Some properties of the ridge regression estimator with survey data

Muhammad Ahmed Shehzad

(in collaboration with Camelia Goga and Hervé Cardot)
IMB, Université de Bourgogne-Dijon,
Muhammad-Ahmed.Shehzad@u-bourgogne.fr
camelia.goga@u-bourgogne.fr
herve.cardot@u-bourgogne.fr

Journée de sondage Dijon 2010

Outline of talk

1. Brief introduction of the estimation of population totals
2. Use of Auxiliary Information in Surveys
3. Principles of ridge regression in statistics
4. Ridge regression with survey data
5. Simulation study

Population, sample.

- a finite population $U = \{1, \ldots, i, \ldots, N\}$
- a sample $s \in \mathcal{S}$ and $s \subset U$
- the sampling design $p(s)$: a probability distribution on the set $\mathcal{S}$; $p(s)$ is controlled by the statistician.
- the inclusion probabilities
  - of first degree : $\pi_i = Pr(i \in s)$
  - of second degree : $\pi_{ij} = Pr(i, j \in s)$ for $i \neq j$ and $\pi_{ii} = \pi_i$

The Horvitz-Thompson Estimator

- the study variable $\mathcal{Y}$,
  $y_i =$ the value of $\mathcal{Y}$ for the i-th individual,
- we know y_i for $i \in s$
- we want to estimate the population total of $\mathcal{Y}$, namely

$$t_y = \sum_U y_i$$

The Horvitz-Thompson Estimator

- the study variable $\mathcal{Y}$,
  $y_i =$ the value of $\mathcal{Y}$ for the i-th individual,
- we know y_i for $i \in s$
- we want to estimate the population total of $\mathcal{Y}$, namely

$$t_y = \sum_U y_i$$

- The Horvitz-Thompson (HT) estimator :

$$\hat{t}_{HT} = \sum_{i \in s} \frac{y_i}{\pi_i} = \sum_{i \in U} \frac{y_i}{\pi_i} I_i$$

$$I_i = \mathbf{1}_{\{i \in s\}} \quad \text{the sample membership indicator}$$

Properties

The estimator HT for a total is

- the HT variance is

$$V(\hat{t}_{HT}) = \sum_{i \in U} \sum_{j \in U} (\pi_{ij} - \pi_i \pi_j) \frac{y_i}{\pi_i} \frac{y_j}{\pi_j}$$

- with its estimator

$$\hat{V}(\hat{t}_{HT}) = \sum_{i \in s} \sum_{j \in s} \frac{\pi_{ij} - \pi_i \pi_j}{\pi_{ij}} \frac{y_i}{\pi_j} \frac{y_j}{\pi_j}$$

Drawbacks :

1. The HT estimator contains little auxiliary information (the π_k) !
2. The variance as well as its estimator contain double sums.

Regression Model and Auxiliary Information

- Think to improve the estimation ?
  Use auxiliary information
- We consider the superpopulation model,$\xi : y = X\beta + \varepsilon$
  $\mathbf{X}_1, ..., \mathbf{X}_p$; we denote by $X = (\mathbf{X}_1, ..., \mathbf{X}_p)$
- $E_\xi(\varepsilon) = 0_N$,
- The error terms ε's are independent of each other.
- $\mathbf{Var}(\varepsilon_i) = \sigma^2\mathbf{V}_{(N\times p)}$
  where $\mathbf{V}_{(N\times p)}$ is a diagonal matrix.

Two estimation approaches

1. **"model assisted" (MA)** : we construct the estimator based on the sampling design and assisted by the super-population model (Särndal, Swensson & Wretman 1992)

$$\hat{t}_{MA} = \sum_s \frac{y_i}{\pi_i} - \left(\sum_s \frac{\mathbf{x}_i'}{\pi_i} - \sum_U \mathbf{x}_i' \right) \beta.$$

Two estimation approaches

1. **"model assisted" (MA)** : we construct the estimator based on the sampling design and assisted by the super-population model (Särndal, Swensson & Wretman 1992)

$$\hat{t}_{MA} = \sum_s \frac{y_i}{\pi_i} - \left(\sum_s \frac{\mathbf{x}_i'}{\pi_i} - \sum_U \mathbf{x}_i' \right) \beta.$$

2. **"model bassed" (MB)** : we predict the population total by using the super-population model without taking into account the sampling design (Royall & Cumberland 1978)

$$\hat{t}_{MB} = \sum_s y_i + \sum_{\bar{s}} \mathbf{x}_i' \beta$$

Both $\hat{t}_{MA}$ and $\hat{t}_{MB}$ rely on the estimation of the β.

Calibration Estimation in Survey Sampling

In survey sampling, we use,

- sampling weights are adjusted to make certain estimators matching with known population totals.
- Let the Horvitz-Thompson estimator of the population total be

$$\hat{t}_{HT} = \sum_{s} d_i y_i$$

  with $d_i = \frac{1}{\pi_i}$ are called design weights because they are obtained from sampling design.
- Let x_i is an auxiliary variable $(x_i)_{i \in s}$ with $X = \sum_U x_i$ then it is possible that

$$\sum_{s} d_i x_i \neq X$$

Calibration Estimation in Survey Sampling

- The class of calibration estimators, calibrated to X, is the class of the estimators of the form

$$\hat{t}_w = \sum_s w_i y_i$$

where weight w_i satisfies

$$\sum_s w_i x_i = X$$

- The weight w_i are allowed to be the function of x_i but not of y_i.

Calibration Estimation in Survey Sampling

- The calibration weights (w_i) minimize some distance $(G(w_i, d_i))$ to the Horvitz-Thompson weights (d_i) (Sarndal 2007).
- Chi-square distance $\sum_s \frac{w_i - d_i}{2 d_i q_i}$ is one such example of distance where q_i is suitably chosen positive scale factor. Normally used $q_i = 1$ for all i. So the task is to

$$\text{minimize} \sum_s \frac{w_i - d_i}{2 d_i q_i}$$

subject to

$$\sum_s w_i x_i = \sum_U x_i$$

Ordinary Least Square and Ridge Estimators

Ordinary least square (OLS) estimator of the regression coefficient β

$$\hat{\beta}_{OLS} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

- what if $\mathbf{X}'\mathbf{X}$ is singular ?
- Hoerl and Kennard (1962) proposed Ridge Estimator of β

$$\hat{\beta}_k = (\mathbf{X}'\mathbf{X} + k\mathbf{I}_p)^{-1}\mathbf{X}'\mathbf{y}, \; k > 0$$

- $\hat{\beta}_k$ is a biased estimator of β

Advantages of the Ridge Estimator

- Serves as a solution to the data problems such as
  - Multicollinearity (Linear correlation among the auxiliary variables)
  - Ill-conditioned data(Near correlation among $\mathbf{X}$ cause $\mathbf{X}'\mathbf{X}$ to be singular $(\mathbf{X}'\mathbf{X})^{-1} = \infty$ resulting larger error and Instability in the data)
- k is added to the diagonal of the data matrix $\mathbf{X}'\mathbf{X}$ to control ill-conditioning or orthogonality
- k is called the ridge parameter or the biasing parameter of the estimator

Some Remarks about the Ridge Estimation

1. $\hat{\beta}_k$ minimizes the residual sum of squares plus a penalty.
2. The variance term is a decreasing function of (an increasing) value of k. That is, the variance decreases with the increase in k. (so as $k \to \infty$, the variance term goes to zero.)
3. $Bias(\hat{\beta}_k)$ decreases with the decrease in the value of k. Hence $(\text{Bias})^2 \to 0$ as $k \to 0$. The squared bias is a continuous, monotonically increasing function of k. (Hoerl and Kennard 1970)
4. The squared bias term dominates the $MSE(\hat{\beta}_k)$ for large values of k. (Izenman2008,)
5. Such a value of k always exists which makes $\text{MSE}(\hat{\beta}_R)$ smaller than $\text{MSE}(\hat{\beta})$. (Existence theorem by Hoerl and Kennard 1970).

Important : The reduction of variance can be achieved through ridge regression by imposing a penalty on the norm of the linear relationship (between covariates $\mathbf{x_i}$ and the study variable y) and then finding a balance between the bias and variance by adjusting the regularization parameter (k).

Ridge regression for model-based estimators

- The sample s selected from population U such that $s \subset U$
- The matrix $\mathbf{X}$ and $\mathbf{y}$ are known only for the sample s
- Let $\mathbf{X}_s = (\mathbf{x}_i')_{i \in s}, \mathbf{y}_s = (\mathbf{y}_i)_{i \in s}$, and
  - $Var_{\xi}(\varepsilon_s) = \sigma^2 \mathbf{V_s}$,
  - $\mathrm{Var}_{\xi}(\varepsilon_{\overline{s}}) = \sigma^2 \mathbf{V}_{\overline{s}}$

    with $\overline{s} = U - s$(non-sampled elements of population), where $\mathbf{V}_s$ is the sample variance matrix with some constant at the diagonal and
    - $\varepsilon_s = (\varepsilon_i)_{i \in s}$
    - $\varepsilon_{\overline{s}} = (\varepsilon_i)_{i \in U - s}$

Model-Based Ridge Estimation of Total

We proceed using the classical method to calculate the estimate of β.

- Under a model-based setting, $\hat{\beta}_{BLUE}$ (Royall, 1970) is the solution of the equation

$$(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)\beta = \mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{y}_s$$

  so that we have,assuming that $(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)$ is non-singular

$$\hat{\beta}_{BLUE} = (\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{y}_s$$

- The ridge estimator of β is

$$\hat{\beta}_{MB,R} = \left(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s + k\mathbf{I}_p\right)^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{y}_s$$

- Thus the MB ridge estimator of the population total becomes,

$$\hat{t}_{MB,R} = \sum_s y_i + \left(\sum_{\bar{s}} \mathbf{x}_i'\right)\hat{\beta}_{MB,R}$$

Model-Assisted Ridge Estimation of Total

- Under the model-assisted approach, the estimate of the regression coefficient β is the solution of the following design-based equation

$$(\mathbf{X}_s'\mathbf{\Pi}_s^{-1}\mathbf{X}_s)\beta = \mathbf{X}_s'\mathbf{\Pi}_s^{-1}\mathbf{y}_s$$

  so that we have if $(\mathbf{X}_s'\mathbf{\Pi}_s^{-1}\mathbf{X}_s)^{-1}$ exists,

$$\hat{\beta}_{MA} = (\mathbf{X}_s'\mathbf{\Pi}_s^{-1}\mathbf{X}_s)^{-1}\mathbf{X}_s'\mathbf{\Pi}_s^{-1}\mathbf{y}_s$$

  where $\mathbf{\Pi}_s = \text{diag}(\pi_k)_{k\in s}$

- The MA ridge estimator of β becomes

$$\hat{\beta}_{MA,R} = \left(\mathbf{X}_s'\mathbf{\Pi}_s^{-1}\mathbf{X}_s + k\mathbf{I}_p\right)^{-1}\mathbf{X}_s'\mathbf{\Pi}_s^{-1}\mathbf{y}_s$$

- The MA ridge estimator of the population total takes the shape

$$\hat{t}_{MA,R} = \sum_s \frac{y_i}{\pi_i} - \left(\sum_s \frac{\mathbf{x}_i'}{\pi_i} - \sum_U \mathbf{x}_i'\right)\hat{\beta}_{MA,R}.$$

Weighted Estimators

- Ridge estimation was first used in survey sampling to eliminate negative or extremely large weights obtained when a too strict condition of unbiasedness is imposed
- Calibrated weights (w_i) taking disproportionate values as compared to the initial weights (d_i) means that either we are calibrating too many population totals, or the sample is particularly unbalanced with respect to the auxiliary variables
- A common practice is to relax some of the calibration constraints using ridge type estimators.

Weighted Estimators (1)

- The goal : find $\mathbf{w} = (w_k)_{k \in s}$ and build

$$t_w = \sum_s w_k y_k$$

- We have two approaches for the calculation of weights :
  1. model-based (MB)
  2. model-assited (MA)
- Model-Based Approach (Bardsley and Chambers, 1984)
- We look for the weights $\mathbf{w} = (w_k)_{k \in s}$ which minimize the MSE_ξ among the class of biased estimators with a bounded bias

$$\mathbf{w}_{MB,R} = \text{argmin}_{\mathbf{w}} (\mathbf{w} - \mathbf{h})' \mathbf{V}_s^{-1} (\mathbf{w} - \mathbf{h}) + \mathbf{B}'\mathbf{C}\mathbf{B}$$

where $\mathbf{B} = \sum_s w_k \mathbf{x}_k - \sum_U \mathbf{x}_k$:the ξ-bias of $\hat{t}_w$
$\mathbf{C}$:diagonal cost matrix and $\mathbf{h} = \mathbf{1_s}$ is the unit vector.

Weighted Estimators (2)

$$\hat{t}_w = \mathbf{w'}_{MB,R}\mathbf{y}_s = \sum_s y_i + \left(\sum_{\bar{s}} \mathbf{x}_i'\right)\hat{\beta}_{w,R}$$

where the weight $\mathbf{w}_{MB,R}$ is as follows,

$$\mathbf{w}_{MB,R} \;=\; h + \mathbf{V}^{-\mathbf{1}}\mathbf{X_s}\left(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s + \mathbf{C}^{-1}\right)^{-1}\left(\sum_{\bar{s}} \mathbf{x}_i'\right)$$

and

$$\hat{\beta}_{w,R} \;=\; \left(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s + C^{-1}\right)^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{y}_s$$

- For $\mathbf{C}^{-1} = k\mathbf{I}_p$, we obtain $\hat{t}_{MB,R}$

Weighted Estimators (3)

- Model-Assisted Approach (Rao and Singh ,1997, 2009) We find weights $\mathbf{w}$ by minimizing the following,

$$\mathbf{w}_{MA,R} = \text{argmin}_{\mathbf{w}}(\mathbf{w}-\mathbf{d})'\boldsymbol{\Pi}_s^{-1}(\mathbf{w}-\mathbf{d}) + \mathbf{B}'\mathbf{D}\mathbf{B}$$

where $\mathbf{d} = \left(\frac{1}{\pi_1},...,\frac{1}{\pi_s}\right)$ and $\boldsymbol{\Pi}_s^{-1} = diag\left(\frac{1}{\pi_1},...,\frac{1}{\pi_s}\right)$.

- the weight $\mathbf{w}_{MA,R}$ is as follows,

$$\mathbf{w_{MA,R}} = \mathbf{d} + \left(\Pi_s^{-1}\mathbf{X}'_s\right)\left(\mathbf{X}'_s\Pi_s^{-1}\mathbf{X}_s + \mathbf{D}^{-1}\right)^{-1}\left(\mathbf{X}'_U\mathbf{1}_U - \mathbf{X}'\mathbf{d}\right)$$

- for $\mathbf{D}^{-1} = k\mathbf{I}_p$, we obtain $\hat{t}_{MA,R}$

Weighted Estimators (4)

- A relationship between $\hat{t}_R$,$\hat{t}_{HT}$ and $\hat{t}_{GREG}$ can be written as,

$$\begin{aligned}\hat{t}_{w,MA,R} &= \mathbf{y'}_s\mathbf{d} + (\mathbf{1'}_U\mathbf{X}_s - \mathbf{d'}\mathbf{X}_s)\left(\mathbf{X'}_s\Pi_s^{-1}\mathbf{X}_s + \mathbf{D}^{-1}\right)^{-1}\left(\mathbf{X'}_s\Pi_s^{-1}\mathbf{y}_s\right) \\ &= (1-\alpha)\hat{t}_{HT} + \alpha\hat{t}_{GREG}\end{aligned}$$

where $\hat{t}_{GREG} = \mathbf{y'}\mathbf{d} + (\mathbf{X'}_U\mathbf{1}_U - \mathbf{X'}_s\mathbf{d})'\left(\mathbf{X'}_s\Pi_s^{-1}\mathbf{X}_s\right)^{-1}\left(\mathbf{X'}_s\Pi_s^{-1}\mathbf{y}_s\right)$ and $\hat{t}_{HT} = \mathbf{y'}\mathbf{d}$. α is given by,

$$\begin{aligned}\alpha &= \left(\left(\mathbf{X'}_s\Pi_s^{-1}\mathbf{X}_s + \mathbf{D}^{-1}\right)^{-1}\left(\mathbf{X'}_s\Pi_s^{-1}\mathbf{y}_s\right)\right) \\ &\quad \left(\left(\mathbf{X'}_s\Pi_s^{-1}\mathbf{X}_s\right)^{-1}\left(\mathbf{X'}_s\Pi_s^{-1}\mathbf{y}_s\right)\right)^{-1}\end{aligned}$$

- Note also that,
  1. if $k \to 0$, this implies that all calibration constraints are satisfied and $\hat{t}_{w,MA,R}$ tends towards $\hat{t}_{GREG}$.
  2. if $k \to \infty$, then $\hat{t}_{w,MA,R}$ tends towards $\hat{t}_{HT}$.

Partially penalized estimators

Suppose that $\mathbf{X}$ is partitioned into two sets $\mathbf{X}_A$ and $\mathbf{X}_B$,

$$\mathbf{X} = (\mathbf{X}_A, \mathbf{X}_B)$$

and we want that the weights $\mathbf{w}$ satisfy **exactly** the calibration equations on $\mathbf{X}_A$:

$$\mathbf{w}'\mathbf{X}_{A,s} = \mathbf{1}'_U\mathbf{X}_A$$

and that are penalized on $\mathbf{X}_B$:

$$|\mathbf{w}'\mathbf{X}_{B,s} - \mathbf{1}'_U\mathbf{X}_B| \leq \mathbf{c}$$

Two ways to obtain such an estimator

- In a model-based approach, Bardsley and Chambers (1984) suggest taking the cost matrix $\mathbf{C}^{-1}$ as follows :

$$\mathbf{C}^{-1} = \begin{pmatrix} \mathbf{C_A}^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_B}^{-1} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C_B}^{-1} \end{pmatrix}$$

in the minimization problem

$$\min_w (\mathbf{w}-\mathbf{h})'\mathbf{V_s}^{-1}(\mathbf{w}-\mathbf{h}) + (\mathbf{w}'\mathbf{X}_s - \mathbf{1}'_U\mathbf{X})\mathbf{C}(\mathbf{w}'\mathbf{X}_s - \mathbf{1}'_U\mathbf{X})'$$

where $\mathbf{h} = \mathbf{1}'_s$.
So, we have an infinite cost matrix $\mathbf{C_A}$ which implies discarding the constraint.

- In a design-based approach, Guggemos and Tillé (2010) suggest considering the following minimization problem :

$$\min_w (\mathbf{w}-\mathbf{d})'\mathbf{\Pi_s}^{-1}(\mathbf{w}-\mathbf{d}) + (\mathbf{w}'\mathbf{X}_{B,s} - \mathbf{1}'_U\mathbf{X_B})\mathbf{C}(\mathbf{w}'\mathbf{X}_{B,s} - \mathbf{1}'_U\mathbf{X}_B)'$$

subject to

$$\mathbf{w}'\mathbf{X}_{A,s} = \mathbf{1}'_U\mathbf{X}_A.$$

Properties of the Ridge Estimator (Model Based)

Result

The ξ-bias of $\hat{t}_{MB,R}$ is given by,

$$Bias_{\xi}\left(\hat{t}_{MB,R}\right) = -k\left(\sum_{U-s}\mathbf{x}_i'\right)\mathbf{P}' diag\left(\frac{1}{\lambda_i + k}\right)\mathbf{P}\beta$$

- with $\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s = \mathbf{P}'(diag\,\lambda_i)\mathbf{P} = \mathbf{P}'\Lambda\mathbf{P}$ where P is a diagonal matrix such that $\mathbf{P}'\mathbf{P} = \mathbf{I}$ and
- $\mathbf{X}_s = \mathbf{V}_s^{\frac{1}{2}}\mathbf{M}\Lambda^{\frac{1}{2}}P'$ where $\mathbf{M}$ is an $n \times p$ matrix of coordinates of the observations along the principal axes of $\mathbf{X}$, standardized in the sense that, $\mathbf{M}'\mathbf{M} = \mathbf{I}$

Properties of the Ridge Estimator (Model Based)

- The relationship between the $\hat{\beta}_{MB,R}$ and $\hat{\beta}_{OLS}$ is as follows,

$$\hat{\beta}_{MB,R} = \mathbf{P}'(diag(\delta_i))\mathbf{P}\hat{\beta}_{OLS}$$

- $\delta_i = \frac{\lambda_i}{\lambda_i + k}$ is the *shrinkage factor*.
- Declining δ_i for increasing k and strictly declining eigenvalues which means that the so-called *shrinkage factor* shrinks coefficient vector for the declining eigenvalues given the fact that $k \in (0, \infty)$.

Properties of the Ridge Estimator (Model Based)

Result

The MSE_ξ of $\hat{t}_{MB,R}$ is given by,

$$E_\xi(\hat{t}_{MB,R} - t_y)^2 = Var_\xi(\hat{t}_{MB,R} - t_y) + (Bias_\xi(\hat{t}_{MB,R}))^2 \text{ where}$$

$$Bias_\xi\left(\hat{t}_{MB,R}\right) = -k\left(\sum_{\bar{s}} \mathbf{x}_i'\right)\mathbf{P}'diag\left(\frac{1}{\lambda_i + k}\right)\mathbf{P}\beta$$

$$Var_\xi(\hat{t}_{MB,R} - t_y) = \sigma^2 \sum_{\bar{s}} v_i + k^2\left(\left(\sum_{\bar{s}} \mathbf{x}_i'\right)\mathbf{P}'diag\left(\frac{1}{\lambda_i + k}\right)\mathbf{P}\beta\right)^2$$

Properties of the Ridge Estimator (Model Assisted)

Result

The $bias_p$ *of* $\hat{t}_{MA,R}$ *is*

$$Bias_p(\hat{t}_{MA,R}) = -Trace\left(Cov_p\left(\sum_s \frac{\mathbf{x}_i}{\pi_i}, \hat{\beta}_{MA,R}\right)\right)$$

Result

The $bias_\xi$ *of* $\hat{t}_{MA,R}$ *is given by,*

$$Bias_\xi(\hat{t}_{MA,R}) = -k\left(\sum_U \mathbf{x}_i' - \sum_s \frac{\mathbf{x}_i'}{\pi_i}\right)\left(\mathbf{X}_s'\mathbf{\Pi}_s^{-1}\mathbf{X}_s + k\mathbf{I}_p\right)^{-1}\beta$$

Properties of the Ridge Estimator (Model Assisted)

Result

Under mild assumptions, the asymptotic variance of $\hat{t}_{MA,R}$,

$$
\begin{aligned}
AV_p(\hat{t}_{MA,R}) &= Var_p\left(\sum_s \frac{y_i - \mathbf{x}_i'\hat{\beta}}{\pi_i}\right) \\
&= \sum_s \sum_s \Delta_{ij}\left(\frac{y_i - \mathbf{x}_i'\hat{\beta}}{\pi_i}\right)\left(\frac{y_j - \mathbf{x}_j'\hat{\beta}}{\pi_j}\right)
\end{aligned}
$$

where,

$$
\hat{\beta} = (\mathbf{X}_U'\mathbf{X}_U + k\mathbf{I_p})^{-1}\mathbf{X}_U'\mathbf{y}_U
$$

The Mediametrie Data for Household Media Users

- The study population : U of $N = 5977$ individuals watching a channel on Monday of the first two weeks of September 2010.
- The study parameter

$$t_y = \sum_U y_i$$

- The true value is $t_y = 230315.8$ (min).
- The variable $\mathcal{Y}$ has many zeros (about 30%).

The Mediametrie Data for Household Media Users

- Auxiliary information :
  - insee : French Regional Code (22 regions divided into 6 geographic locations N,SO,SE,E,Ou,P)
  - sexe : Gender of the individual
  - csp : Profession of the individual
  - age : Age of the individual
  - *chaine.lundi*.1 : Duration of the audience watching a Channel on Monday in the first Week of September 2010
  - Data matrix $\mathbf{X}$: $(5977, 19)$

The Mediametrie Data for Household Media Users

- The eigenvalues of the matrix $\mathbf{X}'\mathbf{X}$ are :

$$\lambda_{max} = 1.220006 \cdot 10^7, \quad 9.437714 \cdot 10^3, \quad 3.543169 \cdot 10^3, \quad \ldots$$

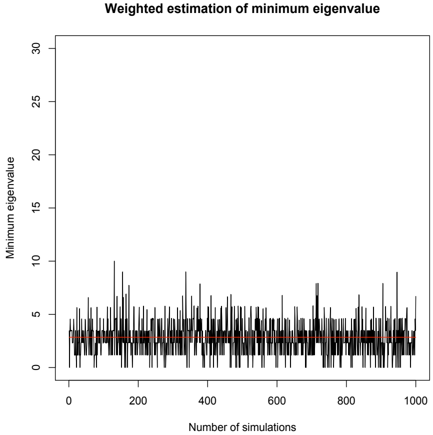
$$188.9848, \quad 143.8786, \quad 114.8046, \quad 91.57, \quad \lambda_{min} = 2.8468$$

  The minimum eigenvalue is very close to zero.
- The conditioning number $\frac{\lambda_{max}}{\lambda_{min}} = 4285535$ is very large, so the matrix $\mathbf{X}'\mathbf{X}$ is ill-conditioned.

Estimation

- We select a sample s of size $n = 500$ according to a simple random sampling without replacement from U.
- We consider $n.sim = 1000$ simulations.
- The GREG estimator happens not to work (the matrix $\mathbf{X_s}'\mathbf{\Pi}_s\mathbf{X_s}$ has sometimes λ_{min} equal to zero).

Plot of the smallest eigenvalue of $\mathbf{X}_s' \mathbf{\Pi}_s \mathbf{X}_s$ with simulation

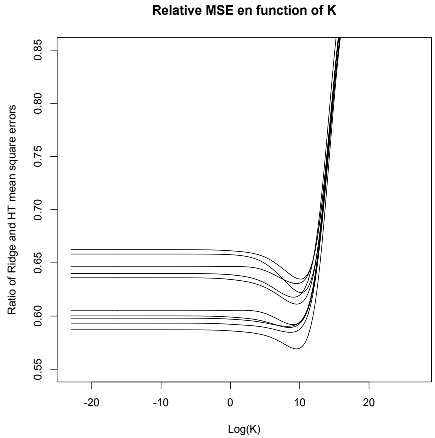


Comparison between $\hat{t}_{ridge}$ and $\hat{t}_{HT}$

- relative bias :

$$RB = \frac{\sum_{j=1}^{1000} \hat{\theta}^{(j)}/1000 - t_y}{t_y}$$

  which is less than 0.2%.
- ratio between the MSE of $\hat{t}_{ridge}$ and that of $\hat{t}_{HT}$.



We repeated 10 times the simulation from above.

- for small values of K, the gain is important ;
- for large values of K, the estimator $\hat{t}_{ridge}$ is similar to $\hat{t}_{HT}$.

Follow Through

- Confidence Intervals (C-I)
- Algorithms for calculating k