Robust Linear Prediction in Finite Populations: A Bayesian Perspective
Author(s): Heleno Bolfarine, Carlos Alberto de Braganca Pereira and Josemar Rodrigues
Source: *Sankhyā: The Indian Journal of Statistics, Series B (1960-2002)*, Vol. 49, No. 1 (Apr., 1987), pp. 23-35
Published by: Springer on behalf of the Indian Statistical Institute
Stable URL: http://www.jstor.org/stable/25052477
Accessed: 06/06/2013 12:44

Sankhyā : The Indian Journal of Statistics
1987, Volume **49**, Series B, Pt. 1, pp. 23–35.

ROBUST LINEAR PREDICTION IN FINITE POPULATIONS: A BAYESIAN PERSPECTIVE

By HELENO BOLFARINE,

CARLOS ALBERTO DE BRAGANCA PEREIRA

and

JOSEMAR RODRIGUES

Universidade de São Paulo, Brasil

SUMMARY. In this article, the multiple regression model is used to describe relationships among quantities associated to finite population units. Postulating normal priors for the regressor parameters and for the error vector, after observing a sample, a posterior distribution for the unsampled part of the population is obtained. The case of noninformative priors is covered as a limit of the normal priors. We describe the general conditions under which omission of additional auxiliary regression variables does not affect the posterior prediction. Some standard situa tions are discussed under this Bayesian approach. A general class of predictors suggested by such robustness conditions is considered and some well known predictors (the ratio estimator for example) are shown to be elements of this class, proving that there are situations where they are robust predictors. This paper may be considered as a Bayesian version of Pereira and Rodrigues (1983) justifying and unifying some results of Royall and Pfeffermann (1982). The concept of robustness treated here follows in Barlow and Wu's (1981).

1. Introduction

This article deals with a finite population $P = \{1, \ldots, N\}$, of N (known) identifiable units. For every unit k of P, there are associated $M+1$ quantities, $y_k, x_{k1}, \ldots, x_{kM}$, where all but y_k are known. A matrix of order $N \times M$ and rank M whose row k $(k = 1, \ldots, N)$ is the row vector $\boldsymbol{X}_k' = (x_{k1}, \ldots, x_{kM})$ is represented by $\boldsymbol{X}$. The column vector of unknown quantities is represented by $\boldsymbol{y}$ whose transpose is $\boldsymbol{y}' = (y_1, \ldots, y_N)$.

In order to describe his uncertainty about $\boldsymbol{y}$ and his prior information about the relationships among $\boldsymbol{y}$ and $\boldsymbol{X}$, a Bayesian, called here Bayesian 1, considers the model

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

AMS (1980) *subject classificagion*: 62D05, 62F15.

Key words and phrases : Bayesian linear prediction, weak robustness set, robustness set, balanced sample, shrinkage factor, class of Bayesian robust predictors.

where (i) $\boldsymbol{\epsilon}$ is a multivariate normal random vector of order N with null mean vector, represented by $E_1\{\boldsymbol{\epsilon}\} = \phi$, and nonsingular covariance matrix, denoted by $V_1\{\epsilon\} = \boldsymbol{V}$,

(ii) $\boldsymbol{\beta}$ is a multivariate normal random vector of order M with mean vector $E_1\{\boldsymbol{\beta}\} = \boldsymbol{b}$ and nonsingular covariance matrix $V_1\{\boldsymbol{\beta}\} = \boldsymbol{B}$.

To gain information about a linear function of $\boldsymbol{y}$, say $\boldsymbol{l}'\boldsymbol{y}$, the statistician, using a sampling plan, selects from P a sample S. Since the data $\boldsymbol{D} = \{(k, y_k); k \in S\}$ is a minimal sufficient statistic (Basu, 1969), S is considered as a subset of P (not as a sequence of elements of P). The effective sample size (the number of elements of S) is represented by n which may take different values for distinct samples (Cassel, Särndal and Wretman, 1977). Note that $\boldsymbol{D}$ is a set of the n elements of S together with their associated quantities y_k's. As in Pereira and Rodrigues (1983), i_k denotes the indicator function that indicates whether k ($k = 1, \ldots, N$) belongs to S; that is, $i_k = 1$ if $k \in S$ and $i_k = 0$ if $k \in P-S$. The distribution of the random vector $(i_1, \ldots, i_N)$ is called sampling design. This article is only devoted to noninformative sampling designs—the vector $(i_1, \ldots, i_N)$ has the same distribution for every fixed $\boldsymbol{y}$. Consequently, by the conditionality principle (Basu, 1975), the sampling design is irrelevant for inferences about $\boldsymbol{y}$ and $(i_1, \ldots, i_N)$ may be considered as known.

Some important aspects of the Bayesian inference are highlighted in the next section. Section 3 introduces the main result of this article which presents a general condition for the Bayesian robustness. Some standard situations are discussed, under this robustness considerations, in Section 4. A class of predictors satisfying the robustness condition is introduced in Section 5. This class includes the ratio estimator which proves that there are situations where this predictor is robust.

2. Bayesian prediction

The quantity of interest, $\boldsymbol{l}'\boldsymbol{y}$, may be partitioned as

$$\boldsymbol{l}'\boldsymbol{y} = \boldsymbol{l}'\boldsymbol{I}_S\boldsymbol{y} + \boldsymbol{l}'(\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{y}$$

where $\boldsymbol{I}_S$ is a diagonal matrix of order N with its k-th diagonal element being i_k ($k = 1, \ldots, N$) and $\boldsymbol{I}$ is the identity matrix of order N. It is clear that, after $\boldsymbol{D}$ (the data) has been observed, $\boldsymbol{l}'\boldsymbol{I}_S\boldsymbol{y}$ becomes known and the part of $\boldsymbol{l}'\boldsymbol{y}$ that remains unknown is $\boldsymbol{l}'(\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{y}$. Hence, a Bayesian naturally describes his uncertainty about $\boldsymbol{l}'(\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{y}$, the unobserved part of $\boldsymbol{l}'\boldsymbol{y}$, by the conditional (posterior) distribution of $\boldsymbol{l}'(\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{y}$ given $\boldsymbol{I}_S\boldsymbol{y}$. As in Pereira and Rodrigues (1983), $\boldsymbol{I}_S\boldsymbol{y}$ is replacing $\boldsymbol{D}$ as data representation.

For simplicity, the following notation is used in the sequel:

$$\boldsymbol{y}_1 = \boldsymbol{I}_S\boldsymbol{y},\ \boldsymbol{y}_2 = (\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{y},\ \boldsymbol{X}_1 = \boldsymbol{I}_S\,\boldsymbol{X},\ \boldsymbol{X}_2 = (\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{X},$$

$$\boldsymbol{V}_1 = \boldsymbol{I}_S\boldsymbol{V}\boldsymbol{I}_S,\ \boldsymbol{V}_2 = (\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{V}(\boldsymbol{I}-\boldsymbol{I}_S),$$

and

$$\boldsymbol{V}_{12} = \boldsymbol{V}_{21}' = \boldsymbol{I}_S\boldsymbol{V}(\boldsymbol{I}-\boldsymbol{I}_S).$$

It is interesting to notice that $\boldsymbol{y}_1'\boldsymbol{y}_2 = \boldsymbol{0}$ and $\boldsymbol{V}_1\boldsymbol{V}_2 = \boldsymbol{\Phi}$ where $\boldsymbol{\Phi}$ is the square null matrix (here of order N). Restricting the study to the case where $\boldsymbol{X}_1$ has rank M and representing the generalized inverse of a square matrix $\boldsymbol{A}$ by $\boldsymbol{A}^-$ and, if $\boldsymbol{A}$ is non singular, the ordinary inverse by $\boldsymbol{A}^{-1}$, the elements below complete the notation

$$\boldsymbol{A}^{-1} = (\boldsymbol{X}_1'\boldsymbol{V}_1^-\boldsymbol{X}_1),$$

$$\boldsymbol{C} = (\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\boldsymbol{A}^{-1},\ \boldsymbol{C}_0 = (\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\boldsymbol{B}^{-1}$$

and

$$\hat{\boldsymbol{\beta}} = \boldsymbol{A}\boldsymbol{X}_1\boldsymbol{V}_1^-\boldsymbol{y}_1.$$

Note that $\boldsymbol{C}+\boldsymbol{C}_0$ is the identity matrix of order M and $\hat{\boldsymbol{\beta}}$ is the classical Gauss-Markov estimator of $\boldsymbol{\beta}$.

The following result introduces the element of work, the posterior distribution, of Bayesian 1.

Theorem 1 : *For Bayesian* 1, *the conditional* (*posterior*) *distribution of* $\boldsymbol{y}_2$ *given* $\boldsymbol{y}_1$ (*the data*) *is multivariate normal with mean vector*

$$E_1\{\boldsymbol{y}_2\,|\,\boldsymbol{y}_1\} = \boldsymbol{X}_2\bar{\boldsymbol{\beta}}+\boldsymbol{V}_{21}\boldsymbol{V}_1^-(\boldsymbol{y}_1-\boldsymbol{X}_1\bar{\boldsymbol{\beta}})$$

and covariance matrix

$$\boldsymbol{V}_1\{\boldsymbol{y}_2\,|\,\boldsymbol{y}_1\} = (\boldsymbol{V}_2-\boldsymbol{V}_{21}\boldsymbol{V}_1^-\boldsymbol{V}_{12})+(\boldsymbol{X}_2-\boldsymbol{V}_{21}\boldsymbol{V}_1^-\boldsymbol{V}_{12})\boldsymbol{C}\boldsymbol{A}(\boldsymbol{X}_2-\boldsymbol{V}_{21}\boldsymbol{V}_1^-\boldsymbol{V}_{12})'$$

where $\bar{\boldsymbol{\beta}} = E_1\{\boldsymbol{\beta}\,|\,\boldsymbol{y}_1\} = \boldsymbol{C}\hat{\boldsymbol{\beta}}+\boldsymbol{C}_0\boldsymbol{b}$ *is the Bayesian estimator of* $\boldsymbol{\beta}$.

Proof: To conclude that $\boldsymbol{C}\hat{\boldsymbol{\beta}}+\boldsymbol{C}_0\boldsymbol{b}$ is the Bayes estimator of $\boldsymbol{\beta}$, apply the lemma of Lindley and Smith (1972) to the model $\boldsymbol{y}_1 = \boldsymbol{X}_1\boldsymbol{\beta}+\boldsymbol{I}_S\boldsymbol{\epsilon}$.

To complete the proof, consider the joint distribution of $\boldsymbol{y}_1$ and $\boldsymbol{y}_2$, apply result 8a.2.11 of Rao (1973) to obtain the conditional distribution of $\boldsymbol{y}_2\,|\,(\boldsymbol{y}_1,\ \boldsymbol{\beta})$, and finally use the properties of conditional expectation (and variance). □

Note that the posterior distribution of $\boldsymbol{l}'\boldsymbol{y}$, the quantity of interest, is normal with mean $\boldsymbol{l}'\boldsymbol{y}_1+\boldsymbol{l}'E_1\{\boldsymbol{y}_2\,|\,\boldsymbol{y}_1\}$ and variance $\boldsymbol{l}'\boldsymbol{V}_1\{\boldsymbol{y}_2\,|\,\boldsymbol{y}_1\}\boldsymbol{l}$.

Following Diaconis and Ylvisaker (1979), from the fact that $\boldsymbol{C}+\boldsymbol{C}_0$ is an identity matrix, the Bayes estimator $\bar{\boldsymbol{\beta}} = \boldsymbol{C}\hat{\boldsymbol{\beta}}+\boldsymbol{C}_0\boldsymbol{b}$ may be viewed as a generalization of the convex combination among the sample contribution ($\hat{\boldsymbol{\beta}}$) and the prior contribution ($\boldsymbol{b}$).

For the case of noninformative prior for $\boldsymbol{\beta}$, the square null matrix, $\boldsymbol{\Phi}$ (here of order M), replaces the matrix $\boldsymbol{B}^{-1}$, the inverse of the covariance matrix of $\boldsymbol{\beta}$. Hence, $\bar{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}$ indicating that there is no prior contribution in the prediction of $\boldsymbol{\beta}$, and $\boldsymbol{CA} = \boldsymbol{A}$ showing that the results of Royall and Pfeffermann (1982) obtain.

The following example illustrates the applicability of Theorem 1,

Example 1 : In the model described, suppose :

(1) $M = 1$; that is, $\boldsymbol{X}' = (x_1, \ldots, x_N)$.

(2) $\boldsymbol{V}$ is a diagonal matrix whose k-th diagonal element is $\boldsymbol{v}\boldsymbol{x}_k$ $(k=1, \ldots, N)$ where $\boldsymbol{v}$ is a known real number (positive if $\boldsymbol{x}_k$'s are positive).

(3) $\boldsymbol{\beta}$ is normally distributed with finite real mean $\boldsymbol{b}$ and variance $\boldsymbol{B} > 0$.

By Theorem 1, $\boldsymbol{y}_2 | \boldsymbol{y}_1$ has a multivariate normal distribution with mean vector

$$\boldsymbol{X}_2 \left(\frac{\Sigma\, i_k x_k}{\boldsymbol{v}} + \frac{1}{\boldsymbol{B}} \right)^{-1} \left(\frac{1}{\boldsymbol{v}}\ \Sigma\, i_k y_k + \frac{\boldsymbol{b}}{\boldsymbol{B}} \right)$$

and covariance matrix $$\boldsymbol{V}_2 + \left(\frac{\Sigma\, i_k x_k}{\boldsymbol{v}} + \frac{1}{\boldsymbol{B}} \right)^{-1} \boldsymbol{X}_2 \boldsymbol{X}_2'$$

where Σ indicates the sum over P and $\boldsymbol{X}_2' = ((1-i_1)x_1, \ldots, (1-i_N)x_N)$.

Suppose now that $\boldsymbol{B}$ is so large that it is not absurd to substitute 0 for $\boldsymbol{B}^{-1}$. Now, let $T = y_1 + \ldots + y_N = (1, \ldots, 1)\boldsymbol{y}$ be the quantity of interest. Hence,

$$E_1\{T \mid \boldsymbol{y}_1\} = \frac{\Sigma x_k}{\Sigma i_k x_k}\ \Sigma i_k y_k = \left(1 + \frac{\bar{x}_2}{\bar{x}_1} \frac{N-n}{n}\right) T_1$$

and $$V_1\{T \mid \boldsymbol{y}_1\} = (N-n)\, v \left(1 + \frac{\bar{x}_2}{\bar{x}_1}\ \frac{N-n}{n}\right) \bar{x}_2$$

where $\bar{x}_1$ and $\bar{\boldsymbol{x}}_2$ are the means of the auxiliary variable for the sampled and the unsampled units, respectively, and $T_1 = (1, \ldots, 1)\boldsymbol{y}_1$ is the sampled part of T. It is interesting that the Bayes predictor $E_1\{T \mid \boldsymbol{y}_1\}$ is the ratio estimator of the classical Survey Sampling Theory.

As a final remark, if the posterior variance is regarded as a good measure of uncertainty, a Bayesian should take $\bar{x}_1$ as large as possible, and hence, $\bar{\boldsymbol{x}}_2$ as small as possible. This suggests that sampling designs be purposive rather than random.

3. Bayesian robustness

The concept of Robustness considered in this paper, was inspired, by Barlow and Wu (1981) rather than by Smith (1983).

Consider a second Bayesian, called Bayesian 2, who believes that Bayesian 1 has omitted, in the regression model, some important auxiliary variables. Although they do not agree on the prior of $\boldsymbol{y}$, they do agree on the distributions of $\boldsymbol{\beta}$ and $\boldsymbol{\epsilon}$. In fact, Bayesian 2 considers the model

$$\boldsymbol{y} = \boldsymbol{X}^*\boldsymbol{\beta}^* + \boldsymbol{\epsilon}$$

where

(i) $\boldsymbol{\epsilon}$ is as described by Bayesian 1;

(ii) $\boldsymbol{X}^*$ is a matrix of order $N \times (M+L)$ whose M first columns are those of $\boldsymbol{X}$ and the last L columns form a matrix $\boldsymbol{Z}$ whose k-th row ($k = 1, \ldots, N$) is $\boldsymbol{Z}_k' = (z_{k1}, \ldots, z_{kL})$. That is, one admits L additional auxiliary variables and the design matrix becomes $\boldsymbol{X}^* = (\boldsymbol{X}; \boldsymbol{Z})$;

(iii) $\boldsymbol{\beta}^{*\prime} = (\boldsymbol{\beta}'; \boldsymbol{\delta}')$ where $\boldsymbol{\beta}$ is as before and $\boldsymbol{\delta}$ is a column vector of order L. Here, $\boldsymbol{\beta}^*$ is a multivariate normal vector of order $M+L$ with mean $\boldsymbol{b}^* = (\boldsymbol{b}'; \boldsymbol{d}')'$ and covariance matrix $\boldsymbol{B}^* = \begin{bmatrix} \boldsymbol{B} & \boldsymbol{D}_0 \\ \boldsymbol{D}_0' & \boldsymbol{D} \end{bmatrix}$. Note that, marginally $\boldsymbol{\beta}$ is as described by Bayesian 1.

A set of conditions, R, on the prior considerations of the two Bayesians is said to be a "Robustness set" if under R these two statisticians reach the same inference about the unknown quantity of interest.

A formalization of this concept is included in the two definitions below. In the sequel $E_2\{\cdot\}$ and $V_2\{\cdot\}$ are the mean and covariance operators for the model considered by Bayesian 2.

Definition 1 (Weak Robustness): A set of conditions R is a "weak robustness set" in relation to any linear function of $\boldsymbol{y}$, say $\boldsymbol{l}'\boldsymbol{y}$, for a class $\mathscr{B}$ of Bayesians, if under R the conditional (posterior) expectations of $\boldsymbol{y}_2$ (of $\boldsymbol{l}'\boldsymbol{y}$) given $\boldsymbol{y}_1$, for all the elements of $\mathscr{B}$, are equal.

Definition 2 (Robustness): A set of conditions R is a "robustness set" in relation to any function of $\boldsymbol{y}$ (to a linear function $\boldsymbol{l}'\boldsymbol{y}$) for a class $\mathscr{B}$ of Bayesians, if under R the conditional (posterior) distributions of $\boldsymbol{y}_2$ (of $\boldsymbol{l}'\boldsymbol{y}$ given $\boldsymbol{y}_1$, are the same for every element of $\mathscr{B}$.

Note that a set R may satisfy Definitions 1 or 2 for a class $\mathscr{B}$, but may not satisfy them for a different class $\mathscr{B}'$.

As before, the sampled and unsampled quantities are denoted by

$$\boldsymbol{X}_1^* = \boldsymbol{I}_S\boldsymbol{X}^*,\ \boldsymbol{X}_2^* = (\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{X}^*,\ \boldsymbol{Z}_1 = \boldsymbol{I}_S\boldsymbol{Z} \quad \text{and} \quad \boldsymbol{Z}_2 = (\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{Z},$$

which implies that

$$\boldsymbol{X}_1^* = (\boldsymbol{X}_1, \boldsymbol{Z}_1) \quad \text{and} \quad \boldsymbol{X}_2^* = (\boldsymbol{X}_2; \boldsymbol{Z}_2).$$

The main contribution of this paper is in the three following results (Theorems 2 and 3, and Corollary 1) because in that way the authors may pin-point the roles of the moments, the prior distributions, and the population quantities of interest. In that way, the authors believe that some obscure points of Royall and Pfeffermann (1982) are clarified.

Theorem 2 : *If a set of conditions, R, envisages that*

(i) $\boldsymbol{\delta}$ *and* $\boldsymbol{\beta}$ *are independent, and*

(ii) $\boldsymbol{l}'\boldsymbol{Z}_2 = \boldsymbol{l}'[(\boldsymbol{X}_2-\boldsymbol{V}_{21}\boldsymbol{V}_1^-\boldsymbol{X}_1)(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\boldsymbol{X}_1'+\boldsymbol{V}_{21}]\boldsymbol{V}_1^-\boldsymbol{Z}_1,$

then R is a "weak robustness set", in relation to the linear function $\boldsymbol{l}'\boldsymbol{y}$, *for any class,* $\mathcal{B}$, *of Bayesians (not necessarily including Bayesians* 1 *and* 2) *whose posterior means of* $\boldsymbol{y}_2$ *are either equal to* $E_1\{\boldsymbol{y}_2 \mid \boldsymbol{y}_1\}$ *or* $E_2\{\boldsymbol{y}_2 \mid \boldsymbol{y}_1\}$.

Proof : First write for Bayesian 2, $\boldsymbol{y}^* = \boldsymbol{y}-\boldsymbol{Z\delta} = \boldsymbol{X\beta}+\boldsymbol{\epsilon}$, $\boldsymbol{y}_1^* = \boldsymbol{I}_S\boldsymbol{y}^*$ and $\boldsymbol{y}_2^* = (\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{y}^*$. Since $\boldsymbol{\beta}$ is independent of $\boldsymbol{\delta}$, from Theorem 1 we have

$$E_2\{\boldsymbol{y}_2^* \mid \boldsymbol{y}_1, \boldsymbol{\delta}\} = E_2\{\boldsymbol{y}_2^* \mid \boldsymbol{y}_1^*, \boldsymbol{\delta}\} = \boldsymbol{X}_2\bar{\bar{\boldsymbol{\beta}}}+\boldsymbol{V}_{21}\boldsymbol{V}_1^-(\boldsymbol{y}_1^*-\boldsymbol{X}_1\bar{\bar{\boldsymbol{\beta}}})$$

where
$$\bar{\bar{\boldsymbol{\beta}}} = \boldsymbol{C}\hat{\bar{\boldsymbol{\beta}}}+\boldsymbol{C}_0\boldsymbol{b} \quad \text{and} \quad \hat{\bar{\boldsymbol{\beta}}} = \boldsymbol{\beta}-\boldsymbol{AX}_1'\boldsymbol{V}_1^-\boldsymbol{Z}_1\boldsymbol{\delta}.$$

Hence,

$$\begin{aligned} E_2\{\boldsymbol{y}_2^* \mid \boldsymbol{y}_1, \boldsymbol{\delta}\} = {} & E_1\{\boldsymbol{y}_2 \mid \boldsymbol{y}_1\}-\boldsymbol{X}_2\boldsymbol{CAX}_1'\boldsymbol{V}_1^-\boldsymbol{Z}_1\boldsymbol{\delta}-\boldsymbol{V}_{21}\boldsymbol{V}_1^-\boldsymbol{Z}_1\boldsymbol{\delta} \\ & +\boldsymbol{V}_{21}\boldsymbol{V}_1^-\boldsymbol{X}_1\boldsymbol{CAX}_1'\boldsymbol{V}_1^-\boldsymbol{Z}_1\boldsymbol{\delta} \end{aligned}$$

and

$$E_2\{\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{\delta}\} = E_1\{\boldsymbol{y}_2 \mid \boldsymbol{y}_1\}+\boldsymbol{Z}_2\boldsymbol{\delta}-[(\boldsymbol{X}_2-\boldsymbol{V}_{21}\boldsymbol{V}_1^-\boldsymbol{X}_1)\boldsymbol{CAX}_1'+\boldsymbol{V}_{21}]\boldsymbol{V}_1^-\boldsymbol{Z}_1\boldsymbol{\delta}.$$

We end the proof by noticing that

$$E_2\{\boldsymbol{y}_2 \mid \boldsymbol{y}_1\} = E_2\{\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{\delta}\} \mid \boldsymbol{y}_1\}. \quad \square$$

The next result shows why the statement "Because the condition in Theorem 1 is true for every fixed $\boldsymbol{\delta}$, it is true when δ has any prior distribution..." of Royall and Pfeffermann (1983) needs a proof. In fact, conditions that are true under the conditional distribution do not need to hold under the marginal distribution. In the particular case of that paper, "... is true for every fixed $\boldsymbol{\delta}$..." means that it is true conditionally on δ which does not necessarily imply that it is true marginally. For some counter examples see Basu and Pereira (1983).

Theorem 3 : *For Bayesians* 1 *and* 2, *the set of conditions* R, *in Theorem* 2, *is actually a "robustness set" in relation to the linear function* $\boldsymbol{l}'\boldsymbol{y}$.

Proof : Since $\boldsymbol{y}_2 \mid \boldsymbol{y}_1$ is normally distributed for both Bayesians and, under R, $E_1\{\boldsymbol{l}'\boldsymbol{y}_2\ \boldsymbol{y}_1\} = E_2\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1\}$, to prove that the two posterior distributions of $\boldsymbol{l}'\boldsymbol{y}_2$ are equal (under R) we need only to prove that

$$\boldsymbol{V}_1\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1\} = \boldsymbol{V}_2\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1\}.$$

Note that

$$\boldsymbol{V}_2\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1\} = \boldsymbol{V}_2\{E_2\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{\delta}\} \mid \boldsymbol{y}_1\} + E_2\{\boldsymbol{V}_2\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{\delta}\} \mid \boldsymbol{y}_1\}$$

and under R, $E_2\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{\delta}\} = E_1\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1\}$ which is independent of $\boldsymbol{\delta}$ that makes $\boldsymbol{V}_2\{E_2\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{\delta}\}\ \boldsymbol{y}_1\}$ null. On the other hand, since

$$\boldsymbol{V}_2\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{\delta}\} = \boldsymbol{V}_2\{\boldsymbol{l}'\boldsymbol{y}_2^* \mid \boldsymbol{y}_1, \boldsymbol{\delta}\} = \boldsymbol{V}_1\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1\},$$

we have that

$$E_2\{\boldsymbol{V}_2\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1, \boldsymbol{\delta}\} = \boldsymbol{V}_1\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1\}.$$

which concludes the proof. □

Corollary 1 : *For Bayesians* 1 *and* 2, *if the set of conditions* R *in Theorem* 2 *includes condition* (ii) *for every vector* $\boldsymbol{l}$ *of order* N, *then* R *is a robustness set in relation to any function of* $\boldsymbol{y}$.

This result is proved by recalling that if $\boldsymbol{l}'\boldsymbol{y}_2$ is normal for every $\boldsymbol{l}$, then $\boldsymbol{y}_2$ is multivariate normal.

It is important to emphasize that the normality of $\boldsymbol{\delta}$ and the independence of $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are essential for Theorem 3 and Corollary 1. However only the independence is essential for Theorem 2.

Another important fact is that conditions (i) and (ii) of Theorem 2 are sufficient conditions for robustness. If in the place of (ii) we consider

$$\text{(ii)}' \quad \boldsymbol{l}'\boldsymbol{Z}_2 E_2\{\boldsymbol{\delta} \mid \boldsymbol{y}_1\} = \boldsymbol{l}'[(\boldsymbol{X}_2 - \boldsymbol{V}_{21}\boldsymbol{V}_1^- \boldsymbol{X}_1)(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\boldsymbol{X}_1' + \boldsymbol{V}_{21}]\boldsymbol{V}_1^- \boldsymbol{Z}_1 \boldsymbol{E}_2\{\boldsymbol{\delta} \mid \boldsymbol{y}_1\}$$

they turn out to be necessary and sufficient conditions for robustness. However, condition (ii)′ depends on the data $\boldsymbol{y}_1$ which may not be feasible since $\boldsymbol{y}_1$ is obtained after S has been chosen. Note that condition (ii)′ implies that

$$\boldsymbol{l}'\boldsymbol{Z}_2\boldsymbol{d} = \boldsymbol{l}'[(\boldsymbol{X}_2 - \boldsymbol{V}_{21}\boldsymbol{V}_1'\boldsymbol{X}_1)(\boldsymbol{A}^{-1}+\boldsymbol{B}^{-1})^{-1}\boldsymbol{X}_1' + \boldsymbol{V}_{21}]\boldsymbol{V}_1^- \boldsymbol{Z}_1\boldsymbol{d}$$

where, as before, $\boldsymbol{d} = E_2\{\boldsymbol{\delta}\}$ is a column vector of order L. From this last equality, for $L = 1$, a strong result for robustness is stated next.

Corollary 2 : *If* δ *is a scalar with finite non-null prior mean,* $E_2\{\delta\}$, *then* R *is a* (*weak*) *robustness set in relation to* $\boldsymbol{l}'\boldsymbol{y}$ (*for the class* $\mathcal{B}$ *of Theorem* 2) *for Bayesians* 1 *and* 2 *if and only if* R *contains* (i) *and* (ii) *of Theorem* 2.

The following example shows that many standard situations may be viewed as applications of the Bayesian robustness results discussed here.

Example 2 : Suppose that $\boldsymbol{V}$ is a positive definite diagonal matrix and define $\boldsymbol{V}_1^- = \boldsymbol{I}_S \boldsymbol{V}^{-1} \boldsymbol{I}_S = \boldsymbol{I}_S \boldsymbol{V}^{-1}$. Suppose in addition that $\boldsymbol{B}^{-1}$ is not far from the null matrix (noninformative prior for $\boldsymbol{\beta}$) and that $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are independent. Then, condition (ii) reduces to

$$\text{(iii)} \qquad \boldsymbol{l}'\boldsymbol{Z}_2 = \boldsymbol{l}'\boldsymbol{X}_2\boldsymbol{A}\boldsymbol{X}_1'\boldsymbol{V}_1^-\boldsymbol{Z}_1$$

or equivalently $\qquad \boldsymbol{l}'(\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{Z} = \boldsymbol{l}'(\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{I}_S\boldsymbol{V}^{-1}\boldsymbol{I}_S\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{I}_S\boldsymbol{V}^{-1}\boldsymbol{Z}.$

Under these robustness conditions $\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1$ is normally distributed with mean and variance given respectively by

$$E_2\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1\} = E_1\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1\}$$
$$= \boldsymbol{l}'(\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{I}_S\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{I}_S\boldsymbol{V}^{-1}\boldsymbol{y}_1$$

and by

$$V_2\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1\} = V_1\{\boldsymbol{l}'\boldsymbol{y}_2 \mid \boldsymbol{y}_1\}$$
$$= \boldsymbol{l}'(\boldsymbol{I}-\boldsymbol{I}_S)[\boldsymbol{V}^{-1}+\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{I}_S\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'](\boldsymbol{I}-\boldsymbol{I}_S)\boldsymbol{l}. \qquad \ldots \quad (1)$$

Condition (iii) above, surprisingly, is the condition that appears in Theorem 2 of Pereira and Rodrigues (1983) which is a paper that discusses robustness under the classical set-up. This particular condition only ensures that the best unbiased estimator of $\boldsymbol{l}'\boldsymbol{y}$ under the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{\epsilon}$ is unbiased under the model $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta}+\boldsymbol{Z}\boldsymbol{\delta}+\boldsymbol{\epsilon}$. This suggests that condition (ii) of Theorem 2 may actually be related only with the expectations of $\boldsymbol{l}'\boldsymbol{y}$.

It seems now clear as to why the main result of this paper is divided into two theorems and two corollaries by way of pin-pointing the role of normality in this linear theory of Bayesian robustness. In fact, if $\boldsymbol{V}_{12}$ is null, Theorem 2 holds for every independent prior distribution (with finite first moment), chosen for $\boldsymbol{\epsilon}$, $\boldsymbol{\beta}$ and δ. Normality is needed for Theorem 3 and Theorem 2 when $\boldsymbol{V}_{12}$ is not null. However, for Bayesians 1 and 2, it would be enough to consider Definition 2 and Theorem 3.

The next section trenats some interesting examples.

4. Special examples

Example 3 : Under the conditions of Example 1, suppose that $\boldsymbol{Z}' = \boldsymbol{l}' = (1, \ldots, 1)$. Condition (ii) of Theorem 2 reduces, in this case to $\bar{x}_2 = \left(\bar{x}_1 + \dfrac{v}{nB}\right)$. This condition guarantees the robustness of

$$\hat{T} = T_1 + \frac{(N-n)}{n}\bar{x}_2\left(\bar{x}_1 + \frac{\boldsymbol{v}}{n\boldsymbol{B}}\right)^{-1}\left(T_1 + \frac{\boldsymbol{v}}{\boldsymbol{B}}\boldsymbol{b}\right)$$

if $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are independent and normally distributed. If 0 be substituted for $\dfrac{\boldsymbol{v}}{\boldsymbol{B}}$, then $\hat{T}$ is the ratio estimator of T which is robust if $\bar{x}_2 = \bar{x}_1$, the balanced sample property.

Note that in Example 1 the prescription was to maximize $\bar{x}_1$ in order to minimize the posterior variance of T. However, Example 3 vindicates a balanced sample. Hence, since the selection of S has to be decided a priori, a Bayesian may want to follow the rules of selection that we present next :

(a) If one does not have any doubt about the non-inclusion of δ in the model, he must decide for a sample that makes $\bar{x}_1$ maximum.

(b) If one is not sure about the value of δ in the prediction of T, he must decide for a balanced sample. In both cases purposive samples are recommended.

Next example has been considered by many authors, see Pereira and Rodrigues (1983) for a complete review.

Example 4 : Consider the following elements, $\boldsymbol{X}' = \boldsymbol{l}' = (1, \ldots, 1)$, $\boldsymbol{V} = \boldsymbol{vI}$, $\boldsymbol{b}$ is a finite real number, $\boldsymbol{B}$ is a positive real number.

$$\boldsymbol{Z} = \begin{bmatrix} x_1 & x_1^2 & \ldots & x_1^L \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & \ldots & \cdot \\ \cdot & \cdot & & \cdot \\ x_N & x_N^2 & \ldots & x_N^L \end{bmatrix},$$

and $\boldsymbol{D}_0 = (0, \ldots, 0)$, the null vector of order L. The prediction of the total $T = y_1 + \ldots + y_N$ for Bayesian 1 is given by

$$\hat{T} = E_1\{T \mid \boldsymbol{y}_1\} = T_1 + \frac{N-n}{n}\left(1 + \frac{\boldsymbol{v}}{n\boldsymbol{B}}\right)^{-1}\left(T_1 + \frac{\boldsymbol{vb}}{\boldsymbol{B}}\right).$$

Condition (ii) of Theorem 2 reduces to the following system of equations

$$\left(1 + \frac{\boldsymbol{v}}{n\boldsymbol{B}}\right)\frac{1}{N-n}\,\Sigma\,(1 - i_k)x_k^j = \frac{1}{n}\,\Sigma\, i_k\, x_k^j, \quad j = 1, \ldots, L,$$

where Σ is the sum from $k = 1$ to $k = N$. If $\dfrac{\boldsymbol{v}}{\boldsymbol{B}}$ is approximately equal to zero, then the condition above is close to the balanced sample property and $\hat{T}$ is approximately equal to $\dfrac{N}{n}\,T_1$, the expansion estimator.

This example shows that "balanced samples may not play any important role for robustness when informative priors are considered".

Note that if a priori $\boldsymbol{\beta}$ is expected to be zero (that is, $b = 0$), then the Bayes predictor of T is $\hat{T} = \left[1+\frac{N-n}{n}\left(1+\frac{\boldsymbol{v}}{n\boldsymbol{B}}\right)^{-1}\right] T_1$ that has some similarity with an estimator proposed by Lindley (1962). Here, the factor $\frac{\boldsymbol{v}}{n\boldsymbol{B}}$ may be called "the Shrinkage factor".

Next we consider a class of predictors that, under noninformative priors, plays an important role in the theory of robust linear prediction.

5, A class of robust predictors

The prediction of a linear function $\boldsymbol{l}'\boldsymbol{y}$ is considered in this section under the assumptions of Example 2. That is, $\boldsymbol{V}$ is a nonsingular diagonal matrix, $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ are independent, and $\boldsymbol{B}^{-1}$, the inverse of the covariance matrix of $\boldsymbol{\beta}$, is approximately equal to the null matrix of order M which corresponds to a noninformative prior for $\boldsymbol{\beta}$.

In addition to conditions (i) and (ii) of Theorem 2 and (iii) of Example 2, the following ones are relevant. Suppose that there is a scalar c such that

$$\text{(iv)} \qquad \boldsymbol{l}'\boldsymbol{X}_2 = c\boldsymbol{l}'\boldsymbol{X}_1,$$

and

$$\text{(v)} \qquad \boldsymbol{l}'\boldsymbol{Z}_2 = c\boldsymbol{l}'\boldsymbol{Z}_1.$$

Finally, to complete the list of conditions consider

$$\text{(vi)} \qquad \boldsymbol{l}'\boldsymbol{X}_1\boldsymbol{A}\boldsymbol{X}\,\boldsymbol{V}_1^- = \boldsymbol{l}'\boldsymbol{I}_S$$

where, as before, $\boldsymbol{A} = (\boldsymbol{X}_1'\boldsymbol{V}_2^-\boldsymbol{X}_1)^{-1}$. Note that since $\boldsymbol{V}$ is diagonal, (vi) is equivalent to $\boldsymbol{l}'\boldsymbol{V}_1 = \boldsymbol{l}'\boldsymbol{X}_1\boldsymbol{A}\boldsymbol{X}_1'$ which satisfies condition L of Royall and Pfeffermann (1982) for V diagonal. On the other hand, (iv), (v), and (vi) together imply (iii) since

$$\boldsymbol{l}'\boldsymbol{X}_2\boldsymbol{A}\boldsymbol{X}_1'\boldsymbol{V}_1^-\boldsymbol{Z}_1 \underset{\text{(iv)}}{=} c\boldsymbol{l}'\boldsymbol{X}_1\boldsymbol{A}\boldsymbol{X}_1'\boldsymbol{V}_1^-\boldsymbol{Z}_1 \underset{\text{(vi)}}{=} c\boldsymbol{l}'\boldsymbol{Z}_1 \underset{\text{(v)}}{=} \boldsymbol{l}'\boldsymbol{Z}_2.$$

This shows that (iv), (v), and (vi) together form a sufficient robustness set of conditions. It is important to notice that conditions (iv), (v), and (vi) are less restrictive than the conditions imposed by Royall and Pfeffermann (1982), since in that paper c is taken equal to $\frac{N-n}{n}$.

In the rest of this section, $\mathbf{1}' = (1, \ldots, 1)$.

Recalling Theorem 1, if (iv) and (vi) hold the posterior mean of T, the population total, may be written as

$$\begin{aligned}\hat{T} &= E_1\{T \mid \boldsymbol{y}_1\} = E_2\{T \mid \boldsymbol{y}_1\} \\ &= \boldsymbol{l}'\boldsymbol{y}_1 + \boldsymbol{l}'\boldsymbol{X}_2\hat{\boldsymbol{\beta}} = \boldsymbol{l}'\boldsymbol{y}_1 + c\boldsymbol{l}'\boldsymbol{y}_1 \\ &= (1+c)\boldsymbol{l}'\boldsymbol{y}_1 = (1+c)T_1 \qquad \ldots \quad (2)\end{aligned}$$

where T_1 is the sample total, and c is the scalar defined in (iv). In the particular case where $c = \frac{(N-n)}{n}$ (balance on $\boldsymbol{X}$ and on $\boldsymbol{Z}$), it follows that $\hat{T} = \frac{N}{n}T_1$, the usual expansion predictor considered by many authors (Royall and Pfeffermann, 1983).

By applying conditions (iv), (v), and (vi) to expression (1) of Example 2 it can be shown that

$$\begin{aligned}V_1\{T \mid \boldsymbol{y}_1\} &= V_2\{T \mid \boldsymbol{y}_1\} = c^2\boldsymbol{l}'\boldsymbol{V}_1\boldsymbol{l} + \boldsymbol{l}'\boldsymbol{V}_2\boldsymbol{l} \\ &= \boldsymbol{l}'\boldsymbol{V}\boldsymbol{l} + (c^2-1)\boldsymbol{l}'\boldsymbol{V}_1\boldsymbol{l}. \qquad \ldots \quad (3)\end{aligned}$$

Note that, efficiency and robustness may be attained if there is a scalar c that, besides satisfying (iv) and (v), minimizes the posterior variance of T given by (3). This emphasizes the interest on the class of predictors defined by expression (2). To illustrate that this class of predictors indeed contains interesting members. the following examples are considered.

Example 5 : If $\boldsymbol{X}$ (or $\boldsymbol{Z}$) includes the unit vector as a column, then $c = \frac{N-n}{n}$ and conditions (iv) and (v) reduce to balanced sample conditions on $\boldsymbol{X}$ and $\boldsymbol{Z}$ and $\hat{T}$ reduces to the expansion estimator.

Example 6 : Suppose that the model considered by Bayesian 1 is as described in Example 1, with noninformative prior for $\boldsymbol{\beta}$. Also assume that for Bayesian 2, $y_1, \ldots, y_N$ are conditionally independent given $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$ whose distributions are $N(\boldsymbol{\beta}x_i + \boldsymbol{\delta}z_i; \boldsymbol{v}x_i)$, $i = 1, \ldots, N$, and $\boldsymbol{\beta}$, $\boldsymbol{\delta}$, and $\boldsymbol{\epsilon}$ are mutually independent where $\boldsymbol{\beta}$ is as described by Bayesian 1.

Here, condition (vi) is easily verified and conditions (iv) and (v) are equivalent to

$$\frac{\bar{x}_2}{\bar{x}_1} = \frac{\bar{z}_2}{\bar{z}_1} = \frac{n}{N-n}\,c \qquad \ldots \quad (4)$$

where $\bar{x}_1$, $\bar{z}_1$, $\bar{x}_2$, and $\bar{z}_2$ are the sample and unsample means for x and z, respectively. Therefore, $\hat{T}$ reduces to the usual ratio predictor which is shown to have the robustness property under condition (4). Note that, if (4) holds for every sample selected, then efficiency demands a sample with $\bar{x}_1$ (or equivalently $\bar{z}_1$) as large as possible minimizing the posterior variance.

As a final remark, it is relevant to note that similar results are obtained when the finite population is partitioned into H strata (by all the Bayesians involved) with different regression coefficients in each stratum. Robustness and efficiency would then be guaranteed by selecting in each stratum h, $h = 1, \ldots, H$, the smallest possible c_h satisfying

$$\frac{\bar{x}_{2h}}{\bar{x}_{1h}} = \frac{\bar{z}_{2h}}{\bar{z}_{1h}} = c_h \frac{n_h}{N_h - n_h}$$

where (a) N and n_h are the sides of stratum h and of the sampled part of stratum h, respectively;

(b) $\bar{x}_{1h}$, $\bar{z}_{1h}$, $\bar{x}_{2h}$, and $\bar{z}_{2h}$ are the sampled and unsampled means for x and z, respectively, in stratum h.

The best predictor of the population total in this case is given by

$$\hat{T} = \Sigma(1+c_h)\boldsymbol{l}'\boldsymbol{y}_h$$

where Σ is the sum from $h = 1$ to $h = H$ and $\boldsymbol{y}_h$ is the column vector of order $N = \Sigma N_h$ whose components are $i_{kh}y_k$ with $i_{kh} = 1$ if $k \in S$ and k is in stratum h and $i_{kh} = 0$ otherwise.

6. Final considerations

Throughout the above sections, the main applicability restrictions are related to the covariance matrix V of ε. First, it was supposed known and later as unknown but a fixed constant.

The analysis with a prior distribution postulated for V is postponed to a future work. The next example shows that a common V is a natural restriction.

Example 7 : Consider the model $\boldsymbol{y} = \boldsymbol{l}\boldsymbol{\beta}+\boldsymbol{\epsilon}$ where $\boldsymbol{l}' = (1, \ldots, 1)$ and $\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{B})$ for both Bayesians. They disagree only on the prior distribution for $\boldsymbol{\epsilon}$. For Bayesian 1, $\boldsymbol{\epsilon} \sim N(\boldsymbol{\phi}, \boldsymbol{I})$ and for Bayesian 2, $\boldsymbol{\epsilon} \sim N(\boldsymbol{\phi}, \boldsymbol{vI})$. Recalling Example 4, we have that

$$E_1\{T \mid \boldsymbol{y}_1\} = T_1 + \frac{N-n}{n}\left(1+\frac{1}{n\boldsymbol{B}}\right)^{-1} T_1$$

and that

$$E_2\{T \mid \boldsymbol{y}_1\} = T_1 + \frac{N-n}{n}\left(1+\frac{\boldsymbol{v}}{n\boldsymbol{B}}\right)^{-1} T_1.$$

Hence, $E_1\{T \mid \boldsymbol{y}_1\} = E_2\{T \mid \boldsymbol{y}_1\}$ if and only if $v = 1$; that is, the two statisticians must have the same prior information.

Note that from Corollary 2 we obtain a necessary and sufficient condition for robustness when we add a new auxiliary variable and the covariance matrix of $\boldsymbol{\epsilon}$ remain the same. Example 7 suggests that equal covariance matrix for $\boldsymbol{\epsilon}$ is a necessary and sufficient condition for robustness when the same model is considered. Hence, by transitivity equal covariance matrix for ε seems to be a necessary condition for the linear robustness considered in this paper.

References

Barlow, R. E. and Wu, A. S. (1981) : Preposterior analysis of Bayes estimators of mean life, *Biometrika*, **68**, 403-10.

Basu, D. (1969) : Role of the sufficiency and likelihood principles in sample survey theory. *Sankhyā. A*, **31**, 441-54.

——— (1975) : Statistical information and likelihood. *Sankhyā A*. **37**, 1-71.

Basu, D. and Pereira, C. A. de B. (1983) : Conditional independence in statistics. *Sankhyā A*, **45**, 324-37.

Cassel, C. M., Sarndal, C. E. and Wretman, J. H. (1977) : *Foundations of Inference in Survey Sampling*, New York, John Wiley.

Diaconis, R. and Ylvisaker, D. (1979) : Conjugate priors for exponential families. *Ann. Statist.* **7**, 269-81.

Lindley, D. V. (1962) : Comments on Stein's paper, *JRSS, B*, **24**, 285-7.

Lindley, D. V. and Smith, A. F. M. (1972) : Bayes estimates for the linear model (with discussion). *J. R. Statist. Soc. B*, **34**, 1-41.

Pereira, C. A. de B. and Rodrigues, J. (1983) : Robust linear prediction in finite populations. *International Statistical Review*, 51, 293-300.

Rao, C. R. (1973) : *Linear Statistical Inference and its Applications*. 2nd edtion, New York J. Wiley.

Royall, R. M. and Pfeffermann, D. (1982) : Balanced samples and robust Bayesian inference in finite population sampling. *Biometrika*, **69**, 401-9.

Smith, A. F. M, (1983) : Bayesian approaches to outliers and robustness, *in Specifying Statistical Models : From Parametric, Using Bayesian or Non-Bayesian Approaches*. 13-35, (*edited by Florens, Mouchart, Raout, Simar, and Smith*—Lecture Notes in Statistics 16, New York. Springer-Verlag).

Paper received : August, 1984.