Two-phase estimation techniques based on HTC

A selection of topics related to generalized regression estimator

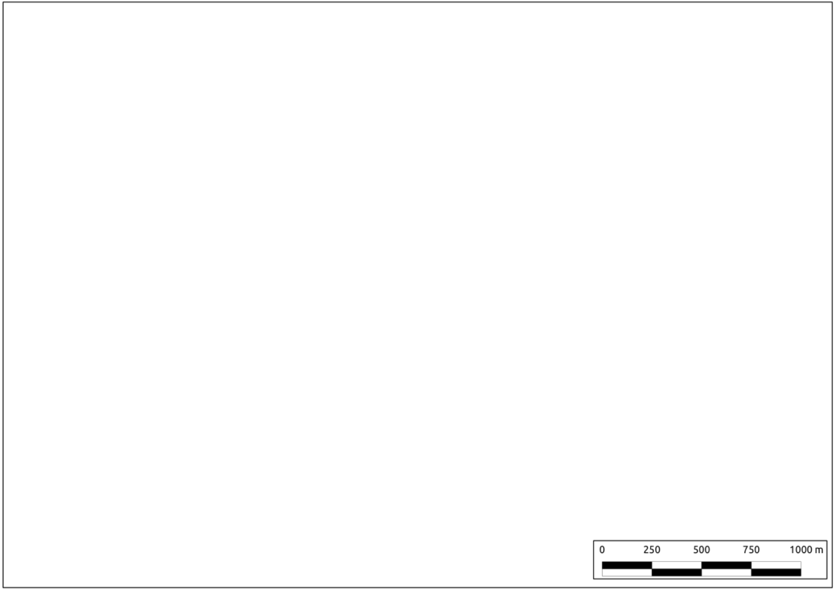
Radim Adolt

ÚHÚL Brandýs nad Labem, Czech Republic

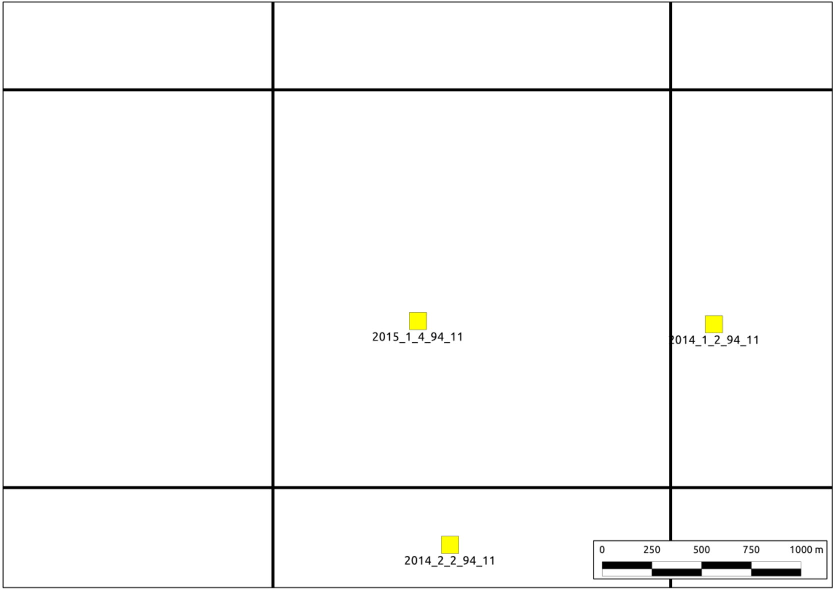
USEWOOD WG2, Training school in Dublin, 16.-19. September 2013

Contents

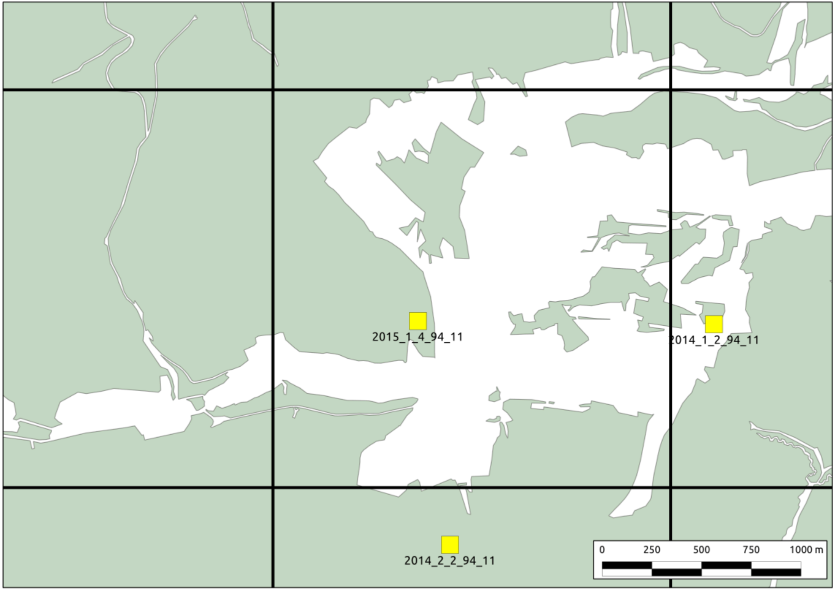
No auxiliary information



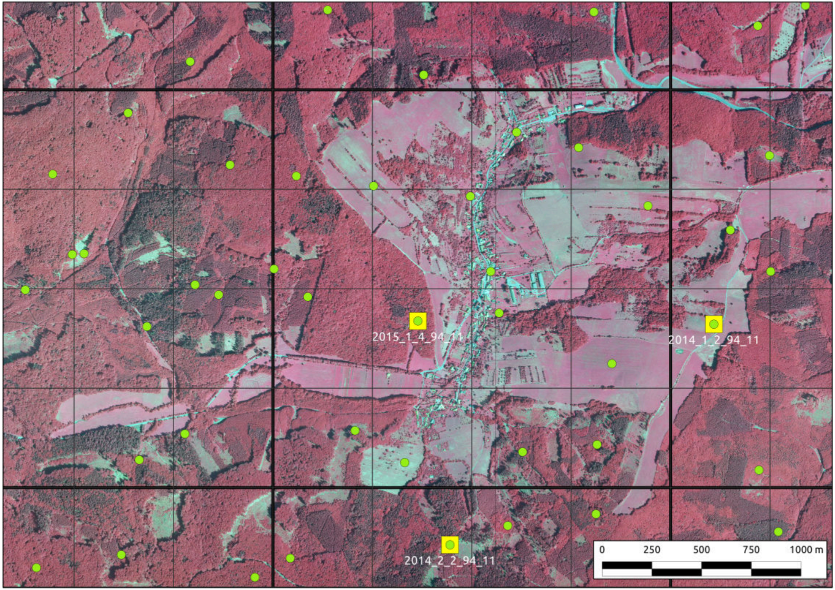
Field survey plots (situation in one-phase sampling)



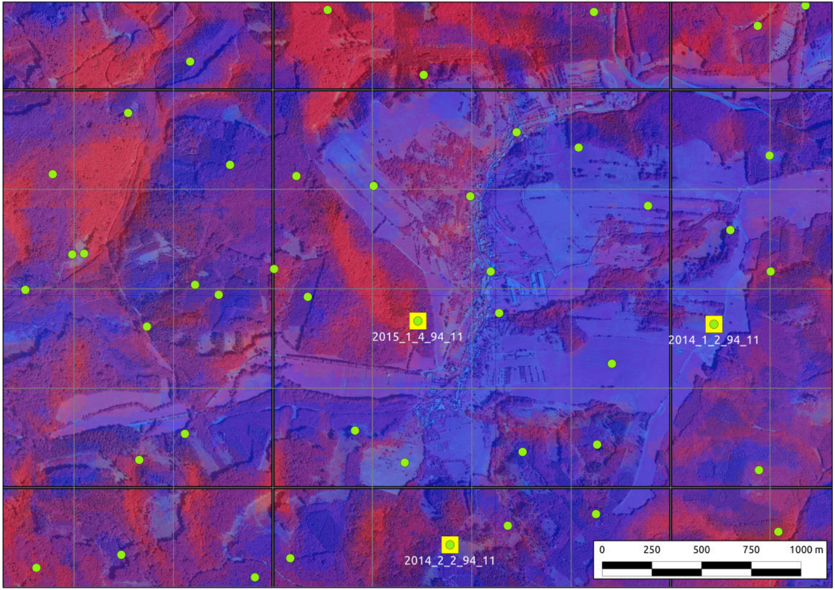
Forest/non-forest map layer (ÚHÚL, aerial imagery ČÚZK)



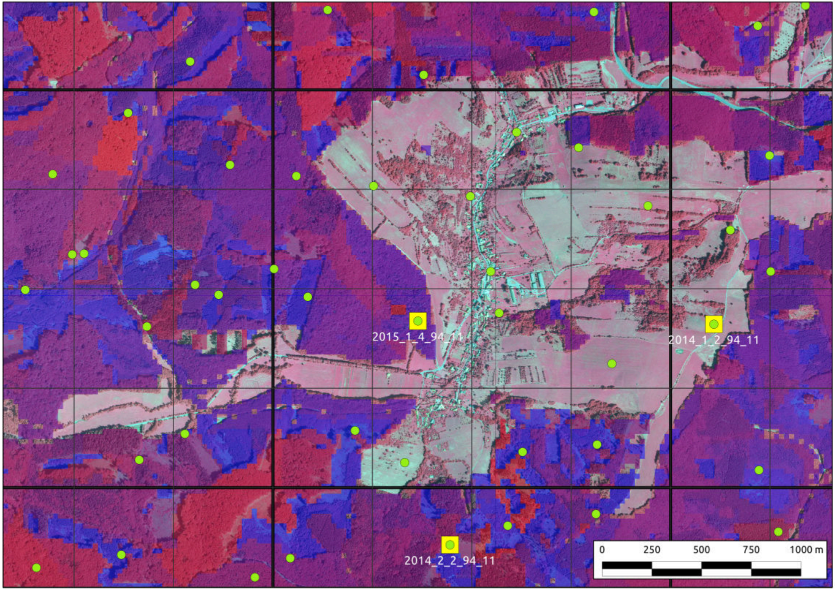
Photogrammetric interpretation (aerial imagery ČÚZK)



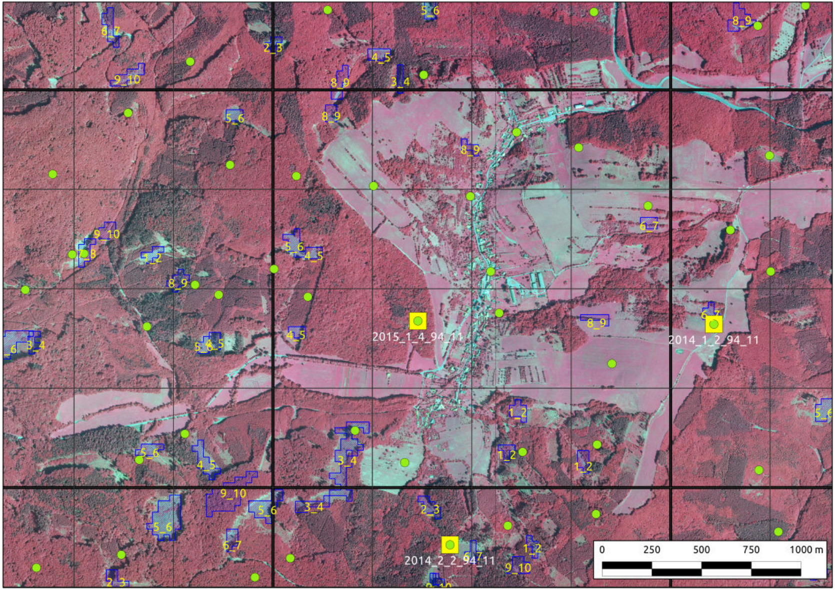
nDSM_SRTM (SRTM NASA, DTM ČÚZK, nDSM ÚHÚL)



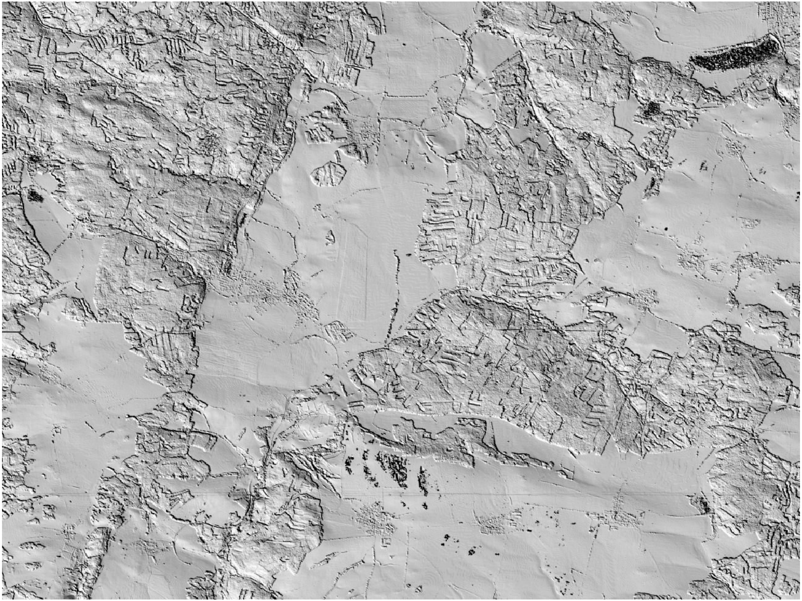
Growing stock, forest management plans (ÚHÚL 2007)



Clearcut areas, (time series using Landsat imagery, ÚHÚL)



nDSM (DSM aerial imagery 2011, DTM ČÚZK 2010)



Two-phase estimators and auxiliary data types

Two-phase estimators according to first-phase data:

- **regression estimators using wall2wall maps** - first-phase is census, the entire auxiliary population is known
- **intensified sample grid** - e. g. photogrammetric interpretation in the first-phase, we have access only to a sample from the auxiliary population, though it is usually much larger that the two-phase sample, **the double-sampling approach**
- **a combination of both** - e. g. wall2wall map plus photogrammetric interpretation

Design-based, model assisted estimators have a big potential to improve accuracy of one-phase estimators **without introducing bias**.

Generalized regression estimator (GREG), definitions I

GREG can be equivalently defined by any of the following formulas:

$$\begin{aligned}
\hat{Y}_{greg,D} &= \hat{Y}_D + (\mathbf{X_D} - \hat{\mathbf{X}}_\mathbf{D})'\hat{\mathbf{B}}_\mathbf{D}, && (1)\\
&= \mathbf{X}_\mathbf{D}'\hat{\mathbf{B}}_\mathbf{D} + \hat{E}_D, && (2)\\
&= \hat{Y}_{pred,D} + \hat{E}_D. && (3)
\end{aligned}$$

- $\mathbf{X_D} = (X_{1,D}, X_{2,D}, \cdots, X_{k,D})'$ is a column vector of known totals of k auxiliary variables in D
- $\hat{\mathbf{X}}_\mathbf{D} = (\hat{X}_{1,D}, \hat{X}_{2,D}, \cdots, \hat{X}_{k,D})'$ is a column vector of one-phase total estimators of the k auxiliary variables in D
- the term $\hat{Y}_D$ is the **one-phase** total estimator of the resource Y_D in D
- $\hat{\mathbf{B}}_\mathbf{D} = (\hat{B}_{1,D}, \hat{B}_{2,D}, \cdots, \hat{B}_{k,D})'$ is a column vector of k regression coefficients (estimators of them)
- the term $\hat{E}_D$ is **one-phase total estimator for model residuals** $E(x)$ in D

Generalized regression estimator (GREG), definitions II

- $\hat{Y}_{pred,D}$ stands for a 'known' total of predictions $\hat{Y}_{pred}(x)$ of the local density $Y(x)$ in D [Mandallaz, 1991, see section 3.1 starting on page 18], defined by (4)

$$\hat{Y}_{pred,D} = \int_D \hat{Y}_{pred}(x)dx \tag{4}$$

Residuals of the local density prediction on sample point x are defined by

$$E(x) = Y(x) - \hat{Y}_{pred}(x) \tag{5}$$

$$E(x) = Y(x) - \mathbf{X}'(x)\hat{\mathbf{B}}_{\mathbf{D}}, \tag{6}$$

where $\mathbf{X}(x) = \left[X_1(x), X_2(x), \cdots, X_k(x)\right]'$ is column vector of the local densities of the auxiliary variables observed on sample point x. We'll suppose that $\mathbf{X}(x)$ is known over the whole area of D.

Estimation of parameter vector (GREG) I

Weighted least squares method, WLS

In a general case the vector of regression parameters $\hat{\mathbf{B}}_{\mathbf{D}} = (\hat{B}_{1,D}, \hat{B}_{2,D}, \cdots, \hat{B}_{k,D})'$ used by GREG (SREG) is estimated by Weighted Least Squares (WLS) method as follows

$$\hat{\mathbf{B}}_{\mathbf{D}} = \left[\sum_{x \in \boldsymbol{s}, x \in D} \frac{\mathbf{X}(x)\mathbf{X}'(x)}{\sigma^2(x)\bar{\pi}_D(x)} \right]^{-1} \sum_{\substack{x \in \boldsymbol{s} \\ x \in D}} \frac{\mathbf{X}(x)Y(x)}{\sigma^2(x)\bar{\pi}_D(x)}, \tag{7}$$

where $\bar{\pi}_D(x)$ is the expected sampling density function at sample point x [Cordy, 1993]. Most typically for NFIs, but not necessarily, it takes a constant value. The term $\sigma^2(x)$ stands for the variance of model errors. Column vector $\mathbf{X}(x) = \left[X_1(x), X_2(x), \cdots, X_k(x)\right]'$ is made up by local densities of k auxiliary variables observed on sample point $x \in D$. $Y(x)$ is the local density of target variable.

Estimation of parameter vector (GREG) II

Ordinary least squares, OLS

We can leave out the local sampling density $\bar{\pi}_D(x)$ from (7) which is considered constant in D (as it is the case for many NFIs). For simplicity let's suppose the error variance $\sigma^2(x)$ is constant (homoscedasticity of errors), so it cancels out as well.

$$\hat{\mathbf{B}}_{\mathbf{D}} = \left[\sum_{x \in \boldsymbol{s},\, x \in D} \mathbf{X}(x)\mathbf{X}'(x) \right]^{-1} \sum_{\substack{x \in \boldsymbol{s} \\ x \in D}} \mathbf{X}(x) Y(x) \tag{8}$$

Equation (8) corresponds to Ordinary Least Squares (OLS) method.

Introduction of g-weights (GREG) I

Substituting (8) into (1) **GREG can be expressed in the form of** (12)

$$\hat{Y}_{greg,D} = \lambda(c) \sum_{\substack{x \in s \\ x \in D}} Y(x) + \left(\mathbf{X_D} - \hat{\mathbf{X}}_\mathbf{D}\right)' \left[\sum_{x \in s, x \in D} \mathbf{X}(x)\mathbf{X}'(x) \right]^{-1} \sum_{\substack{x \in s \\ x \in D}} \mathbf{X}(x) Y(x), \tag{9}$$

$$= \sum_{\substack{x \in s \\ x \in D}} \left\{ \lambda(c) + (\mathbf{X_D} - \hat{\mathbf{X}}_\mathbf{D})' (\mathbf{X}\mathbf{X}')^{-1} \mathbf{X}(x) \right\} Y(x), \tag{10}$$

$$= \sum_{\substack{x \in s \\ x \in D}} \lambda(c) g(x) Y(x), \tag{11}$$

$$= \sum_{\substack{x \in s \\ x \in D}} w^*(x) Y(x). \tag{12}$$

Columns of the auxiliary variable matrix $\mathbf{X}$ are made up by $\mathbf{X}(x)$ (vectors of the local densities of auxiliary variables observed on sample point $x \in D$).

Introduction of g-weights (GREG) II

The term $\lambda(c)$ is **the size of the inventory block** (cell size, supposing there is only one sample point per inventory block), which corresponds to the reciprocal value of mean expected sampling density $\bar{\pi}_D(x) = \bar{\pi}_D$. **Constant sampling density in D is considered**. The estimator (12) is **unbiased unconditionally on sample size in D**.

$$g(x) = 1 + \lambda(c)^{-1}\left(\mathbf{X_D} - \hat{\mathbf{X}}_\mathbf{D}\right)'\left(\mathbf{X}\mathbf{X}'\right)^{-1}\mathbf{X}(x) \tag{13}$$

$$w^*(x) = \lambda(c)g(x) \tag{14}$$

The terms $g(x)$ and $w^*(x)$ are weights defined accoring to (13) (regression weight) and (14) (calibration weight, a product of sampling and regression weights).

Properties of GREG I

Two aspects of GREG's additivity

- **GREG total estimators calculated using identical set of auxiliary variables and model formulation are target variable-additive**
- **calibration weights don't change as the variable changes**, e. g. GREG estimated total growing stock in D equals the sum of GREG estimated growing stock totals for individual species
- **GREG estimator is not geographically additive** in the sense that the GREG total for a variable in D doesn't equal the sum of GREG estimators calculated for geographical partitions of D
- **for different geographical areas a different set of data is used to derive calibration weights**, different sample points enter the calculation of g-weights
- geographical additivity is not reached even if the set of auxiliary variables and the model formulation remain constant between D and its partitions

Properties of GREG II

Equality of GREG and synthetic estimator

If there exists a constant column vector ν fulfilling condition (15)**, the GREG and synthetic estimators are equal.** In other words the estimator $\hat{E}_D$ of the total of residuals in D equals zero.

$$\sigma^2(x) = \nu' \mathbf{X}(x) \tag{15}$$

Proof and further details can be found in [Rao, 2003, chapter 2, page 14] or in [Särndal et al., 2003, chapter 6, pages 231 and 232]. **Condition** (15) **holds for some linear models frequently used by GREG:**

1. models with an intercept term and constant variance $\sigma^2(x) = \sigma^2$
2. **ratio estimator**, i. e. models with just one auxiliary variable $X(x)$ without the intercept and variance $\sigma^2(x) = X(x)\sigma^2$
3. models including one or more auxiliary variables with error variance being linear combination of these, **e. g. post-stratification**

Properties of GREG III

Calibration property of GREG

A theoretically appreciated property of GREG (and also SREG) is expressed by

$$\hat{\mathbf{X}}_{reg,D} = \sum_{\substack{x \in s \\ x \in D}} w(x)\mathbf{X}(x) = \mathbf{X_D}, \tag{16}$$

stating that **the vector of GREG total estimators for the auxiliary variables equals the vector of their true totals**.

GREG is sometimes called a calibration estimator since it minimizes χ^2 differences between calibration $w^*(x)$ and sample weights. It modifies sample weights as little as possible subject to the above constraint, see [Rao, 2003, chapter 2 starting on page 13] or [Särndal et al., 2003, note 6.5.1 on page 234] for further details.

Survey Regression Estimator (SREG), introduction I

SREG properties

- **linear model parametrized on an extended region $D^+ \supseteq D$**
- estimates can be obtained for $D \subseteq D^+$
- **SREG is variable- as well as geographically-additive**
- **SREG fails if D is very different from D^+ and if the structure of auxiliary variables cannot capture specificity of D**
- equality of synthetic and SREG estimators under condition (15) occurs only for $D = D^+$, in this case SREG, GREG and synthetic estimators are all equal
- property (16) holds also for SREG

Survey Regression Estimator (SREG), introduction II

Advantages of SREG over GREG

- **GREG is not suitable for small geographical domains** D,
- the linear model behind **GREG is parametrized using inventory points from inside** D **only**
- if the domain sample size n_D is small, one should not rely on consistency and (approximate) unbiasedness of GREG and its variance
- for small sample sizes, OLS (WLS) parametrization might not be possible due to singularity of $\mathbf{XX}'$
- **SREG addressees all above issues**

SREG estimator **parametrized on the level of an extended region** $D^+ \supseteq D$ is defined as follows:

$$\hat{Y}_{sreg,D} = \lambda(c) \sum_{\substack{x \in s \\ x \in D^+}} I_D(x) Y(x) \tag{17}$$

$$+ \left(\mathbf{X_D} - \hat{\mathbf{X}}_\mathbf{D}\right)' \left[\sum_{x \in s, x \in D^+} \mathbf{X}(x)\mathbf{X}'(x) \right]^{-1} \sum_{\substack{x \in s \\ x \in D^+}} \mathbf{X}(x) Y(x),$$

$$= \sum_{\substack{x \in s \\ x \in D}} \left\{ I_D(x)\lambda(c) + \left(\mathbf{X_D} - \hat{\mathbf{X}}_\mathbf{D}\right)' \left(\mathbf{X}_+ \mathbf{X}_+'\right)^{-1} \mathbf{X}_+(x) \right\} Y(x), \tag{18}$$

$$= \sum_{\substack{x \in s \\ x \in D}} \lambda(c) \tilde{g}(x) Y(x), \tag{19}$$

$$= \sum_{\substack{x \in s \\ x \in D}} \tilde{w}(x) Y(x). \tag{20}$$

Survey regression estimator (SREG), definitions II

Columns of the $\mathbf{X}_+$ matrix are formed by $\mathbf{X}_+(x)$ - the vectors of local densities of auxiliary variables observed on sample point $x \in D^+ \supseteq D$. **Indicator variable** $I_D(x)$ **tells whether the sample point** $x \in D^+$ **belongs also to** D ($I_D(x) = 1$) or not ($I_D(x) = 0$).

$$\tilde{g}(x) = I_D(x) + \lambda(c)^{-1}\left(\mathbf{X_D} - \hat{\mathbf{X}}_\mathbf{D}\right)'\left(\mathbf{X}_+\mathbf{X}'_+\right)^{-1}\mathbf{X}_+(x) \tag{21}$$

$$\tilde{w}(x) = \lambda(c)\tilde{g}(x) \tag{22}$$

The terms $\tilde{g}(x)$ and $\tilde{w}(x)$ are weights defined accoring to (21) (SREG regression weight) and (22) (calibration weight, a product of sampling and regression weights).

Variance of GREG total estimator, URS I

An approximate variance of the GREG estimator of total is given by

$$\hat{\mathbb{V}}_{URS}\left(\hat{Y}_{greg,D}\right) = \frac{\lambda^2(D)}{n_D(n_D - 1)} \sum_{\substack{x \in s \\ x \in D}} \left[E_g(x) - \hat{\bar{E}}_{g,D}\right]^2. \qquad (23)$$

- $E_g(x)$ is a product of residual local density $E(x)$ given by (5) or (6) and regression weights $g(x)$ (13) (GREG) or $\tilde{g}(x)$ (21) (SREG)
- variance **estimator** (23) **is approximate** [Särndal et al., 2003, section 6.6 starting on page 235],
- this variance estimator uses $\pi(x_i, x_j)$ **corresponding to URS with fixed sample size in** D
- it can be used as **an approximation for systematic and spatially stratified sampling designs**

Variance of GREG total estimator, URS II

- due to **decreased spatial correlation of residuals** $E(x)$ we can expect **reduced conservativeness of the URS approximate variance estimator** in comparison to one-phase situation
- besides regression weights (not necessarily involved in the variance estimation) **residuals are the only contributors to variance estimator**

Variance of GREG and SREG total estimators, ST

In analogy to one-phase approach **the variance of $\hat{Y}_{greg,D}$ can be less conservatively estimated** by

$$\hat{\mathbb{V}}_{ST}\left(\hat{Y}_{greg,D}\right) = \frac{\lambda^2(D)}{2kn_D} \sum_{\substack{x \in s \\ x \in D}} \sum_{\substack{\dot{x} \in s \\ \dot{x} \in D}} \left[E_g(x) - E_g(\dot{x})\right]^2. \tag{24}$$

Variance estimator (24) is based on **first order differences that reduce spatial correlation effects present in case of systematic and spatially stratified sampling**.

Estimators $\hat{\mathbb{V}}_{URS}\left(\hat{Y}_{greg,D}\right)$ and $\hat{\mathbb{V}}_{ST}\left(\hat{Y}_{greg,D}\right)$ use products of residuals and regression weights. This approach has been reported better in comparison to using residuals only [Särndal et al., 2003, chapter 6, note 6.6.1 on page 237].

For $\hat{\mathbb{V}}_{URS}\left(\hat{Y}_{sreg,D}\right)$ and $\hat{\mathbb{V}}_{ST}\left(\hat{Y}_{sreg,D}\right)$ **express all quantities on the right hand side of** (23) **and** (24) **in terms of** D^+ **instead of** D.

Definition of a ratio of two regression estimators

There are situations in which we need to estimate a ratio using two GREG (SREG) estimators and to estimate its variance.

GREG (SREG) estimator of a ratio is defined by

$$\hat{R}_{reg,D} = \frac{\hat{Y}_{reg_1,D}}{\hat{Y}_{reg_2,D}}, \tag{25}$$

where $\hat{Y}_{reg_1,D}$ and $\hat{Y}_{reg_2,D}$ are GREG (SREG) total estimators.

Variance of the GREG (SREG) estimator of a ratio

Variance of the GREG (SREG) estimator of a ratio can be expressed as

$$\mathbb{V}\left(\hat{R}_{reg,D}\right) \approx \frac{1}{Y_2^2} \times \\ \times \left[\mathbb{V}\left(\hat{Y}_{reg_1,D}\right) + R_{1,2}^2 \mathbb{V}\left(\hat{Y}_{reg_2,D}\right) - 2R_{1,2}^2 \mathbb{C}\left(\hat{Y}_{reg_1,D}, \hat{Y}_{reg_2,D}\right)\right]. \qquad (26)$$

- it may be lower than the variance of $\hat{R}_{1,2,D}$ (ratio estimator using one-phase totals), provided GREG (SREG) total estimators $\hat{Y}_{reg_1,D}$ and $\hat{Y}_{reg_2,D}$ are more accurate than their one-phase counterparts
- **typically the covariance between $\hat{Y}_{reg_1,D}$ and $\hat{Y}_{reg_2,D}$ is much lower** than for the corresponding one-phase estimators, this decreases the gain of GREG estimation of ratio
- if the accuracy of GREG (SREG) totals is worth the effort, **GREG (SREG) ratio is useful to achieve consistency between estimators of the two totals and the estimator of ratio**

Approximate variance of the GREG (SREG) ratio I

Approximate variance estimators (27) **and** (28) **of the GREG (SREG) ratio** can be derived substituting corresponding variance and covariance estimators to (26).

$$\hat{\mathbb{V}}_{URS}\left(\hat{R}_{greg,D}\right) = \frac{\lambda^2(D)}{n_D(n_D-1)\hat{Y}^2_{2,greg,D}} \sum_{\substack{x\in s \\ x\in D}} Z^2_{greg}(x) \tag{27}$$

$$\hat{\mathbb{V}}_{ST}\left(\hat{R}_{greg,D}\right) = \frac{\lambda^2(D)}{2kn\hat{Y}^2_{2,greg,D}} \sum_{\substack{x\in s \\ x\in D}} \sum_{\substack{\dot{x}\in s \\ \dot{x}\in D}} \left[Z_{greg}(x) - Z_{greg}(\dot{x})\right]^2 \tag{28}$$

$Z_{greg}(x)$ term, **residual density on sample point** x, is defined by

$$Z_{greg}(x) = g_1(x)E_1(x) - g_2(x)E_2(x)\hat{R}_{greg,D}. \tag{29}$$

Approximate variance of the GREG (SREG) ratio II

For $\hat{\mathbb{V}}_{URS}\left(\hat{R}_{sreg,D}\right)$ and $\hat{\mathbb{V}}_{ST}\left(\hat{R}_{sreg,D}\right)$ express all quantities on the right hand side of (27) and (28) in terms of D^+ instead of D. In (29) replace ratio estimator $\hat{R}_{greg,D}$ by $\hat{R}_{sreg,D^+}$ as well as the GREG g-weights $g_1(x)$ and $g_2(x)$ by their SREG counterparts $\tilde{g}_1(x)$ and $\tilde{g}_2(x)$.

Literature

C. B. Cordy. An extension of the horwitz-thompson theorem to point sampling from a continuous universe. *Statistics and Probability Letters*, 18:353–362, 1993.

D. Mandallaz. *An unified approach to sampling theory for forest inventory based on infinite population and superpopulation models*. PhD thesis, Swiss Federal Institute of Technology (ETH), Zurich, 1991.

J. N. K. Rao. *Small Area Estimation*. Wiley, 2003.

C. E. Särndal, B. Swensson, and J. Wretman. *Model Assisted Survey Sampling*. Springer, 2003.