frac{x_i}{\pi_i}} \right)'$$
Example 1 (Cont’d)

- Since $V_\zeta (Y_i) = \lambda' x_i$ for some $\lambda$, we have

$$\hat{Y}_{GREG} = \sum_{i \in U} x'_i \hat{B} = \sum_{g=1}^{G} \left( \sum_{i \in U_g} x_i \right) \frac{\sum_{i \in A_g} y_i / \pi_i}{\sum_{i \in A_g} x_i / \pi_i}.$$ 

This is called separate ratio estimator. If there are differences among but homogeneous within groups in terms of ratio, the separate ratio estimator might be good.

- If $x_i = 1$, two possibilities
  1. Groups = strata: stratification
  2. Groups \neq strata: poststratification
Poststratification

- The finite population decomposed into $G$ groups. The group information is not available at the time of selecting samples. The population size of group $g$, denoted by $N_g$, is known.

- Poststratification estimator

\[
\hat{Y}_{post} = \sum_{g=1}^{G} N_g \frac{\hat{Y}_g}{\hat{N}_g}
\]

- Under SRS, it is

\[
\hat{Y}_{post} = \sum_{g=1}^{G} N_g \left( \frac{\sum_{i \in A_g} y_i}{n_g} \right)
\]

where $n_g$ is the number of sampled elements in group $g$. 
Asymptotic variance (under SRS)

\[ V \left( \hat{Y}_{post} \right) = \sum_{i \in U} \sum_{j \in U} \Delta_{ij} \frac{E_i}{\pi_i} \frac{E_j}{\pi_j} \]

\[ = \frac{N}{n} \left( 1 - \frac{n}{N} \right) \sum_{g=1}^{G} \sum_{i \in U_g} \left( y_i - \bar{Y}_g \right)^2. \]

Thus, it is essentially equal to the variance under stratified sampling with proportional allocation.
Example 2: Two-way ANOVA (additive, no interaction)

- Model

\[ E_\zeta(Y_k) = \alpha_i + \beta_j \]
\[ V_\zeta(Y_k) = \sigma^2 \]

- Setup: Have \( I \times J \) groups or cells. Cell counts \( N_{ij} \) are not known. Marginal counts \( N_i = \sum_{j=1}^{J} N_{ij} \) and \( N_j = \sum_{i=1}^{I} N_{ij} \) are known.

- Example: \( i \) gender, \( j \) age group. (\( I=2, J=3 \))

- Auxiliary variable: Let

\[ \delta_{ijk} = \begin{cases} 
1 & \text{if } k \in U_{ij} \\
0 & \text{otherwise} 
\end{cases} \]

Unfortunately, we do not observe \( \delta_{ijk} \) in the population. Let

\[ x_k = (\delta_{1,k}, \delta_{2,k}, \cdots, \delta_{I,k}, \delta_{1k}, \delta_{2k}, \cdots, \delta_{Jk}) \]

and we know \( \sum_{k=1}^{N} x_k \).
Example 2 (Cont’d)

- If the joint population counts $N_{ij} = \sum_{k=1}^{N} \delta_{ink}$ were known, then post stratified estimator would be used.

- GREG estimator

$$\hat{Y}_{GREG} = \sum_{i \in A} \frac{1}{\pi_i} g_i (A) y_i$$

where

$$g_i (A) = 1 + \left( \sum_{k=1}^{N} x_k - \sum_{k \in A} \frac{1}{\pi_k} x_k \right)' \left( \sum_{k \in A} \frac{1}{\pi_k} x_k x_k' \right)^{-1} \frac{x_i}{\sigma_i^2}.$$ 

Unfortunately, we cannot compute the inverse of $\sum_{k \in A} \frac{1}{\pi_k} x_k x_k'$. 
Alternative method: (Raking ratio estimation)

We want to find \( g_{kA} = g_k(A) \) such that

\[
\sum_{k \in A} \frac{g_{kA}}{\pi_k} \delta_{i,k} = \sum_{k=1}^{N} \delta_{i,k}, \quad i = 1, 2, \ldots, I
\]  

(1)

\[
\sum_{k \in A} \frac{g_{kA}}{\pi_k} \delta_{.jk} = \sum_{k=1}^{N} \delta_{.jk}, \quad j = 1, 2, \ldots, J.
\]  

(2)
Raking ratio estimation

- Do this iteratively (Iterative proportional fitting)

1. Start with $g_{kA}^{(0)} = 1$.
2. For $\delta_{i\cdot k} = 1$,
   \[
   g_{kA}^{(t+1)} = g_{kA}^{(t)} \frac{\sum_{k=1}^{N} \delta_{i\cdot k}}{\sum_{k \in A} g_{kA}^{(t)} \delta_{i\cdot k} / \pi_k}.
   \]
   It satisfies (1), but not necessarily satisfy (2).
3. For $\delta_{.jk} = 1$,
   \[
   g_{kA}^{(t+2)} = g_{kA}^{(t+1)} \frac{\sum_{k=1}^{N} \delta_{.jk}}{\sum_{k \in A} g_{kA}^{(t+1)} \delta_{.jk} / \pi_k}.
   \]
   It satisfies (2), but not necessarily satisfy (1).
4. Set $t \leftarrow t + 2$ and go to Step 2. Continue until convergence.
Example: US Census of housing and population

- Two forms
  - Short Form: 100% sample (obtain basic demographic information)
  - Long form: about 16% sample (obtain other social and economic information as well as demographic information)

- Raking ratio estimation using demographic variable to match known population counts from short form (Deming and Stephan, 1940)
GREG estimator

Optimal Estimation

Summary
Optimal estimation

- Optimal design & estimation: Find a pair of design & estimator \( \{p(\cdot), \hat{\theta}\} \) that minimizes the variance, or MSE, for a given cost (cost \( \equiv \) sample size) among a suitable class of estimators.

- Class of Linear and design-unbiased estimator: Unique solution (HT estimator)

- Non-existence of the UMVUE

Let any noncensus design with \( \pi_k > 0 \) \( (k = 1, 2, \cdots, N) \) be given. Then no uniformly minimum variance estimator exists in the class of all unbiased estimators of \( Y = \sum_{i=1}^{N} y_i \).
Remedy: Change the optimality criterion

- We don't know the variance before sampling
- Use assumption about $Y$: superpopulation model
- Anticipated variance: $AV(\hat{\theta}) = E_\zeta E_p(\hat{\theta} - \theta_N)^2$
- Thus, find a pair of design & estimator $\{p(\cdot), \hat{\theta}\}$ that minimizes the $AV(\hat{\theta})$ for a given cost.
Lemma

If $\hat{\theta}$ is design-unbiased for $\theta_N$, then

$$AV(\hat{\theta}) = E_p V_\zeta(\hat{\theta}) + V_p E_\zeta(\hat{\theta}) - V_\zeta(\theta_N)$$
Result 1

Theorem 1 (Godambe and Joshi, 1965)

Consider a model $\zeta$ with $y_i$’s independent with $V_\zeta(y_i) = \sigma_i^2$. If $p(\cdot)$ is a probability sampling design ($\pi_i > 0, i \in U$) and $\hat{Y}$ is any design unbiased estimator for the total of $y$, then

$$AV(\hat{Y}) \geq \sum_{i \in U} \left( \frac{1}{\pi_i} - 1 \right) \sigma_i^2.$$ 

The right side is called the Godambe-Joshi Lower Bound (GJLB).

For the fixed-size probability sampling design, the GJLB is further minimized if and only if

$$\pi_i \propto \sigma_i.$$
Theorem 2 (Isaki and Fuller, 1982)

Suppose that $p(\cdot)$ is a fixed size probability sampling design and $\zeta$ is a superpopulation model with $y_i$’s independent and $E_{\zeta}(y_i) = x_i'\beta$ and $V_{\zeta}(y_i) = c_i\sigma^2$. Then, the GREG estimator asymptotically attains the GJLB if $c_i = \lambda'x_i$ for some $\lambda$. 
1. GREG estimator

2. Optimal Estimation

3. Summary
GREG estimator can be constructed from a superpopulation model using $x_i$ which is available throughout the population.

- GREG estimator satisfies the calibration constraint.
- GREG estimator is the best in the sense of minimizing the anticipated variance (under the model used for GREG estimation) among the class of all (nearly) design unbiased estimators.