The Annals of Statistics
1987, Vol. 15, No. 4, 1580–1592

ON PRELIMINARY TEST AND SHRINKAGE *M*-ESTIMATION IN LINEAR MODELS

BY PRANAB KUMAR SEN[1] AND A. K. M. EHSANES SALEH[2]

University of North Carolina, Chapel Hill and Carleton University

In a general univariate linear model, M-estimation of a subset of parameters is considered when the complementary subset is plausibly redundant. Along with the classical versions, both the preliminary test and shrinkage versions of the usual M-estimators are considered and, in the light of their asymptotic distributional risks, the relative asymptotic risk-efficiency results are studied in detail. Though the shrinkage M-estimators may dominate their classical versions, they do not, in general, dominate the preliminary test versions.

1. Introduction. Consider the usual linear model

$$\mathbf{X}_n = (X_1, \dots, X_n)' = \mathbf{A}_n \boldsymbol{\beta} + \mathbf{e}_n, \qquad \mathbf{e}_n = (e_1, \dots, e_n)', \tag{1.1}$$

where $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)'$ is a vector of unknown (regression) parameters, $\mathbf{A}_n$ is an $n \times p$ (design) matrix of known regression constants, $n > p \geq 1$, and the errors e_i are independent and identically distributed (i.i.d.) random variables (r.v.) with a distribution function (d.f.) F, defined on the real line R. Without any loss of generality, we may assume that $\mathbf{A}_n$ is of rank p, and we consider the partitioning (where $p = p_1 + p_2$, $p_1 \geq 0$, $p_2 \geq 0$)

$$\boldsymbol{\beta}' = \Big(\underset{p_1 \times 1}{\boldsymbol{\beta}_1'}, \ \underset{p_2 \times 1}{\boldsymbol{\beta}_2'} \Big) \quad \text{and} \quad \mathbf{A}_n = \Big(\underset{n \times p_1}{\mathbf{A}_{n1}}, \ \underset{n \times p_2}{\mathbf{A}_{n2}} \Big), \tag{1.2}$$

so that (1.1) may also be written as $\mathbf{X}_n = \mathbf{A}_{n1}\boldsymbol{\beta}_1 + \mathbf{A}_{n2}\boldsymbol{\beta}_2 + \mathbf{e}_n$. We are primarily interested in the estimation of $\boldsymbol{\beta}_1$ when it is plausible that $\boldsymbol{\beta}_2$ is "close to" $\mathbf{0}$. This situation may arise, for example, in a multifactor design, where $\boldsymbol{\beta}_1$ stands for the *main effects* and $\boldsymbol{\beta}_2$ for the *interactions*: It may be quite likely (though cannot be taken for granted) that all the interactions are insignificant and one may then be mainly interested in the estimation of the main effects. Other examples of this type abound in linear models. Also, instead of the null *pivot* for $\boldsymbol{\beta}_2$, if we have any other specified $\boldsymbol{\beta}_2^0$, then working with $\mathbf{X}_n^0 = \mathbf{X}_n - \mathbf{A}_{n2}\boldsymbol{\beta}_2^0$, we may reduce the pivot to $\mathbf{0}$.

Instead of the classical *least squares estimators* (LSE) (optimal for normal F) or the *maximum likelihood estimators* (MLE) (based on some assumed form of F) treated earlier in Saleh and Sen (1987) and Sen (1986), respectively, we shall

Received November 1985; revised September 1986.

[1] Work supported by National Institutes of Health grant N01-HV-12243-L.

[2] Work supported by NSERC (Canada) grant A3088 and by a Gr-5 grant from the Faculty of Graduate Studies and Research, Carleton University, Ottawa, Canada.

AMS 1980 *subject classifications.* Primary 62C16, 62G05; secondary 62F10, 62H12.

Key words and phrases. Asymptotic distributional risk, asymptotic distributional risk efficiency, James–Stein rule, linear model, local alternatives, M-estimators, minimaxity, preliminary test, robustness, shrinkage estimator.

be more interested in some general robust estimators, namely, the *M-estimators* (which contain both the LSE and MLE as special cases). For the global (unrestrained) model in (1.2), we denote an M-estimator of $\boldsymbol{\beta}$ by $\tilde{\boldsymbol{\beta}}_n = (\tilde{\boldsymbol{\beta}}_{1n}', \tilde{\boldsymbol{\beta}}_{2n}')'$, so that $\tilde{\boldsymbol{\beta}}_{1n}$ is an *unrestrained M-estimator* (UME) of $\boldsymbol{\beta}_1$. For various properties of $\tilde{\boldsymbol{\beta}}_{1n}$, we may refer to Jurečková (1977), Yohai and Maronna (1979) and Singer and Sen (1985), among others. Second, for the restrained model $\mathbf{X}_n = \mathbf{A}_{n1}\boldsymbol{\beta}_1 + \mathbf{e}_n$ (i.e., $\boldsymbol{\beta}_2 = \mathbf{0}$), let $\hat{\boldsymbol{\beta}}_{1n}$ be the corresponding M-estimator of $\boldsymbol{\beta}_1$; $\hat{\boldsymbol{\beta}}_{1n}$ is termed a *restrained M-estimator* (RME) of $\boldsymbol{\beta}_1$. This RME generally performs better than the UME when $\boldsymbol{\beta}_2$ is $\mathbf{0}$ (or very close to $\mathbf{0}$). But, for $\boldsymbol{\beta}_2$ away from the pivot $\mathbf{0}$, the RME may be considerably biased, inefficient and even, possibly, inconsistent, while the UME retains its performance characteristics steadily over the variation of $\boldsymbol{\beta}_2$. For this reason, often to incorporate the rather uncertain prior information on $\boldsymbol{\beta}_2$ in the estimation of $\boldsymbol{\beta}_1$, a suitable (M-) test statistic (for testing H_0: $\boldsymbol{\beta}_2 = \mathbf{0}$) is taken into consideration. In a *preliminary test M-estimation* (PTME) formulation, the PTME $\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{PT}}$ is chosen as the RME or UME, according as this preliminary test leads to the acceptance or rejection of H_0. The *shrinkage M-estimator* (SME), based on the usual James–Stein (1961) rule, incorporates the same test statistic in a smoother manner. When $\boldsymbol{\beta}_2$ is very close to $\mathbf{0}$, generally, both the PTME and SME perform better than the UME, but the RME may still be better than either of them. On the other hand, for $\boldsymbol{\beta}_2$ away from $\mathbf{0}$, the RME may perform rather poorly, while both the PTME and SME are robust. This relative picture on the performance characteristics of all four versions of M-estimators can best be studied in an asymptotic setup similar to that in Sen (1984) or Sen and Saleh (1985). Shrinkage M-estimation of the multivariate location has also been studied in the same vein by Saleh and Sen (1985). The object of the present study is to focus mainly on the linear models. In passing, we may remark that for the particular case of $p_1 = 0$, i.e., $p_2 = p$, we have the classical shrinkage model, while for $p_1 \geq 1$, we have a *partial shrinkage model* not treated in this generality in other places.

The proposed PTME and SME, along with the preliminary notions, are presented in Section 2. The notion of *asymptotic distributional risk* (ADR) is considered in Section 3 and, in this light, the ADR results for the various versions of the M-estimators are formulated in the same section. The main results on the *asymptotic risk efficiency* (ARE) of the different versions of M-estimators are presented in Section 4. The concluding section deals with some general discussions (including the *asymptotic (distributional) minimax* character of these estimators).

2. The proposed PTME and SME. First, we introduce the *score function* $\psi = \{\psi(x),\ x \in R\}$ needed for the definition of M-estimators. We assume that

$$(2.1) \qquad \psi(x) = \psi_1(x) + \psi_2(x), \qquad x \in R,$$

where ψ_1 and ψ_2 are both nondecreasing and skew-symmetric [i.e., $\psi_j(x) + \psi_j(-x) = 0,\ \forall\ x \in R,\ j = 1,2$]; ψ_1 is absolutely continuous on any bounded interval in R and ψ_2 is a step function having only finitely many jumps. Also, we assume that there exists a finite positive constant k, such that $\psi(x) = \psi(k)\operatorname{sign} x$, for $|x| \geq k$, and ψ is nonconstant on $[-k, k]$, so that

$$0 < \sigma_\psi^2 = \int_R \psi^2(x)\, dF(x) < \infty. \tag{2.2}$$

Then let $\mathbf{A}'_n = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ and, for every $\mathbf{b} \in R^p$ and $n \geq 1$, define

$$\mathbf{M}_n(\mathbf{b}) = \big(M_{n1}(\mathbf{b}), \dots, M_{np}(\mathbf{b})\big)' = \sum_{i=1}^n \mathbf{a}_i \psi(X_i - \mathbf{a}'_i\mathbf{b}), \qquad \mathbf{b} \in R^p. \tag{2.3}$$

Also, we assume that the d.f. F (of the e_i) is symmetric about 0, so that

$$\bar{\psi} = \int_R \psi(x)\, dF(x) = 0. \tag{2.4}$$

Further, we let

$$\mathbf{C}_n = \mathbf{A}'_n\mathbf{A}_n = \begin{pmatrix} \mathbf{A}'_{n1}\mathbf{A}_{n1} & \mathbf{A}'_{n1}\mathbf{A}_{n2} \\ \mathbf{A}'_{n2}\mathbf{A}_{n1} & \mathbf{A}'_{n2}\mathbf{A}_{n2} \end{pmatrix} = \begin{pmatrix} \mathbf{C}_{n11} & \mathbf{C}_{n12} \\ \mathbf{C}_{n21} & \mathbf{C}_{n22} \end{pmatrix} \tag{2.5}$$

and assume that there exists a positive definite (p.d.) matrix $\mathbf{C}$, such that as $n \to \infty$,

$$n^{-1}\mathbf{C}_n \to \mathbf{C} = \begin{pmatrix} \mathbf{C}_{11} & \mathbf{C}_{12} \\ \mathbf{C}_{21} & \mathbf{C}_{22} \end{pmatrix}, \tag{2.6}$$

$$n^{-1} \sum_{i=1}^n (\mathbf{a}'_i\mathbf{a}_i)^2 = O(1). \tag{2.7}$$

Note that (2.6) and (2.7) ensure that

$$\max_{1 \leq i \leq n} \left\{\mathbf{a}'_i\mathbf{C}_n^{-1}\mathbf{a}_i\right\} = O(n^{-1/2}) = o(1), \quad \text{as } n \to \infty. \tag{2.8}$$

Now, the UME $\tilde{\boldsymbol{\beta}}_n = (\tilde{\boldsymbol{\beta}}'_{1n}, \tilde{\boldsymbol{\beta}}'_{2n})'$ of $\boldsymbol{\beta}$ is a solution to

$$\mathbf{M}_n(\mathbf{b}) = \mathbf{0}. \tag{2.9}$$

We also write $\mathbf{M}_n(\mathbf{b}) = (\mathbf{M}'_{n1}(\mathbf{b}_1, \mathbf{b}_2), \mathbf{M}'_{n2}(\mathbf{b}_1, \mathbf{b}_2))'$, where for the $\mathbf{M}_n$ and $\mathbf{b}$, we use the same partitioning as in (1.2). Then, the RME $\hat{\boldsymbol{\beta}}_{1n}$ of $\boldsymbol{\beta}_1$ is a solution to

$$\mathbf{M}_{n(1)}(\mathbf{b}_1, \mathbf{0}) = \mathbf{0}. \tag{2.10}$$

For the PTME and SME, we need to introduce a suitable (M-) test statistic for testing the null hypothesis H_0: $\boldsymbol{\beta}_2 \doteq \mathbf{0}$. Toward this, we proceed as in Sen (1982) and Singer and Sen (1985) and let

$$\hat{\mathbf{M}}_{n(2)} = \mathbf{M}_{n(2)}(\hat{\boldsymbol{\beta}}_{1n}, \mathbf{0}), \tag{2.11}$$

where $\hat{\boldsymbol{\beta}}_{1n}$, the RME of $\boldsymbol{\beta}_1$, is defined by (2.10). Also, let

$$S_n^2 = n^{-1} \sum_{i=1}^n \psi^2\big(X_i - \mathbf{a}'_{i(1)}\hat{\boldsymbol{\beta}}_{1n}\big), \qquad \mathbf{a}'_i = \big(\mathbf{a}'_{i(1)}, \mathbf{a}'_{i(2)}\big), \qquad i \geq 1, \tag{2.12}$$

$$\mathbf{C}_{nrr\cdot s} = \mathbf{C}_{nrr} - \mathbf{C}_{nrs}\mathbf{C}_{nss}^{-1}\mathbf{C}_{nsr}, \quad \text{for } r \neq s = 1, 2. \tag{2.13}$$

Then, an appropriate (aligned M-) test statistic is

$$\mathscr{L}_n = S_n^{-2}\left\{\hat{\mathbf{M}}'_{n(2)}\mathbf{C}_{n22\cdot 1}\hat{\mathbf{M}}_{n(2)}\right\}. \tag{2.14}$$

Under H_0, $\mathscr{L}_n$ has asymptotically the chi-square d.f. with p_2 degrees of freedom (DF). Thus, corresponding to a preassigned level of significance α $(0 < \alpha < 1)$, the preliminary test for H_0 may be based on the following:

(2.15) Accept or reject H_0 according as $\mathscr{L}_n$ is $<$ or $\geq$ $\chi^2_{p_2,\alpha}$, where $\chi^2_{t,\alpha}$ is the upper $100\alpha\%$ point of the chi-square d.f. with t DF.

The PTME is then defined by

$$\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{PT}} = \tilde{\boldsymbol{\beta}}_{1n} I\left(\mathscr{L}_n \geq \chi^2_{p_2,\alpha}\right) + \hat{\boldsymbol{\beta}}_{1n} I\left(\mathscr{L}_n < \chi^2_{p_2,\alpha}\right), \tag{2.16}$$

where $I(A)$ stands for the indicator function of the set A. Note that for defining the PTME, it suffices to assume that $p_2 \geq 1$.

To introduce the SME, we consider a (given) positive definite matrix $\mathbf{W}$ (which we adopt in the definition of the risk later on) and let

$$d_n = \mathrm{ch}_{p_1}\left(n\mathbf{W}\mathbf{C}_{n11\cdot 2}^{-1}\right) = \text{smallest characteristic root of } n\mathbf{W}\mathbf{C}_{n11\cdot 2}^{-1}. \tag{2.17}$$

The SME may then be defined as

$$\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{S}} = \hat{\boldsymbol{\beta}}_{1n} + \left(\mathbf{I}_{p_1} - c d_n n^{-1}\mathscr{L}_n^{-1}\mathbf{W}^{-1}\mathbf{C}_{n11\cdot 2}\right)\left(\tilde{\boldsymbol{\beta}}_{1n} - \hat{\boldsymbol{\beta}}_{1n}\right), \tag{2.18}$$

where c is a (positive) *shrinkage factor* to be defined more precisely later on. Note that (2.18) is in line with the general prescription of Berger, Bock, Brown, Casella and Gleser (1977), where the case of the multinormal mean with unknown (and arbitrary) covariance matrix has been treated. To simplify (2.18) further, we assume that the d.f. F has an absolutely continuous density function f (a.e.) with a finite Fisher information $I(f) = \int_R \{f'(x)/f(x)\}^2\, dF(x)$. Also, let $\mathbf{C}_{ii\cdot j} = \mathbf{C}_{ii} - \mathbf{C}_{ij}\mathbf{C}_{jj}^{-1}\mathbf{C}_{ji}$, for $i \neq j = 1, 2$. Then, proceeding as in Singer and Sen (1985), we obtain that

$$n^{1/2}\left(\tilde{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1\right) \sim_{\mathscr{D}} \mathscr{N}_{p_1}\left(\mathbf{0}, \sigma_\psi^2 \gamma^{-2}\mathbf{C}_{11\cdot 2}^{-1}\right), \tag{2.19}$$

where

$$\gamma = \int_R \psi(x)\{-f'(x)/f(x)\}\, dF(x). \tag{2.20}$$

As such, the Mahalanobis distance of $\tilde{\boldsymbol{\beta}}_{1n}$ from $\boldsymbol{\beta}_1$ may be taken as

$$L(\boldsymbol{\beta}_{1n}, \boldsymbol{\beta}_1) = \left\{\left(\tilde{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1\right)'\mathbf{C}_{n11\cdot 2}\left(\tilde{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1\right)\right\}\gamma^2/\sigma_\psi^2. \tag{2.21}$$

With this interpretation of the loss function, it may be quite natural to choose $\mathbf{W} = n^{-1}\mathbf{C}_{n11\cdot 2}$ $(\sim \mathbf{C}_{11\cdot 2})$, in which case, by (2.17), $d_n = 1$ and, hence, (2.18) reduces to

$$\hat{\boldsymbol{\beta}}_{1n}^{\mathrm{S}} = \hat{\boldsymbol{\beta}}_{1n} + \left(1 - \mathscr{L}_n^{-1}\right)\left(\tilde{\boldsymbol{\beta}}_{1n} - \hat{\boldsymbol{\beta}}_{1n}\right). \tag{2.22}$$

In the sequel, we shall mainly use the SME in (2.22), though in the last section we shall comment on the general case in (2.18). Note that in the PTME in (2.16), the indicator functions are 0–1 valued r.v., while in (2.18) or (2.22) we have a smoother version for the SME. Following Sclove, Morris and Radhakrishnan (1972), we may also consider the following *positive-rule SME*:

$$\hat{\boldsymbol{\beta}}_{1,n}^{S(+)} = \hat{\boldsymbol{\beta}}_{1,n} + \left(1 - c\mathscr{L}_n^{-1}\right)^+ \left(\tilde{\boldsymbol{\beta}}_{1,n} - \hat{\boldsymbol{\beta}}_{1,n}\right), \tag{2.23}$$

where a^+ is equal to $a \vee 0$. Note that (2.16) and (2.23) do not agree even if we let $c = \chi^2_{p_2,\alpha}$. However, (2.23) may have some advantage over (2.16) and we shall make some comments on it in the concluding section.

We may note that the M-test based on $\mathscr{L}_n$ in (2.14) is consistent against any (fixed) $\boldsymbol{\beta}_2 \neq \mathbf{0}$, so that the alternative estimators in (2.14), (2.18), (2.22) and (2.23) would be all asymptotically equivalent to the UME $\tilde{\boldsymbol{\beta}}_{1,n}$. Hence, to avoid this asymptotic degeneracy, we consider the case of $\boldsymbol{\beta}_2$ being "close to" $\mathbf{0}$, where these different versions of the M-estimators have nonequivalent performance characteristics.

3. ADR of PTME and SME. In the classical normal theory model, with a loss function defined as in (2.21), the risk is computed as the expected loss. For the classical M-estimators and the PTME, this risk can be computed under the regularity conditions of Section 2 (which ensure the moment convergence of M-estimators). This is also true for the positive-rule M-estimator in (2.23). Thus, in each of these cases, one may compute the asymptotic risk directly by using standard asymptotic results on the actual risk. However, the situation is quite different for the SME in (2.18) or (2.22). A quadratic error [such as in (2.21)] in the SME [say, (2.22)] involves $\mathscr{L}_n^{-1}$ and $\mathscr{L}_n^{-2}$, in addition to the usual factors involving the UME and RME. In the normal distributional models, dealing with the traditional MLE's and the likelihood ratio test statistics, the celebrated *Stein identities* [viz., Stein (1981)] provide access for this computation under additional regularity conditions (on p_1, p_2 and the design matrix $\mathbf{C}_n$). However, sans the multinormality assumption, the ingenuity of the Stein identities may not hold. In fact, it is trivial to construct pathological examples where for quadratic error loss functions, the SME or the classical James–Stein (1961) estimator does not have a finite risk when the underlying F is nonnormal. Dealing with shrinkage U-statistics, Sen (1984) has shown that under additional regularity conditions, ensuring the existence of the negative moments of $\mathscr{L}_n$, this asymptotic risk can be computed. The same result holds for R-estimators of location [cf. Sen and Saleh (1985)]. In the current case, to retain the simplicity of the regularity conditions (of Section 2), we shall compute the risk by reference to the asymptotic distribution of an estimator and term the same as the *asymptotic distributional risk*. For the UME, RME, PTME and the positive rule estimator, the ADR agrees with the corresponding asymptotic risk under the same regularity conditions. In the case of the SME, the ADR can be studied under the same regularity conditions (as in the case with other versions), but the asymptotic risk, to be conformable with the ADR, would demand additional regularity conditions. Since this relative picture is similar to the U-statistics case treated in Sen (1984), we shall omit the details. As such, we shall mainly confine ourselves to the study of the ADR properties of all the versions of M-estimators and comment on their asymptotic dominance in the light of these ADR results.

In the multivariate location model, Sen and Saleh (1985) have pointed out that shrinkage estimation works out well only in a shrinking neighbourhood of the pivot. A similar picture holds here too. Since for β_2 the pivot is taken as $\mathbf{0}$, we consider a shrinking neighbourhood of $\mathbf{0}$ and toward this, we consider the sequence $\{K_n\}$ of alternatives, where

$$K_n: \beta_2 = \beta_{2(n)} = n^{-1/2}\xi, \qquad \xi = (\xi_{p_1+1}, \dots, \xi_p)' \in R^{p_2}, \tag{3.1}$$

so that the null hypothesis H_0 reduces to H_0: $\xi = \mathbf{0}$.

For a suitable estimator β_{1n}^* of β_1, we denote by

$$G^*(\mathbf{x}) = \lim_{n\to\infty} P\{n^{1/2}(\beta_{1n}^* - \beta_1) \le \mathbf{x}|K_n\}, \qquad \mathbf{x} \in R^{p_1}, \tag{3.2}$$

where we assume that G^* is nondegenerate. Then with a quadratic error loss $n(\beta_{1n}^* - \beta_1)'\mathbf{W}(\beta_{1n}^* - \beta_1)$, for a suitable (p.d.) $\mathbf{W}$, the ADR of β_{1n}^* is defined as

$$R(\beta_1^*; \mathbf{W}) = \operatorname{tr}\left\{\mathbf{W}\int_{R^p} \cdots \int \mathbf{x}\mathbf{x}'\, dG^*(\mathbf{x})\right\} = \operatorname{tr}\{\mathbf{W}\mathbf{V}^*\}, \quad \text{say}, \tag{3.3}$$

where $\mathbf{V}^*$ is the dispersion matrix for the asymptotic distribution G^*. We also denote an m-variate normal d.f. with mean vector μ and dispersion matrix Σ by $G_m(\mathbf{x}; \mu, \Sigma)$ and the d.f. of a noncentral chi-square distribution with r DF and noncentrality parameter Δ by $H_r(x; \Delta)$, $x \ge 0$. Then, we have

Theorem 3.1. *Under $\{K_n\}$ and the regularity conditions* (2.1), (2.2), (2.6), (2.7) *and for the density f with finite Fisher information $I(f)$,*

$$\lim_{n\to\infty} P\{\mathscr{L}_n \le x|K_n\} = H_{p_2}(x; \Delta), \qquad \Delta = \sigma_\psi^{-2}(\xi'\mathbf{C}_{22\cdot1}\xi)\gamma^2, \tag{3.4}$$

$$\lim_{n\to\infty} P\{n^{1/2}(\tilde{\beta}_{1n} - \beta_1) \le \mathbf{x}|K_n\} = G_{p_1}(\mathbf{x}; \mathbf{0}, \gamma^{-2}\sigma_\psi^2\mathbf{C}_{11\cdot2}^{-1}), \tag{3.5}$$

$$\lim_{n\to\infty} P\{n^{1/2}(\hat{\beta}_{1n} - \beta_1) \le \mathbf{x}|K_n\} = G_{p_1}(\mathbf{x} + \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\xi; \mathbf{0}, \gamma^{-2}\sigma_\psi^2\mathbf{C}_{11}^{-1}), \tag{3.6}$$

$$\begin{aligned}\lim_{n\to\infty} &P\{n^{1/2}(\hat{\beta}_{1n}^{\mathrm{PT}} - \beta_1) \le \mathbf{x}|K_n\} \\ &= H_{p_2}(\chi^2_{p_2,\alpha}; \Delta)G_{p_1}(\mathbf{x} + \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\xi; \mathbf{0}, \gamma^{-2}\sigma_\psi^2\mathbf{C}_{11}^{-1}) \\ &\quad + \int_{E(\xi)} G_{p_1}(\mathbf{x} - \mathbf{D}_{12}\mathbf{D}_{22}^{-1}\mathbf{z}; \mathbf{0}, \gamma^{-2}\sigma_\psi^2\mathbf{D}_{11\cdot2})\, dG_{p_2}(\mathbf{z}; \mathbf{0}, \gamma^{-2}\sigma_\psi^2\mathbf{D}_{22}),\end{aligned} \tag{3.7}$$

where $\mathbf{D} = \mathbf{C}^{-1}$ and the $\mathbf{D}_{ij}$ and $\mathbf{D}_{ii\cdot j}$ are defined as in (2.6) *and* (2.13), *and* $E(\xi) = \{\mathbf{z}: \sigma_\psi^{-2}\gamma^2(\mathbf{z} + \xi)'\mathbf{C}_{22\cdot1}(\mathbf{z} + \xi) \ge \chi^2_{p_2,\alpha}\}$. *Finally,*

$$n^{1/2}(\hat{\beta}_{1n}^{\mathrm{S}} - \beta_1) \to_{\mathscr{D}} \mathbf{D}_1\mathbf{U} + \frac{\gamma^{-2}\sigma_\psi^2 c\mathbf{C}_{11}^{-1}\mathbf{C}_{12}(\mathbf{D}_2\mathbf{U} + \xi)}{(\mathbf{D}_2\mathbf{U} + \xi)'\mathbf{C}_{22\cdot1}(\mathbf{D}_2\mathbf{U} + \xi)}, \tag{3.8}$$

where

$$\mathbf{U} \sim \mathscr{N}_p(\mathbf{0}, \gamma^{-2}\sigma_\psi^2\mathbf{C}) \quad \text{and} \quad \mathbf{D} = \begin{pmatrix}\mathbf{D}_1 \\ \mathbf{D}_2\end{pmatrix}. \tag{3.9}$$

PROOF. (3.4) follows directly from Singer and Sen [(1985), Section 3], while (3.5) is a restatement of (2.19). Further, (3.6) follows directly from Theorem 3.1 of Jurečková and Sen (1984) and related results in Section 3 of Singer and Sen (1985). Actually, these linearity results [viz., Singer and Sen (1985)] imply that under $\{K_n\}$,

$$\hat{\boldsymbol{\beta}}_{1n} = \tilde{\boldsymbol{\beta}}_{1n} + \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\tilde{\boldsymbol{\beta}}_{2n} + o_p(n^{-1/2}), \tag{3.10}$$

$$\mathscr{L}_n = \left(\tilde{\boldsymbol{\beta}}_{2n}'\mathbf{C}_{n22\cdot1}\tilde{\boldsymbol{\beta}}_{2n}\right)\gamma^2/\sigma_\psi^2 + o_p(1), \tag{3.11}$$

so that the PTME and SME may both be expressed in terms of the UME $\tilde{\boldsymbol{\beta}}_n$. Recall that under $\{K_n\}$,

$$n^{1/2}\left(\left(\tilde{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1\right)', \left(\tilde{\boldsymbol{\beta}}_{2n} - n^{-1/2}\boldsymbol{\xi}\right)'\right)' \sim_{\mathscr{D}} \mathscr{N}_p\left(\mathbf{0}, \gamma^{-2}\sigma_\psi^2\mathbf{C}^{-1}\right). \tag{3.12}$$

Note that by definition, $\mathbf{D}_{11} + \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{D}_{21} = \mathbf{C}_{11}^{-1}$ and $\mathbf{D}_{12} + \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{D}_{22} = \mathbf{0}$. Hence, using (3.10) and (3.11), we may write $\hat{\boldsymbol{\beta}}_{1n} = \mathbf{L}_1\tilde{\boldsymbol{\beta}}_n + \mathbf{o}_p(n^{-1/2})$ and $\mathscr{L}_n = n\tilde{\boldsymbol{\beta}}_n'\mathbf{L}_2\tilde{\boldsymbol{\beta}}_n + o_p(1)$, where $\mathbf{L}_1\mathbf{C}^{-1}\mathbf{L}_2 = \mathbf{0}$. Hence using (3.12), we conclude that under $\{K_n\}$, $n^{1/2}(\hat{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1)$ and $\mathscr{L}_n$ are asymptotically independent, while the joint distribution of $n^{1/2}(\tilde{\boldsymbol{\beta}}_{1n} - \boldsymbol{\beta}_1)$ and $\mathscr{L}_n$ can be obtained from (3.12) by integrating over the proper subspace. As such, following the same arguments as in the proof of Theorem 3.2 of Sen and Saleh (1979), we arrive at (3.7). Note that by Theorem 3.1 of Jurečková and Sen (1984) [and Singer and Sen (1985)], under $\{K_n\}$ and the assumed regularity conditions,

$$n^{1/2}\begin{pmatrix} \boldsymbol{\beta}_{1n} & -\boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_{2n} & -n^{-1/2}\boldsymbol{\xi} \end{pmatrix} = \mathbf{C}^{-1}\gamma^{-1}n^{-1/2}\mathbf{M}_n(\boldsymbol{\beta}) + \mathbf{o}_p(1), \tag{3.13}$$

where $n^{-1/2}\gamma^{-1}\mathbf{M}_n(\boldsymbol{\beta}) \to_{\mathscr{D}} \mathbf{U}$, as defined by (3.9). Since $\mathbf{C}^{-1} = \begin{pmatrix} \mathbf{D}_1 \\ \mathbf{D}_2 \end{pmatrix}$, (3.8) follows readily from (2.22), (3.10), (3.11) and (3.13). □

THEOREM 3.2. *Under the hypothesis of Theorem* 3.1, *the following ADR results hold*:

$$R\left(\tilde{\boldsymbol{\beta}}_1; \mathbf{W}\right) = \left(\sigma_\psi^2\gamma^{-2}\right)\operatorname{tr}\left(\mathbf{W}\mathbf{C}_{11\cdot2}^{-1}\right), \tag{3.14}$$

$$R\left(\hat{\boldsymbol{\beta}}_1; \mathbf{W}\right) = \left(\sigma_\psi^2\gamma^{-2}\right)\operatorname{tr}\left(\mathbf{W}\mathbf{C}_{11}^{-1}\right) + \boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi}, \tag{3.15}$$

$$\begin{aligned} R\left(\hat{\boldsymbol{\beta}}_1^{\mathrm{PT}}; \mathbf{W}\right) = \left(\sigma_\psi^2\gamma^{-2}\right)&\Big\{\operatorname{tr}\left(\mathbf{W}\mathbf{C}_{11\cdot2}^{-1}\right)\left[1 - H_{p_2+2}\left(\chi_{p_2,\alpha}^2; \Delta\right)\right] \\ &+ \operatorname{tr}\left(\mathbf{W}\mathbf{C}_{11}^{-1}\right)H_{p_2+2}\left(\chi_{p_2,\alpha}^2; \Delta\right)\Big\} \\ &+ (\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi})\left[2H_{p_2+2}\left(\chi_{p_2,\alpha}^2; \Delta\right) - H_{p_2+4}\left(\chi_{p_2,\alpha}^2; \Delta\right)\right], \end{aligned} \tag{3.16}$$

$$\begin{aligned} R\left(\hat{\boldsymbol{\beta}}_1^{\mathrm{S}}; \mathbf{W}\right) = \left(\sigma_\psi^2\gamma^{-2}\right)&\Big\{\operatorname{tr}\left(\mathbf{W}\mathbf{C}_{11\cdot2}^{-1}\right) \\ &- c\operatorname{tr}\left(\mathbf{M}\mathbf{C}_{22\cdot1}^{-1}\right)\left[2E\left(\chi_{p_2+2}^{-2}(\Delta)\right) - cE\left(\chi_{p_2+2}^{-4}(\Delta)\right)\right]\Big\} \\ &+ c(c+4)(\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi})E\left(\chi_{p_2+4}^{-4}(\Delta)\right), \end{aligned} \tag{3.17}$$

where Δ *and* $H_q(\cdot)$ *are defined in* (3.4), $E(\chi_q^{-2r}(\Delta)) = \int_0^\infty x^{-r}\, dH_q(x; \Delta)$ *and*

$$\mathbf{M} = \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{W}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}. \tag{3.18}$$

PROOF. Note that (3.14) and (3.15) follow directly from (3.3), (3.5) and (3.6). For the asymptotic distribution in (3.7), we use the results in Section 4 of Sen and Saleh (1979) and obtain the corresponding dispersion matrix $\mathbf{V}^{PT}$ as

$$\begin{aligned} & H_{p_2}\big(\chi^2_{p_2,\alpha}; \Delta\big)\gamma^{-2}\sigma_\psi^2\mathbf{C}_{11}^{-1} + \Big[1 - H_{p_2}\big(\chi^2_{p_2,\alpha}; \Delta\big)\Big]\gamma^{-2}\sigma_\psi^2\mathbf{D}_{11\cdot 2} \\ & + \mathbf{D}_{12}\mathbf{D}_{22}^{-1}\mathbf{D}_{21}\Big[1 - H_{p_2+2}\big(\chi^2_{p_2,\alpha}; \Delta\big)\Big] \\ & + \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\Big[2H_{p_2+2}\big(\chi^2_{p_2,\alpha}; \Delta\big) - H_{p_2+4}\big(\chi^2_{p_2,\alpha}; \Delta\big)\Big] \\ & = \gamma^{-2}\sigma_\psi^2\Big[\mathbf{C}_{11}^{-1}H_{p_2+2}\big(\chi^2_{p_2,\alpha}; \Delta\big) + \mathbf{C}_{11\cdot 2}^{-1}\big(1 - H_{p_2+2}\big(\chi^2_{p_2,\alpha}; \Delta\big)\big)\Big] \\ & + \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\boldsymbol{\xi}\boldsymbol{\xi}'\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\Big[2H_{p_2+2}\big(\chi^2_{p_2,\alpha}; \Delta\big) - H_{p_2+4}\big(\chi^2_{p_2,\alpha}; \Delta\big)\Big]. \end{aligned} \tag{3.19}$$

Then (3.16) follows directly from (3.19), (3.3) and the definition of $\mathbf{M}$ in (3.18). Next, we note that by (3.3), (3.8) and (3.9), the left-hand side of (3.17) is equal to

$$\begin{aligned} & \gamma^{-2}\sigma_\psi^2\operatorname{tr}(\mathbf{W}\mathbf{D}_1\mathbf{C}\mathbf{D}_1') + 2\gamma^{-2}\sigma_\psi^2 cE\left\{\frac{(\mathbf{D}_2U + \boldsymbol{\xi})'\mathbf{C}_{21}\mathbf{C}_{11}\mathbf{W}\mathbf{D}_1\mathbf{U}}{(\mathbf{D}_2\mathbf{U} + \boldsymbol{\xi})'\mathbf{C}_{22\cdot 1}(\mathbf{D}_2\mathbf{U} + \boldsymbol{\xi})}\right\} \\ & + c^2\sigma_\psi^4\gamma^{-4}E\left\{\frac{(\mathbf{D}_2\mathbf{U} + \boldsymbol{\xi})'\mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{W}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}(\mathbf{D}_2\mathbf{U} + \boldsymbol{\xi})}{[(\mathbf{D}_2\mathbf{U} + \boldsymbol{\xi})'\mathbf{C}_{22\cdot 1}(\mathbf{D}_2\mathbf{U} + \boldsymbol{\xi})]^2}\right\}, \end{aligned} \tag{3.20}$$

where $\operatorname{tr}(\mathbf{W}\mathbf{D}_1\mathbf{C}\mathbf{D}_1') = \operatorname{tr}(\mathbf{W}\mathbf{C}_{11\cdot 2}^{-1})$. Using (3.18) and the Stein identity [viz., Appendix B of Judge and Bock (1978)], the last two terms in (3.20) reduce to

$$-2\gamma^{-2}\sigma_\psi^2 c\operatorname{tr}\big(\mathbf{M}\mathbf{C}_{22\cdot 1}^{-1}\big)E\big(\chi_{p_2+2}^{-2}(\Delta)\big) + 4c(\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi})E\big(\chi_{p_2+4}^{-4}(\Delta)\big) \tag{3.21}$$

and

$$\gamma^{-2}\sigma_\psi^2 c^2\operatorname{tr}\big(\mathbf{M}\mathbf{C}_{22\cdot 1}^{-1}\big)E\big(\chi_{p_2+2}^{-4}(\Delta)\big) + c^2(\boldsymbol{\xi}'\mathbf{M}\boldsymbol{\xi})E\big(\chi_{p_2+4}^{-4}(\Delta)\big), \tag{3.22}$$

respectively. (3.20)–(3.22) lead to (3.17). □

In the light of these ADR results, the *asymptotic distributional risk-efficiency* (ADRE) results are considered in the next section.

4. ADRE results. Note that for $\mathbf{C}_{12} = \mathbf{0}$, we have $\mathbf{M} = \mathbf{0}$ and $\mathbf{C}_{11\cdot 2} = \mathbf{C}_{11}$ and, hence, (3.14)–(3.17) all reduce to the common value $\sigma_\psi^2\gamma^{-2}\operatorname{tr}(\mathbf{W}\mathbf{C}_{11}^{-1})$ ($\forall\ \boldsymbol{\xi}$), so that all these versions become ADR-equivalent. Hence in the sequel, it will be assumed that $\mathbf{C}_{12} \neq \mathbf{0}$. Also, note that (3.14) does not depend on $\boldsymbol{\xi}$ and, as argued before (2.22), we may choose $\mathbf{W}$ as proportional to $\mathbf{C}_{11\cdot 2}$. In fact, we let $\mathbf{W} = \gamma^2\sigma_\psi^{-2}\mathbf{C}_{11\cdot 2}$, so that (3.14) reduces to p_1. Then let

$$\mathbf{M}^* = \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{11\cdot 2}\mathbf{C}_{11}^{-1}\mathbf{C}_{12} \quad \text{and} \quad \mathbf{M}^0 = \mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21}\mathbf{C}_{11}^{-1}. \tag{4.1}$$

THEOREM 4.1. *For* $\mathbf{W} = \gamma^2\sigma_\psi^{-2}\mathbf{C}_{11\cdot 2}$, *under the hypothesis of Theorem* 3.2,

$$R(\tilde{\beta}_1;\mathbf{W}) \gtreqless R(\hat{\beta}_1;\mathbf{W}) \quad \textit{according as} \quad \gamma^2\frac{(\xi'\mathbf{M}^*\xi)}{\sigma_\psi^2} \lesseqgtr \operatorname{tr}(\mathbf{M}^0), \tag{4.2}$$

$$R(\tilde{\beta}_1;\mathbf{W}) \gtreqless R(\hat{\beta}_1^{\mathrm{PT}};\mathbf{W}) \quad \textit{according as}$$

$$(\xi'\mathbf{M}^*\xi) \lesseqgtr \frac{\sigma_\psi^2\gamma^{-2}\operatorname{tr}(\mathbf{M}^0)H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right)}{2H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) - H_{p_2+4}\left(\chi^2_{p_2,\alpha};\Delta\right)}. \tag{4.3}$$

The proof follows directly from (3.14)–(3.16) and the particular choice of **W**. Since $\operatorname{tr}(\mathbf{M}^0)$ is positive for $\mathbf{C}_{12} \neq \mathbf{0}$, it follows from (4.2) and (4.3) that the UME fails to dominate the RME or the PTME in the light of their ADR. In a similar manner, it follows that the PTME fails to dominate the UME or RME. Note that at $\xi = \mathbf{0}$ (i.e., under H_0),

$$\begin{aligned} p_1 = R_0(\tilde{\beta}_1;\mathbf{W}) &\geq p_1 - \operatorname{tr}(\mathbf{M}^{**})H_{p_2+2}\left(\chi^2_{p_2,\alpha};\Delta\right) = R_0(\hat{\beta}_1^{\mathrm{PT}};\mathbf{W}) \\ &\geq p_1 - \operatorname{tr}(\mathbf{M}^{**}) = R_0(\hat{\beta}_1;\mathbf{W}), \end{aligned} \tag{4.4}$$

where $R_0(\cdot)$ stands for the ADR under H_0 and

$$\mathbf{M}^{**} = \mathbf{C}_{11}^{-1}\mathbf{C}_{12}\mathbf{C}_{22}^{-1}\mathbf{C}_{21} \quad \left[\text{so that } \operatorname{tr}(\mathbf{M}^{**}) = \operatorname{tr}(\mathbf{M}^0)\right]. \tag{4.5}$$

On the other hand, as ξ moves away from **0** (i.e., $\Delta^* = \xi'\mathbf{M}^*\xi$ increases), $R(\hat{\beta}_1;\mathbf{W})$ monotonically (in Δ^*) increases and it goes to $+\infty$ as $\Delta^* \to +\infty$. The ADR of the PTME also increases as ξ moves away from **0**; in fact, this ADR crosses the line p_1 as ξ moves out of a closed neighbourhood of the pivot and then continues to stay above this line. However, this excess over p_1 is usually small, and as Δ or Δ^* goes to $+\infty$, this ADR approaches the limit p_1. Thus, the PTME has a bounded ADR, and though its value may exceed the ADR of the UME, its smaller values in the neighbourhood of the pivot and its small excess over the asymptote p_1 for large values of the noncentrality parameter make it a very attractive estimator under certain prior information on β_2.

The conclusions derived from Theorem 4.1 do not need any condition on the matrix $\mathbf{M}^0$ (apart from its nonnull character). In the case of the SME, however, we need to introduce extra regularity conditions on the shrinkage factor c as well as the design matrix **C** (and p_1, p_2). To derive a clear picture, we present this in two phases. First, we consider the dominance of the SME over UME (in the light of their ADR) and then we draw a relative picture of the ADR of the PTME and SME. We define $\mathbf{M}^0$ as in (4.1) and let $\operatorname{ch}_1(\cdot)$ stand for the largest characteristic root. Let

$$h = \operatorname{ch}_1(\mathbf{M}^0)/\operatorname{tr}(\mathbf{M}^0), \quad \text{so that } 0 < h \leq 1. \tag{4.6}$$

THEOREM 4.2. *A sufficient condition for the asymptotic dominance of the SME over the UME* [*i.e., for* $R(\hat{\beta}_1;\mathbf{W}) \geq R(\hat{\beta}_1^{\mathrm{S}};\mathbf{W})$ *for all* $\xi \in R^{p_2}$] *is that the* *shrinkage factor c is positive and it satisfies the inequality*

$$2E\{\chi_{p_2+2}^{-2}(\Delta)\} - cE\{\chi_{p_2+2}^{-4}(\Delta)\} - (c+4)h\Delta E\{\chi_{p_2+4}^{-4}(\Delta)\} \geq 0, \quad \forall \Delta \geq 0, \tag{4.7}$$

which, in turn, requires that

$$p_2 \geq 3, \quad 0 < c < 2(p_2 - 2) \quad \text{and} \quad h(c+4) \leq 2. \tag{4.8}$$

PROOF. With $\mathbf{W}$, $\mathbf{M}^*$ and $\mathbf{M}^0$ defined as in Theorem 4.1, we note that by (3.14) and (3.17) we have some simplifications,

$$\begin{aligned} R(\tilde{\boldsymbol{\beta}}_1; \mathbf{W}) - R(\hat{\boldsymbol{\beta}}_1^S; \mathbf{W}) &= c\,\mathrm{tr}(\mathbf{M}^0)\left[2E\left(\chi_{p_2+2}^{-2}(\Delta)\right) - cE\left(\chi_{p_2+2}^{-4}(\Delta)\right)\right] \\ &\quad - c(c+4)\gamma^2\sigma_\psi^{-2}(\boldsymbol{\xi}'\mathbf{M}^*\boldsymbol{\xi})E\left(\chi_{p_2+4}^{-4}(\Delta)\right), \end{aligned} \tag{4.9}$$

where we have made use of the identity that $\mathrm{tr}(\mathbf{M}\mathbf{C}_{22\cdot 1}^{-1}) = \mathrm{tr}(\mathbf{M}^0)$. Note that

$$\begin{aligned} (\boldsymbol{\xi}'\mathbf{M}^*\boldsymbol{\xi})\gamma^2\sigma_\psi^{-2}/\Delta &= (\boldsymbol{\xi}'\mathbf{M}^*\boldsymbol{\xi})/(\boldsymbol{\xi}'\mathbf{C}_{22\cdot 1}\boldsymbol{\xi}) \\ &\leq \mathrm{ch}_1(\mathbf{M}^*\mathbf{C}_{22\cdot 1}^{-1}) = \mathrm{ch}_1(\mathbf{M}^0) = h\,\mathrm{tr}(\mathbf{M}^0), \end{aligned} \tag{4.10}$$

where h is defined by (4.6). Hence, for the positivity of the right-hand side of (4.9), for every $\boldsymbol{\xi}$ (including the null case), we must have c positive and, with that, (4.7) and (4.10) ensure the positiveness of (4.9). Next, we note that

$$E\left(\chi_{p_2+2}^{-2}(\Delta)\right) - (p_2 - 2)E\left(\chi_{p_2+2}^{-4}(\Delta)\right) = \Delta E\left(\chi_{p_2+4}^{-4}(\Delta)\right). \tag{4.11}$$

Using (4.11), we rewrite (4.7) as

$$[2(p_2 - 2) - c]E\left(\chi_{p_2+2}^{-4}(\Delta)\right) + [2 - h(c+4)]\Delta E\left(\chi_{p_2+4}^{-4}(\Delta)\right). \tag{4.12}$$

From the nonnegativeness of (4.12) at $\Delta = 0$, we conclude that c has to be $\leq 2(p_2 - 2)$ (which also needs $p_2 \geq 3$). Further, using the crude inequality that $E(\chi_{p_2+4}^{-4}(\Delta)) \geq (p+2)^{-1}(p-2)E(\chi_{p_2+2}^{-4}(\Delta))$ and allowing Δ to be adequately large, we conclude that for (4.12) to be positive for all Δ, we must have $h(c+4) \leq 2$. This completes the proof of (4.8). □

Let us crucially examine the conditions laid down in (4.8). The quotient h in (4.6) plays a basic role in this context. Note that by (4.1),

$$\begin{aligned} \mathrm{tr}(\mathbf{M}^0) &= \mathrm{tr}\left(\mathbf{I}_{p_1} - \mathbf{C}_{11}^{-1}\mathbf{C}_{11\cdot 2}\right)\left(= \mathrm{tr}\left(\mathbf{I}_{p_2} - \mathbf{C}_{22}^{-1}\mathbf{C}_{22\cdot 1}\right)\right) \leq p^* \\ &= \min(p_1, p_2). \end{aligned} \tag{4.13}$$

Also, for $h \geq \frac{1}{2}$, for any positive c, $h(c+4)$ cannot be $\leq \frac{1}{2}$. Further, $\mathrm{tr}(\mathbf{M}^0)$ is the sum of the positive characteristic roots of $\mathbf{M}^0$, so that if $\mathrm{rank}(\mathbf{M}^0) \leq 2$, then by (4.6), h is $\geq \frac{1}{2}$. Thus, we may reframe (4.8) as

$$\begin{aligned} &p^* = \min(p_1, p_2) \geq 3, \\ &h < \tfrac{1}{2} \quad \text{and} \quad 0 < c \leq \min\{2h^{-1} - 4, 2(p_2 - 2)\}. \end{aligned} \tag{4.14}$$

Fortunately, this condition is verifiable from the given matrix $\mathbf{C}_n$ (or $\mathbf{C}$). For any given linear model, when seeking a partial shrinkage M-estimator, the first condition that needs to be verified is that $h < \frac{1}{2}$ and $p^* \geq 3$. Only then a positive c can be chosen, such that the third condition in (4.14) hǫlds. Thus, the shrinkage factor c depends on the desgn matrix $\mathbf{C}$ and requires bth $p_1, p_2 > 2$ and, moreover, h needs to be strictly $< \frac{1}{2}$. The situation is quite different from the classical location model where we only need that $p \geq 3$. We shall take into account these points in comparing the PTME and SME in the light of their ADR.

THEOREM 4.3. *Under the hypothesis of Theorem 4.2, the PTME fails to dominate the SME. Also, if for α, the level of significance of the PT, we have*

$$H_{p_2+2}\left(\chi^2_{p_2,\alpha}; 0\right) \geq q\{2(p_2 - 2) - q\}/p_2(p_2 - 2), \qquad q = (p_2 - 2) \wedge (2/h - 4), \tag{4.15}$$

then the SME fails to dominate the PTME.

PROOF. With $\mathbf{W}$, $\mathbf{M}^*$ and $\mathbf{M}^0$ defined as in Theorem 4.1, we have

$$\begin{aligned} R(\hat{\boldsymbol{\beta}}_1^{S}; \mathbf{W}) &- R(\hat{\boldsymbol{\beta}}_1^{PT}; \mathbf{W}) \\ &= \operatorname{tr}(\mathbf{M}^0)\left\{H_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right) - c\left[2E\left(\chi^{-2}_{p_2+2}(\Delta)\right) - cE\left(\chi^{-4}_{p_2+2}(\Delta)\right)\right]\right\} \\ &\quad - (\boldsymbol{\xi}'\mathbf{M}^*\boldsymbol{\xi})\sigma_\psi^{-2}\gamma^2\left\{2H_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right) - H_{p_2+4}\left(\chi^2_{p_2,\alpha}; \Delta\right) \right. \\ &\qquad \left. - c(c+4)E\left(\chi^{-4}_{p_2+4}(\Delta)\right)\right\}. \end{aligned} \tag{4.16}$$

We rewrite the right-hand side of (4.16) as

$$\begin{aligned} &\left\{\operatorname{tr}(\mathbf{M}^0)H_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right)\right. \\ &\quad \left. - \sigma_\psi^{-2}\gamma^2(\boldsymbol{\xi}'\mathbf{M}^*\boldsymbol{\xi})\left[2H_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right) - H_{p_2+4}\left(\chi^2_{p_2,\alpha}; \Delta\right)\right]\right\} \\ &\quad + \left\{c(c+4)(\boldsymbol{\xi}'\mathbf{M}^*\boldsymbol{\xi})\sigma_\psi^{-2}\gamma^2E\left(\chi^{-4}_{p_2+4}(\Delta)\right)\right. \\ &\qquad \left. - \operatorname{tr}(\mathbf{M}^0)c\left[2E\left(\chi^{-2}_{p_2+2}(\Delta)\right) - cE\left(\chi^{-4}_{p_2+2}(\Delta)\right)\right]\right\}. \end{aligned} \tag{4.17}$$

By (4.7) and (4.10), the second term of (4.17) is always nonpositive, while the first term is negative whenever

$$\sigma_\psi^{-2}\gamma^2(\boldsymbol{\xi}'\mathbf{M}^*\boldsymbol{\xi}) \geq \operatorname{tr}(\mathbf{M}^0)\left\{2 - H_{p_2+4}\left(\chi^2_{p_2,\alpha}; \Delta\right)\Big/H_{p_2+2}\left(\chi^2_{p_2,\alpha}; \Delta\right)\right\}^{-1}.$$

Hence, (4.16) is negative outside a closed neighborhood of the pivot, so that the PTME fails to dominate the SME when (4.7) [or (4.8)] holds. Next, we note that under H_0: $\boldsymbol{\xi} = \mathbf{0}$, (4.16) reduces to

$$\operatorname{tr}(\mathbf{M}^0)\left\{H_{p_2+2}\left(\chi^2_{p_2,\alpha}; 0\right) - p_2^{-1}c\left(2 - (p_2 - 2)^{-1}c\right)\right\}. \tag{4.18}$$

Now, $cp_2^{-1}(2 - c/(p_2 - 2))$ attains a maximum [over $c \in (0, 2(p_2 - 2))$] at $c =$

$p_2 - 2$, while, by (4.8), $c \le 2h^{-1} - 4$. Hence, defining q as in (4.15), we note that (4.18) is bounded from below by

$$\operatorname{tr}(\mathbf{M}^0)\Big\{H_{p_2+2}\big(\chi^2_{p_2,\alpha};0\big) - q[2(p_2-2) - q]/p_2(p_2-2)\Big\}.$$

Thus, (4.15) ensures the positiveness of (4.18) and this, in turn, implies that the SME fails to dominate the PTME. □

We may remark that (4.15) holds for small α, for a wide range of values of p_2 and q. For example, for $\alpha = 0.10$, (4.15) holds for all $p_2 \le 11$ (and even larger values of p_2 when q is small); for $\alpha = 0.05$, it holds for all $p_2 \le 21$. Further, the maximum excess of the ADR of the PTME over the SME is generally small compared to (4.18). Also, for the asymptotic dominance of the SME (over the UME), we require some stringent conditions on $\mathbf{C}$, p_1 and p_2 (viz., Theorem 4.2) which are not required for the PTME. From these considerations, for small values of α and for p_1, p_2 not too large, we may find the PTME quite comparable to the SME. It is difficult to prescribe a clear-cut recommendation in favor of one over the other.

5. Some general remarks. It follows from the results of Section 4 that both the PTME and SME are robust from the risk-efficiency point of view. Of the two, the SME may have generally the asymptotic minimax character (in the light of the ADR), while, in view of Theorem 4.3, the PTME is not an asymptotically minimax estimator. However, comparing the regularity conditions needed to achieve this asymptotic minimax character for the SME with those for the PTME, we feel that in many practical applications, the SME would not be that attractive, while the PTME may still lead to a robust alternative to the UME or RME.

In Section 4, we have mainly considered the SME in (2.22) and the specific case of $\mathbf{W} \sim \mathbf{C}_{11\cdot 2}$. For the general SME in (2.18) with an arbitrary $\mathbf{W}$, we may virtually repeat the steps in Theorems 3.1 and 3.2 and, denoting by

$$\mathbf{M}^{00} = \mathbf{C}_{21}\mathbf{C}_{11}^{-1}\mathbf{C}_{11\cdot 2}\mathbf{W}^{-1}\mathbf{C}_{11\cdot 2}\mathbf{C}_{11}^{-1}\mathbf{C}_{12}, \tag{5.1}$$

we arrive at the same expression for the ADR in (3.17), where the matrix $\mathbf{M}$ has to be replaced by $\mathbf{M}^{00}$. With this change, the results in Section 4 (viz., Theorems 4.2 and 4.3) also extend readily for the SME in (2.18) and an arbitrary $\mathbf{W}$. For an arbitrary $\mathbf{W}$, the SME in (2.18) would have the asymptotic minimax character [under additional regularity conditions; such as in (4.7) and (4.8)] and the one in (2.22) may not be so. However, as we have already justified the use of the particular $\mathbf{W}$ in Theorems 4.1–4.3 on the ground of the Mahalanobis distance, there is no real demand for the general case in (2.18) and, hence, our discussions in Section 4 remain pertinent for the comparative study made here.

Finally, let us comment on the positive-rule PTME in (2.23). For the multinormal mean problem, Sclove, Morris and Radhakrishnan (1972) have shown that a positive-rule MLE dominates the usual shrinkage MLE (SMLE), although it fails to dominate the usual PTMLE (which may rest on a different choice of c). In view of the asymptotic theory developed in Section 3, for the

M-estimators, we are able to make use of the asymptotic multinormality property and thereby extend their conclusions to the case of the PTME, SME and the positive-rule SME. This suggests that under the specific choice of $\mathbf{W} = n^{-1}\mathbf{C}_{n11\cdot 2}$, the positive-rule SME should be used instead of the usual SME. However, for the PTME, the regularity conditions in (4.7) or (4.8) are not needed, and for small values of α and for p_1, p_2 not large, the PTME in (2.16) may turn out to be a better competitor. For (2.18), one may also define a positive-rule version, but that may not dominate (2.18) (in the light of their ADR) and the appeal for the positive-rule version is thus diminished.

Acknowledgments. Thanks are due to the Associate Editor and the referees for their valuable comments on the manuscript.

REFERENCES

BERGER, J. O., BOCK, M. E., BROWN, L. D., CASELLA, G. and GLESER, L. (1977). Minimax estimation of a normal mean vector for arbitrary quadratic loss and unknown covariance matrix. *Ann. Statist.* **5** 763–771.

JAMES, W. and STEIN, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Probab.* **1** 293–325. Univ. California Press.

JUDGE, G. G. and BOCK, M. E. (1978). *The Statistical Implications of Pretest and Stein-Rule Estimators in Economics*. North-Holland, Amsterdam.

JUREČKOVÁ, J. (1977). Asymptotic relations of M-estimators and R-estimators in linear models. *Ann. Statist.* **5** 664–672.

JUREČKOVÁ, J. and SEN, P. K. (1984). On adaptive scale-equivariant M-estimators in linear models. *Statist. Decisions (Suppl.)* **2** 31–46.

SALEH, A. K. M. E. and SEN, P. K. (1985). On shrinkage M-estimators of location parameters. *Commun. Statist. A—Theory Methods* **14** 2313–2329.

SALEH, A. K. M. E. and SEN, P. K. (1987). Relative performance of Stein-rule and preliminary test estimators in linear models: Least squares theory. *Commun. Statist. A—Theory Methods* **16** 461–476.

SCLOVE, S. L., MORRIS, C. and RADHAKRISHNAN, R. (1972). Non-optimality of preliminary test estimators for the mean of a multi-variate normal distribution. *Ann. Math. Statist.* **43** 1481–1490.

SEN, P. K. (1982). On M-tests in linear models. *Biometrika* **69** 245–248.

SEN, P. K. (1984). A James–Stein detour of U-statistics. *Commun. Statist. A—Theory Methods* **13** 2725–2747.

SEN, P. K. (1986). On the asymptotic distributional risks of shrinkage and preliminary test versions of maximum likelihood estimators. *Sankhyā Ser. A* **48** 355–372.

SEN, P. K. and SALEH, A. K. M. E. (1979). Nonparametric estimation of location parameter after a preliminary test on regression in the multivariate case. *J. Multivariate Anal.* **9** 322–331.

SEN, P. K. and SALEH, A. K. M. E. (1985). On some shrinkage estimators of multivariate location. *Ann. Statist.* **13** 272–281.

SINGER, J. M. and SEN, P. K. (1985). M-methods in multivariate linear models. *J. Multivariate Anal.* **17** 168–184.

STEIN, C. (1981). Estimation of the mean of a multivariate normal distribution. *Ann. Statist.* **9** 1135–1151.

YOHAI, V. J. and MARONNA, R. A. (1979). Asymptotic behavior of M-estimators for the linear model. *Ann. Statist.* **7** 258–268.

DEPARTMENT OF BIOSTATISTICS 201H
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514

DEPARTMENT OF MATHEMATICS AND STATISTICS
CARLETON UNIVERSITY
OTTAWA, ONTARIO K1S 5B6
CANADA