UNIVERSITY OF TARTU

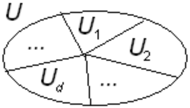
ESTIMATION UNDER RESTRICTIONS BUILT UPON BIASED INITIAL ESTIMATORS

Natalja Lepik

Workshop of Baltic-Nordic-Ukrainian Network on Survey Statistics

24-28 August 2012

Let the finite population U be divided into D non-overlapping domains U_d,
$d \in \mathcal{D} = \{1, 2, ..., D\}$.



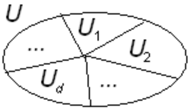
- We are interested in some domain parameters, for example domain totals, $t^d = \sum_{i \in U_d} y_i$ with y_i being the value of study variable for object i.
- It is natural that domain totals sum up to the population total, $\sum_{d=1}^{D} t^d = t = \sum_{i \in U} y_i$.

Relationships between population parameters do not necessarily hold for the estimates in a sample!

=> **inconsistency of estimators**

Introduction

Let the finite population U be divided into D non-overlapping domains U_d, $d \in \mathcal{D} = \{1, 2, ..., D\}$.



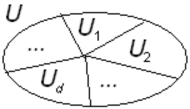
- We are interested in some domain parameters, for example domain totals, $t^d = \sum_{i \in U_d} y_i$ with y_i being the value of study variable for object i.
- It is natural that domain totals sum up to the population total, $\sum_{d=1}^{D} t^d = t = \sum_{i \in U} y_i$.

Relationships between population parameters do not necessarily hold for the estimates in a sample!

=> **inconsistency of estimators**

Introduction

Let the finite population U be divided into D non-overlapping domains U_d, $d \in \mathcal{D} = \{1, 2, ..., D\}$.



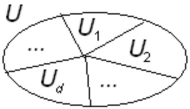
- We are interested in some domain parameters, for example domain totals, $t^d = \sum_{i \in U_d} y_i$ with y_i being the value of study variable for object i.
- It is natural that domain totals sum up to the population total, $\sum_{d=1}^{D} t^d = t = \sum_{i \in U} y_i$.

Relationships between population parameters do not necessarily hold for the estimates in a sample!

=> **inconsistency of estimators**

Introduction

Let the finite population U be divided into D non-overlapping domains U_d, $d \in \mathcal{D} = \{1, 2, ..., D\}$.



- We are interested in some domain parameters, for example domain totals, $t^d = \sum_{i \in U_d} y_i$ with y_i being the value of study variable for object i.
- It is natural that domain totals sum up to the population total, $\sum_{d=1}^{D} t^d = t = \sum_{i \in U} y_i$.

Relationships between population parameters do not necessarily hold for the estimates in a sample!

=> **inconsistency of estimators**

Cases of inconsistency

- Aggregated from registers population statistics and obtained from the sampling survey domain estimates. *Domain estimates from the survey do not sum up to the totals available from the registers.*
- The multi-survey situation: some study variables are common in two or more surveys. *Domain estimates from one survey do not sum up to the estimates of larger domains (or population totals) from another survey.*
- Domains themselves and the population total may be estimated by conceptually different estimators in the same survey. *As a result, the domain totals do not sum up to the population total, or to the relevant larger domains.*

The described inconsistency is annoying from the statistics users viewpoint.

Cases of inconsistency

- Aggregated from registers population statistics and obtained from the sampling survey domain estimates. *Domain estimates from the survey do not sum up to the totals available from the registers.*

- The multi-survey situation: some study variables are common in two or more surveys. *Domain estimates from one survey do not sum up to the estimates of larger domains (or population totals) from another survey.*

- Domains themselves and the population total may be estimated by conceptually different estimators in the same survey. *As a result, the domain totals do not sum up to the population total, or to the relevant larger domains.*

The described inconsistency is annoying from the statistics users viewpoint.

Cases of inconsistency

- Aggregated from registers population statistics and obtained from the sampling survey domain estimates. *Domain estimates from the survey do not sum up to the totals available from the registers.*

- The multi-survey situation: some study variables are common in two or more surveys. *Domain estimates from one survey do not sum up to the estimates of larger domains (or population totals) from another survey.*

- Domains themselves and the population total may be estimated by conceptually different estimators in the same survey. *As a result, the domain totals do not sum up to the population total, or to the relevant larger domains.*

The described inconsistency is annoying from the statistics users viewpoint.

Cases of inconsistency

- Aggregated from registers population statistics and obtained from the sampling survey domain estimates. *Domain estimates from the survey do not sum up to the totals available from the registers.*

- The multi-survey situation: some study variables are common in two or more surveys. *Domain estimates from one survey do not sum up to the estimates of larger domains (or population totals) from another survey.*

- Domains themselves and the population total may be estimated by conceptually different estimators in the same survey. *As a result, the domain totals do not sum up to the population total, or to the relevant larger domains.*

The described inconsistency is annoying from the statistics users viewpoint.

Cases of inconsistency

- Aggregated from registers population statistics and obtained from the sampling survey domain estimates. *Domain estimates from the survey do not sum up to the totals available from the registers.*

- The multi-survey situation: some study variables are common in two or more surveys. *Domain estimates from one survey do not sum up to the estimates of larger domains (or population totals) from another survey.*

- Domains themselves and the population total may be estimated by conceptually different estimators in the same survey. *As a result, the domain totals do not sum up to the population total, or to the relevant larger domains.*

The described inconsistency is annoying from the statistics users viewpoint.

Ideas

- It is known that any auxiliary information incorporated into estimators may increase precision of these estimators
- Known relationships between population parameters is a kind of the auxiliary information.
- If one could use this information in the estimators, one were able to make estimators more accurate and force them to satisfy desired restrictions.

Our aim is the elaboration and study of consistent domain estimators that are more accurate than the initial estimators and satisfy restrictions hold for the population parameters.

Ideas

- It is known that any auxiliary information incorporated into estimators may increase precision of these estimators
- Known relationships between population parameters is a kind of the auxiliary information.
- If one could use this information in the estimators, one were able to make estimators more accurate and force them to satisfy desired restrictions.

Our aim is the elaboration and study of consistent domain estimators that are more accurate than the initial estimators and satisfy restrictions hold for the population parameters.

Ideas

- It is known that any auxiliary information incorporated into estimators may increase precision of these estimators
- Known relationships between population parameters is a kind of the auxiliary information.
- If one could use this information in the estimators, one were able to make estimators more accurate and force them to satisfy desired restrictions.

Our aim is the elaboration and study of consistent domain estimators that are more accurate than the initial estimators and satisfy restrictions hold for the population parameters.

Ideas

- It is known that any auxiliary information incorporated into estimators may increase precision of these estimators
- Known relationships between population parameters is a kind of the auxiliary information.
- If one could use this information in the estimators, one were able to make estimators more accurate and force them to satisfy desired restrictions.

Our aim is the elaboration and study of consistent domain estimators that are more accurate than the initial estimators and satisfy restrictions hold for the population parameters.

Ideas

- It is known that any auxiliary information incorporated into estimators may increase precision of these estimators
- Known relationships between population parameters is a kind of the auxiliary information.
- If one could use this information in the estimators, one were able to make estimators more accurate and force them to satisfy desired restrictions.

Our aim is the elaboration and study of consistent domain estimators that are more accurate than the initial estimators and satisfy restrictions hold for the population parameters.

Ideas

Consistency of domain estimators has been considered for some time.

This work is based on earlier results,

- the general restriction (GR) estimator (*Knottnerus 2003*),
- the GR estimator elaborated for domains (*Sõstra 2007, Sõstra ja Traat 2009*).

These methods produce optimal GR estimators, but they use **unbiased** or **nearly unbiased** initial estimators.

We generalize the approach by allowing **biased** initial estimators.

Ideas

Consistency of domain estimators has been considered for some time.

This work is based on earlier results,

- the general restriction (GR) estimator (*Knottnerus 2003*),
- the GR estimator elaborated for domains (*Sõstra 2007, Sõstra ja Traat 2009*).

These methods produce optimal GR estimators, but they use **unbiased** or **nearly unbiased** initial estimators.

We generalize the approach by allowing **biased** initial estimators.

Ideas

Consistency of domain estimators has been considered for some time.

This work is based on earlier results,

- the general restriction (GR) estimator (*Knottnerus 2003*),
- the GR estimator elaborated for domains (*Sõstra 2007*, *Sõstra ja Traat 2009*).

These methods produce optimal GR estimators, but they use **unbiased** or **nearly unbiased** initial estimators.

We generalize the approach by allowing **biased** initial estimators.

Ideas

Consistency of domain estimators has been considered for some time.

This work is based on earlier results,

- the general restriction (GR) estimator (*Knottnerus 2003*),
- the GR estimator elaborated for domains (*Sõstra 2007, Sõstra ja Traat 2009*).

These methods produce optimal GR estimators, but they use **unbiased** or **nearly unbiased** initial estimators.

We generalize the approach by allowing **biased** initial estimators.

Ideas

Consistency of domain estimators has been considered for some time.

This work is based on earlier results,

- the general restriction (GR) estimator (*Knottnerus 2003*),
- the GR estimator elaborated for domains (*Sõstra 2007*, *Sõstra ja Traat 2009*).

These methods produce optimal GR estimators, but they use **unbiased** or **nearly unbiased** initial estimators.

We generalize the approach by allowing **biased** initial estimators.

Important remarks

- All estimators and their properties are elaborated in general and can be applied for any sampling design.
- Since domain estimation is a multivariate problem, then the matrix technique is used; there are vectors of estimators under consider. The accuracy of these vectors are measured with the MSE matrices.
- The design based approach is used.
- Sample sizes in domains are assumed to be not too small, i.e. small area estimation methods are not considered here.

Important remarks

- All estimators and their properties are elaborated in general and can be applied for any sampling design.
- Since domain estimation is a multivariate problem, then the matrix technique is used; there are vectors of estimators under consider. The accuracy of these vectors are measured with the MSE matrices.
- The design based approach is used.
- Sample sizes in domains are assumed to be not too small, i.e. small area estimation methods are not considered here.

Important remarks

- All estimators and their properties are elaborated in general and can be applied for any sampling design.
- Since domain estimation is a multivariate problem, then the matrix technique is used; there are vectors of estimators under consider. The accuracy of these vectors are measured with the MSE matrices.
- The design based approach is used.
- Sample sizes in domains are assumed to be not too small, i.e. small area estimation methods are not considered here.

Important remarks

- All estimators and their properties are elaborated in general and can be applied for any sampling design.
- Since domain estimation is a multivariate problem, then the matrix technique is used; there are vectors of estimators under consider. The accuracy of these vectors are measured with the MSE matrices.
- The design based approach is used.
- Sample sizes in domains are assumed to be not too small, i.e. small area estimation methods are not considered here.

Important remarks

- All estimators and their properties are elaborated in general and can be applied for any sampling design.
- Since domain estimation is a multivariate problem, then the matrix technique is used; there are vectors of estimators under consider. The accuracy of these vectors are measured with the MSE matrices.
- The design based approach is used.
- Sample sizes in domains are assumed to be not too small, i.e. small area estimation methods are not considered here.

Notations

Let $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)'$ be the parameter vector under study that satisfies linear restrictions:

$$\mathbf{R}\boldsymbol{\theta} = \mathbf{0}, \qquad (1)$$

where $\mathbf{R}$ is an $r \times k$ matrix of rank r.

- For the domain's case,

  $\mathbf{R} = (1, 1, ..., 1, -1) : 1 \times (D + 1)$ and $\boldsymbol{\theta} = (t_y^1, t_y^2, ..., t_y^D, t_y)'$,

  where D domain totals are $t_y^d = \sum_{U_d} y_i$, $d = 1, 2, ..., D$, and the population total is $t_y = \sum_U y_i$.

Notations

Let $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)'$ be the parameter vector under study that satisfies linear restrictions:

$$\mathbf{R}\boldsymbol{\theta} = \mathbf{0}, \qquad (1)$$

where $\mathbf{R}$ is an $r \times k$ matrix of rank r.

- For the domain's case,

  $\mathbf{R} = (1, 1, ..., 1, -1) : 1 \times (D+1)$ and $\boldsymbol{\theta} = (t_y^1, t_y^2, ..., t_y^D, t_y)'$,

  where D domain totals are $t_y^d = \sum_{U_d} y_i$, $d = 1, 2, ..., D$, and the population total is $t_y = \sum_U y_i$.

Let $\boldsymbol{\theta} = (\theta_1, ..., \theta_k)'$ be the parameter vector under study that satisfies linear restrictions:

$$\mathbf{R}\boldsymbol{\theta} = \mathbf{0}, \qquad (1)$$

where $\mathbf{R}$ is an $r \times k$ matrix of rank r.

- For the domain's case,

  $\mathbf{R} = (1, 1, ..., 1, -1) : 1 \times (D + 1)$ and $\boldsymbol{\theta} = (t_y^1, t_y^2, ..., t_y^D, t_y)'$,

  where D domain totals are $t_y^d = \sum_{U_d} y_i$, $d = 1, 2, ..., D$, and the population total is $t_y = \sum_U y_i$.

GR for the unbiased $\hat{\boldsymbol{\theta}}$

The GR estimator for the unbiased initial estimators $\hat{\theta} = (\hat{\theta}_1, ..., \hat{\theta}_k)'$ (Knottnerus, 2003) that satisfies (1) has the form

$$\hat{\theta}_{GR} = (\mathbb{I} - \mathbf{KR})\hat{\theta} \qquad (2)$$

with

$$\mathbb{C}\mathrm{ov}(\hat{\theta}_{GR}) = (\mathbb{I} - \mathbf{KR})\mathbf{V}, \qquad (3)$$

where

$$\mathbf{K} = \mathbf{VR}'(\mathbf{RVR}')^{-1} \qquad (4)$$

and $\mathbf{V} = \mathbb{C}\mathrm{ov}(\hat{\theta})$ such that $\mathbf{RVR}'$ can be inverted.

The estimator $\hat{\theta}_{GR}$ is **unbiased** and has the **minimum variance** property.

GR for the unbiased $\hat{\boldsymbol{\theta}}$

The GR estimator for the unbiased initial estimators $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, ..., \hat{\theta}_k)'$ (Knottnerus, 2003) that satisfies (1) has the form

$$\hat{\boldsymbol{\theta}}_{GR} = (\mathbb{I} - \mathbf{KR})\hat{\boldsymbol{\theta}} \qquad (2)$$

with

$$\mathbb{C}\text{ov}(\hat{\boldsymbol{\theta}}_{GR}) = (\mathbb{I} - \mathbf{KR})\mathbf{V}, \qquad (3)$$

where

$$\mathbf{K} = \mathbf{VR}'(\mathbf{RVR}')^{-1} \qquad (4)$$

and $\mathbf{V} = \mathbb{C}\text{ov}(\hat{\boldsymbol{\theta}})$ such that $\mathbf{RVR}'$ can be inverted.

The estimator $\hat{\theta}_{GR}$ is **unbiased** and has the **minimum variance** property.

GR for the unbiased $\hat{\boldsymbol{\theta}}$

The GR estimator for the unbiased initial estimators $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, ..., \hat{\theta}_k)'$ (Knottnerus, 2003) that satisfies (1) has the form

$$\hat{\boldsymbol{\theta}}_{GR} = (\mathbb{I} - \mathbf{KR})\hat{\boldsymbol{\theta}} \qquad (2)$$

with

$$\mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}}_{GR}) = (\mathbb{I} - \mathbf{KR})\mathbf{V}, \qquad (3)$$

where

$$\mathbf{K} = \mathbf{VR}'(\mathbf{RVR}')^{-1} \qquad (4)$$

and $\mathbf{V} = \mathbb{C}\mathrm{ov}(\hat{\boldsymbol{\theta}})$ such that $\mathbf{RVR}'$ can be inverted.

The estimator $\hat{\boldsymbol{\theta}}_{GR}$ is **unbiased** and has the **minimum variance** property.

- Knottnerus' GR estimator is **not optimal** with **biased** initial estimators.
- Three new GR estimators **handling bias** $b = \mathbb{E}(\hat{\theta}) - \theta$ are proposed.
- All of them satisfy the restriction equation $\mathbf{R}\hat{\theta}_{GR} = \mathbf{0}$.

- Knottnerus' GR estimator is **not optimal** with **biased** initial estimators.
- Three new GR estimators **handling bias** $\boldsymbol{b} = \mathbb{E}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}$ are proposed.
- All of them satisfy the restriction equation $\mathbf{R}\hat{\boldsymbol{\theta}}_{GR} = \mathbf{0}$.

- Knottnerus' GR estimator is **not optimal** with **biased** initial estimators.
- Three new GR estimators **handling bias** $\boldsymbol{b} = \mathbb{E}(\hat{\boldsymbol{\theta}}) - \boldsymbol{\theta}$ are proposed.
- All of them satisfy the restriction equation $\mathbf{R}\hat{\boldsymbol{\theta}}_{GR} = \mathbf{0}$.

Estimator GR1

Proposition 2.2

The estimator

$$\hat{\boldsymbol{\theta}}_{GR1} = (\mathbb{I} - \mathbf{KR})(\hat{\boldsymbol{\theta}} - \boldsymbol{b}), \qquad (5)$$

with $\mathbf{K} = \mathbf{VR}'(\mathbf{RVR}')^{-1}$ is unbiased for $\boldsymbol{\theta}$. Its variance is

$$\mathbb{C}\text{ov}(\hat{\boldsymbol{\theta}}_{GR1}) = (\mathbb{I} - \mathbf{KR})\mathbf{V}, \qquad (6)$$

and it is the optimal estimator among all linear estimators in $(\hat{\boldsymbol{\theta}} - \boldsymbol{b})$ that satisfy restriction equation (1).

Estimator GR2

Proposition 2.3

The estimator, satisfying restrictions (1), is

$$\hat{\theta}_{GR2} = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\hat{\theta}, \tag{7}$$

where $\mathbf{K}^* = \mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}$ and $\mathbf{M} = \mathbb{MSE}(\hat{\theta})$. The bias of the $\hat{\theta}_{GR2}$ is

$$\boldsymbol{b}(\hat{\theta}_{GR2}) = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\boldsymbol{b}, \tag{8}$$

and the mean square error matrix is

$$\mathbb{MSE}(\hat{\theta}_{GR2}) = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{M}. \tag{9}$$

Furthermore, $$\mathbb{MSE}(\hat{\theta}_{GR2}) \leq \mathbf{M} \tag{10}$$
in the sense of Löwner ordering.

Proposition 2.3

The estimator

$$\hat{\boldsymbol{\theta}}_{GR3} = (\mathbb{I} - \mathbf{K}^*\mathbf{R})(\hat{\boldsymbol{\theta}} - \boldsymbol{b}) \qquad (11)$$

with $\mathbf{K}^* = \mathbf{M}\mathbf{R}'(\mathbf{R}\mathbf{M}\mathbf{R}')^{-1}$ satisfies restrictions (1) and is unbiased for $\boldsymbol{\theta}$. It's MSE is

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR3}) = (\mathbb{I} - \mathbf{K}^*\mathbf{R})\mathbf{V}(\mathbb{I} - \mathbf{K}^*\mathbf{R})'. \qquad (12)$$

Furthermore,

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR3}) \leq \mathbf{M}. \qquad (13)$$

$$\hat{\boldsymbol{\theta}}_{GR1} = (\mathbb{I} - \mathsf{KR})(\hat{\boldsymbol{\theta}} - \boldsymbol{b}),$$

$$\hat{\boldsymbol{\theta}}_{GR2} = (\mathbb{I} - \mathsf{K}^*\mathsf{R})\hat{\boldsymbol{\theta}},$$

$$\hat{\boldsymbol{\theta}}_{GR3} = (\mathbb{I} - \mathsf{K}^*\mathsf{R})(\hat{\boldsymbol{\theta}} - \boldsymbol{b})$$

Proposition 2.4

The mean square error matrices of the restriction estimators $\hat{\boldsymbol{\theta}}_{GR1}$, $\hat{\boldsymbol{\theta}}_{GR2}$, $\hat{\boldsymbol{\theta}}_{GR3}$ and the initial estimator $\hat{\boldsymbol{\theta}}$ can be ordered (in the sense of Löwner ordering) as following:

$$\mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR1}) \leq \mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR3}) \leq \mathbb{MSE}(\hat{\boldsymbol{\theta}}_{GR2}) \leq \mathbb{MSE}(\hat{\boldsymbol{\theta}}). \qquad (14)$$

Remarks to GR1-GR3

- GR estimators are general:
  - applicable for any initial estimator-vector $\hat{\theta}$,
  - applicable for any restriction matrix $\mathbf{R}$ (not only summation),
  - applicable for any sampling design.

Remarks to GR1-GR3

- GR estimators are general:
  - applicable for any initial estimator-vector $\hat{\boldsymbol{\theta}}$,
  - applicable for any restriction matrix $\mathbf{R}$ (not only summation),
  - applicable for any sampling design.

Simulation study

Population:	Data of the healthcare personnel of Estonia, $N = 21764$
Study variables:	Hourly wage (continuous) Physician (binary)
Auxiliary information:	ID of healthcare institution Age Sex Education level (1..5) Domain indicator d ($d = 1, 2, 3, 4$)

Data description

The population is divided into 4 domains by the type of the healthcare institution:

Table: Population and domain sizes

Domain	no. of laborers	%
1	10863	49.9
2	6742	31.0
3	3139	14.4
4	1020	4.7
Population	21764	100

Data description

- Two sample designs were applied for the population frame, SI and MN.
- For both designs 5000 independent samples were drawn with $n = 400$.
- There is no any empty domain sample through simulations.

Data description

- Two sample designs were applied for the population frame, SI and MN.
- For both designs 5000 independent samples were drawn with $n = 400$.
- There is no any empty domain sample through simulations.

Data description

- Two sample designs were applied for the population frame, SI and MN.
- For both designs 5000 independent samples were drawn with $n = 400$.
- There is no any empty domain sample through simulations.

The choice of $\boldsymbol{\theta}$

The restriction matrix $\mathbf{R}$ and the restriction equation for the vector of true totals $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_P)'$, are:

$$\mathbf{R} = (1, 1, 1, 1, -1), \mathbf{R}\boldsymbol{\theta} = 0.$$

The initial estimator vector:

$$\hat{\boldsymbol{\theta}} = (\hat{t}^1_{greg\text{-}P}, \hat{t}^2_{greg\text{-}D}, \hat{t}^3_{syn\text{-}P}, \hat{t}^4_{syn\text{-}P}, \hat{t}_{greg})'.$$

The choice of $\boldsymbol{\theta}$

The restriction matrix $\mathbf{R}$ and the restriction equation for the vector of true totals $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_P)'$, are:

$$\mathbf{R} = (1, 1, 1, 1, -1), \mathbf{R}\boldsymbol{\theta} = 0.$$

The initial estimator vector:

$$\hat{\boldsymbol{\theta}} = (\hat{t}^1_{greg\text{-}P}, \hat{t}^2_{greg\text{-}D}, \hat{t}^3_{syn\text{-}P}, \hat{t}^4_{syn\text{-}P}, \hat{t}_{greg})'.$$

The choice of $\boldsymbol{\theta}$

The restriction matrix $\mathbf{R}$ and the restriction equation for the vector of true totals $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_P)'$, are:

$$\mathbf{R} = (1, 1, 1, 1, -1), \mathbf{R}\boldsymbol{\theta} = 0.$$

The initial estimator vector:

$$\hat{\boldsymbol{\theta}} = (\hat{t}^1_{greg\text{-}P}, \hat{t}^2_{greg\text{-}D}, \hat{t}^3_{syn\text{-}P}, \hat{t}^4_{syn\text{-}P}, \hat{t}_{greg})'.$$

The choice of $\boldsymbol{\theta}$

The restriction matrix **R** and the restriction equation for the vector of true totals $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_P)'$, are:

$$\mathbf{R} = (1, 1, 1, 1, -1), \mathbf{R}\boldsymbol{\theta} = 0.$$

The initial estimator vector:

$$\hat{\boldsymbol{\theta}} = (\hat{t}^1_{greg-P}, \hat{t}^2_{greg-D}, \hat{t}^3_{syn-P}, \hat{t}^4_{syn-P}, \hat{t}_{greg})'.$$

The choice of $\boldsymbol{\theta}$

The restriction matrix $\mathbf{R}$ and the restriction equation for the vector of true totals $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_P)'$, are:

$$\mathbf{R} = (1, 1, 1, 1, -1), \mathbf{R}\boldsymbol{\theta} = 0.$$

The initial estimator vector:

$$\hat{\boldsymbol{\theta}} = (\hat{t}^1_{greg\text{-}P}, \hat{t}^2_{greg\text{-}D}, \hat{t}^3_{syn\text{-}P}, \hat{t}^4_{syn\text{-}P}, \hat{t}_{greg})'.$$

The choice of $\boldsymbol{\theta}$

The restriction matrix **R** and the restriction equation for the vector of true totals $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_P)'$, are:

$$\mathbf{R} = (1, 1, 1, 1, -1), \mathbf{R}\boldsymbol{\theta} = 0.$$

The initial estimator vector:

$$\hat{\boldsymbol{\theta}} = (\hat{t}^1_{greg\text{-}P}, \hat{t}^2_{greg\text{-}D}, \hat{t}^3_{syn\text{-}P}, \hat{t}^4_{syn\text{-}P}, \hat{t}_{greg})'.$$

The choice of $\boldsymbol{\theta}$

The restriction matrix $\mathbf{R}$ and the restriction equation for the vector of true totals $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_P)'$, are:

$$\mathbf{R} = (1, 1, 1, 1, -1), \mathbf{R}\boldsymbol{\theta} = 0.$$

The initial estimator vector:

$$\hat{\boldsymbol{\theta}} = (\hat{t}^1_{greg\text{-}P}, \hat{t}^2_{greg\text{-}D}, \hat{t}^3_{syn\text{-}P}, \hat{t}^4_{syn\text{-}P}, \hat{t}_{greg})'.$$

Consistency of GR, SI

Estimator	Sample	Domain d				Population	$\mathbf{R}\hat{\theta} =$
		$d = 1$	$d = 2$	$d = 3$	$d = 4$		
Initial	1	2450.3	1428.3	884.1	178.3	4706.2	234.9
	2	2291.8	1878.7	1003.3	191.0	5160.7	204.1
	3	2777.1	1460.7	917.8	192.5	5019.6	328.6
GR1	1	2523.2	1459.8	544.0	151.2	4678.3	0.0
	2	2382.3	1917.8	662.2	163.7	5126.0	0.0
	3	2796.2	1469.0	580.7	166.4	5012.3	0.0
GR2	1	2425.6	1417.6	707.4	165.0	4715.7	0.0
	2	2270.2	1869.4	849.8	179.5	5168.9	0.0
	3	2742.5	1445.8	670.7	173.9	5032.9	0.0
GR3	1	2463.7	1434.1	643.6	159.6	4701.0	0.0
	2	2308.4	1885.9	785.9	174.1	5154.3	0.0
	3	2780.6	1462.3	606.8	168.6	5018.2	0.0

Consistency of GR, SI

Estimator	Sample	Domain d				Population	$\mathbf{R}\hat{\theta} =$
		$d=1$	$d=2$	$d=3$	$d=4$		
Initial	1	2450.3	1428.3	884.1	178.3	4706.2	234.9
	2	2291.8	1878.7	1003.3	191.0	5160.7	204.1
	3	2777.1	1460.7	917.8	192.5	5019.6	328.6
GR1	1	2523.2	1459.8	544.0	151.2	4678.3	0.0
	2	2382.3	1917.8	662.2	163.7	5126.0	0.0
	3	2796.2	1469.0	580.7	166.4	5012.3	0.0
GR2	1	2425.6	1417.6	707.4	165.0	4715.7	0.0
	2	2270.2	1869.4	849.8	179.5	5168.9	0.0
	3	2742.5	1445.8	670.7	173.9	5032.9	0.0
GR3	1	2463.7	1434.1	643.6	159.6	4701.0	0.0
	2	2308.4	1885.9	785.9	174.1	5154.3	0.0
	3	2780.6	1462.3	606.8	168.6	5018.2	0.0

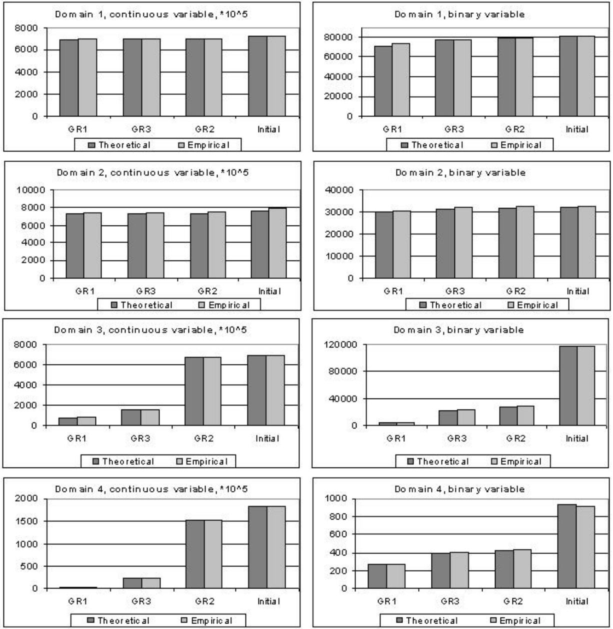
Consistency of GR, MN

Estimator	Sample	Domain d				Population	$\mathbf{R}\hat{\theta} =$
		$d = 1$	$d = 2$	$d = 3$	$d = 4$		
Initial	1	2666.0	1676.8	952.5	178.1	4883.7	589.7
	2	2419.3	1423.5	931.1	169.9	4725.0	218.8
	3	2674.4	1241.9	908.7	164.7	4604.8	384.8
GR1	1	2533.9	1613.3	622.7	154.0	4923.8	0.0
	2	2502.3	1463.4	591.2	142.9	4699.8	0.0
	3	2661.1	1235.5	573.3	139.0	4608.8	0.0
GR2	1	2600.0	1645.1	513.6	145.0	4903.7	0.0
	2	2394.8	1411.7	768.3	157.6	4732.4	0.0
	3	2631.3	1221.1	622.4	143.1	4617.9	0.0
GR3	1	2640.5	1664.6	446.9	139.5	4891.4	0.0
	2	2435.3	1431.2	701.5	152.1	4720.1	0.0
	3	2671.8	1240.6	555.6	137.5	4605.6	0.0

Consistency of GR, MN

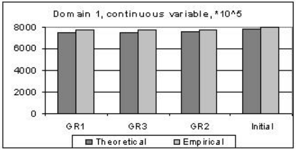
Estimator	Sample	Domain d				Population	$\mathbf{R}\hat{\theta} =$
		$d = 1$	$d = 2$	$d = 3$	$d = 4$		
Initial	1	2666.0	1676.8	952.5	178.1	4883.7	589.7
	2	2419.3	1423.5	931.1	169.9	4725.0	218.8
	3	2674.4	1241.9	908.7	164.7	4604.8	384.8
GR1	1	2533.9	1613.3	622.7	154.0	4923.8	0.0
	2	2502.3	1463.4	591.2	142.9	4699.8	0.0
	3	2661.1	1235.5	573.3	139.0	4608.8	0.0
GR2	1	2600.0	1645.1	513.6	145.0	4903.7	0.0
	2	2394.8	1411.7	768.3	157.6	4732.4	0.0
	3	2631.3	1221.1	622.4	143.1	4617.9	0.0
GR3	1	2640.5	1664.6	446.9	139.5	4891.4	0.0
	2	2435.3	1431.2	701.5	152.1	4720.1	0.0
	3	2671.8	1240.6	555.6	137.5	4605.6	0.0

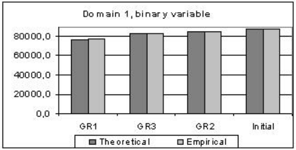
MSEs, SI case

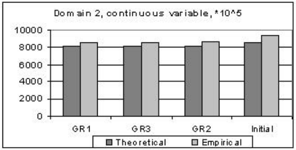


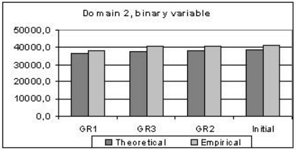
Domain 1, continuous variable, *10^5
8000
6000
4000
2000
0
GR1
GR3
GR2
Initial
Theoretical
Empirical
Domain 1, binary variable
80000
60000
40000
20000
0
GR1
GR3
GR2
Initial
Theoretical
Empirical
Domain 2, continuous variable, *10^5
10000
8000
6000
4000
2000
0
GR1
GR3
GR2
Initial
Theoretical
Empirical
Domain 2, binary variable
40000
30000
20000
10000
0
GR1
GR3
GR2
Initial
Theoretical
Empirical
Domain 3, continuous variable, *10^5
8000
6000
4000
2000
0
GR1
GR3
GR2
Initial
Theoretical
Empirical
Domain 3, binary variable
120000
80000
40000
0
GR1
GR3
GR2
Initial
Theoretical
Empirical
Domain 4, continuous variable, *10^5
2000
1500
1000
500
0
GR1
GR3
GR2
Initial
Theoretical
Empirical
Domain 4, binary variable
1000
800
600
400
200
0
GR1
GR3
GR2
Initial
Theoretical
Empirical

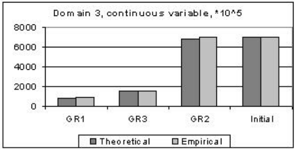
MSEs, MN case

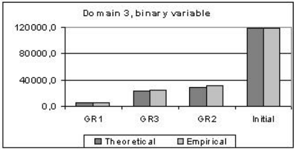

Domain 1, continuous variable, *10^5
8000
6000
4000
2000
0
GR1
GR3
GR2
Initial
Theoretical
Empirical

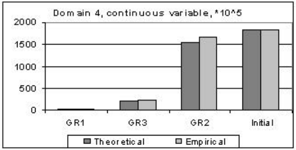

Domain 1, binary variable
80000,0
60000,0
40000,0
20000,0
0,0
GR1
GR3
GR2
Initial
Theoretical
Empirical

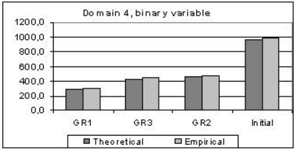

Domain 2, continuous variable, *10^5
10000
8000
6000
4000
2000
0
GR1
GR3
GR2
Initial
Theoretical
Empirical


Domain 2, binary variable
50000,0
40000,0
30000,0
20000,0
10000,0
0,0
GR1
GR3
GR2
Initial
Theoretical
Empirical


Domain 3, continuous variable, *10^5
8000
6000
4000
2000
0
GR1
GR3
GR2
Initial
Theoretical
Empirical


Domain 3, binary variable
120000,0
80000,0
40000,0
0,0
GR1
GR3
GR2
Initial
Theoretical
Empirical


Domain 4, continuous variable, *10^5
2000
1500
1000
500
0
GR1
GR3
GR2
Initial
Theoretical
Empirical


Domain 4, binary variable
1200,0
1000,0
800,0
600,0
400,0
200,0
0,0
GR1
GR3
GR2
Initial
Theoretical
Empirical

Final conclusion

- Three GR estimators satisfying linear restrictions, generalized to allow biased initial estimators, were proposed.
- The mean square error (MSE) and the bias expressions for the GR estimators were derived and studied.
- The ordering of the GR estimators was established, they all were more accurate than the initial estimator.

Final conclusion

- Three GR estimators satisfying linear restrictions, generalized to allow biased initial estimators, were proposed.
- The mean square error (MSE) and the bias expressions for the GR estimators were derived and studied.
- The ordering of the GR estimators was established, they all were more accurate than the initial estimator.

Final conclusion

- Three GR estimators satisfying linear restrictions, generalized to allow biased initial estimators, were proposed.
- The mean square error (MSE) and the bias expressions for the GR estimators were derived and studied.
- The ordering of the GR estimators was established, they all were more accurate than the initial estimator.

Final conclusion

- Three GR estimators satisfying linear restrictions, generalized to allow biased initial estimators, were proposed.
- The mean square error (MSE) and the bias expressions for the GR estimators were derived and studied.
- The ordering of the GR estimators was established, they all were more accurate than the initial estimator.

Final conclusion

- Three GR estimators satisfying linear restrictions, generalized to allow biased initial estimators, were proposed.
- The mean square error (MSE) and the bias expressions for the GR estimators were derived and studied.
- The ordering of the GR estimators was established, they all were more accurate than the initial estimator.

THANK YOU!