SAE course Spring 2015
Risto Lehtonen, University of Helsinki

TOPIC 5 Model-based SAE: SYN and EBLUP

Model-based estimators of domain totals

$$t_d = \sum_{k \in U_d} y_k \,, \; d = 1,...,D \qquad (1)$$

Synthetic estimators
EBLUP estimators
Example: SAS macro EBLUPGREG

We assume access to **unit-level population data**

$\mathbf{x}_k = (1, \mathrm{x}_{1k}, x_{2k}, ..., x_{jk}, ..., x_{Jk})'$ for every unit $k \in U$

"Traditional" synthetic estimator SYN of (1)

$$\hat{t}_{dSYN} = \sum_{k \in U_d} \hat{y}_k \,, \;\; d = 1,...,D \qquad (2)$$

with predicted (fitted) values

$$\hat{y}_k = \mathbf{x}_k' \hat{\boldsymbol{\beta}}, \;\; k \in U$$

Model specification: **Fixed-effects linear model**

$$E_m(Y_k) = \mathbf{x}_k' \boldsymbol{\beta} = \beta_0 + \beta_1 x_{1k} + \beta_2 x_{2k} + ... + \beta_J x_{Jk} \qquad (3)$$

Estimation of model parameters of (3) by OLS

The SYN estimator (2) is an **indirect model-based estimator**
It borrows strength from other domains for the estimation for the given domain *d*

SYN estimators $\hat{t}_{dSYN}$ of domain totals t_d are **design biased**

Bias$(\hat{t}_{dSYN}) = E(\hat{t}_{dSYN}) - t_d$ in domain *d* can be large if the model does not fit well in that domain

Bias can be small if the model fits well in the domain *d*

Alternative fixed-effects model specification

For example a fixed-effects model with fixed intercepts for domains (Note: here $\beta_0 = 0$)

$$\begin{aligned} E_m(Y_k) &= \mathbf{x}_k'\boldsymbol{\beta} \\ &= \beta_{01} I_{1k} + \ldots + \beta_{0D} I_{Dk} + \beta_1 x_{1k} + \beta_2 x_{2k} + \ldots + \beta_J x_{Jk} \end{aligned} \quad (4)$$

where

$\mathbf{x}_k = (I_{1k}, \ldots, I_{Dk}, x_{1k}, \ldots, x_{Jk})'$, $\boldsymbol{\beta} = (\beta_{01}, \ldots, \beta_{0D}, \beta_1, \ldots, \beta_J)'$

Domain indicators: $I_{dk} = 1$ if $k \in U_d$

$I_{dk} = 0$ if $k \notin U_d$

Variance estimation for SYN (2)

$$\hat{v}\left(\hat{t}_{dSYN}\right) = \sum_{k \in U_d} \mathbf{x}_k' \mathrm{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{x}_k , \quad d = 1, 2, \ldots, D \qquad (5)$$

For vector $\hat{\mathbf{t}}_{SYN}$ of estimated domain totals we have:

$$\hat{v}\left(\hat{\mathbf{t}}_{SYN}\right) = \mathbf{t}_x \mathrm{Cov}(\hat{\boldsymbol{\beta}}) \mathbf{t}_x'$$

where

$\mathrm{Cov}(\hat{\boldsymbol{\beta}})$ denotes covariance matrix estimator of $\hat{\boldsymbol{\beta}}$

$\mathbf{t}_x$ denotes vector of auxiliary variable totals in population

Standard errors: $\mathrm{s.e}\left(\hat{t}_{dSYN}\right) = \sqrt{\hat{v}\left(\hat{t}_{dSYN}\right)}$

NOTE: Use of SYN estimators is not necessarily recommended

Reasons

- In traditional SYN (2), the model is specified as a common model (no accounting for possible domain differences)
- SYN estimator for given domain *d* is very sensitive to the correctness of the model in the domain
- Recall: ”*All models are wrong (but some are useful*”)
- Variance estimates tend to be too small

SYN computation in SAS macro EBLUPGREG

Option: macro parameter SYN=1 (Synthetic estimator)

In EBLUPGREG, a **modified SYN estimator**

$$\hat{t}_{dSYN} = \sum\nolimits_{k \in U_d} \hat{y}_k$$

is based on the fixed part of **mixed model**

$$E_m(Y_k \mid u_{0d}) = (\beta_0 + u_{0d}) + \beta_1 x_{1k} + ... + \beta_J x_{Jk} \qquad (6)$$

where the model is fitted without design weights

Predictions $\hat{y}_k$ are calculated by model (6) but without the contribution of the random intercept estimates $\hat{u}_{0d}$:

$$\hat{y}_k = \hat{\beta}_0 + \hat{\beta}_1 x_{1k} + ... + \hat{\beta}_J x_{Jk} \qquad (7)$$

NOTE: Components $\hat{u}_{0d}$ enter on the scene in EBLUP estimators

Empirical Best Linear Unbiased Predictor EBLUP

"**Traditional" EBLUP estimator** of domain total t_d:

$$\hat{t}_{dEBLUP} = \sum_{k \in s_d} y_k + \sum_{k \in U_d;\, k \notin s_d} \hat{y}_k , \quad d = 1, 2, ..., D \qquad (8)$$

is based on linear mixed model with domain-level random effects (intercepts and slopes)

$$E_m(y_k \mid \mathbf{u}_d) = \mathbf{x}'_k(\boldsymbol{\beta} + \mathbf{u}_d) \qquad (9)$$

with predictions

$$\hat{y}_k = \mathbf{x}'_k(\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_d) \qquad (10)$$

calculated for non-sampled elements $k \in U_d; k \notin s_d$

EXAMPLE: Linear mixed model with random intercepts

$$\begin{aligned} y_k &= \mathbf{x}'_k \boldsymbol{\beta} + u_{0d} \\ &= (\beta_0 + u_{0d}) + \beta_1 x_{1k} + ... + \beta_J x_{Jk} + \varepsilon_k \end{aligned} \qquad (11)$$

where $u_{0d} \sim N(0, \sigma_u^2)$, $\varepsilon_k \sim N(0, \sigma^2)$, u_{0d} and ε_k are independent

We estimate $\hat{\boldsymbol{\beta}}$ and $\hat{u}_{0d}$, $d = 1, ..., D$, and calculate fitted values

$$\hat{y}_k = \mathbf{x}'_k \hat{\boldsymbol{\beta}} + \hat{u}_{0d}, \quad k \in U_d,\ d = 1, ..., D \qquad (12)$$

Alternative EBLUP estimator uses fitted values (12):

$$\hat{t}_{dEBLUP} = \sum_{k \in U_d} \hat{y}_k , \quad d = 1, 2, ..., D \qquad (13)$$

NOTE: (13) is the standard EBLUP estimator

NOTE: In domain *d*, EBLUP estimator (13) is close to SYN estimator (2) and EBLUP estimator (8) if domain sample size n_d is small and if the model underlying SYN fits well in domain *d*

Estimation of mixed model parameters

Macro EBLUPGREG: Manual

Combination of GLS (Generalized least squares) and REML (Restricted ML) or ML (Maximum likelihood) estimation

Model is of form (11)
EBLUP estimator is of form (13)
MSE is of form (14)
See EBLUPGREG Manual p. 18–20

MSE (Mean Squared Error) estimation of EBLUP (13)

Macro EBLUPGREG:

We calculate MCPE = *Mean Cross Product Error matrix* (Saei and Chambers, 2004, Chapter. 3.3)

$$\text{MCPE} = g_1 + g_2 + 2g_3 + g_4 \qquad (14)$$

Components in MCPE:

- g_1 (general estimate of variation),
- g_2 (uncertainty of estimating the beta coefficients),
- g_3 (uncertainty of estimating variance components)
- g_4 (uncertainty of estimating the model)

MSE estimates are obtained from the main diagonal of the estimated MCPE matrix

EBLUPGREG computes all four MCPE components

Output: The MSE estimate and the components g_1, g_2, g_3 ja g_4 and MSE estimate without component g_3

Protection against instability of g_3 in certain cases (in the case of grossly wrong model specification)

EXAMPLE
SAS MACRO EBLUPGREG
COMPARISON: GREG, SYN AND EBLUP

EBLUP: Formula (13)
MCPE: Kaava (14)
Malli: Lineaarinen sekamalli (kiinteä osa ja satunnaisosa)

Macro call: see manual for EBLUPGREG

```
%eblupgreg
    (sample=sample1,
    population=pop,
    y=y,
    xlist=x,
    regionIdentifier=domain,
    test=1,
    estimateMeans=0,
    weights=samplingweight,
    convergenceCrit=1e-8,
    maxiterations=200,
    initialSigma2=1,
    modules=modules.eurarea,
    parametersEstimatedBy='REML', eblup=1,
    greg=1,
    synthetic=1,
    stratified=0,
    output=out1
   );
```

OUTPUT

SAS Macro EBLUPGREG / **GREG and SYN estimation**

SAS Macro EBLUPGREG / GREG and SYN

domain	n	GREG	SYN synthetic	sqrt MSEsyn	stdGREG	true value
1	8	1292.93	1302.88	16.3148	24.8442	1299.27
2	13	2472.54	2560.54	24.7205	42.2263	2532.79
3	14	1816.11	1921.33	19.4337	32.3286	1839.14
4	5	1913.76	1880.74	18.1961	41.4054	1864.56
5	7	1747.49	1750.69	17.8414	53.4452	1737.94
6	19	4745.43	4595.54	45.7247	78.6909	4662.57
7	8	874.72	831.64	12.0799	31.9516	835.20
8	6	1026.49	1053.62	10.4176	31.4930	1022.06
9	6	939.52	922.79	9.5419	32.7146	884.18
10	14	3630.78	3632.68	35.6288	48.2893	3593.91

sqrtMSEsyn Root MSE of SYN estimator
stdGREG Standard error s.e of GREG estimator

NOTE:
GREG is design-based model assisted estimator
SYN is model-based estimator

SAS Macro EBLUPGREG / **EBLUP estimation**

SAS Macro EBLUPGREG / EBLUP

domain	n	EBLUP	sqrtMSE	sqrt MSENoG3	sqrtg1	sqrtg2	sqrtg3	sqrtg4	True value
1	8	1299.00	30.0123	24.9866	19.8890	9.3051	11.7560	11.9236	1299.27
2	13	2515.35	45.9894	36.9976	31.4816	11.3290	19.3158	15.7919	2532.79
3	14	1879.40	34.5820	28.1648	23.1123	8.5218	14.1891	13.6549	1839.14
4	5	1898.53	40.4164	34.0563	28.4264	12.7666	15.3892	13.7400	1864.56
5	7	1749.16	38.3467	31.6833	26.3662	11.1594	15.2748	13.5693	1737.94
6	19	4697.90	70.5633	56.5358	49.2219	18.5020	29.8570	20.7648	4662.57
7	8	843.73	19.6283	16.6735	12.3899	5.9942	7.3234	9.4110	835.20
8	6	1045.11	21.3437	18.1871	14.0229	6.2106	7.8988	9.7754	1022.06
9	6	927.46	18.4963	16.0095	11.6288	6.4678	6.5502	8.9019	884.18
10	14	3631.15	66.2352	52.6922	46.2247	16.3351	28.3781	19.3109	3593.91

sqrtMSE Root MSE of EBLUP estimator

MSENoG3 MSE excluding component g3

Components g1, g2, g3 ja g4 of MSE of EBLUP
sqrtg1 sqrtg2 sqrtg3 sqrtg4

See EBLUPGREG manual p. 7 plus examples

NOTE:
EBLUP is model-based estimator