Freie Universität Berlin
Survey Sampling Group of the Italian Statistical Society
UNIVERSIDADE DA CORUÑA

Variable and Transformation Selection for Linear Mixed Models with Application to Small Area Estimation

Natalia Rojas-Perilla
María José Lombardía
Esther López Vizcaíno
Cristina Rueda Sabate

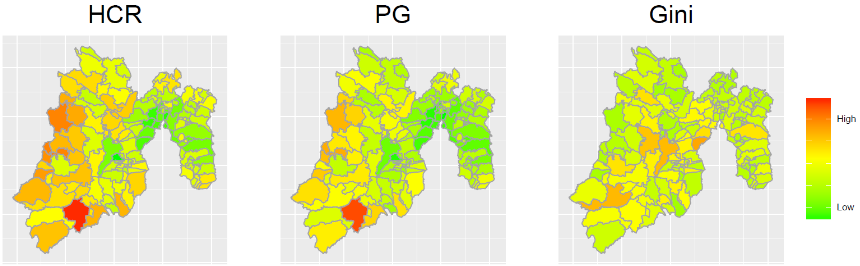
ITACOSM2019
Florence-Italy
Juni 2019

Freie Universität Berlin · Survey Sampling Group of the Italian Statistical Society · UNIVERSIDADE DA CORUÑA

Contents

Freie Universität Berlin · S2G Survey Sampling Group of the Italian Statistical Society · UNIVERSIDADE DA CORUÑA

Motivation: Poverty Mapping



Freie Universität Berlin | Survey Sampling Group of the Italian Statistical Society | UNIVERSIDADE DA CORUÑA

Recent Unit-level Methodologies

- The World Bank method
  (Elbers et al., 2003, Econometrica)
- The Empirical Best Predictor (EBP) method
  (Molina & Rao, 2010, CJS)
- EBP based on normal mixtures
  (Lahiri and Gershunskaya, 2011; Elbers & Van der Weide, 2014)
- Methods based on M-Quantiles
  (Marchetti et al., 2012, CSDA)

EBP Approach for Poverty Indicators

- Point of departure: Nested error regression model (Battese et. al. (1988, JASA))

  Notation: (k =domain, i =individual)

$$y_{ik} = \boldsymbol{x}_{ik}^T \beta + \boldsymbol{z}_{ik}^T u_k + \epsilon_{ik}, i = 1, ..., n_k, k = 1, ..., D$$

- Use sample data to estimate β, σ_u, σ_ϵ, u_k
- Generate $u_k^* \sim \ N(0, \hat{\sigma}_u^2 * (1 - \gamma_k))$ & $\epsilon_{ik}^* \sim \ N(0, \hat{\sigma}_\epsilon^2)$

Micro-simulating a synthetic population:

- Generate a synthetic population under the model a large number of times each time estimating the target parameter
- Linear and non-linear indicators can be computed

Model selection approaches

Freie Universität Berlin · S2G Survey Sampling Group of the Italian Statistical Society · UNIVERSIDADE DA CORUÑA

Principles and Approaches

1. Information criteria based on likelihood functions: Applicable to parametric model-based problems
   - Akaike information criterion (AIC)
   - Finite-sample corrected AIC (CAIC)
   - Bayesian information criterion (BIC)
   - Hannan and Quinn criterion (HQ)
   - Bridge criterion (BC)
2. Methods that do not require parametric assumptions
   - Cross-validation (CV)
3. Methods from other perspectives

Freie Universität Berlin · Survey Sampling Group of the Italian Statistical Society · UNIVERSIDADE DA CORUÑA

Akaike Information Criterion (AIC)

$$AIC(M) = -2\log(l(M)) + 2P$$

- $l(M)$ is the model likelihood $\Longleftarrow$ Loss function
- P measurement of model complexity $\Longleftarrow$ Penalty term
- The model with the lowest value of AIC is selected.

The Log-likelihood for Mixed Models

Marginal approach:

$$\begin{aligned} E(\mathbf{Y}) &= \boldsymbol{\theta} = \mathbf{x}\boldsymbol{\beta} \\ Var(\mathbf{Y}) &= \mathbf{V}_y = \mathbf{Z}'\boldsymbol{\Sigma}_u\mathbf{Z} + \boldsymbol{\Sigma}_e \end{aligned}$$

$$\log(l_m(M)) = -\frac{1}{2}D\log(2\pi) - \frac{1}{2}\log|\mathbf{V}_y| - \frac{1}{2}(\mathbf{Y}-\boldsymbol{\theta})'\mathbf{V}_y^{-1}(\mathbf{Y}-\boldsymbol{\theta})$$

Conditional approach:

$$\begin{aligned} E(\mathbf{Y}|\mathbf{u}) &= \boldsymbol{\mu} = \mathbf{x}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} \\ Var(\mathbf{Y}|\mathbf{u}) &= \mathbf{V}_{y|u} = \Sigma_e \end{aligned}$$

$$\log(l_c(M)) = -\frac{1}{2}D\log(2\pi) - \frac{1}{2}\log|\mathbf{V}_{y|u}| - \frac{1}{2}(\mathbf{Y}-\boldsymbol{\mu})'\mathbf{V}_{y|u}^{-1}(\mathbf{Y}-\boldsymbol{\mu})$$

The Penalty Term

Different concepts or terms have been used:

- The number of parameters in the model.
- The Degrees of Freedom.
- The Divergence.
- The Effective Degrees of Freedom.
- The Generalized Degrees of Freedom

Most of these five terms coincides for simple models as the normal linear regression model but no for complex models as constrained, lasso or mixed models (Kato, 2009; Rueda, 2013; Tibshirani and Taylor, 2012).

The Calculation of GDF

- The Generalized Degrees of Freedom (GDF) is a measure of the sensitivity of each fitted value, $\widehat{m}_d$ to perturbation in the corresponding observed value Y_d, applicable to complex modeling procedures where it is assumed that $EY_d = m_d$.

$$GDF = \sum_{d=1}^{D} \frac{\partial E(\widehat{m}_d)}{\partial m_d}$$

(Ye, 1998;Vaida and Blanchard, 2005;You et al., 2016)

Criteria for the Model Selection

$$GAIC(M) = -2\log(l(\widehat{M})) + \widehat{GDF}$$

- $l(\widehat{M})$: $l(M)$ depends on the $m(=EY)$ and the variance parameters which are estimated under model M.
- $\widehat{GDF}$: although a known quantity in simple models, it is unknown in complex modeling procedures. GDF is estimated by bootstrap: $\widehat{GDF}$.
- Different combinations of conditional and marginal log-likelihood and expectations to estimate GDF have been defined in the literature.
- The model with the lowest value of $GAIC$ is selected.

$GAIC(M)$ for Linear Mixed Models

- $mGAIC(M) = -2\log(l_m(\widehat{M})) + m\widehat{GDF}$,
- $cGAIC(M) = -2\log(l_c(\widehat{M})) + c\widehat{GDF}$,
- $yGAIC(M) = -2\log(l_c(\widehat{M})) + y\widehat{GDF}$,

$$yGDF = \sum_{d=1}^{D}\sum_{j=1}^{n_i}\frac{\partial E_Y(\widehat{\mu}_{ij})}{\partial\theta_{ij}}$$

  $yGDF$ is defined from the conditional mean estimator and the marginal expectation (You et al. (2016)).

- Pferfferman (2013) and Rao and Molina (2015) for a SAE context!

Freie Universität Berlin · S2G Survey Sampling Group of the Italian Statistical Society · UNIVERSIDADE DA CORUÑA

Our Proposals: $l_x(M)$ and $xGDF$

We consider the quasi-loglikelihood (M.J. Lombardía et. al. 2017) from a normal model with mean $\boldsymbol{\mu}$ and Variance $\mathbf{V}_y$:

$$\log(l_x(M)) = -\frac{1}{2} D \log(2\pi) - \frac{1}{2} \log|\mathbf{V}_y| - \frac{1}{2}(\mathbf{Y} - \boldsymbol{\mu})' \mathbf{V}_y^{-1} (\mathbf{Y} - \boldsymbol{\mu}).$$

and $xGDF = yGDF$ defined from the conditional mean estimator and the marginal expectation

$$\begin{aligned} xGDF &= \sum_{d=1}^{D} \sum_{j=1}^{n_i} \frac{\partial E_Y(\widehat{\mu}_{ij})}{\partial \theta_{ij}} \\ &= \sum_{i=1}^{D} \sum_{j=1}^{n_i} \sum_{k=1}^{n_i} V_y^{ijk} Cov_Y(\widehat{\mu}_{ij}, y_{ik}) \end{aligned}$$

Freie Universität Berlin · S2G Survey Sampling Group of the Italian Statistical Society · UNIVERSIDADE DA CORUÑA

Approximation of xGDF

1. Fit the model and obtain $\hat{\beta}$, $\hat{\sigma}_u^2$ and $\hat{\sigma}_e^2$.
2. Repeat B times $(b = 1, \ldots, B)$
   2.1 Generate u_{ij}^* and e_{ij}^* from $N(0, \widehat{\sigma}_u^2)$ and $N(0, \widehat{\sigma}_e^2)$ respectively. Construct the bootstrap model $y_{ij}^{*(b)} = \mu_{ij}^{*(b)} + e_{ij}^{*(b)}$, with $\mu_{ij}^{*(b)} = \mathbf{x}_{ij}'\hat{\beta} + u_{ij}^{*(b)}$, and the variance-covariance matrix $\widehat{\mathbf{V}}_y$.
   2.2 From each bootstrap sample $\{y_{ij}^{*(b)}, \mathbf{x}_{ij}\}$, calculate $\widehat{\mu}_{ij}^{*(b)} = \widehat{\mu}_d(y_{ij}^{*(b)}, \mathbf{x}_{ij}) = \mathbf{x}_{ij}'\hat{\beta}^{*(b)} + \hat{u}_d^{*(b)}$.
3. Approximate *xGDF* by Monte Carlo,

$$\widehat{xGDF} = \frac{1}{B-1}\sum_{b=1}^{B}\sum_{i=1}^{D}\sum_{j=1}^{n_i}\sum_{k=1}^{n_i}\hat{V}_y^{ijk}(\widehat{\mu}_{ij}^{*(b)} - \bar{\widehat{\mu}}_{ij}^*)(y_{ik}^{*(b)} - \bar{y}_{ik}^*)$$

where $\bar{\widehat{\mu}}_{ij}^* = \frac{1}{B}\sum_{b=1}^{B}\widehat{\mu}_{ij}^{*(b)}$ and $\bar{y}_{ik}^* = \frac{1}{B}\sum_{b=1}^{B} y_{ik}^{*(b)}$, $\hat{V}_y^{ijk}$ is the $jk-$element of the inverse of $\hat{\mathbf{V}}_i = \mathbf{Z}_i'\hat{\mathbf{\Sigma}}_u\mathbf{Z}_i + \hat{\sigma}_e^2\mathbf{I}_{n_i}$.

Freie Universität Berlin · S2G Survey Sampling Group of the Italian Statistical Society · UNIVERSIDADE DA CORUÑA

Proposed xGAIC

We define the generalized Akaike Information Criteria ($GAIC$) for a small area model as follows:

$$M: \; y_{ij} = \mathbf{x}_{ij}'\beta + \mathbf{u}_i + \mathbf{e}_i, \qquad i = 1, \ldots, D; j = 1, \ldots, n_i$$

$$GAIC(M) = -2\log(l(M)) + GDF.$$

Here, we propose to combine $l_x(M)$ and $x\widehat{GDF}$ by considering the random effect and the variability between areas to define:

$$xGAIC \;\; = \;\; -2\log(l_x(\widehat{M})) + x\widehat{GDF}.$$

Use of transformations

Model Requirements

- EBP relies on Gaussian assumptions :
  - ✓ $u_k \overset{iid}{\sim} N(0, \sigma_u^2)$, the random area-specific effects
  - ✓ $\epsilon_{ik} \overset{iid}{\sim} N(0, \sigma_\epsilon^2)$, the unit-level error terms, $u_k \perp \epsilon_{ik}$
- What if these fail?
  - Option 1: EBP formulation under an alternative distribution (Molina, I et al., 2015)
  - Option 2: Explore the use of transformations

First Selection of Transformations

- Shifted transformations
  - Log-shift
- Power transformations
  - Box-Cox
  - Exponential
  - Sign power
  - Modulus
  - Dual power
  - Folder power
  - Convex-to-concave
- Multi-parameter transformations
  - Johnson
  - Sinh-arcsinh

Freie Universität Berlin · S2G Survey Sampling Group of the Italian Statistical Society · UNIVERSIDADE DA CORUÑA

Transformations for the EBP Method

Log-Shift Transformation (λ) (Royston et al. 2011)

$$T_\lambda(y_{ij}) = \log(y_{ij} + \lambda),$$

Box-Cox Transformation (λ) (Box & Cox, 1964)

$$T_\lambda(y_{ij}) = \begin{cases} \frac{(y_{ij}+s)^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log(y_{ij} + s), & \lambda = 0 \end{cases},$$

Dual Power Transformation (λ) (Yang, 2006)

$$T_\lambda(y_{ij}) = \begin{cases} \frac{(y_{ij}+s)^\lambda - (y_{ij}+s)^{-\lambda}}{2\lambda} & \text{if } \lambda > 0; \\ \log(y_{ij} + s) & \text{if } \lambda = 0. \end{cases}$$

with λ the transformation parameter and $y_{ij} + s > 0$

The EBP Approach under Transformations

1. Select $T_\lambda(y_{ij}) = y^*_{ij}$ and obtain the transformed sample data
2. Use transformed sample data to fit the model and obtain $\hat{\beta}, \hat{\sigma}^2_u, \hat{\sigma}^2_e$ and predict $\bar{u}_j = E(u_j|y_s)$ for in-sample domains
3. Use census data to micro-simulate L synthetic populations by

$$y^*_{ij} = \mathbf{x}^T_{ij}\hat{\beta} + \bar{u}_j + u_j + e_{ij},$$

   where $u_j \sim N(0, \hat{\sigma}^2_u * (1 - \gamma_j))$ and $e_{ij} \sim N(0, \hat{\sigma}^2_e)$
4. Back-transform to the original scale
5. Calculate the target linear and non-linear indicators in each replication and average over L

Combined estimators

Some Ideas

- Model-based estimators are derived from models defined using the auxiliary information.
- In SAE, we can talk about estimation selection instead of model selection.
- Obtaining different small areas estimators from a collection of models.
- Selecting the best estimator: xGAIC measure combined with REML for transformatios (for mixed models).

Estimation Method (λ)

Residual Maximum Likelihood (REML) (Gurka et al., 2006)

- Using a scaled version of the transformation
- This allows for applying standard maximum likelihood theory

$$
\begin{aligned}
L_{\text{REML}}(T_\lambda, \lambda|\boldsymbol{\theta}) &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\sum_{i=1}^{D}\log|\boldsymbol{V}_i| \\
&- \frac{1}{2}\log\left|\sum_{i=1}^{D}\boldsymbol{X}_i'\boldsymbol{V}_i^{-1}\boldsymbol{X}_i'\right| \\
&- \frac{1}{2}\sum_{i=1}^{D}[T_\lambda(\boldsymbol{y}_i) - \boldsymbol{X}_i\beta]^T\boldsymbol{V}_i^{-1}[T_\lambda(\boldsymbol{y}_i) - \boldsymbol{X}_i\beta]
\end{aligned}
$$

Estimation Algotithm (λ)

REML Algorithm for the EBP Method

1. Choose a transformation
2. Define a parameter interval for λ
3. Set λ to a value inside the interval
4. Find the best model by using xGAIC
5. Maximize the residual log-likelihood function with respect to θ conditional on the fixed λ
6. Repeat 3 and 4 until a maximum $\hat{\lambda}$ is found
7. Apply the EBP method (under optimal model)

Remarks and Future Research Direction

Remarks:

- Use of transformations improves the performance of the EBP and fit of the model (Rojas et. al.)
- xGAIC outperforms other GAIC measures in the SAE context and works for non linear models

Next steps:

- Proposed bootstrap approach taking into account uncertainty from the multi-parameter parameter estimation algorithm.

Essential Bibliography

[1] Rao, J.N.K. (2003) *Small area estimation.* John Wiley & Sons.

[2] Molina, I., & Rao, J.N.K. (2010) *Small area estimation of poverty indicators.* Canadian Journal of Statistics, 38(3), 369–385.

[3] Gurka, M. J., Edwards, L. J., Muller, K. E. & Kupper, L. L. (2006) *Extending the Box–Cox Transformation to the Linear Mixed Model.* Journal of the Royal Statistical Society. Series A (Statistics in Society), 26(2), 211–252.

[4] Royston, P. & Lambert, P. C. (2011) *Flexible Parametric Survival Analysis Using Stata: Beyond the Cox Model.* Stata Press

Thank you very much for your attention.

Natalia Rojas-Perilla (natalia.rojas@fu-berlin.de)