HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Small Area Estimation

Spring 2015

Topic 4: GREG and calibration estimators
PART III: Extended family of GREG estimators

Risto Lehtonen, University of Helsinki

Topic 4 Part III

- **GREG and calibration estimators**
  PART III: Extended family of GREG estimators
  - Assisting models
  - GREG estimator assisted by linear mixed model
  - GREG estimator assisted by logistic mixed model
  - Additional model-assisted methods

  - See (available at course webpage):
    Supplement: Extended family of GREG estimators

Indirect estimators and models - 1

- **Model-assisted and model-based estimators**
  - In both approaches the model underlying an indirect estimator extends beyond the domain of interest (i.e. the model is NOT domain-specific as is the case in direct estimators)

Indirect estimators and models - 2

- **The role of model differs in model-assisted and model-based estimators**
  - GREG uses models as assisting tools
    This is to avoid design bias
    Cost to be paid is poor accuracy in small domains
  - SYN, EBLUP and EBP rely solely on models
    A benefit is better accuracy in small domains
    Cost to be paid is the risk of design bias

Models: Examples

- **Members of the generalized linear mixed models (GLMM) family**
  - Linear fixed-effects models
  - Linear mixed models
  - Logistic fixed-effects models
  - Logistic mixed models
- Large literature
  - E.g. Demidenko (2005). Mixed Models: Theory and Applications. Wiley.

Linear GREG and extensions

- Linear GREG – “Traditional” GREG estimator
  - GREG estimator assisted by a linear fixed-effects model
  - Särndal, Swensson and Wretman (1992)

- Members of extended GREG family
  - GREG estimators assisted by more complex models
  - Logistic fixed-effects model

    LGREG, Lehtonen and Veijanen (1998)
  - Linear mixed model

    MGREG, Lehtonen, Särndal and Veijanen (2003)
  - Logistic mixed model

    MLGREG, Lehtonen, Särndal and Veijanen (2005)

Assisting models: GLMM formulation

GLMM formulation with **domain - specific random terms**

$$E_m(y_k | \mathbf{u}_d) = f(\mathbf{x}'_k(\boldsymbol{\beta} + \mathbf{u}_d)), \text{ where}$$

$f(.)$ refers to the chosen functional form

$\mathbf{x}_k = (1, x_{1k}, ..., x_{Jk})'$ predictor variables (for all $k \in U$)

$\boldsymbol{\beta} = (\beta_0, \beta_1, ..., \beta_J)'$ fixed effects

$\mathbf{u}_d = (u_{0r}, ..., u_{Jd})'$ domain-spefic random effects

Predictions: $\hat{y}_k = f(\mathbf{x}'_k(\hat{\boldsymbol{\beta}} + \hat{\mathbf{u}}_d)), \ k \in U_d, \ d = 1, ..., D$

Models - 1

Linear mixed model for continuous study variable y

Domain-level random intercepts u_d

$$y_k = \mathbf{x}_k'\boldsymbol{\beta} + u_d + \varepsilon_k,$$

where $\mathbf{x}_k = (1, x_{1k}, \ldots, x_{pk})'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_p)'$

$u_d \sim N(0, \sigma_u^2)$, $\varepsilon_k \sim N(0, \sigma^2)$, u_d and ε_k independent

Estimate $\boldsymbol{\beta}$ and σ_u^2 from the data

Calculate estimates $\hat{u}_d$, $d = 1, \ldots, D$

Calculate fitted values

$$\hat{y}_k = \mathbf{x}_k'\hat{\boldsymbol{\beta}} + \hat{u}_d, \quad k \in U_d, \ d = 1, \ldots, D$$

Used in MGREG estimator

Models - 2

Logistic fixed - effects model

for binary response variable *y*

$$E_m(y_k) = \frac{\exp(\mathbf{x}_k'\boldsymbol{\beta})}{1+\exp(\mathbf{x}_k'\boldsymbol{\beta})}$$

Estimate $\boldsymbol{\beta}$ from the data

Calculate fitted values $\hat{y}_k = \dfrac{\exp(\mathbf{x}_k'\hat{\boldsymbol{\beta}})}{1+\exp(\mathbf{x}_k'\hat{\boldsymbol{\beta}})}, \quad k \in U$

Used in LGREG estimator

Models - 3

Logistic mixed model for binary response variable y

Domain-level random intercepts u_d

$$E_m(y_k | u_d) = \frac{\exp(\mathbf{x}_k' \boldsymbol{\beta} + u_d)}{1 + \exp(\mathbf{x}_k' \boldsymbol{\beta} + u_d)} \qquad \text{with } u_d \sim N(0, \sigma_u^2)$$

Estimate $\boldsymbol{\beta}$ and σ_u^2 from the data

Calculate estimates $\hat{u}_d$, $d = 1, \ldots, D$

Calculate fitted values

$$\hat{y}_k = \frac{\exp(\mathbf{x}_k' \hat{\boldsymbol{\beta}} + \hat{u}_d)}{1 + \exp(\mathbf{x}_k' \hat{\boldsymbol{\beta}} + \hat{u}_d)}, \quad k \in U_d,\ d = 1, \ldots, D$$

Used in MLGREG estimator

Estimation of the model

- GLMMs can be fitted for example by:
  - R packages `nlme` or `lme4 (glmer` function) using maximum likelihood
  - SAS procedures GLIMMIX (using ML) or MIXED (using REML or ML)

- Some methodological references
  - Datta (2009)
  - Jiang and Lahiri (2006)
  - Rao (2003)

GREG estimator

- In all three cases:
  - MGREG – Mixed model assisted GREG
  - LGREG – Logistic fixed-effects model assisted GREG
  - MLGREG – Logistic mixed model assisted GREG

the formulation of the GREG estimators for domain total and mean remain the same!

$$\hat{t}_{dGREG} = \sum_{k \in U_d} \hat{y}_k + \sum_{k \in s_d} a_k (y_k - \hat{y}_k) \ , \ d = 1,\ldots,D$$

$$\hat{\bar{y}}_{dGREG} = \hat{t}_{dGREG} / N_d \ \text{ or } \ \hat{\bar{y}}_{dGREG} = \hat{t}_{dGREG} / \hat{N}_d$$

Data requirements

- Traditional linear GREG estimator
  - Unit-level x-vectors not necessarily needed
  - Known domain totals of x-variables only are needed
  - Applicable in ”survey” countries in particular
- Extended GREG family
  - MGREG, LGREG, MLGREG
  - Unit-level x-data are needed for all units in population
  - Applicable in ”register” countries
  - Applicable also in ”survey” countries if census data can be merged with sample survey data at the unit level

EXAMPLES FROM LITERATURE

- **Numerical (simulation) results on relative performance (design bias and accuracy) of the extended family GREG estimators**

- Lehtonen, Särndal and Veijanen (2003)
  - MGREG estimation assisted by linear mixed model
- Lehtonen, Särndal and Veijanen (2005)
  - LGREG estimation assisted by logistic mixed model
- Lehtonen and Veijanen (2009)
  - MLGREG estimation assisted by linear mixed model
- Lehtonen, R., Veijanen, A., Myrskylä, M. and Valaste, M. (2011)
  - AMELI project; Applications to poverty indicators

EMPIRICAL EXAMPLES

- CASE STUDY 1: Empirical example on estimation of regional means of perceived income by HT, linear GREG (assisted by fixed-effects linear model) and mixed-model assisted MGREG methods. SILC data (Finland) will be used
- CASE STUDY 2: Simulation example in accounting for unequal probability sampling in model-based EBLUP
- (To be discussed on Tuesday 17 Feb.)

Other model-assisted methods

- **Model calibration**
  - Wu and Sitter (2001), Montanari and Ranalli (2005, 2009)
- **Model calibration for domains**
  - Lehtonen and Veijanen (2012, 2014)
- **Assisting models**
  - Members of the GLMM family, for example:
  - Linear mixed models
  - Logistic mixed models

- NOTE: Traditional GREG estimator is restricted to linear fixed-effects assisting model for continuous variables

Calibration methods - Summary

	CALIBRATION METHODS		
	Model-free (linear) calibration MFC	Model calibration MC	Hybrid calibration HC
Weight calibration	Calibration to reproduce the known population totals of the auxiliary variables	Calibration to the population total of the predictions derived via the specified model	Combination of MC and MFC, depending on coherence requirements
Typical study variable	Continuous	Continuous, binary, polytomous, count	Same as MC
Auxiliary data	Aggregate level	Unit level	Unit level
Model specification	No explicit model statement	Generalized linear models family	Same as MC
Main aims	Coherence of estimates with published statistics, "Multi-purpose" weighting, Accuracy improvement	Accuracy improvement, extension of calibration to nonlinear relationships	Accuracy improvement, extension of calibration to nonlinear relation-ships, Coherence of estimates with published statistics
Literature	Deville and Särndal (1992) Särndal (2007) Lehtonen and Veijanen (2009)	Wu and Sitter (2001) Montanari and Ranalli (2005) Lehtonen and Veijanen (2012)	Montanari and Ranalli (2009) Lehtonen and Veijanen (2014)

Additional references

- Deville, J.-C. and C.-E. Särndal (1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association* 87, 376-382.
- Lehtonen, R. and Veijanen, A. (2012). Small area poverty estimation by model calibration. *Journal of the Indian Society of Agricultural Statistics*, 66, 125-133.
- Lehtonen and Veijanen (2014). Small area estimation of poverty rate by model calibration and "hybrid" calibration. NORDSTAT 2014 Conference, June 2014, Turku.
- Lehtonen, R., Veijanen, A., Myrskylä, M. and Valaste, M. (2011). Small Area Estimation of Indicators on Poverty and Social Exclusion. AMELI WP2 Deliverable 2.2. Available at:
- http://www.uni-trier.de/index.php?id=24676&L=2

Additional references (contd.)

- Montanari, G. E. and M. G. Ranalli (2005). Nonparametric model calibration estimation in survey sampling. *Journal of the American Statistical Association* 100, 1429-1442.

- Montanari, G.E. and Ranalli, M.G. (2009). Multiple and ridge model calibration. Proceedings of Workshop on Calibration and Estimation in Surveys 2009. Statistics Canada.

- Wu, C. and Sitter, R.R. (2001). A model-calibration approach to using complete auxiliary information from survey data. *JASA* 96, 185-193.