

Simulation Comparison Between an Outlier Resistant Model-Based Finite Population Estimator and Design-Based Estimators under Contamination

Bogong Li

U.S. Bureau of Labor Statistics

There are two approaches to finite population estimation. One assumes the finite population elements are fixed quantities. The randomness associated with estimators comes from the random selection of samples. Each sampled element has a weight determined by the sample design, see Cochran (1977) [2]. Most Bureau of Labor Statistics surveys rely on this theory, for example, the Occupational Employment Statistics program which was discussed in depth in Li (2002) [10] about methods used in small area estimation. The other approach assumes the population elements are a random draw from a larger population or “super-population”, in a way similar to taking random samples from a random variable. The random variable has its own mean and variance structure, usually expressed in terms of linear models. Finite population summaries of the variable under investigation, such as mean, total are estimated through fitting sample data to the linear model. The design of the sample selection is relevant only to the selection of a suitable linear model under the “super-population”, but is irrelevant to how the estimates are derived from the model. The first approach is known as “design-based”, the second “model-based”. Särndal, Swensson and Wretman (1992) [14] discusses this distinction in greater detail. Table 1 lists some familiar design-based estimators and their model-based equivalents. A broader review on this connection is in Li (2001) [9]. Since least squares linear estimators are very vulnerable to outlying observations, that is, a small variation in outlying observation produces larger variation in the estimate than non-outlying observations, we desire a model-based estimator that is less responsive to outliers while being efficient. This is especially necessary for survey data since processing error occur prominently in surveys. This study aims to investigate the statistical quality of such a model-based, outlier insensitive finite population estimator

we propose. Limited simulation result shows that this estimator is as efficient as other model-based estimators when there is no significant violation from model assumptions and is more efficient than least squares types of estimators when there are contaminating symmetrical outlying observations.

1. Outlier-Insensitive Linear Regression

In the past several decades, research on influence of outlying observations on linear model estimates yielded classes of estimators with controllable levels of reaction to outlying observations. We briefly summarize some of them here, in particularly the ones that are implemented in popular statistical software packages such as R, S-plus and SAS. These are the M-estimator, least median of squares estimator (LMS), the least trimmed means estimator (LTS). Also included are other methods that are applicable to more complex linear model structures, the Mallows’, Schweppe’s methods, and the broader class of the Generalized M-estimator (GM).

- Huber M-Estimator for Regression: Estimator $\hat{\beta}_n$ for the regression parameter β (location) is defined by

$$\min_{\hat{\beta}_n} \sum_{i=1}^n \rho((y_i - \mathbf{x}'_i \hat{\beta}_n)/\sigma)$$

or

$$\sum_{i=1}^n \psi((y_i - \mathbf{x}'_i \hat{\beta})/\sigma) \mathbf{x}_i = \mathbf{0}$$

for some symmetric function $\rho : \mathbb{R} \rightarrow \mathbb{R}^+$ and for a fixed σ . $\psi(t)$ is the derivative of $\rho(t)$. Examples of $\rho(t)$ are probability density function $f(t)$, t^2 , $|t|$ etc. defined within certain distance from 0. Optimal $\hat{\beta}$ is selected to minimize asymptotic variance while allowing as high a breakdown points as it could be possible, see Huber (1981) [7] for more details.

The views expressed in this article are those of the author and do not constitute the policy of the Bureau of Labor Statistics. The author can be contacted at: U.S. Bureau of Labor Statistics, Postal Square Building, Suite 1950, 2 Massachusetts Ave. NE, Washington, DC 20212-0001, or E-Mail: li.t@bls.gov.

| Estimators | Finite Population \hat{T} | Model Equivalent |
|--------------------------------|--|--|
| Expansion estimator | NY_s | $Y_i = \mu + \epsilon_i, \epsilon_i \sim (0, \sigma^2)$ |
| Linear regression estimator | $N[\bar{Y}_s + \hat{b}_{ls}(\bar{x} - \bar{x}_s)]$ | $Y_i = \alpha + x_i\beta + \epsilon_i, \epsilon_i \sim (0, \sigma^2)$ |
| Ratio estimator | $N\bar{Y}_s\bar{x}/\bar{x}_s$ | $Y_i = x_i\beta + x_i^{\frac{1}{2}}\epsilon_i, \epsilon_i \sim (0, \sigma^2)$ |
| Stratified expansion estimator | $\sum_h N_h \bar{Y}_{hs}$ | $Y_{hi} = \mu_h + \epsilon_{hi}, \epsilon_{hi} \sim (0, \sigma_h^2)$ |
| Stratified ratio estimator | $\sum_h N_h \bar{Y}_{hs} \bar{x}_h / \bar{x}_{hs}$ | $Y_{hi} = x_{hi}\beta + x_{hi}^{\frac{1}{2}}\epsilon_{hi}, \epsilon_{hi} \sim (0, \sigma_h^2)$ |
| Simple Proportion Estimator | $\sum_s Y_i/n$ | $Y_i \sim \text{ber}(p_i), \text{logit}(p_i) = \mathbf{x}'_i\boldsymbol{\beta}$ |

Table 1: Finite population estimators and their model-based equivalents

- Least Median of Squares Regression (LMS): Regression parameters $\hat{\boldsymbol{\beta}}$ are implicitly defined by

$$\min_{\hat{\boldsymbol{\beta}}} \text{Med} \left\{ (y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}})^2, i = 1, 2, \dots, n \right\},$$

where only the median of the squared residuals are relevant in defining the best $\hat{\boldsymbol{\beta}}$, see Rousseeuw and Leroy (1987) [13] for details.

- Least Trimmed Squares Regression(LTS): Similar to least squares estimators, but only the smallest h residuals are taken to minimize summed squared residuals,

$$\min_{\hat{\boldsymbol{\beta}}} \sum_{i=1}^h (y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}})_{i:n}^2 = \min_{\hat{\boldsymbol{\beta}}} \sum_{i=1}^h (r^2)_{i:n}$$

, where $(r^2)_{1:n} \leq \dots \leq (r^2)_{h:n} \leq \dots \leq (r^2)_{n:n}$ are the ordered squared residuals, see Rousseeuw and Leroy (1987) [13] for more details on this.

- Mallows' Method: A weight $w(\mathbf{x}_i)$ is part of the estimating equation in Huber M-estimator which limits the influence of outlying \mathbf{x} rows,

$$\sum_{i=1}^n w(\mathbf{x}_i) \psi((y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}^{(M)})/\sigma) \mathbf{x}_i = \mathbf{0},$$

see Mallows (1975) [11].

- Scheppe's Method: Weights $w(\mathbf{x}_i)$ depending on \mathbf{x}_i are as well attached to scale the residual in response to outlying influential observation in \mathbf{x} :

$$\sum_{i=1}^n w(\mathbf{x}_i) \psi((y_i - \mathbf{x}'_i \hat{\boldsymbol{\beta}}^{(S)})/(\sigma w(\mathbf{x}_i))) \mathbf{x}_i = \mathbf{0}$$

, see Hill (1982) [6].

- Generalized M-estimator for Mixed Models: Estimating equations written in matrix form to accommodating more complex covariance structure and weight matrices \mathbf{V} , \mathbf{W} and \mathbf{U} . This is

the method we used to obtain outlier-insensitive model parameter estimates,

$$\min_{\hat{\boldsymbol{\beta}}, \hat{\mathbf{V}}} \left\{ \mathbf{W}^{-1/2} \psi(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \right\}' \mathbf{U}^{-1/2'} \hat{\mathbf{V}}^{-1} \mathbf{U}^{-1/2} \left\{ \mathbf{W}^{-1/2} \psi(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}) \right\}, \quad (1)$$

see Krasker and Welsch (1982) [8]. Solutions for $\hat{\boldsymbol{\beta}}$ and $\hat{\mathbf{V}}$ need numerical algorithm similar to that for ML and REML, see Harville (1977) [5] and Searle, Casella and McCulloch (1992) [15] for methods to obtain ML and REML.

2. Outlier-Resistant Best Linear Unbiased Estimation

Linear model with mixed-effects is gradually popularizing with survey data usage, for example Battese, Hater and Fuller (1987) [1]. We propose a super-population structure incorporates mixed-effects:

$$\begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \mathbf{Y}_g \end{bmatrix}_{N \times 1} = \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \\ \vdots \\ \mathbf{x}_g \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{z}_1 & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}_2 & \dots & \mathbf{0} \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{z}_g \end{bmatrix} \begin{bmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_g \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_g \end{bmatrix},$$

where

$$\begin{cases} \nu_i \stackrel{\text{iid}}{\sim} (0, \sigma_\nu^2) \\ \epsilon_i \stackrel{\text{iid}}{\sim} (0, \sigma_\epsilon^2 \mathbf{I}) \end{cases}$$

and \mathbf{Y}_i and \mathbf{x}_i are population and auxiliary information vectors. ν_i is random (can be constant) stratum-specific effects. ϵ_i is independent random error vector. If we rearrange the Y by observed (s) and non-observed (r) population elements:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_s \\ \mathbf{Y}_r \end{bmatrix} = \begin{bmatrix} \mathbf{X}_s \\ \mathbf{X}_r \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \mathbf{Z}_s \\ \mathbf{Z}_r \end{bmatrix} \boldsymbol{\nu} + \boldsymbol{\epsilon}',$$

and

$$\text{Var}(\mathbf{Y}) = \begin{bmatrix} \mathbf{V}_{ss} & \mathbf{V}_{sr} \\ \mathbf{V}_{rs} & \mathbf{V}_{rr} \end{bmatrix}, \quad \text{then}$$

the best linear unbiased predictor (BLUP) of

$$T = \mathbf{I}'_{N \times 1} \mathbf{Y} = \mathbf{I}'_s \mathbf{Y}_s + \mathbf{I}'_r \mathbf{Y}_r$$

which could be the mean, total or contrasts, is

$$\tilde{T} = \mathbf{1}'_s \mathbf{y}_s + \mathbf{1}'_r [\mathbf{X}_s \hat{\boldsymbol{\beta}}_{\text{LS}} + \hat{\mathbf{V}}_{rs} \hat{\mathbf{V}}_{ss}^{-1} (\mathbf{y}_s - \mathbf{X}_s \hat{\boldsymbol{\beta}}_{\text{LS}})]. \quad (2)$$

Valliant (2000) [17] provides detailed derivation and proofs of unbiasedness and variance minimization of \tilde{T} . We propose an outlier resistant RBLUP, based on above formulation:

$$\begin{aligned} \tilde{T}^{(R)} = & \mathbf{1}'_s \mathbf{y}_s + \mathbf{1}'_r \left[\mathbf{W}_s \mathbf{X}_s \hat{\boldsymbol{\beta}}^{(R)} \right. \\ & \left. + \hat{\mathbf{V}}_{rs}^{(R)} \hat{\mathbf{V}}_{ss}^{(R)-1} \psi(\mathbf{y}_s - \mathbf{X}_s \hat{\boldsymbol{\beta}}^{(R)}) \right] \quad (3) \end{aligned}$$

where $\hat{\boldsymbol{\beta}}^{(R)}$, $\hat{\mathbf{V}}^{(R)}$ are any bounded influence estimators from the mixed-effects linear model; \mathbf{W}_s are weight functions depend on \mathbf{X}_s , ψ is the Huber ψ -function. In practice, $\hat{\boldsymbol{\beta}}^{(R)}$, $\hat{\mathbf{V}}^{(R)}$ are generalized M-estimators given in section 1. Strictly speaking RBLUP is neither “best” in the sense of minimizing variance nor unbiased with respect to the model among all linear unbiased estimators. The name given simply indicates its predecessor.

Influence of RBLUP is necessarily bounded. By definition, *influence function* of a statistical functional $T(F) \in \mathbb{R}^k$ is the first derivative of $T(F)$:

$$\text{IF}_{T,F}(z) := \lim_{\epsilon \rightarrow 0} \frac{T(F_\epsilon) - T(F)}{\epsilon} = \left. \frac{\partial T(F_\epsilon)}{\partial \epsilon} \right|_{\epsilon \rightarrow 0},$$

where $F_\epsilon := (1 - \epsilon)F + \epsilon \delta_z$, $\epsilon \in \mathbb{R}^1$, with δ_z the point mass at z . $\text{IF}_{T,F}(z)$ reflects the influence on T of an infinitesimal change occurred at z . Bounded influence estimators retain finite influence from outlying observations as well as any other observations. A sample version of the influence function is

$$\hat{\text{IF}}_i := (n - 1) \left\{ T(\hat{F}) - T(\hat{F}_{(i)}) \right\},$$

where $\hat{F}_{(i)}$ is the sample functional without i th observation. *Cook's Distance* is another form that expresses the magnitude of $\hat{\text{IF}}_i$ by calculating a weighted norm:

$$D_i = (\hat{\text{IF}}_i)' M (\hat{\text{IF}}_i) / c$$

where M is $k \times k$ positive or semipositive definite matrix and c is a scalar to be defined, see Hampel, Ronchetti, Rousseeuw, and Stahel (1986) [4] for greater details.

The influence function of RBLUP, $\text{IF}_{\text{RBLUP},F}$ is then necessarily bounded. Now if we rewrite RBLUP as

$$\tilde{T}^{(R)} = \eta(\hat{\boldsymbol{\tau}}),$$

where $\hat{\boldsymbol{\tau}}$ is the bounded influence estimator (with $\text{IF}_{\hat{\boldsymbol{\tau}},F} < \infty$) of the vector of all unknown parameters in the model,

$$\boldsymbol{\tau} = (\boldsymbol{\beta}', v_{11}, \dots, v_{ij}, \dots, v_{NN})'.$$

then $\hat{\boldsymbol{\tau}}$ is asymptotically normally distributed, Maronna and Yohai (1981) [12] discusses its asymptotic mean and variance; η is a finite projection from $\hat{\boldsymbol{\tau}}$ to $\tilde{T}^{(R)}$. Given such projection exists, by the *Chain Rule* defined on Gâteaux differentials, the influence function of RBLUP is

$$\text{IF}_{\text{RBLUP},F} = \eta' \text{IF}_{\hat{\boldsymbol{\tau}},F}.$$

Since both $\text{IF}_{\hat{\boldsymbol{\tau}},F}$ and η' are finite, $\text{IF}_{\text{RBLUP},F}$ is necessarily bounded.

3. Simulation Study

We study the statistical quality of proposed RBLUP through a simulation. First we produce an artificial population according to the model:

$$\left\{ \begin{array}{l} \begin{pmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \mathbf{Y}_3 \\ \mathbf{Y}_4 \end{pmatrix} = (\mathbf{1}_{1000} \quad \mathbf{x}) \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} \\ + \begin{pmatrix} \mathbf{1}_{50} & \cdots & \mathbf{0} \\ & \mathbf{1}_{150} & \vdots \\ \vdots & & \mathbf{1}_{250} \\ \mathbf{0} & \cdots & & \mathbf{1}_{550} \end{pmatrix} \begin{pmatrix} \nu_1 \\ \vdots \\ \nu_4 \end{pmatrix} + \begin{pmatrix} \epsilon_1 \\ \vdots \\ \vdots \\ \vdots \\ \epsilon_{1000} \end{pmatrix}, \\ \left\{ \begin{array}{l} \nu_k \stackrel{\text{iid}}{\sim} (0, \sigma_\nu^2) \\ \epsilon_i \stackrel{\text{iid}}{\sim} (0, \sigma_\epsilon^2) \end{array} \right., \\ \text{Var}(\mathbf{Y}_k) = \begin{pmatrix} \sigma_\epsilon^2 + \sigma_\nu^2 & \sigma_\epsilon^2 & \cdots & \sigma_\epsilon^2 \\ \sigma_\epsilon^2 & \sigma_\epsilon^2 + \sigma_\nu^2 & \cdots & \sigma_\epsilon^2 \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_\epsilon^2 & \sigma_\epsilon^2 & \cdots & \sigma_\epsilon^2 + \sigma_\nu^2 \end{pmatrix} \\ \nu_k, e_i \stackrel{\text{taken from}}{\iff} (1 - \lambda)\mathcal{N}(0, 1) + \lambda\mathcal{N}(0, 41) \end{array} \right.$$

Then contamination is introduced into this population in four ways: no contamination, contamination in the random effects, in the error term and in both. Contamination of percentage λ are taken from normal distribution with large variance (40 vs. 1 of clean data). Notice in this case the contamination is symmetrically centered at the central location of clean data. This approach is similar to de Jongh, Wet and Welsh (1988) [3].

| | ν_k | e_i | |
|-----------|---------|-------|--|
| λ | 0.0 | 0.0 | : no contamination |
| | 0.1 | 0.0 | : only random effect contamination |
| | 0.0 | 0.1 | : only error contamination |
| | 0.1 | 0.1 | : both random effect and error contamination |

$\boldsymbol{\beta}$ and \mathbf{V} are estimated through Generalized M-Estimator, as those used by Krasker and Welsh (1982) [8] and Stahel and Welsh (1992) [16].

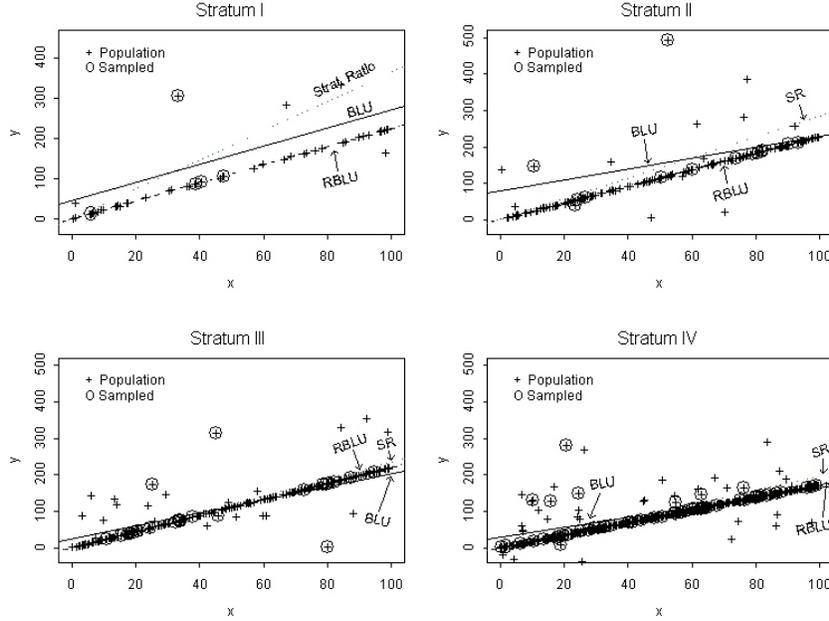


Figure 1: Fitted super-population models under three estimation methods (Contamination Patter: 0.1:0.1)

We then repeatedly draw stratified random samples of size 100 from this artificial population, 100 times total. Sample sizes are allocated to four strata first by the method of sample proportional to size then the Neyman allocation. Results show little difference between the two methods therefore all results listed later are based on Neyman allocation method. For each draw, we calculate an estimate of the population total using RBLUP. In addition we also calculate estimates using the following alternative estimators:

1. Stratified Direct estimator (SD),

$$\hat{T}_{SD} = 1/N \sum_{i=1}^4 N_i/n_i \sum_{j=1}^{n_i} y_{ij}$$

with standard deviation

$$SD(\hat{T}_{SD}) = \left(\sum_{i=1}^4 (N_i/N)^2 (1 - f_i)/n_i \right)^{1/2}$$

, where S_i^2 is the stratum variance and f_i the sampling fraction in stratum i .

2. Stratified Linear Regression estimator (SLR),

$$\hat{T}_{SLR} = \sum_{i=1}^4 N_i/N \sum_{j=1}^{n_i} y_{ij}/n_i - \sum_{i=1}^4 \hat{b}_i N_i/N \left(\sum_{j=1}^{n_i} x_{ij}/n_i - \sum_{j=1}^{N_i} x_{ij}/N_i \right),$$

where

$$\hat{b}_i = \frac{\sum_{j=1}^{n_i} (y_{ij} - \sum_{j=1}^{n_i} y_{ij}/n_i)(x_{ij} - \sum_{j=1}^{n_i} x_{ij}/n_i)}{\sum_{j=1}^{n_i} x_{ij}/n_i - \sum_{i=1}^4 \left(x_{ij} - \sum_{j=1}^{n_i} x_{ij}/n_i \right)^2}$$

, with standard deviation

$$SD(\hat{T}_{SLR}) = \left(\sum_{i=1}^4 (1 - f_i) S_{y_i}^2 (1 - \rho_i^2) \right)^{1/2}.$$

$\rho_i = S_{xyi}/S_{xi}S_{yi}$ is the population correlation in stratum i , S_{yi}^2 , S_{xi}^2 , and S_{xyi}^2 are variance of \mathbf{y} , \mathbf{x} and covariances between \mathbf{y} and \mathbf{x} in stratum i , and

3. Best Linear Unbiased estimator (BLUP) (2).

The estimates were then compared to the actual population total yielding values of root mean squared error, a measure for deviation from the true value. Spread of the deviation were estimated through the estimates spread based on the 100 independent repetitions. The median root mean square errors and a nominal 95% confidence region based on the 100 independent samples are listed in Table 2. Figure 1 displays the model fits from one of the random draws. Different dash lines indicate the method used. Notice both the SD and BLU are “pulled” in larger degree than RBLUP by the outliers in the sample. RBLUP tend to stay closer to the majority of the population in all four strata when outlier in \mathbf{y} directions is mild, as in stratum IV where all estimates produces similar model fits.

4. Conclusion

- ✓ Outlier-resistant predictor RBLUP under linear mixed-effects model is less influenced by outliers (in \mathbf{y} or rows of \mathbf{X}) than LS predictors such as BLUP.
- ✓ When there is no significant deviations from model assumptions and/or there are not outlying \mathbf{y} or rows of \mathbf{X} , RBLUP is relatively efficient.
- ✓ Given there is no contamination and presence of asymmetric outlying observations, RBLUP is biased. However it reduces variability as compared to other estimators.

| Contamination | | Root Mean Squared Error | | | |
|---------------|------|-------------------------|-------------|-------------|-------------|
| Pattern | T | SD | SLR | $BLUP$ | $RBLUP$ |
| 0.0 : 0.0 | 95.2 | 0.65 | 1.30 | 1.31 | 1.31 |
| | | (0.61-0.68) | (0.54-1.61) | (0.79-1.78) | (0.79-1.78) |
| 0.1 : 0.0 | 95.2 | 1.23 | 0.97 | 1.18 | 1.18 |
| | | (0.59-7.83) | (0.12-1.79) | (.05-2.31) | (.05-2.31) |
| 0.0 : 0.1 | 95.2 | 5.41 | 2.68 | 2.72 | 1.61 |
| | | (3.29-7.97) | (1.12-3.62) | (1.25-3.25) | (0.81-2.76) |
| 0.1 : 0.1 | 95.2 | 5.12 | 3.20 | 2.89 | 1.52 |
| | | (1.98-7.25) | (1.75-4.21) | (1.60-4.25) | (0.16-2.86) |

Table 2: Median root mean squared error and 95% confidence region from 100 independent samples

- ✓ Under contamination and presence of outlying observations, RBLUP reduces bias and standard error.
- ✓ Differences between RBLUP and BLUP decrease as the sampling rate increases.
- ✓ Effect of the outlying observation on the RBLUP depends on the outlier resistant estimators used.
- ✓ Computation is slow, algorithm complex, solution is not guaranteed especially with large datasets.

References

- [1] Battese, G. E., R. M. Harter and W. A. Fuller (1987), "An Error-Components Model for Prediction of County Crop Areas using Survey and Satellite Data," *Journal of the American Statistical Association*, 83, 28-36.
- [2] Cochran, W. G. (1977), *Sampling Techniques*, 3rd edition. Wiley, New York, 3rd edition, 1977.
- [3] de Jongh, P. J., de Wet, T., and Welsh, A. H. (1988), "Mallow-Type Bounded-Influence-Regression Trimmed Means," *Journal of the American Statistical Association*, 83, 805-810.
- [4] Hampel, F. R., Ronchetti, E. M., Rousseeuw, P.J. and Stahel, W. A. (1986), *Robust Statistics, the Approach Based on Influence Functions*. Wiley, New York, 1986.
- [5] Harville, D. A. (1977), "Maximum Likelihood Approaches to Variance Component Estimation and to Related Problems," *Journal of the American Statistical Association*, 72, 320-338.
- [6] Hill, R. W. (1982), "Robust Regression When There are Outliers in the Carriers: The Univariate Case," *Communications in Statistics, Part A-Theory and Methods*, 11, 849-868.
- [7] Huber, P.J. (1981) *Robust Statistics*, Wiley, New York, 1981.
- [8] Krasker, W. S. and Welsch, R. E. (1982), "Efficient Bounded Influence Regression Estimation," *Journal of the American Statistical Association*, 77, 595-604.
- [9] Li, B. (2001) *Basic Sample Survey Principles & Methods*. SSA Course Note on Survey Sampling Methodology, Sacramento, CA.
- [10] Li, B. (2002) "Small Area Research for the Occupational Employment Statistics Survey," *2002 Proceedings of the American Statistical Association*, Government Statistics Section: 2091-2096 [CD-ROM], Alexandria, VA: American Statistical Association.
- [11] Mallows, C. L. (1975), *On Some Topics in Robustness*, Unpublished memorandum, Bell Telephone Laboratories, Murray Hill, NJ.
- [12] Maronna, R. A., and Yohai, V. J. (1981), "Asymptotic Behaviour of General M -Estimates for Regression and Scale with Random Carriers," *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete/Probability and Related Fields*, 58, 7-20.
- [13] Rousseeuw, P. J. and A. M. Leroy (1987), *Robust Regression and Outlier Detection*. Wiley, New York, 1987.
- [14] Särndal, C. E., Swensson, E. and Wretman, J. (1992), *Model Assisted Survey Sampling*.
- [15] Searle, S. R., Casella, G. and McCulloch, C. E. (1992), *Variance Components*. Wiley, New York, 1992.
- [16] Stahel, W. A., Welsh, A. H. (1992), "Robust Estimation of Variance Components," *Research Report 69*, ETH, Zürich.
- [17] Valliant, R., A. H. Dorfman and R.M. Royall (2000), *Robust Regression and Outlier Detection*. Wiley, New York, 2000.