*Journal of the Korean Statistical Society (1998),* **27**: *3, pp 331-348*

# Robust Bayes and Empirical Bayes Analysis in Finite Population Sampling with Auxiliary Information[†]

Dal Ho Kim[1]

ABSTRACT

In this paper, we have proposed some robust Bayes estimators using ML-II priors as well as certain empirical Bayes estimators in estimating the finite population mean in the presence of auxiliary information. These estimators are compared with the classical ratio estimator and a subjective Bayes estimator utilizing the auxiliary information in terms of "posterior robustness" and "procedure robustness". Also, we have addressed the issue of choice of sampling design from a robust Bayesian viewpoint.

*Keywords:* Finite population; Mean; Bayes; Robust Bayes; Empirical Bayes; $\varepsilon$-contamination; ML-II prior; Posterior robustness; Procedure robustness

## 1. INTRODUCTION

We consider a finite population $\mathcal{U}$ with units labeled $1, 2, \ldots, N$. Let $y_i$ denote the value of a single characteristic attached to the unit $i$ ($i = 1, 2, \ldots, N$). The vector $y = (y_1, \cdots, y_N)^T$ is the unknown state of nature, and is assumed to belong to $\Theta = R^N$. A subset $s$ of $\{1, 2, \ldots, N\}$ is called a sample. Let $n(s)$ denote the number of elements belonging to $s$. The set of all possible samples is denoted by $S$. A design is a *function* $p$ on $S$ such that $p(s) \in [0, 1]$ for all $s \in S$ and $\sum_{s \in S} p(s) = 1$. Given $y \in \Theta$ and $s = \{i_1, \cdots, i_{n(s)}\}$ with $1 \le i_1 < \cdots < i_{n(s)} \le N$, let $y(s) = \{y_{i_1}, \cdots, y_{i_{n(s)}}\}$. One of the main objectives in sample surveys is to draw inference about $y$ or some function (real- or vector-valued) $\gamma(y)$ of $y$ on the basis of $s$ and $y(s)$. In this article, we will be concerned exclusively with $\gamma(y) = N^{-1} \sum_{i=1}^{N} y_i$, the finite population mean, although the general methods to be described later are applicable to other parameters of interest as well.

[†]This paper was supported by Kyungpook National University Research Fund, 1997.
[1]Department of Statistics, Kyungpook National University, Taegu, 702-701, Korea.

In many sample surveys, for every unit $i$ in the finite population, information is available for one or more auxiliary characteristics, characteristics other than the one of direct interest. For example, if the characteristic of direct interest is the yield of a particular crop, the auxiliary characteristic could be the area devoted that crop by different farms in the list. We consider the simplest situation when for every unit $i$ in the population, value of a certain auxiliary characteristic, say $x_i(> 0)$ is known $(i = 1, 2, \ldots, N)$.

One time-honored estimator of the finite population mean in this situation is the classical ratio estimator which seems to incorporate the auxiliary information in a very natural manner. Moreover, this estimator can be justified both from the model and design based approach. While Cochran (1977) provides many design-based properties of the ratio estimator, Royall (1970, 1971) justifies this estimator based on certain superpopulation models.

In model-based inference approach for such problems, the finite population was viewed as a sample from a superpopulation. From a Bayesian perspective, this amounts to putting a prior distribution on $y$. A unified and elegant formulation of Bayes estimation in finite population sampling was given by Ericson (1969). Since then, there are many papers in the area of Bayes estimation in finite population sampling.

Most of the Bayesian literature in survey sampling deals with subjective Bayesian analysis in that the inference procedure is based on a single completely specified prior distribution. The subjective Bayesian approach has frequently criticized on the ground that it presumes an ability to completely and accurately quantify subjective information in terms of a single prior distribution. We shall see in Section 3 that failure to specify accurately one or more of the parameters of a prior distribution has a serious consequence when calculating the Bayes risk and often protection is needed against the possibility of such occurrence.

A robust Bayesian viewpoint assumes that subjective information can be quantified only in terms of a class $\Gamma$ of possible distributions. Inferences and decisions should be relatively insensitive to deviations as the prior distribution varies over $\Gamma$. The robust Bayesian idea can be traced back to Good as early as in 1950 (see for example Good (1965)), and the topic has enjoyed wide popularity in recent years sparked by the stimulating article of Berger (1984).

The need for robust Bayesian analysis in survey sampling has also been felt by some authors. Godambe and Thompson (1971) adapted a framework whereby the prior information could only be quantified up to a class of prior distributions. For estimating the population total in the presence of auxiliary information, they came up with the usual ratio and difference estimators, justifying these on the ground of location invariance. The model assumption there played a very minimal role, and the main idea was that model-based inference statements could be replaced in the case of model-failure by design-based inference. In a later study, Godambe (1982) considered more common phenomenon of specific departures from the assumed model. Royall and Pfefferman (1982) have addressed a different robustness issue. Their main concern is to find out conditions under which the Bayes estimators under the assumed model remain the same under departures from the model.

The present article considers an $\varepsilon$-contamination class of priors following the lines of Berger and Berliner (1986). In Section 2, we develop some robust Bayes (RB) estimators employing ML-II priors as well as certain empirical Bayes (EB) estimators of the finite population mean when certain auxiliary information is present. In Section 3, we provide analytic expressions for the indices of posterior and procedure robustness (cf Berger (1984)) of the proposed RB and EB estimators, and compare these indices with similar indices for the classical ratio estimator as well as the subjective Bayes estimator. We have proved the asymptotic optimality in the sense of Robbins (1955) of the RB and EB estimators. In Section 4, the issue of choice of sampling designs is addressed from a robust Bayesian viewpoint. In Section 5, a numerical example is provided to illustrate the results of the preceding section.

For simplicity, we shall assume that $n(s) \neq n \Rightarrow p(s) = 0$, that is, we effectively consider only samples of fixed size $n$. Also, for notational simplicity, we shall, henceforth, assume that $s = \{i_1, \cdots, i_n\}$ where $1 \leq i_1 < \cdots < i_n \leq N$. Let $\bar{s} = \{1, 2, \ldots, N\} - s = \{j_1, \cdots, j_{N-n}\}$ (say), where $1 \leq j_1 < \cdots < j_{N-n} \leq N$. We shall write $y(s) = (y_{i_1}, \cdots, y_{i_n})^T$, $y(\bar{s}) = (y_{j_1}, \cdots, y_{j_{N-n}})^T$. Also, throughout the loss is assumed to be squared error.

## 2. ROBUST BAYES AND EMPIRICAL BAYES ESTIMATORS

We consider the superpopulation model $y_i = \beta x_i + \varepsilon_i (i = 1, 2, \cdots, N)$, where $\beta, \varepsilon_1, \cdots, \varepsilon_N$ are independently distributed with $\beta \sim \mathrm{N}(\beta_0, \sigma_0^2)$ and $\varepsilon_1, \cdots, \varepsilon_N$ are iid $\mathrm{N}(0,\ \tau^2 h(x_i))$ for any positive valued function $h$. Let $x = (x_1, x_2, \cdots, x_N)^T$, $x(s)$ the column vector with elements equal to $x_i (i \in s)$ arranged in ascending order of suffixes, i.e., if $s = (i_1, \cdots, i_n)$ with $1 \leq i_1 < i_2 < \cdots, i_n \leq N$, $x(s) = (x_{i_1}, \cdots, x_{i_n})^T$; similarly $x(\bar{s})$ is the column vector with its elements equal to $x_i (i \in \bar{s})$ arranged in ascending order of the suffixes. Further, let $D = D(x) =$

Diag$(h(x_1),\cdots,h(x_N))$, $D(s)$ is a diagonal matrix with its elements equal to $h(x_i)(i\in s)$ arranged in ascending order of suffixes, i.e., if $s=(i_1,\cdots,i_n)$ with $1\le i_1<i_2<\cdots,i_n\le N$, $D(s)=\text{Diag}(h(x_{i_1}),\cdots,h(x_{i_n}))$. Similarly, $D(\bar{s})$ is a diagonal matrix with elements equal to $h(x_i)(i\in\bar{s})$ arranged in ascending order of suffixes. The following theorem gives the distribution of $y(s)$, and the conditional distribution of $y(\bar{s})$ given $(s,y(s))$ under the assumed model.

**Theorem 2.1.** *Under the assumed model,*
*(i) marginally* $y(s)\sim N[\beta_0 x(s),\tau^2 D(s)+\sigma_0^2 x(s)x^T(s)]$;
*(ii) the conditional distribution of* $y(\bar{s})$ *given* $s,y(s)$ *is*

$$N\left[\frac{\beta_0\tau^2+\sigma_0^2\sum_{i\in s}y_ix_i/h(x_i)}{\tau^2+\sigma_0^2\sum_{i\in s}x_i^2/h(x_i)}x(\bar{s}),\ \tau^2[D(\bar{s})+\frac{\sigma_0^2x(\bar{s})x^T(\bar{s})}{\tau^2+\sigma_0^2\sum_{i\in s}x_i^2/h(x_i)}]\right].$$

The proof of the theorem is postponed to the Appendix. It follows from this theorem that the Bayes estimator of $\gamma(y)=N^{-1}\sum_{i=1}^N y_i$ is

$$e_0(s,y(s))=N^{-1}[n\bar{y}(s)+\frac{\beta_0\tau^2+\sigma_0^2\sum_{i\in s}y_ix_i/h(x_i)}{\tau^2+\sigma_0^2\sum_{i\in s}x_i^2/h(x_i)}\sum_{i\in\bar{s}}x_i]. \tag{2.1}$$

The case $h(x_i)=x_i$ $(i=1,\cdots,N)$ is of particular interest. In this case writing $\bar{x}(s)=n^{-1}\sum_{i\in s}x_i$, $\bar{x}(\bar{s})=(N-n)^{-1}\sum_{i\in\bar{s}}x_i$, $M_0=\tau^2/\sigma_0^2$, and $B_0(s)=M_0/(M_0+n\bar{x}(s))$, the subjective Bayes estimator $e_0$ reduces to

$$e_0(s,y(s))=f\bar{y}(s)+(1-f)\bar{x}(\bar{s})\{(1-B_0(s))\bar{y}(s)/\bar{x}(s)+B_0(s)\beta_0\} \tag{2.2}$$

where $f=n/N$ and $\bar{y}(s)=n^{-1}\sum_{i\in s}y_i$.

The classical ratio estimator can be obtained as a limiting Bayes estimator from (2.2) by making $\sigma_0^2\to\infty$. An alternative approach to derive the ratio estimator (as mentioned in Royall and Pfefferman (1982)) is to assume the model that conditional on $\beta,y_1,\cdots,y_N$ are independently distributed with $y_i\sim\text{N}(\beta x_i,\ \tau^2x_i)$, while $\beta$ is uniform $(-\infty,\ \infty)$.

In the remainder of this section, we take $h(x_i)=x_i$ $(i=1,\cdots,N)$. In this case the classical ratio estimator is given by $e_R(s,y(s))=(\bar{y}(s)/\bar{x}(s))\bar{x}$, where $\bar{x}=N^{-1}\sum_{i=1}^N x_i$.

To derive the robust Bayes estimator of $\gamma(y)$, we consider $\varepsilon$-contaminated priors. Specifically, let the prior distribution of $\beta$ be the same as that of $\pi=(1-\varepsilon)\pi_0+\varepsilon q$, where $\varepsilon\in[0,1)$ is known, $\pi_0$ is the $\text{N}(\beta_0,\sigma_0^2)$ distribution, and

$q \in Q$, the class of all possible distributions. Then the marginal pdf of $y(s)$ under the prior $\pi$ is given by

$$m(y(s)|\pi) = (1-\varepsilon)m(y(s)|\pi_0) + \varepsilon m(y(s)|q), \tag{2.3}$$

where $m(y(s)|\pi_0)$ denotes the pdf of $\mathrm{N}(\beta_0 x(s), \tau^2 D(s) + \sigma_0^2 x(s)x^T(s))$, while

$$m(y(s)|q) = \int (2\pi\tau^2)^{-\frac{n}{2}}(\prod_{i\in s} x_i^{-\frac{1}{2}}) \exp[-\frac{1}{2\tau^2}\sum_{i\in s}(y_i - \beta x_i)^2/x_i] q(d\beta). \tag{2.4}$$

Now we find the ML-II prior within the given class of priors. Since $\sum_{i\in s}(y_i - \beta x_i)^2/x_i$ is minimized with respect to $\beta$ at $\widehat{\beta}_s = \bar{y}(s)/\bar{x}(s)$, from (2.3) and (2.4), the ML-II prior which maximizes the marginal likelihood $m(y(s)|\pi)$ with respect to $q \in Q$ is given by

$$\widehat{\pi}_{ML}(\beta) = (1-\varepsilon)\pi_0(\beta) + \varepsilon\widehat{q}_s(\beta), \tag{2.5}$$

where $\widehat{q}_s(\beta)$ is degenerate at $\beta = \widehat{\beta}_s$. Then we have the following theorem.

**Theorem 2.2.** *Under the ML-II prior $\widehat{\pi}_{ML}$, marginally*

$$y(s) \sim (1-\varepsilon)N[\beta_0 x(s), \tau^2 D(s) + \sigma_0^2 x(s)x^T(s)] + \varepsilon F_s$$

*where $F_s$ has the (improper) pdf*

$$f_s(y(s)) = (2\pi\tau^2)^{-n/2}(\prod_{i\in s} x_i^{-1/2}) \exp[-\sum_{i\in s}(y_i - \hat{\beta}_s x_i)^2/(2\tau^2 x_i)].$$

*Also, the conditional distribution of $y(\bar{s})$ given $(s, y(s))$ is*

$$\begin{aligned} &\widehat{\pi}_{ML}(y(\bar{s})|s, y(s)) \\ &= \widehat{\lambda}_{ML}(\bar{y}(s))N\Big[\Big((1 - B_0(s))\frac{\bar{y}(s)}{\bar{x}(s)}\bar{x}(s) + B_0(s)\beta_0\Big)x(\bar{s}), \tau^2\Big(D(\bar{s}) + \frac{x(\bar{s})x^T(\bar{s})}{M_0 + n\bar{x}(s)}\Big)\Big] \\ &+ (1 - \widehat{\lambda}_{ML}(\bar{y}(s)))N[(\bar{y}(s)/\bar{x}(s))x(\bar{s}), \tau^2 D(\bar{s})] \end{aligned} \tag{2.6}$$

*where for $0 \le \varepsilon < 1$,*

$$\begin{aligned} &\widehat{\lambda}_{ML}^{-1}(\bar{y}(s)) \\ &= 1 + \varepsilon(1-\varepsilon)^{-1}B_0^{-\frac{1}{2}}(s)\exp[nB_0(s)(\bar{y}(s) - \beta_0\bar{x}(s))^2/(2\tau^2\bar{x}(s))]. \end{aligned}$$

The proof of this theorem involves some heavy algebra, and is postponed to the Appendix. It follows from this theorem that the robust Bayes estimator of $\gamma(y)$ under the ML-II prior $\widehat{\pi}_{ML}$ simplifies to

$$\begin{aligned} e_{RB}(s,y(s)) &= f\bar{y}(s)+(1-f)\bar{x}(\bar{s})\{(1-\widehat{\lambda}_{ML}(\bar{y}(s))B_0(s)\bar{y}(s)/\bar{x}(s) \\ &\quad +\widehat{\lambda}_{ML}(\bar{y}(s))B_0(s)\beta_0\}. \end{aligned} \tag{2.7}$$

We now turn to the empirical Bayes analysis. The empirical Bayes analysis is closely related to the robust Bayes analysis in the sense that in the former analysis, the prior distribution is assumed to belong to some class of distributions. Contrary to the robust Bayes anaysis where $\varepsilon$ is typically taken to be very small, in an empirical Bayes analysis $\varepsilon$ is taken as 1. This point is very clearly brought out in Berger and Berliner (1984).

To derive an empirical Bayes estimator, we consider the model given in the robust Bayes analysis with $\varepsilon = 1$, but assume that $Q$ is the class of $\{\mathrm{N}(\beta_0,\sigma^2),\sigma^2 > 0\}$ priors. Observe that if $\beta$ has the $N(\beta_0,\sigma^2)$ prior, then calculations similar to (2.2) lead to the Bayes estimator

$$e_B(s,y(s)) = f\bar{y}(s)+(1-f)\bar{x}(\bar{s})\{(1-B(s))\bar{y}(s)/\bar{x}(s)+B(s)\beta_0\} \tag{2.8}$$

of $\gamma(y)$. In the above $B(s) = M/(M+n\bar{x}(s))$ and $M = \tau^2/\sigma^2$. In an empirical Bayes analysis, we estimate $B(s)$ from the marginal distribution of $y(s)$. We may note that marginally $\bar{y}(s)$ is sufficient for $\sigma^2$ and $\bar{y}(s) \sim \mathrm{N}[\beta_0\bar{x}(s), \frac{\tau^2\bar{x}(s)}{nB(s)}]$. Hence, $E[n(\bar{y}(s)-\beta_0\bar{x}(s))^2/(\tau^2\bar{x}(s))] = B^{-1}(s)$. Thus, $B(s)$ is estimated by

$$\hat{B}(s) = \min\left\{1, \frac{\tau^2\bar{x}(s)}{n(\bar{y}(s)-\beta_0\bar{x}(s))^2}\right\}. \tag{2.9}$$

Accordingly, an empirical Bayes estimator of $\gamma(y)$ is given by

$$e_{EB}(s,y(s)) = f\bar{y}(s)+(1-f)\bar{x}(\bar{s})\{(1-\hat{B}(s))\bar{y}(s)/\bar{x}(s)+\hat{B}(s)\beta_0\}. \tag{2.10}$$

In practice, however, $\tau^2$ is not usually known. In such situations, one can conceive of an inverse gamma prior for $\tau^2$ independent of the prior for $\beta$ to derive a Bayes estimator of $\gamma(y)$. In a robust Bayes approach, if one assumes a mixture of a normal-gamma prior for $(\beta,\sigma^2)$, then the ML-II prior puts its mass on the point $(\bar{y}(s)/\bar{x}(s), n^{-1}\sum_{i\in s}(y_i - x_i\bar{y}(s)/\bar{x}(s))^2/x_i)$. In an empirical Bayes approach, one estimate $\tau^2/B(s)$ by MSB $= n(\bar{y}(s)-\beta_0\bar{x}(s))^2/\bar{x}(s)$, and $\tau^2$ by MSW $= \sum_{i\in s}\{(y_i-\bar{y}(s)/\bar{x}(s)x_i)^2/x_i\}/(n-1)$. The latter can be justified on the ground that $\sum_{i\in s}(y_i-\bar{y}(s)/\bar{x}(s)x_i)^2/x_i \sim \tau^2\chi^2_{n-1}$ so that $E[\mathrm{MSW}] = \tau^2$. Hence, an estimator of $B(s)$ is given in this case by $\hat{B}_*(s) = \min\{1, \text{MSW}/\text{MSB}\}$, and the corresponding empirical Bayes estimator is obtained by substituting $\hat{B}_*(s)$ for $\hat{B}(s)$ in (2.10). We have not, however, pursued here the resulting Bayesian analysis.

## 3. COMPARISON OF THE ESTIMATORS BASED ON BAYESIAN ROBUSTNESS

In this section, we compare the performance of the subjective Bayes estimator $e_0$, the classical ratio estimator $e_R$, the robust Bayes estimator $e_{RB}$ and the empirical Bayes estimator $e_{EB}$ of $\gamma(y)$ from the robustness perspective. The main idea is that we want examine whether these estimators perform satisfactorily over a broad class of priors.

With this end, we consider the $\{N(\beta_0, \sigma^2), \sigma^2 > 0\}$ class of priors. The choice of this class of priors may be justified as follows. Very often, based on prior elicitation, one can take a fairly accurate guess at the prior mean. However, the same need not necessarily be true for the prior variance, where there is a greater chance of vagueness. Note that when $\sigma^2 \neq \sigma_0^2$, none of the estimators $e_R$, $e_0$, $e_{RB}$, or $e_{EB}$ is the optimal (Bayes) estimator. For a typical member $N(\beta_0, \sigma^2)$ of this class, the Bayes estimator of $\gamma(y)$ is given by (2.8).

We now introduce the general definition of posterior robustness. For a given prior $\xi$, denote by $\rho(\xi, (s, y(s)), e)$ the posterior risk of any arbitrary estimator $e(s, y(s))$ of $\gamma(y)$ under squared error loss. The following definition is taken from Berger (1984).

**Definition 3.1.** *An estimator $e_0(s, y(s))$ of $\gamma(y)$ is $\zeta$-posterior robust with respect to $\Gamma$ if for the observed $(s, y(s))$,*

$$POR_\Gamma(e_0) = \sup_{\xi \in \Gamma} |\rho(\xi, (s, y(s)), e_0) - \inf_{a \in \mathcal{A}} \rho(\xi, (s, y(s)), a)| \leq \zeta. \tag{3.1}$$

We shall, henceforth, refer to the left hand side of (3.1) as the posterior robustness index of the estimator $e_0(s, y(s))$ of $\gamma(y)$ under the class of priors $\Gamma$. $\text{POR}_\Gamma(e_0)$ in a sense is the sensitivity index of the estimator $e_0$ of $\gamma(y)$ as the prior varies over $\Gamma$. For any given $\zeta > 0$, it is very clear that whether or not posterior robustness exits will often depend on which $(s, y(s))$ is observed. This will be revealed in the subsequent calculations.

With this end, first note that under the $N(\beta_0, \sigma^2)$ prior denoted by $\xi_{\sigma^2}$, the posterior risk of any estimator $e$ of $\gamma(y)$ is

$$\rho(\xi_{\sigma^2},(s,y(s)),e)=\rho(\xi_{\sigma^2},(s,y(s)),e_B)+(e-e_B)^2 \tag{3.2}$$

where $e_B$ is given in (2.8). Using (3.1) and (3.2) one gets for the class $\Gamma=\{\xi_{\sigma^2}:\sigma^2>0\}$ of priors

$$\begin{aligned}\mathrm{POR}_\Gamma(e_R) &= \sup_{0<B(s)<1}(1-f)^2\bar{x}^2(\bar{s})B^2(s)[\bar{y}(s)/\bar{x}(s)-\beta_0]^2\\ &= (1-f)^2\bar{x}^2(\bar{s})[\bar{y}(s)/\bar{x}(s)-\beta_0]^2;\end{aligned} \tag{3.3}$$

$$\begin{aligned}\mathrm{POR}_\Gamma(e_0) &= \sup_{0<B(s)<1}(1-f)^2\bar{x}^2(\bar{s})(B(s)-B_0(s))^2[\bar{y}(s)/\bar{x}(s)-\beta_0]^2\\ &= (1-f)^2\bar{x}^2(\bar{s})\max[B_0^2(s),(1-B_0(s))^2]\\ &\qquad\times[\bar{y}(s)/\bar{x}(s)-\beta_0]^2;\end{aligned} \tag{3.4}$$

$$\begin{aligned}\mathrm{POR}_\Gamma(e_{RB}) &= \sup_{0<B(s)<1}(1-f)^2\bar{x}^2(\bar{s})(\widehat{\lambda}_{ML}(\bar{y}(s))B_0(s)-B(s))^2\\ &\qquad\times[\bar{y}(s)/\bar{x}(s)-\beta_0]^2\\ &= (1-f)^2\bar{x}^2(\bar{s})\max[\widehat{\lambda}^2_{ML}(\bar{y}(s))B_0^2(s),(1-\widehat{\lambda}_{ML}(\bar{y}(s))B_0(s))^2]\\ &\qquad\times[\bar{y}(s)/\bar{x}(s)-\beta_0]^2;\end{aligned} \tag{3.5}$$

$$\begin{aligned}\mathrm{POR}_\Gamma(e_{EB}) &= \sup_{0<B(s)<1}(1-f)^2\bar{x}^2(\bar{s})(\hat{B}(s)-B(s))^2[\bar{y}(s)/\bar{x}(s)-\beta_0]^2\\ &= (1-f)^2\bar{x}^2(\bar{s})\max[\hat{B}^2(s),(1-\hat{B}(s))^2]\\ &\qquad\times[\bar{y}(s)/\bar{x}(s)-\beta_0]^2.\end{aligned} \tag{3.6}$$

Note from (3.3) - (3.6) that if we allow all possible distributions $N(\beta,\sigma^2)$, where $\beta$ is widely different from $\beta_0$ as our priors, all POR indices can become prohibitively large because the supremum over $\beta$ (real) becomes $+\infty$. Hence, we have considered the $\{N(\beta_0,\sigma^2),\sigma^2>0\}$ class of priors only. From (3.3) - (3.6), it is clear that $\max[\mathrm{POR}_\Gamma(e_0),\ \mathrm{POR}_\Gamma(e_{RB}),\ \mathrm{POR}_\Gamma(e_{EB})]\le\mathrm{POR}_\Gamma(e_R)$ so that the subjective Bayes, robust Bayes and empirical Bayes estimators all achieve a greater degree of posterior robustness than the classical ratio estimator. However, there is no clear winner among $e_0$, $e_{RB}$ and $e_{EB}$ in terms of posterior robustness. That is, the ratio of $\mathrm{POR}_\Gamma$ among $e_0$, $e_{RB}$ and $e_{EB}$ can take values both larger and smaller than 1 depending on the particular $(s, y(s))$.

Although the Bayesian thinks conditionally on $(s, y(s))$, it seems quite sensible to use the overall Bayes risk as a suitable robustness criteria, at least at a preexperimental stage. This issue is also addressed in Berger (1984) who introduced also the criterion of procedure robustness. He gives the following definition.

**Definition 3.2.** *An estimator $e_0(s,\ y(s))$ of $\gamma(y)$ is said to be $\zeta$-procedure robust with respect to $\Gamma$ if*

$$PR_\Gamma(e_0) = \sup_{\xi\in\Gamma} |r(\xi, e_0) - \inf_{a\in\mathcal{A}} r(\xi, a)| < \zeta, \tag{3.7}$$

*where $r$ denotes the Bayes risk.*

We shall, henceforth, refer to $\mathrm{PR}_\Gamma(e_0)$ as the procedure robustness index of $e_0$. Simple calculations yield for the class $\Gamma = \{\xi_{\sigma^2} : \sigma^2 > 0\}$ of priors

$$\begin{aligned} \mathrm{PR}_\Gamma(e_R) &= \sup_{0<B(s)<1} (1-f)^2\bar{x}^2(\bar{s})\tau^2(n\bar{x}(s))^{-1}B(s) \\ &= (1-f)^2\bar{x}^2(\bar{s})\tau^2(n\bar{x}(s))^{-1}; \end{aligned} \tag{3.8}$$

$$\begin{aligned} \mathrm{PR}_\Gamma(e_{B_0}) &= \sup_{0<B(s)<1} (1-f)^2\bar{x}^2(\bar{s})\tau^2(n\bar{x}(s))^{-1}(B_0(s)-B(s))^2/B(s) \\ &= +\infty; \end{aligned} \tag{3.9}$$

$$\mathrm{PR}_\Gamma(e_{RB}) = \sup_{0<B(s)<1} (1-f)^2\bar{x}^2(\bar{s})E[(B_0(s)\widehat{\lambda}_{ML}(\bar{y}(s)) - B(s))^2(\bar{y}(s)/\bar{x}(s)-\beta_0)^2] \tag{3.10}$$

$$\mathrm{PR}_\Gamma(e_{EB}) = \sup_{0<B(s)<1} (1-f)^2\bar{x}^2(\bar{s})E[(\hat{B}(s) - B(s))^2(\bar{y}(s)/\bar{x}(s)-\beta_0)^2]. \tag{3.11}$$

It is thus clear that the subjective Bayes estimator $e_0$ lacks procedure robustness, while the ratio estimator $e_R$ is quite procedure robust. The procedure robustness of $e_{RB}$ can be examined on the basis of the following theorem.

**Theorem 3.1.** *$E[(B_0(s)\widehat{\lambda}_{ML}(\bar{y}(s)) - B(s))^2(\bar{y}(s)/\bar{x}(s) - \beta_0)^2] = O_e(B^{1/2}(s))$, for every $\varepsilon > 0$, where $O_e$ denotes the exact order.*

The proof of this theorem is technical, and is deferred to the Appendix. In view of (3.10) and Theorem 3.1. it appears that $e_R$ has distinct advantage over $e_{RB}$ in terms of procedure robustness, especially for small B. This is not surprising though since small B signifies small $M = \tau^2/\sigma^2$ which amounts to greater instability in the assessment of the prior distribution of $\beta$ relative to the superpopulation model. It is not surprising that in such circumstances, it is safer to use $e_R$ for estimating $\gamma(y)$ if one is seriously concerned about the long-run performance of the estimator.

To examine the procedure robustness of $e_{EB}$, we proceed as follows.

**Theorem 3.2.** *$E[(\hat{B}(s)-B(s))^2(\bar{y}(s)/\bar{x}(s)-\beta_0)^2] = O(B^{1/2-\eta}(s))$, where $\eta(0 < \eta < 1/2)$ can be made arbitrarily small.*

The proof of this theorem is also technical, and is deferred to the Appendix. Theorem 3.1 and Theorem 3.2 clearly demonstrate the procedure robustness of $e_{RB}$ and $e_{EB}$ as $B(s) \to 0$. It follows from the above theorems that as $n \to \infty$, i.e., $B(s) \to 0$, under the $\xi_{\sigma^2}$ prior, the robust Bayes and empirical Bayes procedures are asymptotically optimal in the sense of Robbins (1955).

The summary of our findings of this section is as follows. The subjective Bayes, robust Bayes and empirical Bayes estimators are all more posterior robust than the classical ratio estimator. However, the subjective Bayes estimator fails miserably according to the criterion of procedure robustness, while the robust Bayes and empirical Bayes estimators do not suffer from that drawback. The main point is that both the robust Bayes and empirical Bayes estimators are strong contenders of the classical as well as the subjective Bayes estimators.

## 4. CHOICE OF SAMPLING DESIGN

From a Bayesian point of view, one selects those units $i$ for which the posterior risk of the Bayes estimator of $\gamma(y)$ is minimized. Denote $\xi_0$ by the $N(\beta_0, \sigma_0^2)$ prior. For the subjective Bayes estimator as given in (2.2), the posterior risk is given by

$$\rho(\xi_0, (s, y(s)), e_0) = N^{-2}[\sum_{i\in\bar{s}} \tau^2(x_i + \frac{x_i^2}{M_0 + n\bar{x}(s)}) + \sum_{\substack{i \in \bar{s}, i' \in \bar{s} \\ i \neq i'}} \frac{\tau^2 x_i x_i'}{M_0 + n\bar{x}(s)}], \quad (4.1)$$

which is minimized by selecting those units $i$ in the sample for which the $x_i$ values are the largest. Royall and Pfefferman (1982) have considered the special case of (4.1) with $M_0 = 0$, i.e., when $\beta$ has the uniform $(-\infty, \infty)$ prior.

Next, we compute the posterior risk of the robust Bayes estimator under the $\varepsilon$-contaminated prior $\pi$. Denoting this prior by $\xi_{ML}$, it follows that the posterior risk of $e_{RB}$ is given by

$$\rho(\xi_{ML}, (s, y(s)), e_{RB}) = N^{-2}[\sum_{i\in\bar{s}} V_{\xi_{ML}}(y_i|s, y(s)) + \sum_{\substack{i \in \bar{s}, i' \in \bar{s} \\ i \neq i'}} Cov_{\xi_{ML}}(y_i, y_i'|s, y(s))] \quad (4.2)$$

Generalizing the formula (1.8) given in Berger and Berliner (1983), one gets for $i \in \bar{s}, i' \in \bar{s}, i \neq i'$,

$$V_{\xi_{ML}}(y_i|s, y(s)) = \tau^2[x_i + x_i^2 \frac{\widehat{\lambda}_{ML}(\bar{y}(s))}{M_0 + n\bar{x}(s)}] + \widehat{\lambda}_{ML}(\bar{y}(s))(1 - \widehat{\lambda}_{ML}(\bar{y}(s)))x_i^2 B_0^2(s)\left(\frac{\bar{y}(s)}{\bar{x}(s)} - \mu_0\right)^2. \quad (4.3)$$

and

$$Cov_{\xi_{ML}}(y_i, y_i'|s, y(s)) = \tau^2 x_i x_i' \frac{\widehat{\lambda}_{ML}(\bar{y}(s))}{M_0 + n\bar{x}(s)} + \widehat{\lambda}_{ML}(\bar{y}(s))(1 - \widehat{\lambda}_{ML}(\bar{y}(s)))x_i x_i' B_0^2(s)\left(\frac{\bar{y}(s)}{\bar{x}(s)} - \beta_0\right)^2. \quad (4.4)$$

From (4.2) - (4.4), one gets

$$\rho(\xi_{ML}, (s, y(s)), e_{RB}) = N^{-2}[\tau^2 \left((N-n)\bar{x}(\bar{s}) + (N-n)^2\bar{x}^2(\bar{s})\frac{\widehat{\lambda}_{ML}(\bar{y}(s))}{M_0 + n\bar{x}(s)}\right) + \widehat{\lambda}_{ML}(\bar{y}(s))(1 - \widehat{\lambda}_{ML}(\bar{y}(s)))(N-n)^2\bar{x}^2(\bar{s})B_0^2(s)\left(\frac{\bar{y}(s)}{\bar{x}(s)} - \mu_0\right)^2. \quad (4.5)$$

If we want to select those units $i$ in the sample which minimize (4.5), the objective is clearly unachievable without knowing all the $y_i$'s along with all the $x_i$'s. A more sensible approach will be to minimize

$$r(\xi_{ML}, e_{RB}) = E_\pi \rho(\xi_{ML}, (s, y(s)), e_{RB})$$

where the expectation is taken over the marginal distribution of $y(s)$, when $\pi$ is the prior. Noting that

$$\widehat{\lambda}_{ML}(\bar{y}(s)) = (1-\varepsilon)m(y(s)|\pi_0)/\{(1-\varepsilon)m(y(s)|\pi_0) + \varepsilon m(y(s)|\hat{q}_s)\}, \tag{4.6}$$

where $m(y(s)|q)$ denotes the marginal pdf of $y(s)$ under the prior $q$, it follows that

$$\begin{aligned} &r(\xi_{ML}, e_{RB}) \\ &= N^{-2}\tau^2[(N-n)\bar{x}(\bar{s}) + (N-n)^2\bar{x}^2(\bar{s})\frac{1-\varepsilon}{M_0 + n\bar{x}(s)}] \\ &+(1-\varepsilon)N^{-2}(N-n)^2\bar{x}^2(\bar{s})B_0^2(s)E_{\pi_0}\left[(1-\widehat{\lambda}_{ML}(\bar{y}(s)))\frac{(\bar{y}(s)-\beta_0\bar{x}(s))^2}{\bar{x}^2(\bar{s})}\right]. \end{aligned} \tag{4.7}$$

It follows after some simplifications that under the prior $\pi_0$, marginally $\bar{y}(s) \sim N[\mu_0\bar{x}(s), \frac{\tau^2\bar{x}(s)}{nB_0(s)}]$. Accordingly, under the prior $\pi_0$, marginally $nB_0(s)[\bar{y}(s) - \beta_0\bar{x}(s)]^2/(\tau^2\bar{x}(s)) \sim \chi_1^2$. Then

$$\begin{aligned} &E_{\pi_0}\left[(1-\widehat{\lambda}_{ML}(\bar{y}(s)))\frac{(\bar{y}(s)-\beta_0\bar{x}(s))^2}{\bar{x}^2(\bar{s})}\right] \\ &= \frac{\tau^2}{n\bar{x}(s)B_0(s)}E\left[\frac{\frac{\varepsilon}{1-\varepsilon}\exp(U/2)}{B_0^{\frac{1}{2}}(s) + \frac{\varepsilon}{1-\varepsilon}\exp(U/2)}U\right] \end{aligned} \tag{4.8}$$

where $U \sim \chi_1^2$. Hence,

$$\begin{aligned} &r(\xi_{ML}, e_{RB}) \\ &= N^{-2}\tau^2[(N-n)\bar{x}(\bar{s}) + (N-n)^2\bar{x}^2(\bar{s})\frac{1-\varepsilon}{M_0 + n\bar{x}(s)}] \\ &+(1-\varepsilon)N^{-2}(N-n)^2\bar{x}^2(\bar{s})\frac{B_0(s)}{n\bar{x}^{1/2}(s)}E\left[\frac{\frac{\varepsilon}{1-\varepsilon}\exp(U/2)}{(\bar{x}(s)B_0(s))^{1/2} + \frac{\varepsilon}{1-\varepsilon}\exp(U/2)}U\right]. \end{aligned} \tag{4.9}$$

Since $B_0(s)/\bar{x}(s) = M_0/\{\bar{x}(s)(M_0+n\bar{x}(s))\}$ is decreasing in $\bar{x}(s)$ and $B_0(s)\bar{x}(s) = M_0\bar{x}(s)/(M_0 + n\bar{x}(s))$ is increasing in $\bar{x}(s)$, it follows from (4.9) that even from a robust Bayesian viewpoint, one should select those units $i$ for which the $x_i$ values are the largest. This suggests some robustness in the choice of designs even in a subjective Bayesian analysis.

## 5. AN EXAMPLE

The example in this section considers one of the six real populations which are used in Royall and Cumberland (1981) for an empirical study of the ratio estimator and estimates of its variance. Our population consists of the 1960 and 1970 population, in millions, of 125 US cities with 1960 population between 100,000 and 1,000,000. Here the auxiliary information is the 1960 population.

The problem is to estimate the mean (or total) number of inhabitants in those 125 cities in 1970. For the complete population in 1970, we find that the population mean is 0.29034. We select 20% simple random sample without replacement from this population. So the sample size is n=25. Also, we are using $\sigma^2 = (N-1)^{-1}\sum_{i=1}^{N}(y_i - \beta x_i)^2 = 4.84844\times10^{-3}$ which is assumed to be known. We can obtain easily the ratio estimate. To do a Bayesian analysis, we use both 1950 and 1960 populations in 125 cities to elicit the base prior $\pi_0$ for $\beta$. The elicited prior $\pi_0$ is the N(1.15932, $1.21097\times10^{-3}$) distribution based on prior information. Under this elicited prior $\pi_0$, we use formulas (2.2) to obtain the subjective Bayes estimate. But we have some uncertainty in $\pi_0$ and the prior information, so we choose $\varepsilon = .1$ and we get the robust Bayes estimate using formula (2.7). Also we can obtain easily the empirical Bayes estimate. A number of samples are tried and we have reported our analysis for one sample for illustration purpose. Table 5.1 provides the classical ratio estimate $e_R$, the subjective Bayes estimate $e_0$, the robust Bayes estimate $e_{RB}$ and the empirical Bayes estimate $e_{EB}$. Table 1 also provides the posterior robustness index for each estimate which in a sense the sensitivity index of the estimate as the prior varies over the class $\{N(\beta_0, \sigma^2), \sigma^2 > 0\}$.

Table 5.1: Estimates, Closeness and Posterior Robustness Index

| | Predictor | $\lvert\gamma(y) - e\rvert$ | POR |
|---|---|---|---|
| $e_R$ | 0.28426 | $6.08452\times10^{-3}$ | $5.38660\times10^{-4}$ |
| $e_0$ | 0.29336 | $3.01418\times10^{-3}$ | $4.57488\times10^{-4}$ |
| $e_{RB}$ | 0.28660 | $3.74880\times10^{-3}$ | $4.35696\times10^{-4}$ |
| $e_{EB}$ | 0.29032 | $5.83622\times10^{-5}$ | $3.29786\times10^{-4}$ |

An inspection of Table 1 reveals that the subjective Bayes estimator $e_0$ is closest to $\gamma(y)$, but not much good in the posterior robustness. The robust Bayes estimator $e_{RB}$ and the empirical Bayes estimator $e_{EB}$ are well behaved in the sense that they are closer to $\gamma(y)$ than at least the classical ratio estimate $e_R$ and good in the posterior robustness index. The classical ratio estimate $e_R$ is the worst in terms of both the closeness to $\gamma(y)$ and the posterior robustness index.

## APPENDIX

**Proof of Theorem 2.1.** First observe that the joint distribution of $y(s)$ and $y(\bar{s})$ is given by

$$\begin{pmatrix} y(s) \\ y(\bar{s}) \end{pmatrix} = \mathrm{N}\left[\begin{pmatrix} \beta_0 x(s) \\ \beta_0 x(\bar{s}) \end{pmatrix}, \begin{pmatrix} \tau^2 D(s)+\sigma_0^2 x(s)x^T(s) & \sigma_0^2 x(s)x^T(\bar{s}) \\ \sigma_0^2 x(\bar{s})x^T(s) & \tau^2 D(\bar{s})+\sigma_0^2 x(\bar{s})x^T(\bar{s}) \end{pmatrix}\right]. \tag{A.1}$$

Part (i) of Theorem 2.1 follows immediately from (A.1). Again, from (A.1), it follows that $y(\bar{s})|s, y(s) \sim \mathrm{N}(m, V)$, where

$$m = \beta_0 x(\bar{s}) + \sigma_0^2 x(\bar{s})x^T(s)[\tau^2 D(s) + \sigma_0^2 x(s)x^T(s)]^{-1}(y(s) - \beta_0 x(s)) \tag{A.2}$$

and

$$V = \tau^2 D(\bar{s}) + \sigma_0^2 x(\bar{s})x^T(\bar{s}) - \sigma_0^2 x(\bar{s})x^T(s)[\tau^2 D(s) + \sigma_0^2 x(s)x^T(s)]^{-1}\sigma_0^2 x(s)x^T(\bar{s}). \tag{A.3}$$

Now,

$$[\tau^2 D(s) + \sigma_0^2 x(s)x^T(s)]^{-1} = \tau^{-2}\left[D^{-1}(s) - \frac{\sigma_0^2 D^{-1}(s)x(s)x^T(s)D^{-1}(s)}{\tau^2 + \sigma_0^2 x^T(s)D^{-1}(s)x(s)}\right]. \tag{A.4}$$

Hence, using (A.4), it follows from (A.2) that

$$\begin{aligned} m &= \beta_0 x(\bar{s}) + \sigma_0^2 x(\bar{s})\tau^{-2}\left(1 - \frac{\sigma_0^2 x^T(s)D^{-1}(s)x(s)}{\tau^2 + \sigma_0^2 x^T(s)D^{-1}(s)x(s)}\right) \\ &\qquad \times x^T(s)D^{-1}(s)(y(s) - \beta_0 x(s)) \\ &= \beta_0 x(\bar{s}) + \frac{\sigma_0^2 x^T(s)D^{-1}(s)(y(s) - \beta_0 x(s))}{\tau^2 + \sigma_0^2 x^T(s)D^{-1}(s)x(s)} x(\bar{s}) \\ &= \frac{\beta_0\tau^2 + \sigma_0^2 \sum_{i\in s} y_i x_i/h(x_i)}{\tau^2 + \sigma_0^2 \sum_{i\in s} x_i^2/h(x_i)} x(\bar{s}); \end{aligned} \tag{A.5}$$

$$\begin{aligned} V &= \tau^2 D(\bar{s}) + \sigma_0^2 \left(1 - \frac{\sigma_0^2 x^T(s) D^{-1}(s) x(s)}{\tau^2 + \sigma_0^2 x^T(s) D^{-1}(s) x(s)}\right) x(\bar{s}) x^T(\bar{s}) \\ &= \tau^2 \left[D(\bar{s}) + \frac{\sigma_0^2 x(\bar{s}) x^T(\bar{s})}{\tau^2 + \sigma_0^2 \sum_{i\in s} x_i^2/h(x_i)}\right]. \qquad (A.6) \end{aligned}$$

Part (ii) of Theorem 2.1 follows now from (A.5) and (A.6).

**Proof of Theorem 2.2.** The proof of the first part is straightforward, and is omitted. The conditional distribution of $y(\bar{s})$ given $(s, y(s))$ is

$$f(y(\bar{s})|s, y(s)) = \int_{-\infty}^{\infty} f(y(\bar{s})|\beta)\hat{\pi}_{ML}(\beta|s, y(s))d\beta \qquad (A.7)$$

where

$$\begin{aligned} &\hat{\pi}_{ML}(\beta|s, y(s)) \\ &= \frac{(1-\varepsilon)m(y(s)|\pi_0)}{m(y(s)|\hat{\pi}_{ML})}\pi_0(\beta|s, y(s)) + \frac{\varepsilon m(y(s)|\hat{q}_s)}{m(y(s)|\hat{\pi}_{ML})}\hat{q}_s(\beta|s, y(s)) \\ &= \hat{\lambda}(y(s))\pi_0(\beta|s, y(s)) + [1 - \hat{\lambda}(y(s))]\hat{q}_s(\beta|s, y(s)), \text{ (say)}. \end{aligned}$$

Then,

$$\begin{aligned} f(y(\bar{s})|s, y(s)) &= \hat{\lambda}(y(s)) \int_{-\infty}^{\infty} f(y(\bar{s})|\beta)\pi_0(\beta|s, y(s))d\beta \\ &\quad +(1 - \hat{\lambda}(y(s))) \int_{-\infty}^{\infty} f(y(\bar{s})|\beta)\hat{q}_s(\beta|s, y(s))d\beta. \qquad (A.8) \end{aligned}$$

By Theorem 2.1(ii), $\int_{-\infty}^{\infty} f(y(\bar{s})|\beta)\pi_0(\beta|s, y(s))d\beta$ is $\mathrm{N}[(B_0(s)\beta_0 + (1-B_0(s))\frac{\bar{y}(s)}{\bar{x}(s)})$ $x(\bar{s}), \tau^2(D(\bar{s}) + \frac{x(\bar{s})x^T(\bar{s})}{M_0 + n\bar{x}(s)})]$ pdf, while $\int_{-\infty}^{\infty} f(y(\bar{s})|\beta)\hat{q}_s(\beta|s, y(s))d\beta$ gives $\mathrm{N}[\frac{\bar{y}(s)}{\bar{x}(s)}x(\bar{s})$ $\tau^2 D(\bar{s})]$ pdf. Now,

$$\begin{aligned} &\hat{\lambda}^{-1}(y(s)) = 1 + \frac{\varepsilon}{1-\varepsilon}\{\tau^{-n}(\prod_{i\in s} x_i^{-1/2})\exp[-\sum_{i\in s}(y_i - \frac{\bar{y}(s)}{\bar{x}(s)}x_i)^2/(2\tau^2 x_i)]\} \\ &\{|\tau^2 D(s) + \sigma_0^2 x(s) x^T(s)|^{-1/2} \\ &\exp[-\frac{1}{2}(y(s) - \beta_0 x(s))^T(\tau^2 D(s) + \sigma_0^2 x(s)x^T(s))^{-1}(y(s) - \beta_0 x(s))]\}. \end{aligned} \qquad (A.9)$$

Using (A.4) and

$$\begin{aligned} |\tau^2 D(s) + \sigma_0^2 x(s) x^T(s)| &= \begin{vmatrix} \tau^2 D(s) & \sigma_0 x(s) \\ -\sigma_0 x(s) & 1 \end{vmatrix} \\ &= |\tau^2 D(s)||1 + \frac{\sigma_0^2}{\tau^2} x^T(s) D^{-1}(s) x(s)| = \frac{\tau^{2n}\prod_{i\in s} x_i}{B_0(s)}, \end{aligned}$$

it follows from (A.9) after some simplifications that $\hat{\lambda}^{-1}(y(s)) = \hat{\lambda}_{ML}^{-1}(\bar{y}(s))$.

**Proof of Theorem 3.1.** Note that $n[\bar{y}(s) - \beta_0\bar{x}(s)]^2 \sim \frac{\tau^2\bar{x}(s)}{B(s)}\chi_1^2$. Then

$$E[\{B_0(s)\hat{\lambda}_{ML}(\bar{y}(s)) - B(s)\}^2\}(\bar{y}(s) - \beta_0\bar{x}(s))^2/\bar{x}^2(s)]$$
$$= \int_0^\infty \{\frac{B_0(s)}{1 + g(s)\exp\{\frac{B_0(s)u}{2B(s)}\}} - B(s)\}^2 \frac{\tau^2}{n\bar{x}(s)B(s)} u\exp(-\frac{u}{2})\frac{u^{\frac{1}{2}-1}}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})}du \tag{A.10}$$

where $g(s) = (\varepsilon/(1-\varepsilon))B_0^{-1/2}(s)$. Next observe that

$$\text{rhs of (A.10)}$$
$$\leq \int_0^\infty [\frac{B_0^2(s)}{(1 + g(s)\exp\{\frac{B_0(s)u}{2B(s)}\})^2} + B^2(s)]\frac{\tau^2}{n\bar{x}(s)B(s)}\exp(-\frac{u}{2})\frac{u^{\frac{3}{2}-1}}{2^{\frac{3}{2}}\Gamma(\frac{3}{2})}du$$
$$\leq \frac{\tau^2}{n\bar{x}(s)}[\int_0^\infty \frac{B_0^2(s)B^{-1}(s)}{2g(s)\exp\{\frac{B_0(s)u}{2B(s)}\}}\exp(-\frac{u}{2})\frac{u^{\frac{3}{2}-1}}{2^{\frac{3}{2}}\Gamma(\frac{3}{2})}du + B(s)]$$
$$= \frac{\tau^2}{n\bar{x}(s)}[\frac{B_0^2(s)}{2g(s)}B^{-1}(s)\{1 + \frac{B_0(s)}{B(s)}\}^{-\frac{3}{2}} + B(s)] = O(B^{1/2}(s)). \tag{A.11}$$

Also, writing $g'(s) = \max\{1, g(s)\}$,

$$\text{lhs of (A.10)}$$
$$\geq \frac{\tau^2}{n\bar{x}(s)}\int_0^\infty [\frac{B_0^2(s)B^{-1}(s)}{[2g'(s)\exp\{\frac{B_0(s)u}{2B(s)}\}]^2} - \frac{2B_0(s)}{g(s)\exp\{\frac{B_0(s)u}{2B(s)}\}} + B(s)]$$
$$\times \exp(-\frac{u}{2})\frac{u^{\frac{3}{2}-1}}{2^{\frac{3}{2}}\Gamma(\frac{3}{2})}du$$
$$\geq \frac{\tau^2}{n\bar{x}(s)}[(\frac{B_0(s)}{2g'(s)})^2\frac{1}{B(s)}\{1 + \frac{2B_0(s)}{B(s)}\}^{-\frac{3}{2}} - \frac{2B_0(s)}{g(s)}\{1 + \frac{B_0(s)}{B(s)}\}^{-\frac{3}{2}} + B(s)]$$
$$= O(B^{1/2}(s)). \tag{A.12}$$

Conbining (A.11) and (A.12), Theorem 3.1 follows.

**Proof of Theorem 3.2.** Recall that MSB $= n[\bar{y}(s) - \beta_0\bar{x}(s)]^2/\bar{x}(s) \sim \frac{\tau^2}{B(s)}\chi_1^2$.

Noting that $\hat{B}(s) = \min\{1, \tau^2/MSB\} \stackrel{d}{=} \min\{1, B(s)/\chi_1^2\}$, one gets

$$E[(\hat{B}(s) - B(s))^2 \frac{(\bar{y}(s) - \beta_0\bar{x}(s))^2}{\bar{x}^2(s)}] = \frac{1}{n\bar{x}(s)} E[(\hat{B}(s) - B(s))^2 MSB]$$
$$\leq \frac{\tau^2}{n\bar{x}(s)}\{(1 - B(s))^2 P(\chi_1^2 < B(s)) + B(s)E[(\frac{1}{\chi_1^2} - 1)^2 \chi_1^2 I_{[\chi_1^2 > B(s)]}\}.$$

Now,

$$\begin{aligned} P(\chi_1^2 < B(s)) &= \int_0^{B(s)} \exp(-u/2) \frac{u^{1/2-1}}{2^{1/2}\Gamma(1/2)} du \\ &\leq \int_0^{B(s)} \frac{u^{1/2-1}}{2^{1/2}\Gamma(1/2)} = (2/\pi)^{1/2} B^{1/2}(s). \end{aligned}$$

Moreover,

$$E[(\frac{1}{\chi_1^2} - 1)^2 \chi_1^2 I_{[\chi_1^2 > B(s)]}$$
$$= \int_{B(s)}^{\infty} \frac{1}{u} e^{-\frac{u}{2}} \frac{u^{\frac{1}{2}-1}}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} du - 2P(\chi_1^2 > B(s)) + \int_{B(s)}^{\infty} u e^{-\frac{u}{2}} \frac{u^{\frac{1}{2}-1}}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} du$$
$$\leq B^{-\frac{1}{2}-\eta}(s) \int_{B(s)}^{\infty} e^{-\frac{u}{2}} \frac{u^{-\frac{1}{2}-1}}{2^{\frac{1}{2}}\Gamma(\frac{1}{2})} du + \int_{B(s)}^{\infty} e^{-\frac{u}{2}} \frac{u^{\frac{3}{2}-1}}{2^{\frac{3}{2}}\Gamma(\frac{3}{2})} du$$
$$\leq B^{-\frac{1}{2}-\eta}(s)\Gamma(\eta)/\sqrt{2\pi} + 1.$$

Combining (A.13) - (A.15), Theorem 3.2 follows.

## REFERENCES


Berger, J. O. (1984). "The robust Bayesian viewpoint (with discussion)," *Robustness of Bayesian Analysis*, Ed. J. Kadane. North-Holland, Amsterdam, 63-124.

Berger, J. O. and Berliner, L. M. (1986). "Robust Bayes and empirical Bayes analysis with $\varepsilon$-contaminated priors," *The Annals of Statistics*, 14, 461-486.

Cochran, W. G. (1977). *Sampling Techniques.* (3rd edn.). Wiley, New York.

Ericson, W. A. (1969). "Subjective Bayesian models in sampling finite populations (with discussion)," *The Journal of the Royal Statistical Society*, Ser. B, 31, 195-233.

Godambe, V. P. (1982). "Estimation in survey sampling: Robustness and optimality (with discussion)" *Journal of the American Statistical Association*, 77, 393-406.

Godambe, V. P. and Thompson, M. E. (1971). "The specification of prior knowledge by classes of prior distributions in survey sampling estimation," *Foundations of Statistical Inference*. Eds. V. P. Godambe and D. A. Sprott. Holt, Rinehart and Winston, Toronto, 243-258.

Good, I. J. (1965). *The Estimation of Probabilities: An Essay on Modern Bayesian Methods*. M. I. T. Press, Cambridge.

Robbins, H. (1955). "The empirical Bayes approach to statistics," *Proc. Third Berkely Symp. Math. Statist. Probab.*, 1, 157-164. University of California Press, Berkely.

Royall, R. M. (1970). "On finite population sampling theory under certain linear regression models," *Biometrika*, 57, 377-387.

Royall, R. M. (1971). "Linear regression models in finite population sampling theory," *Foundations of Statistical Inference*, Eds. V. P. Godambe and D. A. Sprott. Holt, Rinehart and Winston, Toronto, 259-279.

Royall, R. M. and Cumberland, W. G. (1981). "An empirical study of the ratio estimator and estimators of its variance (with discussion)," *Journal of the American Statistical Association*, **76**, 66-88.

Royall, R. M. and Pfeffermann, D. (1982). "Balanced samples and robust Bayesian inference in finite population sampling," *Biometrika*, 69, 401-409.