SEMIPARAMETRICS, NONPARAMETRICS AND EMPIRICAL BAYES PROCEDURES IN LINEAR MODELS

by

Pranab Kumar Sen

Department of Biostatistics
University of North Carolina

Institute of Statistics
Mimeo Series No. 2187

February 1998

SEMIPARAMETRICS, NONPARAMETRICS AND EMPIRICAL BAYES PROCEDURES IN LINEAR MODELS *†

PRANAB KUMAR SEN
University of North Carolina at Chapel Hill

In a classical parametric setup, a key factor in the implementation of the Empirical Bayes methodology is the incorporation of a suitable prior that is compatible with the parametric setup and yet lends to the estimation of the Bayes (shrinkage) factor in an empirical manner. The situation is more complex in semiparametric and (even more in) nonparametric models. Although the Dirichlet priors have been considered in some simple nonparametric models, in a general linear model there are certain limitations for such procedures, and alternative semiparametric methods have gained popularity in practice. Using first-order asymptotic representations for semiparametric and nonparametric estimators it is shown that a general Gaussian prior on the regression parameters can be readily adopted to formulate suitable empirical Bayes estimators that are essentially related to robust Stein-rule versions of such estimators which were introduced in the statistical lierature in a somewhat different perspective. Properties of such robust empirical Bayes estimators are studied.

1. INTRODUCTION

In the 1930's and 1940's, *nonparametrics* evolved mostly around some specific hypothesis testing problems, under the disguise of *quick and dirty* methods. They used to be referred to as *distribution-free* methods. Most of these developements in nonparametrics were due to researchers in sociology, psychometry, economics and allied fields, who advocated the use of such tools to emphasize the need for making less stringent regularity assumptions in order to enhance the scope of statistical conclusions that could be drawn from their (qualitative to quantitative) observational studies. During the past fifty years, there has been a spectacular growth of statistical research literature, focusing novel theory and methodology, as well as contemplating useful applications in various directions. In this context, Bayes and empirical Bayes procedures (mostly having a dominant parametric flavor) have received considerable attention from theoretical as well as applications points of view. It has been well recognised that in various interdisciplinary

*AMS *Subject classifications:* 62C12, 62F35, 62G35, 62H99, 62J99

†*Key words and phrases:* ADR; asymptotics; BAN estimators; FOADR; generalized Pitman closeness; L-, M- and R-estimators; Least squares estimators; regression quantiles; regression rank scores; robustness; shrinkage factor; Stein-rule estimator; Zellner prior.

scientific investigations, the scope for strictly parametric model based statistical analysis may be somewhat limited due to possible departures from such model assumptions, and often nonparametrics fare well in some respects. Though this feature has been repeatedly elaborated in contemporary texts and research publications, there seems to be some ambiguity regarding a clearcut demarcation of relative merits and demerits of parametric and nonparametric approaches. More recent advents in this domain include the so called *robust* procedures as well as *semiparametric* models. Therefore, in depicting this composite picture, we need to take into account the complexities arising from all corners. In the present study, we are particularly interested in semiparametrics and nonparametrics approaches that pertain to the empirical Bayes methodology.

In the current statistical literature (on theory, methodology and applications), regression (linear) models and their statistical analysis schemes have received the utmost attention. This has been primarily due to the fact that in most scientific, socio-economic and clinical studies, the basic theme of regression analysis has emerged as most appealing. In agricultural, physical, laboratory and other experimentation, it is generally contemplated that a causal relationship exists between the input and output variables, though that relation may be distorted to a certain extent due to measurement errors and other chance variations associated with the experimental schemes. The situation is far more complex for biomedical, clinical and environmental studies where experimentation may not be conductable in a precise controlled setup. In observational studies arising mostly in epidemiological investigations the scenario is even more vulnerable to various uncontrollable factors and calls for more complex statistical modeling and analysis schemes. All these observations pertain to two basic questions: *To what extent a simple parametric (such as a normal theory) linear model can be adopted in practice? Moreover, to what extent the empirical Bayes methodology percolates beyond the conventional parametric setups?* Often, the Box-Cox type of transformations are incorporated to induce linearity of regression relationships to a greater extent, though in that process, the adherence to normality of the error components may be partially confounded. Thus, even if a linear regression model is contemplated, a semiparametric model that allows the error distribution to be rather arbitrary may have a greater appeal from practical applications perspectives. The development of nonparametrics in the context of estimation theory has its genesis in rank based statistics [viz., Puri and Sen (1971, 1985), while semiparametrics have greater affinity to M-statistics that put more emphasis on robustness aspects [viz., Huber (1973)]. On the other hand, in an empirical Bayes setup, there is a subdued *testimator* flavor arising mainly due to the formulation of the shrinkage factor with an (empirical) Bayesian interpretation. The finite sample justifications for the estimation of such a Bayes (shrinkage) factor that can be made in a parametric setup may no longer be totally tenable in a semiparametric or nonparametric setup. Even in *generalized linear models* (GLM) that pertain to suitable densities belonging to the so called *exponential family* such interprepretations may not be universally true (viz., McCullagh and Nelder, 1989). In all these procedures, there may be a need to deemphasize the *likelihood principle* to a certain extent and emphasize robustness and efficiency aspects through alternative formulations. In such a setup, there may be some asymptotics flavor so as to rationalize suitable interpretations and to acieve some of the desired goals. Viewed from this perspective, we are tempted to examine the adoptability of the classical empirical Bayes approach in linear models when robust and/or nonparametric estimators are used instead of the classical *least squares* estimators (LSE) or normal theory *maximum likelihood estimators* (MLE).

Section 2 deals with a very brief review of the normal theory MLE and their Stein-rule versions, and in Section 3, following Ghosh, Saleh and Sen (1989), an empirical Bayes formulation is recapitulated. This provides the access to the incorporation of the Zellner (1986) g-prior for the formulation of empirical Bayes methodology for BAN (best asymptotically normal) estimators, and this is presented in Section 4. The main results on empirical Bayes estimators for semiparametric linear models are then discussed in Section 5. The concluding section deals with some general remarks and some related nonparametric models.

2. PRELIMINARY NOTION

Let us consider the classical (univariate) linear model:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}; \ \mathbf{e} = (e_1, \ldots, e_n)', \tag{2.1}$$

where the e_i are independent and identically distributed (i.i.d.) random variables (r.v.), $\boldsymbol{\beta}$ is an unknown parameter (p-vector, for some $p \geq 1$), $\mathbf{X}$ is a known (design) matrix of order $n \times p$ (whose row vectors are denoted by $\mathbf{x}_1, \ldots, \mathbf{x}_n$ respectively), and the observation vector $\mathbf{Y}$ has n independent elements $Y_1, \ldots, Y_n$. We are primarily intertested in the estimation of the parameter $\boldsymbol{\beta}$ when there may be some uncertain contraints on this parameter (vector).

In a classical parametric (normal theory) setup, it is assumed that

$$\mathbf{e} \sim \mathcal{N}_n(\mathbf{0}, \ \sigma^2\mathbf{I}_n), \quad 0 < \sigma^2 < \infty. \tag{2.2}$$

In this setup, the LSE and MLE of $\boldsymbol{\beta}$ are the same, and is given by

$$\hat{\boldsymbol{\beta}}_{LS} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}, \tag{2.3}$$

where $\mathbf{A}^-$ stands for a generalized inverse of $\mathbf{A}$. For simplicity of presentation (and without any loss of generality), we assume that $\mathbf{X}'\mathbf{X}$ is of full rank (p), so that the generalized inverse may be replaced by the usual inverse $(\mathbf{X}'\mathbf{X})^{-1}$. It is easy to verify that

$$E(\hat{\boldsymbol{\beta}}_{LS}) = \boldsymbol{\beta}; \ \ \mathbf{D}(\hat{\boldsymbol{\beta}}_{LS}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}. \tag{2.4}$$

For the above model, the log-likelihood function is given by

$$l_n(\boldsymbol{\beta}) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{1}{2\sigma^2}||\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}||^2, \tag{2.5}$$

so the information matrix (on $\boldsymbol{\beta}$) is given by

$$\mathcal{I}_n(\boldsymbol{\beta}) = \sigma^{-2}(\mathbf{X}'\mathbf{X}). \tag{2.6}$$

Thus, $\hat{\boldsymbol{\beta}}_{LS}$ is unbiased, efficient and sufficient for $\boldsymbol{\beta}$. However, led by the remarkable observation of Stein (1956), we may conclude that for $p \geq 3$, with respect to a *quadratic risk* function, $\hat{\boldsymbol{\beta}}_{LS}$ is not *admissible*; it can be dominated by a class of *shrinkage* estimators, now referred to as the Stein-rule estimators. Similarly, with respect to a *generalized Pitman closeness criterion* (GPCC), $\hat{\boldsymbol{\beta}}_{LS}$ is not admissible, for $p \geq 2$. In this case also, the Stein-rule estimators dominate the scenario. Motivated by this, we outline the genesis of Stein-rule (shrinkage) estimators relevant to the linear model considered in (2.1).

We rewrite (2.1) as

$$\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{e}, \tag{2.7}$$

where $\mathbf{X}$ is partitioned into $\mathbf{X}_1$ and $\mathbf{X}_2$ of order $n \times p_1$ and $n \times p_2$ respectively (so that $p = p_1 + p_2$), and similarly $\boldsymbol{\beta}$ is partitioned into $\boldsymbol{\beta}_1$ and $\boldsymbol{\beta}_2$ of order p_1 and p_2 respectively. We are primarily interested in estimating $\boldsymbol{\beta}_1$ when it is plausible that $\boldsymbol{\beta}_2$ is *close to* a specified $\boldsymbol{\beta}_2^o$ (that we can take without any loss of generality as $\mathbf{0}$). For example, in a factorial design, $\boldsymbol{\beta}_1$ may refer to the main-effects of the factors and $\boldsymbol{\beta}_2$ for the interaction-effects of various orders. A similar situation is encountered in a polynomial regression equation where $\boldsymbol{\beta}_1$ refers to the linear effects while the other component for the higher degree polynomial effects. Actually, in a canonical reduction of the parameter such a partition can always be posed as an alternative to setting a null hypothesis as $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$ where $\mathbf{C}$ is a $p_2 \times p$ matrix of known constants (and of rank p_2). Here, in particular, we have $\mathbf{C} = (\mathbf{0}, \mathbf{I}_{p_2})$. Then under the restraint that $\boldsymbol{\beta}_2 = \mathbf{0}$, the restricted MLE/LSE of $\boldsymbol{\beta}_1$ is given by

$$\hat{\boldsymbol{\beta}}_{RLS,1} = (\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'\mathbf{Y}, \tag{2.8}$$

and we parition the unrestricted estimator as

$$\hat{\boldsymbol{\beta}}_{LS} = (\hat{\boldsymbol{\beta}}'_{LS,1}, \boldsymbol{\beta}'_{LS,2})'. \tag{2.9}$$

Consider then the hypothesis testing problem for

$$H_0 : \boldsymbol{\beta}_2 = \mathbf{0} \quad \text{against} \quad H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}. \tag{2.10}$$

Let $\mathcal{L}_n$ be the conventional ANOVA (analysis of variance) test statistic, standardized in such a way that asymptotically under the null hypothesis it has the central chi squared distribution with p_2 degrees of freedom (DF). Then, typically a Stein-rule estimator of $\boldsymbol{\beta}_1$ can be posed as

$$\hat{\boldsymbol{\beta}}^S_{LS,1} = \hat{\boldsymbol{\beta}}_{RLS,1} + (1 - k/\mathcal{L}_n)(\hat{\boldsymbol{\beta}}_{LS,1} - \hat{\boldsymbol{\beta}}_{RLS,1}), \tag{2.11}$$

where $k(\geq 0)$ is a suitable shrinkage factor, often taken equal to $p_2 - 2$ when $p_2 \geq 2$. Side by side, we may also list the so called *positive-rule shrinkage* estimator:

$$\hat{\boldsymbol{\beta}}^{S+}_{LS,1} = \hat{\boldsymbol{\beta}}_{RLS,1} + (1 - k/\mathcal{L}_n)^+(\hat{\boldsymbol{\beta}}_{LS,1} - \hat{\boldsymbol{\beta}}_{RLS,1}), \tag{2.12}$$

where $a^+ = \max(a, 0)$. There are many other variants of such Stein-rule estimators. Among these, the *preliminary test estimator* (PTE) deserves mention. This can be posed as

$$\hat{\boldsymbol{\beta}}^{PT}_{LS,1} = \hat{\boldsymbol{\beta}}_{RLS,1} I(\mathcal{L}_n < \mathcal{L}_{n,\alpha}) + \hat{\boldsymbol{\beta}}_{LS,1} I(\mathcal{L}_N \geq \mathcal{L}_{n,\alpha}), \tag{2.13}$$

where $\mathcal{L}_{n,\alpha}$ stands for the critical value of $\mathcal{L}_{n,\alpha}$ at the level of significance $\alpha(0 < \alpha < 1)$. Technically, the Stein-rule estimator does not invovle the level of significance (α), and has some theoretical advantages over the PTE, though operationally a PTE may be more intuitive and convenient.

3. EMPIRICAL BAYES INTERPRETATIONS

We mainly follow the line of attack of Ghosh, Saleh and Sen (1989). In the usual Bayesian setup, we assume that

$$\mathbf{Y} \mid \boldsymbol{\beta} \sim \mathcal{N}_n(\mathbf{X}\boldsymbol{\beta},\ \sigma^2\mathbf{I}_n), \quad 0 < \sigma^2 < \infty, \tag{3.1}$$

and the *prior* Π on $\boldsymbol{\beta}$ is given by

$$\boldsymbol{\beta} \sim \mathcal{N}_p(\boldsymbol{\nu},\ \tau^2\mathbf{V}), \tag{3.2}$$

where $\mathbf{V}$ is positive definite (p.d.) and $0 < \tau^2 < \infty$. Then the *posterior* distribution of $\boldsymbol{\beta}$, given $\mathbf{Y}$, is

$$\mathcal{N}_p(\boldsymbol{\nu} + (\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\tau^2}\mathbf{V}^{-1})^{-1}\mathbf{X}'(\mathbf{Y} - \mathbf{X}\boldsymbol{\nu}), \boldsymbol{\Gamma}), \tag{3.3}$$

where

$$\boldsymbol{\Gamma} = \sigma^2(\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\tau^2}\mathbf{V}^{-1})^{-1}. \tag{3.4}$$

For a p.d. $\mathbf{Q}$, consider the (nonnegative) quadratic loss function

$$L(\mathbf{b}, \boldsymbol{\beta}) = (\mathbf{b} - \boldsymbol{\beta})'\mathbf{Q}(\mathbf{b} - \boldsymbol{\beta}). \tag{3.5}$$

Then the Bayes estimator of $\boldsymbol{\beta}$ is given by the posterior mean, i.e.,

$$\begin{aligned} \hat{\boldsymbol{\beta}}_B &= \boldsymbol{\nu} + (\mathbf{X}'\mathbf{X} + \frac{\sigma^2}{\tau^2}\mathbf{V}^{-1})^{-1}\mathbf{X}'(\mathbf{Y} - \mathbf{X}\boldsymbol{\nu}) \\ &= \boldsymbol{\nu} + (\mathbf{I} + \frac{\sigma^2}{\tau^2}\mathbf{V}^{-1}(\mathbf{X}'\mathbf{X})^{-1})^{-1}(\hat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\nu}). \end{aligned} \tag{3.6}$$

Under the so called Zellner (1986) *g-prior* (see Arnold, 1981),

$$\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}, \tag{3.7}$$

we have

$$\mathbf{I} + \frac{\sigma^2}{\tau^2}\mathbf{V}^{-1}(\mathbf{X}'\mathbf{X})^{-1} = (1 + \frac{\sigma^2}{\tau^2})\mathbf{I}, \tag{3.8}$$

so that the Bayes estimator simplifies to

$$\hat{\boldsymbol{\beta}}_B = \boldsymbol{\nu} + (1 - B)(\hat{\boldsymbol{\beta}}_{LS} - \boldsymbol{\nu}), \tag{3.9}$$

where the Bayes (shrinkage) factor is given by

$$B = \sigma^2(\sigma^2 + \tau^2)^{-1} \ (\geq 0). \tag{3.10}$$

Recall that as in Section 2, our primary interest lies in the estimation of the component parameter $\boldsymbol{\beta}_1$. With this end in mind, we write $\boldsymbol{\nu} = (\boldsymbol{\nu}_1', \boldsymbol{\nu}_2')'$, and consider the posterior distribution of $\boldsymbol{\beta}_1$, given $\mathbf{Y}$, which is given by

$$\mathcal{N}_{p_1}(\ \boldsymbol{\nu}_1 + (1 - B)(\hat{\boldsymbol{\beta}}_{LS,1} - \boldsymbol{\nu}_1),\ \sigma^2(1 - B)\mathbf{C}_{11.2}^{-1}\ ), \tag{3.11}$$

where we let

$$\begin{aligned} \mathbf{X}'\mathbf{X} &= \mathbf{C} = ((\mathbf{C}_{ij}))_{i,j=1,2}; \\ \mathbf{C}_{ii.j} &= \mathbf{C}_{ii} - \mathbf{C}_{ij}\mathbf{C}_{jj}^{-1}\mathbf{C}_{ji}, \text{ for } i,j = 1,2. \end{aligned} \tag{3.12}$$

Note that $\boldsymbol{\nu}_1$, σ^2, τ^2, and hence, B are all unknown quantities. In an empirical Bayes approach, in order to deemphasize the uncertainty of the prior Π, we substitute their estimates derived from the marginal distribution of $\mathbf{Y}$, (given $\mathbf{X}$). In the current context, in line with Section 2, we set

$$\boldsymbol{\nu}' = (\boldsymbol{\nu}_1', \mathbf{0}') \ \text{ i.e., } \boldsymbol{\nu}_2 = \mathbf{0},\ \boldsymbol{\nu}_1 \text{ arbitrary}. \tag{3.13}$$

Also note that the RLSE of $\boldsymbol{\beta}_1$ is $\hat{\boldsymbol{\beta}}_{RLS,1} = \mathbf{C}_{11}^{-1}(\mathbf{X}_1'\mathbf{Y})$, while the marginal distribution of $\mathbf{Y}$ is

$$\mathcal{N}_n(\mathbf{X}_1\boldsymbol{\nu}_1,\ \sigma^2\mathbf{I}_n + \tau^2\mathbf{P}_{\mathbf{X}}), \tag{3.14}$$

where the *projection matrix* is defined as

$$\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'. \tag{3.15}$$

Further, noting that $\mathbf{X}_1'\mathbf{Y} = \mathbf{C}_{11}\hat{\boldsymbol{\beta}}_{RLS,1} = \mathbf{C}_{11}\hat{\boldsymbol{\beta}}_{LS,1} + \mathbf{C}_{12}\hat{\boldsymbol{\beta}}_{LS,2}$, we obtain that

$$\hat{\boldsymbol{\beta}}_{LS,1} = \hat{\boldsymbol{\beta}}_{RLS,1} - \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\hat{\boldsymbol{\beta}}_{LS,2}. \tag{3.16}$$

Ghosh, Saleh and Sen (1989) invoked the completeness and sufficiency properties of $(\hat{\boldsymbol{\beta}}_{RLS,1}, \hat{\boldsymbol{\beta}}_{LS,2}'\mathbf{C}_{22.1}\hat{\boldsymbol{\beta}}_{LS,2}, ||\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{LS}||^2)$, noticed that

$$\begin{aligned} & E(\hat{\boldsymbol{\beta}}_{RLS,1}) = \boldsymbol{\nu}_1,\ ||\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{LS}||^2/\sigma^2 \sim \chi^2_{n-p}, \\ & (\tau^2 + \sigma^2)^{-1}(\hat{\boldsymbol{\beta}}_{LS,2}\mathbf{C}_{22.1}\hat{\boldsymbol{\beta}}_{LS,2}) \sim \chi^2_{p_2}, \end{aligned} \tag{3.17}$$

and thereby considered the estimator $\hat{\boldsymbol{\beta}}_{RLS,1}$ for $\boldsymbol{\nu}_1$, $(p_2-2)/(\hat{\boldsymbol{\beta}}'_{LS,2}\mathbf{C}_{22.1}\hat{\boldsymbol{\beta}}_{LS,2})$, (UMV) estimator, for $(\sigma^2+\tau^2)^{-1}$, and $S_\epsilon^2 = ||\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{LS}||^2/(n-p+2)$, best scale invariant estimator, for σ^2. These led them to the following empirical Bayes estimator of $\boldsymbol{\beta}_1$:

$$\hat{\boldsymbol{\beta}}_{EB} = \hat{\boldsymbol{\beta}}_{RLS,1} + \Big\{1 - \frac{(p_2-2)S_\epsilon^2}{\hat{\boldsymbol{\beta}}'_{LS,2}\mathbf{C}_{22.1}\hat{\boldsymbol{\beta}}_{LS,2}}\Big\}\{\hat{\boldsymbol{\beta}}_{LS,1} - \hat{\boldsymbol{\beta}}_{RLS,1}\}. \tag{3.18}$$

A positive-rule empirical Bayes estimator can then be easily obtained from (3.18) by replacing the shrinkage factor {...} by its nonnegative part $\{...\}^+$. Similarly, replacing $p_2 - 2$ by a constant k, a general shrinkage estimator can be interpreted in an empirical Bayes fashion. However, it follows from Ghosh, Saleh and Sen (1989) that under the quadratic loss mentioned before, the risk of such an empirical Bayes estimator is minimized at $k = p_2 - 2$. Hence, we prefer to use the estimator in (3.18).

With the quadratic loss defined in (3.5), we consider the risk (i.e., the expected loss) $r(.)$ of these estimators. It follows from Ghosh, Saleh and Sen (1989) that

$$r(\hat{\boldsymbol{\beta}}_{LS,1}) - r(\hat{\boldsymbol{\beta}}_{EB,1}) = \mathbf{B}\frac{(n-p)(p_2-2)}{p_2(n-p+2)}\text{trace}(\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{Q}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{C}_{22.1}^{-1}) \tag{3.19}$$

and this is nonnegative for all $p_2 > 2$. It also suggests that whenever $\mathbf{C}_{12}$ is a null matrix, there is no reduction in risk even if p_2 is greater than 2. Also, we have

$$r(\hat{\boldsymbol{\beta}}_{RLS,1}) - r(\hat{\boldsymbol{\beta}}) \geq 0 \text{ iff } (\frac{1-B}{B})^2 \geq \frac{2(n-p+p_2)}{p_2(n-p+2)}. \tag{3.20}$$

Whenever τ^2 is small compared to σ^2, $B \sim 1$, and hence, $\hat{\boldsymbol{\beta}}_{RLS,1}$ performs better than the empirical Bayes estimator. Since the right hand side of (3.20) converges to $2/p_2$ as n becomes large (and $p_2 > 2$), the empirical Bayes estimator dominates the RLSE for a range of τ^2/σ^2.

Let us look into this relative picture in the light of the generalized Pitman closeness criterion (GPCC). For two rival estimators, say, $\hat{\boldsymbol{\theta}}_1$ and $\hat{\boldsymbol{\theta}}_2$, of a common parameter $\boldsymbol{\theta}$, and a loss function $L(\mathbf{a}, \mathbf{b})$, as defined in (3.5), the GPCC of $\hat{\boldsymbol{\theta}}_1$ with respect to $\hat{\boldsymbol{\theta}}_2$ is defined as

$$\begin{aligned} P(\hat{\boldsymbol{\theta}}_1, \hat{\boldsymbol{\theta}}_2|\boldsymbol{\theta}) &= P\{L(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}) < L(\hat{\boldsymbol{\theta}}_2, \boldsymbol{\theta})|\boldsymbol{\theta}\} \\ &+ \frac{1}{2}P\{L(\hat{\boldsymbol{\theta}}_1, \boldsymbol{\theta}) = L(\hat{\boldsymbol{\theta}}_2, \boldsymbol{\theta})|\boldsymbol{\theta}\}. \end{aligned} \tag{3.21}$$

Then $\hat{\boldsymbol{\theta}}_1$ is regarded as closer to $\boldsymbol{\theta}$ than $\hat{\boldsymbol{\theta}}_2$ if the right hand side of (3.21) is $\geq 1/2$ for all $\boldsymbol{\theta}$, with the strict inequality sign holding for some $\boldsymbol{\theta}$. We refer to Keating, Mason and Sen (1993) for some discussion of the GPCC.

Ghosh and Sen (1991) extended the GPCC concept to a Bayesian setup and introduced the idea of the *Posterior Pitman closeness* (PCC) measure. Let $\Pi(\boldsymbol{\theta})$ be a prior distribution defined on $\boldsymbol{\Theta}$, and let $\boldsymbol{\delta}_1, \boldsymbol{\delta}_2$ be two Bayes estimators of $\boldsymbol{\theta}$ under the prior $\Pi(\boldsymbol{\theta})$. Then

$$\begin{aligned} PPC_{\Pi}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2|\mathbf{Y}) &= P_{\Pi}\{L(\boldsymbol{\delta}_1, \boldsymbol{\theta}) < L(\boldsymbol{\delta}_2, \boldsymbol{\theta})\} \\ &+ \frac{1}{2}P_{\Pi}\{L(\boldsymbol{\delta}_1, \boldsymbol{\theta}) = L(\boldsymbol{\delta}_2, \boldsymbol{\theta})\}. \end{aligned} \tag{3.22}$$

Thus, $\boldsymbol{\delta}_1$ is said to be posterior Pitman closer to $\boldsymbol{\theta}$ than $\boldsymbol{\delta}_2$, under the prior $\Pi(\boldsymbol{\theta})$, provided

$$PPC_{\Pi}(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2|\mathbf{Y}) \geq \frac{1}{2}, \ \forall \ \mathbf{Y} \text{ a.e.}, \tag{3.23}$$

with strict inequality holding for some $\mathbf{Y}$.

For real valued θ, $\mathcal{M}(\theta|\mathbf{Y})$, the *poterior median* of θ, given $\mathbf{Y}$, has the PPC_{Π} property. For vector $\boldsymbol{\theta}$, characterizations of multivariate posterior medians have been considered by a host of researchers in the past few years. In particular, if the posterior distribution of $\boldsymbol{\theta}$, given $\mathbf{Y}$, is diagonally symmetric about (a location) $\boldsymbol{\delta}(\mathbf{Y})$, then $\boldsymbol{\delta}(\mathbf{Y})$ is a posterior median of $\boldsymbol{\theta}$, and it has the PPC_{Π}-property (Sen, 1991). It follows from the above results that the dominance of the empirical Bayes estimator over the LSE holds under the PPC criteion as well.

4. BAN ESTIMATORS: EMPIRICAL BAYES VERSIONS

We consider the same linear model as in Section 2, but deemphasize the normality of the errors. Thus, we let as in (2.1), $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}$ where $\mathbf{e} = (e_1, \ldots, e_n)'$, and the e_i are i.i.d.r.v.'s with a probability density function $f(e)$, where the form of f is free from $\boldsymbol{\beta}$; typically a location-scale family of density is contemplated here, and the normal density assumed in Section 2 is an important member of this class. Under suitable regularity assumption (on f,$\mathbf{X}$ and $\boldsymbol{\beta}$), $\hat{\boldsymbol{\beta}}_{ML}$, the MLE of $\boldsymbol{\beta}$, is BAN in the sense that asymptotically (as $n \to \infty$)

$$\sqrt{n}(\hat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta}) \ \sim \mathcal{N}_p(\mathbf{0}, \ [\mathcal{I}(f)]^{-1}\mathbf{C}^{-1} \ ), \tag{4.1}$$

where $\mathbf{C} = \lim_{n\to\infty} n^{-1}\mathbf{X}'\mathbf{X}$ has been defined before, and $\mathcal{I}(\{)$ is the Fisher information for the pdf f, defined as

$$\mathcal{I}(f) = E\{[-f'(e)/f(e)]^2\}; \tag{4.2}$$

note that by assumption, $\mathcal{I}(f)$ does not depend on $\boldsymbol{\beta}$. The *estimating equation* (EE) for the MLE is given by

$$\sum_{i=1}^{n} \mathbf{x}_i\{-f'(Y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}})/f(Y_i - \mathbf{x}_i\hat{\boldsymbol{\beta}})\} = \mathbf{0}'. \tag{4.3}$$

This asymptotic normality of the MLE, and attainment of the information limit for its asymptotic covariance matrix are shared by a large class of estimators that are known as the BAN or first-order efficient estimators. Following the line of attack of Jurečková and Sen (1996), we may assume that for such a BAN estimator of $\boldsymbol{\beta}$, denoted by $\hat{\boldsymbol{\beta}}_n$, under suitable regularity assumptions, the following *first-order asymptotic distributional representation* (FOADR) result holds:

$$\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = \sum_{i=1}^{n} \mathbf{c}_{ni}\phi(e_i) \ + \ \mathbf{R}_n, \tag{4.4}$$

where the score function $\phi(.)$ is given by

$$\phi(x) = -f'(x)/f(x), \quad (-\infty < x < \infty), \tag{4.5}$$

the $\mathbf{c}_{ni}$ depend on the matrix $\mathbf{X}$, and

$$n^{1/2}||\mathbf{R}_n|| \rightarrow 0, \text{ in a suitable mode;} \tag{4.6}$$

for the empirical Bayes approach, we assume that this convergence holds in the second mean, while for the GPCC approach, convergence in probability suffices.

In the context of such BAN estimators, we conceive of the same Zellner g-prior, namely that

$$\boldsymbol{\beta} \sim \mathcal{N}_p(\boldsymbol{\nu}, \ \tau^2\mathbf{V}\ ), \tag{4.7}$$

where as in before, we let $\mathbf{V} = (\mathbf{X}'\mathbf{X})^{-1}$. Then writing $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\nu} = \hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} + \boldsymbol{\beta} - \boldsymbol{\nu}$, we obtain by convolution a FOADR for $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\nu}$ that yields the following (asymptotic) marginal law:

$$\mathcal{N}_p(\mathbf{0}, \ ([\mathcal{I}(f)]^{-1} + \tau^2)(\mathbf{X}'\mathbf{X})^{-1}\ ). \tag{4.8}$$

As a result, in the FOADR for the posterior distribution of $\boldsymbol{\beta}$, given $\hat{\boldsymbol{\beta}}_n$, the pricipal component has the following law:

$$\mathcal{N}_p(\boldsymbol{\nu} + (1 + \frac{1}{\mathcal{I}(f)\tau^2})^{-1}(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\nu}), \ \boldsymbol{\Gamma}\ ), \tag{4.9}$$

where

$$\begin{aligned} \boldsymbol{\Gamma} &= \frac{1}{\mathcal{I}(f)}(1 + \frac{1}{\mathcal{I}(f)\tau^2})^{-1}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \{\tau^2/(1 + \mathcal{I}(f)\tau^2)\}(\mathbf{X}'\mathbf{X})^{-1}. \end{aligned} \tag{4.10}$$

As such, we may proceed as in Section 2, and by some standard arguments, conclude that in a FOADR representation for the posterior mean (as well as generalized median) of $\boldsymbol{\beta}$, the principal term is

$$\hat{\boldsymbol{\beta}}_{nB} = \boldsymbol{\nu} + (1 - B)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\nu}\ ), \tag{4.11}$$

where the Bayes (shrinkage) factor B is given by

$$B = \{1 + \mathcal{I}(f)\tau^2\}^{-1}, \tag{4.12}$$

and the other terms in this FOADR are all negligible in a suitable norm.

Given this asymptotic representation for the Bayes BAN-estimator, noting the structural analogy with the LSE, we may proceed as in Section 3, and formulate the following empirical Bayes estimators. Let $\breve{\boldsymbol{\beta}}_{n,1}$ be the BAN estimator of $\boldsymbol{\beta}_1$ under the reduced model: $\mathbf{Y} = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{e}$. Further, partition the BAN estimator $\hat{\boldsymbol{\beta}}_n$ as $(\hat{\boldsymbol{\beta}}'_{n,1}, \hat{\boldsymbol{\beta}}'_{n,2})'$. Then note by the FOADR for $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\nu}$ considered in (4.8),

$$\frac{\mathcal{I}(f)}{1 + \mathcal{I}(f)\tau^2}\{\hat{\boldsymbol{\beta}}'_{n,2}\mathbf{C}_{22.1}\hat{\boldsymbol{\beta}}_{n,2}\} \sim \chi^2_{p_2}. \tag{4.13}$$

As such, by analogy with Section 3, we estimate

$$\frac{\mathcal{I}(f)}{1+\tau^2\mathcal{I}(f)} \quad \text{by} \quad \frac{p_2-2}{\hat{\boldsymbol{\beta}}'_{n,2}\mathbf{C}_{22.1}\hat{\boldsymbol{\beta}}_{n,2}}. \tag{4.14}$$

Moreover, by the FOADR for $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\nu}$, considered in (4.8),

$$\begin{aligned} V_n^* &= \frac{-1}{n}\sum_{i=1}^{n}\frac{\partial^2}{\partial x^2}\log f(x)\big|_{x=Y_i-\mathbf{x}_i\hat{\boldsymbol{\beta}}_n} \\ &\xrightarrow{\mathcal{P}} \mathcal{I}(f), \text{ as } n\to\infty. \end{aligned} \tag{4.15}$$

As a result, we estimate

$$\{1+\mathcal{I}(f)\tau^2\}^{-1} \quad \text{by} \quad \frac{(p_2-2)}{V_n^*(\hat{\boldsymbol{\beta}}'_{n,2}\mathbf{C}_{22.1}\hat{\boldsymbol{\beta}}_{n,2})}. \tag{4.16}$$

At this stage, we denote the conventional log-lilelihood ratio type test statistic (or its asymptotically equivalent form based on the Wald or the Rao score statistics), for testing $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ vs. $H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$, by $\mathcal{L}_{n2}$. Then, using the FOADR results stated above, it can be shown that under the null as well as contiguous alternatives,

$$|V_n^*\{\hat{\boldsymbol{\beta}}'_{n,2}\mathbf{C}_{22.1}\hat{\boldsymbol{\beta}}_{n,2} \; - \; \mathcal{L}_{n2} \mid \xrightarrow{\mathcal{P}} 0, \text{ as } n\to\infty. \tag{4.17}$$

Consequently, using the above approximation, we arrive at the following emprical Bayes version of a BAN estimator:

$$\hat{\boldsymbol{\beta}}_{nEB.1} = \breve{\boldsymbol{\beta}}_{n,1} + \{1-\frac{p_2-2}{\mathcal{L}_{n2}}\}(\hat{\boldsymbol{\beta}}_{n,1}-\breve{\boldsymbol{\beta}}_{n,1}). \tag{4.18}$$

In this form, the empirical Bayes version agrees with the shrinkage MLE version considered in detail in Sen (1986), though the empirical Bayes interpretation was not explored there.

In this context, if we want to justify the asymptotic risk computations in a conventional sense (i.e., as the limit of the actual risk when n is made to increase indefinitely), then we need to impose the regularity assumptions for the Hájek-LeCam-Inagaki *regular estimators*, as have been displayed in detail in LeCam (1986). Some of these stringent regularity assumptions can be avoided to a certain extent by adopting the measure *asymptotic distributional risk* (ADR) that is based directly on the FOADR itself (wherein the remainder term is neglected). We refer to Sen (1986) where the ADR concept has ben elaborated and incorporated in the study of asymptotic properties of shrinkage MLE's. In that setup, in the FOADR, it suffices to show that the remaider term is $o_p(n^{-1/2})$ (while the others need the same order in quadratic mean). In this respect, the situation is much more handy with the GPCC, as formulated in the preceding section; there the limits involve only the probability distributions, and hence, $o_p(n^{-1/2})$ characterization for the remainder term in the FOADR suffices. For certain general isomorphism of GPCC and quadratic risk dominance results, we may refer to

Sen (1994), and these findings based on suitable FOADR results, pertain to such Bayes and empirical Bayes estimators as well. Further, the GPCC dominance results hold for a broader range of p_2 values.

5. SEMIPARAMETRIC EMPIRICAL BAYES PROCEDURES

The BAN estimation methodology provides the access to semiparametrics in a very natural way, and empirical Bayes versions can be worked out in an analogous manner. The main concern on the unrestricted use of parametic empirical Bayes estimators is their vulnerability or nonrobustness to possible model departures that can either arise due to nonlinearity of the model and/or heteroscedaticity of the errors , or due to a plausible departure from the assumed form of the error density function. In a semiparametric setup, the error distribution may be taken as quite arbitrary, albeit the linearity of the model is presumed. In this way, we retain the primary emphasis on the linearity of the model, and want to draw conclusions on the regression parameter without making necessarily stringent distributional assumption on the error component. In that scheme, we may allow for some *local* or *global* departures from an assumed form of the error density; in the former case, there is a stronger emphasis on robustness to local departures (or infinitesimal robustness) without much compromise on the (asymptotic) efficiency. M-estimators and M-statistics (viz., Huber 1973, Hampel et al. 1986) are particularly useful in this context. L-estimators have also found their way to robust estimation in linear models, though they have generally computational complications in other than location and scale models (viz., Jurečková and Sen (1996)). In the case of global robustness, however, rank based procedures have greater appeal. R-estimators, and their siblings: *regression rank scores* estimators (Gutenbrunner and Jurečková, 1992) , and *regression quantiles* (Koenker and Bassett, 1978) are more popular in this context. On the other hand, if the basic linearity of the model is questionable, then nonparametric regression function formulation may be more appealing, wherein the form of the regression function is allowed to be smooth but rather arbitrary. Thus, a nonparametric regression formulation may be comparatively more robust. However, there is a price that we may need to pay for this option. The finite dimensional parameter vector (i.e., $\boldsymbol{\beta}$), considered in earlier sections, has to be replaced by a functional parameter, i.e., the regression function, and its estimation in a nonparametric fashion entails slower rates of convergence, as well as, a possible lack of optimality properties even in an asymptotic setup. Development of empirical Bayes estimators of such functional parameters is not contemplated in the current study, and we shall confine ourselves only to semiparametric linear models. The procedures to be considered here are generally more robust than standard parametric ones (referred to in earlier sections) as long as the postulated linearity of the model holds , although they may not be robust to possible departures from the assumed linearity of the model. There are other semiparametric models, such as the multiplicative intensity process models that are basically related to the Cox (1972) *proportional hazards model* (PHM). Such PHM's generally relate to the *hazard regression* problem wherein the baseline hazard function is treated as arbitrary (that is nonparametric) while the regression on the covariates is treated as parametric, and the statistical methodology evolved with the development of the *partial likelihood* principle that has been extended to a more general context wherein *matrix valued counting processes* are incorporated to facilitate statistical analysis that has a predominant asymptotic flavor. Here also the hazard regression need not be a finite-dimensional parametric function, and in a more general setup, the regression parameters may be *time-dependent* that results in a functional parameter space [viz., Murphy and Sen, 1991]. For the finite dimensional parametric hazard function formulations, we may refer to the monograph of Andersen et al. (1993) for a unified, up-to-date treatment, and Pedroso de Lima and Sen (1997) for some generalization to some multivariate cases. Development of empirical Bayes procedures for such matrix valued counting process needs a somewhat different (and presumably more complex) approach than the one for linear models that are contemplated here; we shall not enter into these discussions in the current study. Keeping this scenario in mind, we consider specifically the following semiparametric linear model (wherein we use the same notations as introduced in Section 2):

$$\begin{aligned} Y_i &= \mathbf{x}_i\boldsymbol{\beta} + e_i, \; i = 1, \ldots, n, \\ e_i &\sim \text{pdf } f(e), \; e \in \mathcal{R}, \end{aligned} \tag{5.1}$$

where the form of f, though unknown, is assumed to be free from $\boldsymbol{\beta}$. It is also assumed that f has a finite Fisher information (with respect to location) $\mathcal{I}(f)$. As in earlier section, we assume that $\boldsymbol{\beta}$ has the Zellner g-prior, namely

$$\begin{aligned} \boldsymbol{\beta} &\sim \mathcal{N}_p(\boldsymbol{\nu}, \; \tau^2\mathbf{V} \;), \\ \mathbf{V} &= (\mathbf{X}'\mathbf{X})^{-1}; \; \tau^2 = c\{\mathcal{I}(f)\}^{-1}, \; c > 0. \end{aligned} \tag{5.2}$$

Note that here we have a parametric regression model, a parametric prior on $\boldsymbol{\beta}$, while a nonparametric pdf $f(e)$. That's why, we term it a semiparametric linear model. We show that empirical Bayes estimators exist for such semiparametric models, and they correspond to the so called Stein-rule or shrinkage estimators (based on robust statistics) that have been studied extensively in the literature.

In a frequentist setup, let $\hat{\boldsymbol{\beta}}_n$ be a suitable estimator of $\boldsymbol{\beta}$. Such an estimator can be chosen from a much wider class of *robust* estimators: $M, L-$ and R-estimators of regression (and location) parameters, regression quantile estimators, regression rank scores estimators, and many other robust estimators belong to this class. We refer to Jurečková and Sen (1996), chapters 3 to 7, for a detailed coverage, where a basic FOADR result has been exploited in a systematic manner. Based on this exploitation, we assume that the following FOADR result holds for $\hat{\boldsymbol{\beta}}_n$:

$$\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta} = \sum_{i=1}^{n} \mathbf{c}_{ni}\phi(e_i) + \mathbf{R}_n, \tag{5.3}$$

where $\phi(.)$ is a suitable score function, which may generally depend on the unknown density $f(.)$, is so normalized that

$$\int \phi(\epsilon)dF(\epsilon) = 0, \int \phi^2(\epsilon)dF(\epsilon) = \sigma^2_\phi(< \infty), \tag{5.4}$$

the regression vectors $\mathbf{c}_{ni} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}'_i$ depend on $\mathbf{X}$ and satisfy the generalized Noether condition:

$$\max_{\{1 \leq i \leq n\}} \mathbf{c}'_{ni} (\sum_{j=1}^{n} \mathbf{c}_{nj}\mathbf{c}'_{nj})^{-1}\mathbf{c}_{ni} \to 0; \tag{5.5}$$

the remaider term $\mathbf{R}_n$ satisfies the condition

$$n^{1/2}||\mathbf{R}_n|| \xrightarrow{\mathcal{P}} 0, \text{ as } n \to \infty; \tag{5.6}$$

we may need to strengthen the mode of convergence to the mean-square norm if we are to deal with the conventional quadratic risk criterion, while for the GPCC, the convergence in probability suffices. Note that whenever the density $f(.)$ has a finite Fisher information $\mathcal{I}(f)$, by the Cramér-Rao information inequality,

$$\sigma^2_\phi \geq \{\mathcal{I}(f)\}^{-1}, \tag{5.7}$$

where the equality sign holds when $\hat{\boldsymbol{\beta}}_n$ is a BAN estimator of $\boldsymbol{\beta}$. For the class of estimators, referred to above, there is a subclass that comprise the BAN estimators, and this facilitates the incorporation of the methodology presented in the preceding section.

By virtue of (5.1), (5.2) and the above asymptotic representation, we may proceed as in the case of BAN estimators, treated in the preceding section, and by (asymptotic) convolution (for $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$ and $\boldsymbol{\beta} - \boldsymbol{\nu}$) obtain the marginal (asymptotic) distribution of $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\nu}$ (that is multinormal with null mean vector and dispersion matrix as the sum of the two dispersion matrices that appear in (5.2) and in the FOADR in (5.3), that is,

$$\mathcal{N}_p(\mathbf{0},\ (\tau^2 + \sigma^2_\phi)(\mathbf{X}'\mathbf{X})^{-1}\ ). \tag{5.8}$$

As a result, the asymptotic posterior distribution of $\boldsymbol{\beta}$, given $\hat{\boldsymbol{\beta}}_n$, is multinormal with the following form:

$$\mathcal{N}_p(\boldsymbol{\nu} + \frac{\tau^2}{\tau^2 + \sigma^2_\phi}(\hat{\boldsymbol{\beta}}_N - \boldsymbol{\nu}),\ \frac{\tau^2\sigma^2_\phi}{\tau^2 + \sigma^2_\phi}(\mathbf{X}'\mathbf{X})^{-1}\ ). \tag{5.9}$$

To pose the empirical Bayes versions, we consider the reduced model:

$$Y_i = \mathbf{x}_{i(1)}\boldsymbol{\beta}_1 + e_i,\ i = 1, \ldots, n, \tag{5.10}$$

where we partition $\boldsymbol{\beta}$ and $\mathbf{x}_i$ as in before, and the pdf $f(.)$ of the e_i is defined as in the case of the full model. Let $\breve{\boldsymbol{\beta}}_{n,1}$ be the corresponding estimator for $\boldsymbol{\beta}_1$. For this estimator, under the reduced model, we have a similar FOADR result where the $\mathbf{c}_{ni}$ are to be replaced by

$$\mathbf{c}^{*(1)}_{ni} = (\mathbf{X}'_1\mathbf{X}_1)^{-1}\mathbf{x}'_{i(1)},\ i = 1, \ldots, n. \tag{5.11}$$

Further, for testing $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$ against $H_1 : \boldsymbol{\beta}_2 \neq \mathbf{0}$, let $\mathcal{L}_{n2}^*$ be a suitable test statistic that has asymptotically under the null hypothesis a central chi square distribution with p_2 degrees of freedom. Actually, in some cases (such as dealing with $L-$estimators,) we deal with a quadratic form in $\hat{\boldsymbol{\beta}}_{n,2}$ while in some other cases (such as R-and M-estimators), we deal with suitable aligned (rank or M-statistics) as test statistics that are asymptotically equivalent to a quadratic form in the estimators $\hat{\boldsymbol{\beta}}_{n,2}$.

As in Section 3, we may write $B = \sigma_\phi^2\{\sigma_\phi^2 + \tau^2\}^{-1}$ so that the (asymptotic) Bayes version can be written as

$$\hat{\boldsymbol{\beta}}_{B,n} = \boldsymbol{\nu} + (1 - B)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\nu}), \tag{5.12}$$

where on $\boldsymbol{\nu}$ we adopt the same simplifications as outlined in Section 3. As such, we need to restrict ourselves to empirical bayes estimates of $\boldsymbol{\nu}_1$ (treating $\boldsymbol{\nu}_2 = \mathbf{0}$), and the Bayes (shrinkage) factor B. We may analogously estimate $\boldsymbol{\nu}_1$ by the reduced model based estimator $\breve{\boldsymbol{\beta}}_{n,1}$. The estimation of B can be carried out in several ways: either estimating $\{\sigma_\phi^2 + \tau^2\}^{-1}$ from the marginal distribution of $\hat{\boldsymbol{\beta}}_{n,2}$ (using a similar quadratic form as in Section 3) and σ_ϕ^2 from the distribution of $\hat{\boldsymbol{\beta}}$, or using the test statistic $\mathcal{L}_{n2}^*$ to estimate B directly. While these alternative approaches are asymptotically equivalent, from computational point of view, the second approach seems to be more adoptable. Hence, as in the case of BAN estimators, we consider the following empirical Bayes version of the estimator of $\boldsymbol{\beta}$:

$$\hat{\boldsymbol{\beta}}_{EB,n,1} = \breve{\boldsymbol{\beta}}_{n,1} + (1 - \frac{p_2 - 2}{\mathcal{L}_{n2}})(\hat{\boldsymbol{\beta}}_{n,1} - \breve{\boldsymbol{\beta}}_{n,1}). \tag{5.13}$$

This corresponds to the usual Stein-rule or shrinkage versions of such robust estimators that have been extensively worked out in the literature during the past fifteen years, and reported in a systematic manner in Jurečková and Sen (1996); detailed references to the relevant literature are also cited there. PTE versions can also be posed along the same line, though their empirical Bayes interpretations may not be so apparent.

The empirical Bayes interpretation of such semiparametric shrinkage estimators, as outlined above, enables us to incorporate the vast literature of shrinkage robust estimators (viz., Jurečková and Sen (1996)) in the study of the properties of such estimators. In particular, the relative picture of (asymptotic) risks of various empirical Bayes estimators (based on diverse robust statistics) as well as their GPCC comparisons remains comparable to the one portrayed in Sections 3 and 4; the basic difference comes in the related noncentrality parameters, and these are related to each other by the usual *Pitman asymptotic relative efficiency* (PARE) measure. For this reason, we skip the details here, but mention the following main results that are pertinent in the current context:

(i) If the influence function in the FOADR of $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$, i.e., the score $\phi(e)$ in (5.3), agrees with the conventional Fisher score function, i.e., $\phi_f(e) = -f'(e)/f(e)$, then the corresponding empirical version (as considered above) is an empirical Bayes BAN estimator, and hence, shares the properties as ascribed to such estimators.

(ii) By virtue of (i), it is possible to choose a robust empirical Bayes estimator such that it becomes stochastically equivalent to an empirical Bayes BAN estimator (considered in Section 4) for a specific type of underlying density. This property may particularly be important in the context of local robustness aspects when an anticipated density $f(.)$ is in the picture.

(iii) In particular, if we use a rank based (R- or regression rank scores) estimator of $\boldsymbol{\beta}$ that incorporates the so called the *normal score* generating function in its formulation. Such an estimator is (asymptotically) at least as efficient as the LSE (considered in Section 2) for a large class of underlying densities, so that the corresponding empirical Bayes version would perform better than the ones considered in Section 3 when the underlying density is not normal; at the sametime, it retains robustness to a certain extent.

(iv) Along the sameline as in (iii), an adaptive empirical Bayes version of robust estimators of $\boldsymbol{\beta}$ can be formulated (using the results of Hušková and Sen (1985)) that combines the robustness and asymptotic efficiency properties to a greater extent. However, this generally entails a slightly slower rate of convergence, and will not be pursued here.

Modern computational facilities make it possible to advocate such alternative robust empirical Bayes estimators in practice too.

6. SOME CONCLUDING REMARKS

The foundation of empirical Bayes estimators based on robust statistics and their asymptotic isomorphism to shrinkage or Stein-rule versions rest on two basic factors: (i) FOADR results that permit asymptotic Gaussian laws under suitable regularity assumptions, and (ii) the incorporation of the Zellner (1986) g-prior that permits the convolution result in a manner compatible with the standard parametric cases. The g-prior that incorporates the reciprocal of the design matrix ($\mathbf{X}'\mathbf{X}$) in its dispersion matrix clearly convey the following message:

Whenever with the increase in n, $n^{-1}\mathbf{X}'\mathbf{X}$ converges to a p.d. matrix ($\mathbf{C}$), as has been assumed in this study (as well as in Ghosh, Saleh and Sen (1989) and elsewhere in the literature), the prior has increasing concentration around $\boldsymbol{\nu}$. By having the rate of convergence of this prior comparable with that of $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}$, the convolution result comes up in a handy nondegenerate (Gaussian) form, and this facilitates the computation and simplification of the posterior distributions. This is also comparable to the general asymptotics for shrinkage robust estimators of regression parameters, where the dominance results relate to a small neighborhood of the pivot that has a diameter of the order $n^{-1/2}$; we refer to Jurečková and Sen (1996) for some discussion of such dominance results for shrinkage robust estimators. Thus, in reality, we adopt a sequence of Zellner priors matches the rate of convergence of the frequentist estimators, and in that sense, our findings relate to a *local* empirical Bayes setup.

The FOADR approach advocated here has the main advantage of identifying the influence curves (IC) of the estimators in a visible manner. It is, of course, not necessary to force the use of such FOADR results. For example, for various types of rank based statistics, $L-, M-$, and R- estimators, U-statistics, von Mises functionals, and Hadamard differentiable statistical functionals, asymptotic normality results have been derived by alternative approaches by a host of researchers; we refer to Sen (1981) where, in particular, a unified *martingale* approach has been advocated. Such results also permit the convolution result for the marginal law of $\hat{\boldsymbol{\beta}}_n - \boldsymbol{\nu}$, though in a less visible manner. The main advantage of using a FOADR approach is that whenever a second-order asymptotic distributional representation (SOADR) holds, a precise order for the remainder term can be studied, and this can be incorporated in the study of the rate of approach to the desired asymptotic results. For such SOADR results for some important members of the family of robust estimators, we may again refer to Jurečková and Sen (1996). More work along this line is under way, and it is anticipated that they would be of considerable importance in the study of the asymptotic properties of robust empirical Bayes estimators.

Our findings also pertain to general models (not necessarily linear ones) provided we have a finite dimensional parameter and we incorporate an appropriate Zellner-type prior to obtain the convolution law in a simple way. In particular, if we work with BAN estimators then such a prior can be essentially related to the information matrix related to the parameter as may be assumed to exist. However, the simplification we have in the linear model that estimates of $\boldsymbol{\beta}$ have dispersion matrices that are scalar multiples of the matrix $(\mathbf{X}'\mathbf{X})^{-1}$ may not be generally true in such a case, and that may call for some further adjustments and additional regularity assumptions to obtain the convolution distribution in a natural form that permits a Gaussian posterior distribution. These need to be worked out on a case by case basis.

Finally, we have not made use of the conventional Dirichlet priors (on the d.f. F) that is generally used in nonparametric (empirical) Bayes estimation of suitable parameters based on i.i.d. sample observations. But, our findings may also be linked to such priors under appropriate differentiability conditions on the parameters that are regarded as functionals of the d.f. F. The details are to be provided in a separate communication.

REFERENCES

Anderson, P. K., Borgan, O., Gill, R. D., and Keiding, N. (1993). *Statistical Models Based on Counting Processes*, Springer-Verlag, New York.

Arnold, S. F. (1981). *The Theory of Linear Models and Multivariate Analysis.* John Wiley, New York.

Cox, D. R. (1972). Regression models and life tables (with discussion). *J. Roy. Statist. Soc.* **B34**, 187- 220.

Ghosh, M., Saleh, A. K. M. E., and Sen, P. K. (1989). Empirical Bayes subset estimation in regression models. *Statist. Decs.* *7*, 15-35.

Gutenbrunner, C., and Jurečková, J. (1992). Regression rank scores and regression quantiles. *Ann. Statist.* **20**, 305-330.

Hampel, F. R., Rousseeuw, P. J., Ronchetti, E., and Stahel, W. (1986). *Robust Statistics- The Approach Based on Influence Functions.* John Wiley, New York.

Huber, P. J. (1973). Robust regression: Asymptotivs, conjectures and Monte Carlo. *Ann. Statist.* **1**, 799-821.

Hušková, M., and Sen, P. K. (1985). On sequentially adative asymptotically efficient rank statistics. *Sequen. Anal. 4*, 125-151.

Jurečková, J., and Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations.* John Wiley, New York.

Keating, J. P., Mason, R. L., and Sen, P. K. (1993). *Pitman's Measure of Closeness: A Comparison of Statistical Estimators.* SIAM, Philadelphia.

Koenker, R., and Bassett, G. (1978). Regression quantiles. *Econometrica* **46**, 33-50.

LeCam, L. (1986). *Asymptotic Methods in Statistical Decision Theory.* Springer-Verlag, New York.

McCullagh, P., and Nelder, J. A. (1989). *Generalized Linear Models*, 2nd Ed., Chapman-Hall, London.

Murphy, S. A., and Sen, P. K. (1991). Time-dependent coefficients in a Cox-type regression model. *Stochas. Proc. Appl.* **39**, 153-180.

Pedroso de Lima, A. C., and Sen, P. K. (1997). A matrix-valued counting process with first-order interactive intensities. *Ann. Appl. Probab.* **7**, 494-507.

Pitman, E. J. G. (1937). The closest estimates of statistical parameters. *Proc. Cambridge Phil. Soc.* **33**, 212-222.

Puri, M. L., and Sen, P. K. (1971). *Nonparametric Methods in Multivariate Analysis* John Wiley, New York.

Puri, M. L., and Sen, P. K. (1985). *Nonparametric Methods in General Linear Models.* John Wiley, New York.

Sen, P. K. (1981). *Sequential Nonparametrics: Invariance Principles and Statistical Inference.* John Wiley, New York.

Sen, P. K. (1986). On the asymptotic distributional risk of shrinkage and preliminary test versions of maximum likelihood estimators. *Sankhyā, Ser. A 48*, 354-371.

Sen, P. K. (1991). Pitman closeness of statistical estimators: Latent years and the renaissance. In *Issues in Statistics: D. Basu Volume.* (eds. M. Ghosh and P. K. Pathak), Inst. Math. Statist. Lect. Notes Sr. 17, pp. 52-74.

Sen, P. K. (1993). Perspectives in multivariate nonparametrics: Conditional functionals and ANOCOVA models. *Sankhyā, Ser A, 55*, 516-532.

Sen, P. K. (1994). Isomorphism of quadratic norm and PC ordering of estimators admitting first order AN representation. *Sankhyā, Ser. A 56*, 465-475.

Sen, P. K. (1996a). Robust and nonparametric methods in linear models with mixed-effects. *Tetra Mount Math. Publ.* **7**, pp. 331-343.

Sen, P. K. (1996b). Regression rank scores estimation in ANOCOVA. *Ann. Statist.* **24**, 1486-1502

Stein, C. (1956). Inadmissibility of the usual estimator of the mean of a multivariate normal distribution. In *Proc. Third Berkeley Symp. Math. Statist. Probability. 1* Univ. Calif. Press, Berkeley, pp. 197-206.

Zellner, A. (1986). On assessing prior distributions and Bayesian regression analysis with g-prior distributions. In *Bayesian Inference and Decision Techniques* (eds. P. K. Goel and A. Zellner), North Holland, Amsterdam, pp. 233-243.