

Admissible, consistent multiple testing with applications including variable selection

Chuanwen Chen^{*}, Arthur Cohen[†] and Harold B. Sackrowitz[†]

*Department of Statistics,
Rutgers University, Piscataway, NJ 08854*
e-mail: tranwen@stat.rutgers.edu; artcohen@rci.rutgers.edu
sackrowi@rci.rutgers.edu

Abstract: For multivariate normal models and some exponential family models a multiple testing stepwise method is offered that is both admissible and consistent. The method is readily adaptable to selecting variables in linear regression models where it is akin to the forward selection method plus a screening stage plus a sign compatibility stage.

AMS 2000 subject classifications: Primary 62F03; secondary 62C15, 62J05.

Keywords and phrases: backward method, exponential family, forward method, linear regression, step-down procedures, step-up procedures, variable selection.

Received March 2009.

Contents

1	Introduction	633
2	Multivariate normal models	636
3	MRDSS: Multivariate normal model	637
4	Admissibility and consistency of MRDSS	639
5	Equivalence of MRD and a forward selection method	643
6	Exponential family models	644
A	Proof of main theorem	646
	References	649

1. Introduction

Multiple testing in general linear models is a long standing statistical practice. The subject has been given great impetus over the last two decades with applications arising in fields where the numbers of hypotheses to be tested can be extremely large (namely thousands). Such applications are needed in microarrays, astronomy, mutual fund evaluations, proteomics, disclosure risk,

^{*}This represents part of the Ph.D thesis of C. Chen

[†]Supported by NSF Grant DMS 0804547 and NSA GRANT H-98230-06-0076

cytometry, imaging and others. Traditional methods, usually classified as single step methods, are deemed too conservative, particularly when the number of hypotheses to be tested is very large. To compensate, stepwise procedures were introduced, which enable more rejection of hypotheses. See, for example, [Dudoit and Van Der Laan \(2008\)](#) and [Lehmann and Romano \(2005\)](#) for description and some properties of single step and stepwise procedures. Many stepwise procedures are based on P-values derived from the marginal distributions of relevant test statistics that test an individual hypothesis. Oftentimes, these and other multiple testing procedures (MTPs) are focused on controlling some type of error rate, namely, the familywise error rate (FWER), weak or strong, or the false discovery rate (FDR). Oftentimes the error rate controlling procedures seek to have good average power. While focusing on error rate control of the overall procedure, sometimes the properties of the ensuing tests of each individual hypothesis is not given enough attention. It is true that the resulting test for an individual hypothesis resulting from an MTP can be complicated. Nevertheless, examining the properties of the individual tests can be important. In fact, a decision theory approach to multiple testing, focusing on individual tests, can suggest procedures that do well when evaluated by the expected number of type I errors and expected numbers of type II errors. This represents another way to evaluate MTPs. See [Dudoit and Van Der Laan \(2008\)](#) where attention is given to type I and type II errors as well as to error rate controlling procedures.

In a series of papers, [Cohen and Sackrowitz \(2005a,b, 2007, 2008\)](#) and [Cohen, Kolassa, and Sackrowitz \(2007\)](#) have shown that many of the standard stepwise procedures under a wide variety of assumptions, often turn out to be inadmissible for a variety of risk functions that involve both expected type I and expected type II errors. For exponential family models the inadmissibility is often shown to follow because in testing an individual hypothesis, relevant acceptance sections are not intervals. This represents a disturbing practical shortcoming of many of the usual procedures.

In response to the inadmissibility property and the fact that many stepwise procedures are based on the marginal distributions of test statistics, even when they are statistically dependent, [Cohen, Sackrowitz, and Xu \(2009\)](#) recommend a new MTP method called maximum residual down (MRD). MRD takes correlation into account and is admissible (in many exponential family models) for a risk function that focuses on expected type I and type II errors.

For some models the MRD method is not consistent. See [Section 2](#) for a formal definition of consistency. Informally, consistency means that the probability of correctly testing each hypothesis tends to 1 as the sample size tends to infinity and as critical values tend to infinity at a certain rate.

In this paper we introduce a modification of MRD which offers a procedure that is admissible and consistent. The modification, discussed in [Section 3](#), adds a screening stage following MRD, and a sign stage following the screening stage. The screening stage, which is intuitively desirable, enables the overall procedure to be consistent. Its addition however can alter the admissibility property of MRD. The sign stage restores the admissibility and retains the consistency property. Call the new procedure MRDSS.

For the most part we assume a multivariate normal model with unknown mean vector and a covariance matrix that is known or known up to a scalar multiple. The tests of hypotheses concern the components of the mean vector. The results also apply, with minor modifications, to many, but not all, exponential family models. For example, the results apply to testing differences of binomial parameters, Poisson parameters, or scale parameters of exponential distributions. Furthermore the MRD, screening, and sign stages suggest an approach with the potential to yield an admissible and consistent MTP for more complicated exponential family models. One such model entails testing whether all local log odds ratios are zero in an $R \times C$ contingency table. This latter problem is treated in the companion paper by [Chen, Cohen, and Sackrowitz \(2009\)](#).

In that paper all three stages require the development of nontrivial analogues of the MRD, screen, and sign stages that enable an admissible, consistent MTP to ensue. Furthermore, as in the multivariate normal case and other straightforward exponential family cases, the resulting MTP has other desirable properties. These include computational feasibility and as justified by simulation results, they do considerably better than typical stepwise procedures in terms of type I and type II errors over large portions of the parameter space.

We will apply the MRDSS method to selection of variable models in regression. For such models, the MRD method is shown to be equivalent to some forward method of selecting variables. See [Miller \(2002\)](#) for a discussion of forward methods. Both MRD and the forward method are stepwise methods based on what we define as residuals. Forward methods and hence MRD are typically not consistent. See, for example, [An and Gu \(1985\)](#). We show that MRDSS, and thus a modification of a forward method, with a screen and sign stage added, is admissible and consistent under modest and typical conditions on the critical values and design matrix.

The results will be proved assuming a full rank multivariate normal model with unknown mean vector and known covariance matrix while testing components of the mean vector. However, extension to the following cases, sometimes with modifications are possible: covariance equal to an unknown variance times a known positive definite matrix; testing functions of natural parameters of exponential family densities; regression models with suitable design matrices.

Through the first five sections the focus will be on the multivariate normal model. Definitions and other preliminaries are given in [Section 2](#). MRDSS is described in [Section 3](#). [Section 4](#) is concerned with results on admissibility and consistency. A small simulation is also reported in [Section 4](#). The simulation shows the improvement of MRDSS relative to the FDR controlling step-up procedure over a large portion of the parameter space. In [Section 5](#) we demonstrate that MRD is equivalent to a forward method of variable selection in a linear regression model. Some remarks about the backward method are also made in [Section 5](#). [Section 6](#) contains extensions to some exponential family models. The proof of admissibility of MRDSS is in the Appendix.

2. Multivariate normal models

Let \mathbf{X}_α , $\alpha = 1, \dots, n$, be a random sample of $M \times 1$ vectors from a multivariate normal distribution whose mean vector is $\boldsymbol{\mu}$ and whose covariance matrix is Γ . Sometimes Γ is known and sometimes $\Gamma = \sigma^2 \Sigma$ where Σ is known and σ^2 is unknown. In the later case, $s^2 = \sum_{\alpha=1}^n (\mathbf{X}_\alpha - \bar{\mathbf{X}})' \Sigma^{-1} (\mathbf{X}_\alpha - \bar{\mathbf{X}}) / M(n-1)$ is an unbiased estimator of σ^2 , which is independent of $\bar{\mathbf{X}} = \sum_{\alpha=1}^n \mathbf{X}_\alpha / n$. For this model with Γ known and nonsingular the density of $\bar{\mathbf{X}}$ is

$$\begin{aligned} f_{\bar{\mathbf{X}}}(\bar{\mathbf{x}} | \boldsymbol{\mu}, \Gamma) &= (n^{1/2} / (2\pi)^{M/2} |\Gamma|^{1/2}) \exp \{-n/2(\bar{\mathbf{x}} - \boldsymbol{\mu})' \Gamma^{-1} (\bar{\mathbf{x}} - \boldsymbol{\mu})\} \\ &= \beta(\boldsymbol{\mu}, \Gamma) h(\bar{\mathbf{x}}) \exp n \bar{\mathbf{x}}' \Gamma^{-1} \boldsymbol{\mu} \end{aligned} \tag{2.1}$$

We wish to test $H_i : \mu_i = 0$ vs $K_i : \mu_i \neq 0, i = 1, 2, \dots, M$. Sometimes one sided alternatives $K_i : \mu_i > 0$, are of interest.

The general linear model has $\mathbf{Y} = A\boldsymbol{\beta} + \boldsymbol{\epsilon}$ where \mathbf{Y} is an $n \times 1$ vector which is multivariate normal with mean vector $A\boldsymbol{\beta}$ and covariance matrix $\sigma^2 I$. $A = (\mathbf{a}_1, \dots, \mathbf{a}_M)$ is an $n \times M$ fixed design matrix and $\boldsymbol{\beta}$ is an $M \times 1$ vector of parameters. If A has rank M , then $\hat{\boldsymbol{\beta}} = (A'A)^{-1} A' \mathbf{Y} \sim N(\boldsymbol{\beta}, \sigma^2 S^{-1})$ where $S^{-1} = (A'A)^{-1}$. An unbiased estimator of σ^2 is $s^2 = (\mathbf{y}'\mathbf{y} - \hat{\boldsymbol{\beta}}' A' \mathbf{y}) / (n - M)$, which is called mean squared error and which is independent of $\hat{\boldsymbol{\beta}}$. Thus this general linear model reduces to our multivariate normal model with $\Gamma = \sigma^2 S^{-1}$. We wish to test $H_i : \beta_i = 0$ vs $K_i : \beta_i \neq 0, i = 1, 2, \dots, M$.

For notational convenience we consider the model in (2.1) with $n = 1$ for all matters not concerning consistency. This is done without loss of generality. An MTP $\Phi(\mathbf{x}) = (\phi_1(\mathbf{x}), \dots, \phi_M(\mathbf{x}))'$ where $\phi_i(\mathbf{x})$ is a test function for testing H_i vs $K_i, i = 1, \dots, M$. That is, $\phi_i(\mathbf{x})$ is the probability of rejecting H_i for the sample point \mathbf{x} . A type I error is made if H_i is rejected when $\mu_i = 0$ and a type II error is made if H_i is not rejected and $\mu_i \neq 0$. The expected type I error is $E_{\boldsymbol{\mu}} \phi_i(\mathbf{x})$ when $\mu_i = 0$. The expected type II error is $E_{\boldsymbol{\mu}} (1 - \phi_i(\mathbf{x}))$ when $\mu_i \neq 0$. The risk function for $\phi_i(\mathbf{x})$ is

$$R_i(\phi_i(\mathbf{x}, \boldsymbol{\mu})) = (1 - \nu_i) E_{\boldsymbol{\mu}} \phi_i(\mathbf{x}) + \nu_i E_{\boldsymbol{\mu}} (1 - \phi_i(\mathbf{x})) \tag{2.2}$$

where $\nu_i = 1$ if $\mu_i \neq 0$ and $\nu_i = 0$ if $\mu_i = 0$.

The risk function for the procedure $\Phi(\mathbf{x})$ is defined as the vector risk function

$$(R_1(\phi_1, \boldsymbol{\mu}), R_2(\phi_2, \boldsymbol{\mu}), \dots, R_M(\phi_M, \boldsymbol{\mu})) \tag{2.3}$$

An MTP $\Phi(\mathbf{x})$ is inadmissible if there exists another MTP, say $\Phi^*(\mathbf{x})$ such that for all $i = 1, \dots, M, R_i(\phi_i^*, \boldsymbol{\mu}) \leq R_i(\phi_i, \boldsymbol{\mu})$ with strict inequality holding for some i and some $\boldsymbol{\mu}$.

Now we return to the model in (2.1) for the case of $n > 1$. Consistency is defined in An and Gu (1985) and in Bunea, Wegkamp, and Auguste (2006). Let $I = \{i : \mu_i \neq 0\}$ and $\hat{I} = \{i : \Phi_i(\bar{\mathbf{x}}) \neq 0, i = 1, \dots, M\}$. We say Φ is consistent when $P\{\hat{I} = I\}$ tends to 1 when n tends to infinity.

3. MRDSS: Multivariate normal model

We start with the multivariate normal model with $\Gamma = \Sigma = (\sigma_{ij})$ known and nonsingular. Without loss of generality for now we let $n = 1$. (For $n > 1$, $\bar{\mathbf{X}}$ would replace \mathbf{X} where \mathbf{X} is given in the next line.)

Let $\mathbf{X} = (X_1, \dots, X_M)'$, $\mathbf{X}^{(i_1, \dots, i_r)}$ be a subvector of \mathbf{X} consisting of all components except X_{i_1}, \dots, X_{i_r} . $\Sigma_{(i_1, \dots, i_r)}$ is the $(M - r) \times (M - r)$ covariance matrix of $\mathbf{X}^{(i_1, \dots, i_r)}$, for any $j \neq i_1, \dots, i_r$, $\sigma_{(j)}^{(i_1, \dots, i_r)}$ is the $(M - r - 1) \times 1$ vector of covariances between X_j and all variables except X_{i_1}, \dots, X_{i_r} and X_j .

$$\sigma_{(j \cdot i_1, \dots, i_r)}^2 = \sigma_{jj} - \sigma_j^{(i_1, \dots, i_r)'} \Sigma_{(i_1, \dots, i_r, j)}^{-1} \sigma_j^{(i_1, \dots, i_r)} \tag{3.1}$$

is the conditional variance of X_j , given all variables except $X_{i_1}, \dots, X_{i_r}, X_j$.

Now define the j th normalized residual at step m to be

$$U_{mj}(\mathbf{X}^{(i_1, \dots, i_{m-1})}) = \frac{X_j - \sigma_j^{(i_1, \dots, i_{m-1})'} \Sigma_{(i_1, \dots, i_{m-1}, j)}^{-1} \mathbf{X}^{(i_1, \dots, i_{m-1}, j)}}{[\sigma_{(j \cdot i_1, \dots, i_{m-1})}^2]^{1/2}} \tag{3.2}$$

$m = 1, \dots, M$. At step m , we let \mathbf{U}_m be the $(M - m + 1) \times 1$ vector of U_{mj} 's. We note that if D_m is the diagonal matrix of the terms in (3.1), then

$$\mathbf{U}_m = D_m^{1/2} \Sigma_{(i_1, \dots, i_{m-1})}^{-1} \mathbf{X}^{(i_1, \dots, i_{m-1})}$$

Let $C_1 > C_2 > \dots > C_M > 0$ be a given set of constants. Then the MTP, called maximum residual down (MRD), is determined as follows:

At step 1, consider $U_{1j}(\mathbf{X})$, $j \in 1, \dots, M$. Let $j_1 = j_1(\mathbf{X})$ be such that $|U_{1j_1}(\mathbf{X})| = \max_j |U_{1j}(\mathbf{X})|$. If $|U_{1j_1}(\mathbf{X})| < C_1$, stop and accept all H_i . Otherwise reject H_{j_1} and continue to step 2.

At step 2, consider the $(M - 1)$ functions $U_{2j}(\mathbf{X}^{(j_1)})$, $j \in \{1, \dots, M\} \setminus \{j_1\}$. Let $j_2 = j_2(\mathbf{X}^{(j_1)})$ be such that $|U_{2j_2}(\mathbf{X}^{(j_1)})| = \max \{|U_{2j}(\mathbf{X}^{(j_1)})| : j \in 1, \dots, M \setminus \{j_1\}\}$. If $|U_{2j_2}| < C_2$ stop and accept all remaining null hypotheses. Otherwise reject H_{j_2} and continue to step 3.

In general, at step m , $m = 1, \dots, M$, consider $M - m + 1$ functions $U_{mj}(\mathbf{X}^{(j_1, \dots, j_{m-1})})$, $j \in \{1, \dots, M\} \setminus \{j_1, \dots, j_{m-1}\}$. Let $j_m = j_m(\mathbf{X}^{(j_1, \dots, j_{m-1})})$ be such that $|U_{mj_m}| = \max \{|U_{mj}(\mathbf{X}^{(j_1, \dots, j_{m-1})})| : j \in 1, \dots, M \setminus \{j_1, \dots, j_{m-1}\}\}$. If $|U_{mj_m}| < C_m$, stop and accept all remaining null hypotheses. Otherwise reject H_{j_m} and continue to step $m + 1$ (unless $m = M$, in which case, stop).

If and when H_i is rejected, say at step r , record the sign of U_{ri} .

To add a screening stage to MRD, let $C_U > C_L > 0$ be two additional constants. Typically $C_L \leq C_M < C_U$. After MRD is done, each hypothesis is accepted or rejected.

Let H_{j_1}, \dots, H_{j_p} be those hypotheses that are rejected. Should any

$$|X_{j_i}|/\sqrt{\sigma_{j_i j_i}} < C_L, \tag{3.3}$$

$i = 1, \dots, p$, then reverse the reject decision to an accept decision.

For those hypotheses that MRD accepted, say, $H_{j_{p+1}}, \dots, H_{j_M}$, reject any for which

$$|X_{j_i}|/\sqrt{\sigma_{j_i j_i}} > C_U, \tag{3.4}$$

$i = p + 1, \dots, M$.

To add a sign stage first consider only those hypotheses rejected by MRD, i.e., H_{j_1}, \dots, H_{j_p} , and such that $C_L < |X_{j_i}|/\sqrt{\sigma_{j_i j_i}} < C_U, i = 1, \dots, p$. Then switch those to accept whenever the sign of X_{j_i} is different from the sign of $U_{m^* j_i}$ where m^* is the step when H_{j_i} was rejected by MRD. The MRD procedure followed by a screening stage and a sign stage is called MRDSS.

We will illustrate MRDSS in the following example:

Example. Let $\mathbf{X} = (X_1, X_2, X_3, X_4)'$ be such that $\mathbf{X} \sim N(\boldsymbol{\mu}, \Sigma)$ where

$$\Sigma = \begin{pmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{pmatrix}.$$

We wish to test $H_i : \mu_i = 0$ vs $K_i : \mu_i \neq 0, i = 1, \dots, 4$.

For the purpose of demonstrating the application of MRDSS we will let $C_1 = 2.3, C_2 = 2.0, C_3 = 1.7, C_4 = 1.4, C_L = .6$, and $C_U = 3.6$. Now suppose we observe $X_1 = 1, X_2 = 5, X_3 = 4.7$ and $X_4 = 5.3$.

First we apply MRD. At step 1 the residuals are:

$$\begin{aligned} U_{11} &= \left[X_1 - (1, 1, 1) \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}^{-1} (X_2, X_3, X_4)' \right] / \\ &\quad \left[2 - (1, 1, 1) \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} \right]^{1/2} \\ &= [X_1 - (X_2 + X_3 + X_4)/4](\frac{4}{5})^{1/2} = -2.46 \\ U_{12} &= [X_2 - (X_1 + X_3 + X_4)/4](\frac{4}{5})^{1/2} = 2.01 \\ U_{13} &= [X_3 - (X_1 + X_2 + X_4)/4](\frac{4}{5})^{1/2} = 1.68 \\ U_{14} &= [X_4 - (X_1 + X_2 + X_3)/4](\frac{4}{5})^{1/2} = 2.35 \end{aligned} \tag{3.5}$$

Since $|U_{11}|$ is the largest and $|U_{11}| > 2.3 = C_1$, MRD rejects H_1 and goes to the

next step. Now with X_1 eliminated,

$$\begin{aligned}
 U_{22} &= \left[X_2 - (1, 1) \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^{-1} (X_2, X_4)' \right] / \left[2 - (1, 1) \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}^{-1} \begin{pmatrix} 1 \\ 1 \end{pmatrix} \right]^{1/2} \\
 &= [X_2 - (X_3 + X_4)/3] / (\frac{3}{4})^{1/2} = 1.44 \\
 U_{23} &= [X_3 - (X_2 + X_4)/3] / (\frac{3}{4})^{1/2} = 1.10 \\
 U_{24} &= [X_4 - (X_2 + X_3)/3] / (\frac{3}{4})^{1/2} = 1.79
 \end{aligned}$$

Since $|U_{22}|$, $|U_{23}|$ and $|U_{24}|$ are all less than $C_2 = 2.0$, MRD stops and accepts H_2, H_3 , and H_4 .

Next we do the screen stage by comparing each $|X_i/\sqrt{2}|$ to $C_L = .6$ and $C_U = 3.6$. We see that $C_L < |X_i/\sqrt{2}| < C_U$ for $i = 1, 2, 3$ but $|X_4/\sqrt{2}| = 3.75 > 3.6 = C_U$. Thus the MRD decision to accept H_4 is reversed to reject H_4 .

Lastly we perform the sign stage. This applies only to hypotheses H_i that were rejected by MRD and for which $C_L < |X_i/\sqrt{2}| < C_U$. In this example only H_1 falls into this category. Upon examination we see that H_1 was rejected at the first step of MRD stage with $U_{11} = -2.46$ while $X_1/\sqrt{2} = .707$. Since these signs are opposite we reverse the MRD decision to reject H_1 .

In conclusion MRDSS will accept H_1, H_2, H_3 and reject H_4 .

For the multivariate normal model where $\Gamma = \sigma^2 \Sigma$, σ^2 is unknown and $n > 1$, a modification of U_{mj} is called for. The numerator would use the component of $\bar{\mathbf{X}}$ instead of \mathbf{X} , and the denominator of U_{mj} would be multiplied by the square root of an estimator of σ^2 divided by \sqrt{n} . The estimator of σ^2 could be s^2 defined in Section 2 or it could be $V = \sum_{\alpha=1}^n \mathbf{x}_\alpha \Sigma^{-1} \mathbf{x}_\alpha / n$, which is an unbiased estimator of σ^2 when $\boldsymbol{\mu} = \mathbf{0}$. Similar modifications would also be made in the screen stage determined by (3.3) and (3.4). The advantage of using $V^{1/2}$ is that both admissibility and consistency of the overall procedure are easily established. The advantage of using s is that the overall procedure is both location and scale invariant.

4. Admissibility and consistency of MRDSS

Let $\Phi_{MRDSS}(\mathbf{X})$ be the MRDSS MTP for the multivariate normal model with σ^2 known. When studying admissibility, without loss of generality, $n = 1$.

Theorem 4.1. *The MTP $\Phi_{MRDSS}(\mathbf{X})$ is admissible for the vector risk function.*

Proof. See Appendix. □

Note that MRD has been proven to be admissible in [Cohen, Sackrowitz, and Xu \(2009\)](#).

Let Σ be positive definite and let \mathbf{g}_i be the i -th column of Σ . Then define the class of MTPs with:

Property N: Suppose $\Psi(\mathbf{X})$ is an MTP with individual test functions $\Psi_i(\mathbf{X})$. Suppose $\Psi_i(\mathbf{X})$, for every i , is such that the acceptance section along the line $\mathbf{x} + r\mathbf{g}_i$ is an interval. Then the MTP is said to have *Property N*. It is noted in the Appendix that *Property N* is necessary and sufficient for an MTP to be admissible.

Next we address the issue of consistency of MRDSS. We first note that MRDSS is consistent if and only if for each i , the individual test of hypothesis $H_i : \mu_i = 0$ vs $K_i : \mu_i \neq 0$ is consistent.

In discussing consistency the sample size $n > 1$ and the random vector \mathbf{X} is replaced by $\bar{\mathbf{X}}$. The screen step of MRDSS becomes accept H_i if

$$\sqrt{n}|\bar{X}_i|/\sigma_{ii}^{1/2} < C_L(n) ,$$

and reject if

$$\sqrt{n}|\bar{X}_i|/\sigma_{ii}^{1/2} > C_U(n) ,$$

where now the critical values change with n .

Theorem 4.2. Assume $C_L(n) < C_U(n)$, assume $C_L(n) \rightarrow \infty$ as $n \rightarrow \infty$ in such a way that $C_L(n) = o(\sqrt{n})$. Also $C_U(n) \rightarrow \infty$ as $n \rightarrow \infty$ and $C_U(n) = o(\sqrt{n})$. Then Φ_{MRDSS} is consistent.

Proof. Note first that when $\mu_i = 0$

$$\begin{aligned} P(\text{MRDSS accepts } H_i) &\geq P\{\sqrt{n}|\bar{X}_i|/\sigma_{ii}^{1/2} \leq C_L\} \\ &= P\left\{-C_L - \sqrt{n}\mu_i/\sigma_{ii}^{1/2} \leq \sqrt{n}(\bar{X}_i - \mu_i)/\sigma_{ii}^{1/2} \leq C_L - \sqrt{n}\mu_i/\sigma_{ii}^{1/2}\right\} \end{aligned} \quad (4.1)$$

Since $C_L \rightarrow \infty$ and $\mu_i = 0$, (4.1) $\rightarrow 1$ as n tends to ∞ .

Next note that when $\mu_i \neq 0$,

$$\begin{aligned} P_{\boldsymbol{\mu}}(\text{MRDSS rejects } H_i) &\geq P_{\boldsymbol{\mu}}(\sqrt{n}|\bar{X}_i|/\sigma_{ii}^{1/2} \geq C_U) \\ &= 1 - P_{\boldsymbol{\mu}}\left\{-C_U - \sqrt{n}\mu_i/\sigma_{ii}^{1/2} \leq \sqrt{n}(\bar{X}_i - \mu_i)/\sigma_{ii}^{1/2} \leq C_U - \sqrt{n}\mu_i/\sigma_{ii}^{1/2}\right\} \\ &= 1 - \left[\Phi(C_U - \sqrt{n}\mu_i/\sigma_{ii}^{1/2}) - \Phi(-C_U - \sqrt{n}\mu_i/\sigma_{ii}^{1/2})\right] \end{aligned} \quad (4.2)$$

where Φ is the CDF of a standard normal distribution. Since $C_U = o(\sqrt{n})$ and $\mu_i \neq 0$, (4.2) $\rightarrow 1$. □

Remark 4.3. For the model where \mathbf{X}_α , $\alpha = 1, \dots, n$ is multivariate normal with mean $\boldsymbol{\mu}$ and covariance matrix $\sigma^2\Sigma$ with σ^2 unknown the screen step of MRDSS replaces $\sqrt{n}|\bar{X}_i|/\sigma_{ii}^{1/2}$ by $\sqrt{n}|\bar{X}_i|/(s\sigma_{ii}^{1/2})$. Recall s^2 is defined in Section 2.

The proof of Theorem 4.2, with appropriate modification would establish consistency of MRDSS in this case. That is, the statistic for the screen step

would be $t_i = \sqrt{n}|\bar{X}_i|/s\sigma_{ii}^{1/2}$. In (4.1) we would replace $\sqrt{n}(\bar{X}_i - \mu_i)/\sigma_{ii}^{1/2}$ by $(t_i - Et_i)/\sqrt{\text{var}(t_i)}$. Using properties of the noncentral t distribution as given in Johnson and Kotz (1970), and analogues of (4.1) and (4.2) would establish consistency. Consistency also follows if $\sqrt{n}|\bar{X}_i|/V^{1/2}\sigma_{ii}^{1/2}$ is used in the screen stage.

Remark 4.4. In the next section we will show that MRD is the same as a forward method in selection of variables in a multiple linear regression model. For this application the statistic used for the screening step is $T_i = |\hat{\beta}_i|/s\lambda_{ii}^{1/2}$ where λ_{ii} is the i -th diagonal element of S^{-1} . Then for consistency we require that $\max \lambda_{ii}(n) \rightarrow 0$ as $n \rightarrow \infty$ in such a way that $\lambda_{ii} = O(1/n)$. A different requirement for λ_{ii} can be made if C_L and C_U are chosen to tend to ∞ at a rate different from the one given in Theorem 4.2.

At this point we note that MRD alone is oftentimes not consistent. For a simple example we need only consider the model of the illustrative example of Section 3. We need only begin by thinking of the X_i as sample means based on n independent observations and change the covariance matrix to $\frac{1}{n}\Sigma$. Next we study the distribution of $(U_{11}, U_{12}, U_{13}, U_{14})$ when the parameter point is $\boldsymbol{\mu} = (0, r, r, r)'$. From (3.5) it is clear the covariance matrix is symmetric and the mean vector is $\sqrt{4/5}(-3r/4, r/2, r/2, r/2)$. Therefore, regardless of n , $P_{\boldsymbol{\mu}}(|U_{11}| = \max(|U_{11}|, |U_{12}|, |U_{13}|, |U_{14}|)) > 1/4$. Thus, with probability $> 1/4$, the decision at the first step of MRD will be based on $|U_{11}|$. If $|U_{11}| \geq C_1$, H_1 is rejected and a type I error is made. If $|U_{11}| < C_1$, the process stops and all hypotheses, including H_2, H_3 and H_4 are accepted. In this case three type II errors are made. Here for any n , the probability that some error will be made is $> 1/4$.

We conclude this section with the results of a simulation. As mentioned in Cohen, Sackrowitz, and Xu (2009), the MRD procedure can be viewed as a family of procedures that is parameterized by a set of constants C_1, \dots, C_M . MRDSS has two more parameters: C_L and C_U .

The simulation concerns a multivariate normal model with a mean vector of order 1000×1 and a covariance matrix which is intraclass with equal diagonal elements and equal off-diagonal elements. The parameter points are such that at most 25 percent are non-zero. This is not atypical for many practical models when the number of hypotheses to be tested is very large. We perform MRDSS and an FDR controlling step-up procedure at FDR level .05. Type I and type II errors are compared as well as total number of mistakes. For MRD and MRDSS, we compute the percent decrease of total number of mistakes relative to the step-up procedure. For all simulations, we used 2000 iterations.

The simulation results are listed in Table 4.1. The step-up procedure is the Benjamini and Hochberg (1995) FDR controlling procedure where $\text{FDR} = 0.05$. For MRD and MRDSS we use the same set of constants (C_1, \dots, C_M) as in Cohen, Sackrowitz, and Xu (2009), i.e., for $\alpha = 0.05$, $M = 1000$, $C_1 = \Phi^{-1}(1 - \alpha/2M)$, $C_i = 0.71\Phi^{-1}(1 - \alpha/2(M - i + 1))$, $i = 2, \dots, M$. We set $C_L = \Phi^{-1}(1 - \alpha/2)$, $C_U = C_1$. For MRD and MRDSS, there is (usually substantial) improvement in both the expected number of type I and II errors.

TABLE 4.1
Comparison of MRDSS and SU for treatments vs. control

Number of means equal to	Expected # of type I errors		Expected # of type II errors		FDR		Total errors		% improvement			
	-4	-2	2	4	MRDSS	SU	MRDSS	SU	MRDSS	SU		
1000	0	0	0	0	0.2	3.9	0.070	0.026	0.2	3.9	93	94
920	0	80	0	0	2.6	4.4	0.122	0.032	54.4	77.5	38	30
920	80	0	0	0	2.6	4.6	0.031	0.036	4.4	17.7	76	75
840	0	80	80	0	2.3	4.2	0.037	0.030	101.8	150.1	43	32
840	0	0	160	0	2.7	5.4	0.078	0.031	110.1	144.0	30	24
840	80	0	80	0	2.4	6.3	0.022	0.037	55.7	79.2	39	30
840	0	0	80	80	2.5	4.3	0.023	0.031	55.9	79.6	39	30
840	80	0	0	80	2.4	7.3	0.015	0.037	5.8	21.8	81	73
840	0	0	0	160	2.5	7.5	0.015	0.039	5.8	24.0	82	76
760	0	80	160	0	2.3	6.3	0.026	0.033	154.4	218.0	40	29
760	0	0	240	0	3.0	7.6	0.077	0.032	176.0	207.8	19	15
760	80	0	160	0	2.5	7.3	0.020	0.033	113.2	137.9	26	18
760	0	0	160	80	2.6	5.7	0.022	0.027	114.1	141.4	28	19
760	80	0	80	80	2.3	7.6	0.012	0.034	57.5	79.4	39	28
760	0	0	80	160	2.3	6.6	0.013	0.033	56.9	80.6	40	29
760	80	0	0	160	2.3	9.6	0.010	0.035	7.4	26.7	84	72
760	0	0	0	240	2.3	9.3	0.010	0.037	7.9	29.3	86	73

5. Equivalence of MRD and a forward selection method

For this section we assume the general linear model described in Section 2 with σ^2 known. We test $H_i : \beta_i = 0$ vs $K_i : \beta_i \neq 0$, $i = 1, \dots, M$. Let $C_1 > C_2 > \dots > C_M > 0$ be a given set of constants.

Recall $A = (\mathbf{a}_1, \dots, \mathbf{a}_n)$ and let A be partitioned as (A_1, A_2) where A_1 is $n \times (m - 1)$ and A_2 is $n \times (M - m + 1)$. Let $V = R(A)$ be the linear space spanned by the columns of A and $V_1 = R(A_1)$ be the linear space spanned by the columns of A_1 . Furthermore, let P_V be the projection matrix for space V and let P_1 be the projection matrix for space V_1 . Let $V_2 = V \cap V_1^\perp = R(A_{2.1})$, where $A_{2.1} = A_2 - P_1 A_2 = (\mathbf{a}_{2.1m}, \dots, \mathbf{a}_{2.1M})$. Finally let P_2 denote the projection matrix for V_2 so that $P_V = P_1 + P_2$.

Forward selection methods are studied in Miller (2002) and many other texts studying linear regression. They are methods that select fixed variables to enter a linear regression model in steps. In the first step one considers the random variables $(\mathbf{a}'_i \mathbf{y}) / \sigma \|\mathbf{a}_i\|$, $i = 1, \dots, M$. The forward method selects the fixed variable \mathbf{a}_{i_1} provided that

$$U_{1i_1} = \max_{1 \leq i \leq M} (|\mathbf{a}'_i \mathbf{y}| / \sigma \|\mathbf{a}_i\|) > C_1. \tag{5.1}$$

If $U_{1i_1} \leq C_1$, no variables are chosen and the procedure stops. Now assume $(m - 1)$ variables have been selected. Without loss of generality say they are $(\mathbf{a}_1, \dots, \mathbf{a}_{m-1}) = A_1$. Then at stage m , the method selects \mathbf{a}_{i_m} provided

$$U_{mi_m} = \max_{m \leq j \leq M} (|\mathbf{a}'_{2.1j} \mathbf{y}| / \sigma \|\mathbf{a}_{2.1j}\|) > C_m. \tag{5.2}$$

If $U_{mi_m} \leq C_m$, the procedure stops. Thus at step m , the forward method is tantamount to rejecting $H_{i_m} : \beta_{i_m} = 0$ by virtue of (5.2).

Note that the vector of components $\mathbf{a}'_{2.1j} \mathbf{y}$ in (5.2) can be written as $A'_{2.1} \mathbf{y}$.

Now recall MRD applied to $\hat{\beta} \sim N(\beta, \sigma^2 S^{-1})$. At step one the numerator of the residual is proportional to $S\hat{\beta} = A'\mathbf{y}$, which is the vector of components in the expression in the numerator in (5.1). At step m , assuming $H_i : \beta_i = 0$, $i = 1, \dots, m - 1$ have been rejected, MRD deletes the first $m - 1$ rows and first $m - 1$ columns of S^{-1} and considers $\hat{\beta}_2 \sim N(\beta_2, \Lambda_{22})$ where $\beta_2 = (\beta_m, \dots, \beta_M)'$ and $S^{-1} = \begin{pmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{pmatrix}$, with Λ_{11} $(m - 1) \times (m - 1)$ and Λ_{22} $(M - m + 1) \times (M - m + 1)$.

The residuals \mathbf{U}_m at this step are proportional to $\Lambda_{22}^{-1} \hat{\beta}_2$. That is,

$$\mathbf{U}_m \propto \Lambda_{22}^{-1} \hat{\beta}_2. \tag{5.3}$$

However, note

$$A_1 \hat{\beta}_1 + A_2 \hat{\beta}_2 = A \hat{\beta} = P_V \mathbf{y}. \tag{5.4}$$

Multiply both sides of (5.4) by P_2 to get

$$A_{2.1} \hat{\beta}_2 = P_2 P_V \mathbf{y} = P_2 \mathbf{y}. \tag{5.5}$$

Since $\Lambda_{22}^{-1} = A'_2 A_2 - A'_2 P_1 A_2 = A'_{2.1} A_{2.1}$, we find using (5.3) and (5.5) that

$$\mathbf{U}_m \propto A'_{2.1} P_2 \mathbf{y} = A'_{2.1} \mathbf{y}. \tag{5.6}$$

Note that the components of (5.6) are $(\mathbf{a}'_{2.1}\mathbf{y})$, the unnormalized terms in the expression (5.2). Thus we have proved

Theorem 5.1. *For the assumption of the general linear regression model, the forward selection method is equivalent to MRD.*

Remark 5.2. Whereas the forward method and MRD are oftentimes not consistent procedures, MRDSS and thus the forward method plus screening, plus sign is consistent. It is interesting to note that Christensen (1987) p.288, claims that the precise screening step we recommend is an improvement on the forward method.

The backward method starts by assuming a linear regression model with M parameters. At step 1, one can decide that $\beta_{i_1} = 0$ if $T_{i_1} = \min_{1 \leq i \leq M} T_i(\hat{\boldsymbol{\beta}}) = \min_{1 \leq i \leq M} (|\hat{\beta}_i|/\sigma_{\hat{\beta}_i}) \leq C_m$. Otherwise the procedure stops and no variables are eliminated.

At stage m , assuming without loss of generality that variables $\mathbf{a}_1, \dots, \mathbf{a}_{m-1}$ have been eliminated, one considers residuals of the form

$$(\hat{\beta}_j - \boldsymbol{\lambda}_j^{(1, \dots, m-1)} \Lambda_{11(m-1)}^{-1} (\hat{\beta}_1, \dots, \hat{\beta}_{m-1})') / \lambda_{j \cdot (1, \dots, m-1)}^{1/2}. \quad (5.7)$$

where $\boldsymbol{\lambda}_j^{(1, \dots, m-1)}$ are the elements in the j^{th} row of S^{-1} that represent the covariance of $\hat{\beta}_j$ and $(\hat{\beta}_1, \dots, \hat{\beta}_{m-1})'$; $\Lambda_{11(m-1)}^{-1}$ represents the covariance matrix of $(\hat{\beta}_1, \dots, \hat{\beta}_{m-1})'$ and $\lambda_{j \cdot (1, \dots, m-1)}$ is the conditional variance of $\hat{\beta}_j$ given $\hat{\beta}_1, \dots, \hat{\beta}_{m-1}$.

If $\min_{m \leq j \leq M}$ of the terms in (5.7) $\leq C_{M-m+1}$, and the index is i_m then decide $\beta_{i_m} = 0$ and continue. Otherwise stop.

Remark 5.3. The backward method is consistent under mild conditions (see An and Gu, 1985). Nevertheless the backward method is akin in some sense to the step-up method of multiple testing. They are identical at the first step. As previously mentioned step-up methods are often inadmissible in normal and exponential family models and it can be shown that the backward method would sometimes be inadmissible for a risk function whose components are expected number of type I errors and expected number of type II errors.

6. Exponential family models

There are many models involving distributions that belong to the exponential family. For example, hypothesis testing problems involving differences of binomial parameters, Poisson parameters, and scale parameters of exponential distributions. Such exponential family models lend themselves to a somewhat straightforward extension of the MRDSS method and result in admissible and

consistent MTPs. In this section we describe the extension. In other exponential family models it can be non-trivial to find (if such exist) functions with the properties needed to play the roles of the analogues of MRD, screen and sign stage. One such case is an $R \times C$ contingency table where the hypotheses of interest concern all local log odds ratios. Such a case will be treated in the companion paper by [Chen, Cohen, and Sackrowitz \(2009\)](#).

For exponential families, where the hypotheses of interest may be differences or contrasts among natural parameters, we describe MRDSS.

Let $\mathbf{Z}_\alpha = (Z_{1\alpha}, \dots, Z_{(m+1)\alpha})'$, $\alpha = 1, \dots, n$ be independent $(M + 1) \times 1$ random vectors representing independent one dimensional exponential family distributions with natural parameter ω_i , $i = 1, \dots, M + 1$. Thus the joint density of the sufficient statistic $\bar{\mathbf{Z}} = \sum_{\alpha=1}^n \mathbf{Z}_\alpha/n$ is

$$f_{\bar{\mathbf{Z}}}(\bar{\mathbf{z}} | \boldsymbol{\omega}) = \beta(\boldsymbol{\omega})h(\bar{\mathbf{z}})e^{n\bar{\mathbf{z}}'\boldsymbol{\omega}}$$

where $\boldsymbol{\omega} = (\omega_1, \dots, \omega_{(M+1)})'$. Let G be a full rank $M \times (M + 1)$ matrix of contrasts, let $B = \begin{pmatrix} G \\ F \end{pmatrix}$, where the row F is a $1 \times (M + 1)$ vector of 1. Now let $\boldsymbol{\theta} = B\boldsymbol{\omega}$ and note

$$\bar{\mathbf{z}}'\boldsymbol{\omega} = \bar{\mathbf{z}}'(B'B^{-1})B^{-1}\boldsymbol{\theta} = \bar{\mathbf{z}}'B'(BB')^{-1}\boldsymbol{\theta}$$

If we let $\mathbf{X} = nB\bar{\mathbf{Z}}$, we find the density of \mathbf{X} is

$$f_{\mathbf{X}}(\mathbf{x} | \boldsymbol{\theta}) = \beta^*(\boldsymbol{\theta})h^*(\mathbf{x})e^{\mathbf{x}'\Sigma^{-1}\boldsymbol{\theta}} \tag{6.1}$$

where $\Sigma = BB'$. Further note that

$$\Sigma = \begin{pmatrix} GG' & 0 \\ 0 & M + 1 \end{pmatrix} \quad \text{and} \quad \Sigma^{-1} = \begin{pmatrix} (GG')^{-1} & 0 \\ 0 & 1/(M + 1) \end{pmatrix}.$$

We wish to test $H_i : \theta_i = 0$ vs $K_i : \theta_i \neq 0$, $i = 1, \dots, M$. Note that the first M components of \mathbf{X} are contrasts among the components of $\bar{\mathbf{Z}}$ while $X_{M+1} = \sum_{i=1}^{M+1} \bar{Z}_i$. Also note that the density in (6.1) is similar to the density (2.1). The MRD stage for this model will entail residuals W_{mj} defined as follows: The numerator of W_{mj} , the analogue of U_{mj} of (3.2), is defined exactly as in (3.2). The denominator is defined as an estimator of the standard deviation of the numerator assuming all ω_i are equal. The estimator will be a function of $X_{M+1}/(M + 1)$ which is the $(M + 1)$ st component of $\mathbf{x}'\Sigma^{-1}$. The term $X_{M+1}/(M + 1)$ is the average of the \bar{Z}_i , $i = 1, \dots, M$. Such a definition of W_{mj} will ensure that MRD will be admissible in this case. The screen stage will be determined by statistics whose numerators are the absolute values of $|X_i|$, $i = 1, \dots, M$ and whose denominator is also a function of $X_{M+1}/(M + 1)$. Such statistics along with a sign stage based on the signs of W_{mj}^* and X_{j_i} will ensure that the resulting procedure will be admissible and consistent under conditions similar to those assumed in the normal case.

Appendix A: Proof of main theorem

Consider the setup where \mathbf{X} , defined in Section 2, has density (2.1) with $n = 1$ and we wish to test $H_i : \theta_i = 0$ vs $K_i : \theta_i \neq 0$. Let $\psi_i(\mathbf{X})$ be a test function for H_i vs K_i . Let $\mathbf{W} = \Sigma^{-1}\mathbf{X}$. A lemma concerning admissibility of any single test, say $\psi_i(\mathbf{X})$, is as follows:

Lemma A.1. *A necessary and sufficient condition for $\psi_i(\mathbf{X}) = \psi_i^*(\mathbf{W})$ to be admissible is that for almost every fixed $\{W_j, j = 1, \dots, M; j \neq i\}$, the acceptance section of the test is convex in W_i .*

Proof. See Matthes and Truax (1967) □

Suppose, without loss of generality, we study the admissibility of $\psi_1(\mathbf{x})$. According to Lemma A.1 we will then regard W_2, \dots, W_M to be fixed while W_1 varies. Note that one way to study $\psi_1^*(\mathbf{W})$ is to study $\psi_1(\mathbf{x} + r\mathbf{g})$, where \mathbf{g} is the first column of Σ , as r varies. This is true since \mathbf{W} is a function of \mathbf{x} , and so \mathbf{W} evaluated at $\mathbf{x} + r\mathbf{g}$ is equal to $\Sigma^{-1}(\mathbf{x} + r\mathbf{g}) = \mathbf{W} + (r, 0, \dots, 0)' = (W_1 + r, W_2, \dots, W_M)'$.

Note that the components of \mathbf{W} are constants times the numerators of the components of \mathbf{U}_1 defined in Section 3. Note further, assuming H_1 has not yet been rejected, that $|U_{m1}|$ as a function of r is decreasing and then increasing. Here the function $U_{m1} \propto W_1 + r$ is strictly increasing in r . All this implies that if $|U_{m1}|$ is decreasing in r then U_{m1} is negative, while when $|U_{m1}|$ is increasing in r then U_{m1} is positive.

Next we define

Property L: A test is said to have *Property L* if there exists three points $\mathbf{x}, \mathbf{x}^* = \mathbf{x} + r_1\mathbf{g}, \mathbf{x}^{**} = \mathbf{x} + r_2\mathbf{g}$, with $0 < r_1 < r_2$, such that \mathbf{x} and \mathbf{x}^{**} are accept points and \mathbf{x}^* is a reject point when testing H_1 vs K_1 .

Note that if a procedure has *Property L* (for a set of positive measure) then it is inadmissible by virtue of Lemma A.1.

Proof of Theorem 4.1: The proof is by contrapositive. That is, we will assume that Ψ_1 of the MRDSS has *Property L* and then show that this is impossible. That is, we assume that for ψ_1 there exists 3 sample points \mathbf{x}, \mathbf{x}^* and \mathbf{x}^{**} for which we accept, reject, accept (ARA) in that order with $\mathbf{x}^* = \mathbf{x} + r_1\mathbf{g}, \mathbf{x}^{**} = \mathbf{x} + r_2\mathbf{g}$, $0 < r_1 < r_2$, we must show this leads to a contradiction.

We will consider every possible way that the three stages of ψ_1 might yield an ARA sequence for the points \mathbf{x}, \mathbf{x}^* and \mathbf{x}^{**} as defined before. Note that MRD is admissible so that the sequence ARA is impossible at stage 1. The other 7 possible (first stage) test results for MRD are AAA, AAR, RAA, ARR, RAR, RRA and RRR. Recall that screening (stage 2) can, potentially, reverse any stage 1 decision. However, the sign stage (stage 3) is applicable only when the stage 1 action was to Reject and no change was made at stage 2. As an example of

the reasoning used to identify the possible paths suppose stage 1 results in the sequence AAR. By definition stage 3 cannot impact the first two actions. Thus the only two paths possible, beginning with AAR, are $AAR \rightarrow ARR \rightarrow ARA$ and $AAR \rightarrow ARA \rightarrow ARA$. Following these rules we find that there are 18 paths (listed in Table A.1) that must be studied. We will show that each of these paths leads to contradiction.

Note that the fourth row in each cell of Table A.1 indicates how $|X_i|$ relates to the screen stage. This will help in showing that each of the 18 cases cannot happen.

TABLE A.1
Possible paths leading to ARA of MRDSS

	case 1			case 2			case 3		
MRD	A	A	A	A	A	R	A	A	R
MRD+Sc	A	R	A	A	R	A	A	R	R
MRDSS	A	R	A	A	R	A	A	R	A
$ X_1 $	$< C_U$	$> C_U$	$< C_U$	$< C_U$	$> C_U$	$< C_L$	$< C_U$	$> C_U$	$> C_L$
	case 4			case 5			case 6		
MRD	A	R	A	A	R	R	A	R	R
MRD+Sc	Cannot occur			A	R	A	A	R	R
MRDSS				A	R	A	A	R	A
$ X_1 $				$< C_U$	$> C_L$	$< C_L$	$< C_U$	$> C_L$	$> C_L$
	case 7			case 8			case 9		
MRD	R	A	A	R	A	A	R	A	R
MRD+Sc	A	R	A	R	R	A	A	R	A
MRDSS	A	R	A	A	R	A	A	R	A
$ X_1 $	$< C_L$	$> C_U$	$< C_U$	$> C_L$	$> C_U$	$< C_U$	$< C_L$	$> C_U$	$< C_L$
	case 10			case 11			case 12		
MRD	R	A	R	R	A	R	R	A	R
MRD+Sc	R	R	A	R	R	R	A	R	R
MRDSS	A	R	A	A	R	A	A	R	A
$ X_1 $	$> C_L$	$> C_U$	$< C_L$	$> C_L$	$> C_U$	$> C_L$	$< C_L$	$> C_U$	$> C_U$
	case 13			case 14			Type8, case 15		
MRD	R	R	A	R	R	A	R	R	R
MRD+Sc	A	R	A	R	R	A	A	R	A
MRDSS	A	R	A	A	R	A	A	R	A
$ X_1 $	$< C_L$	$> C_L$	$< C_U$	$> C_L$	$> C_L$	$< C_U$	$< C_L$	$> C_L$	$< C_L$
	case 16			case 17			case 18		
MRD	R	R	R	R	R	R	R	R	R
MRD+Sc	R	R	R	R	R	A	A	R	R
MRDSS	A	R	A	A	R	A	A	R	A
$ X_1 $	$> C_L$	$> C_L$	$> C_L$	$> C_L$	$> C_L$	$< C_L$	$< C_L$	$> C_L$	$> C_L$
	$< C_U$	$< C_U$	$< C_U$	$< C_U$	$< C_U$	$< C_U$	$< C_U$	$< C_U$	$< C_U$

At this point we note a fact.

F1: Consider $\mathbf{x}, \mathbf{x}^* = \mathbf{x} + r_1\mathbf{g}, \mathbf{x}^{**} = \mathbf{x} + r_2\mathbf{g}$, with $0 < r_1 < r_2$. Then $|x_1^*| \leq \max(|x_1|, |x_1^{**}|)$.

Now for the 18 cases listed in Table A.1. Cases 1, 2, 7, 9 and 15 violate F1 as seen from the fourth row of the cells in Table A.1.

Case 3: From Table A.1 we see that a sign change occurred at \mathbf{x}^{**} when both \mathbf{x}_1^{**} and $U_{m^*1}(\mathbf{x}^{**})$ were positive. This contradicts the sign stage rule.

Case 4: Cannot occur since MRD is admissible.

Case 5: From Table A.1 we see that at \mathbf{x}^{**} the screen stage took a reject to an accept implying $|x_1^{**}| < C_L$. But $|x_1^*| > C_L$ implying $x_1^* < 0$. But MRD is ARR implying $U_{m^*1}(\mathbf{x}^*) > 0$. Hence MRDSS being ARA is contradicted at \mathbf{x}^* .

Case 6: From the first row in this cell we see that $U_{m^*1}(\mathbf{x}^{**}) > 0$. From the sign change at \mathbf{x}^{**} , we find $x_1^{**} < 0$ which implies $x_1^* < 0$. But this violates the sign rule.

Case 8: From the fourth row we find $x_1^* < 0$ which implies $x_1 < 0$. Since MRD is RAA, $U_{m^*1}(\mathbf{x}) < 0$. However from row 3 we have a sign change, implying a contradiction.

Case 10: Again both $U_{m^*1}(\mathbf{x}) < 0$ and $x_1 < 0$ contradicting a sign change.

Case 11: Both x_1^{**} and $U_{m^*1}(\mathbf{x}^{**})$ are positive, contradicting a sign change.

Case 12: Same reasoning as Case 11.

Case 13: From the fourth row $x_1^* > 0$ and between C_U and C_L . From the first row $U_{m^*1}(\mathbf{x}) < 0$. This contradicts row 3 at \mathbf{x}^* .

Case 14: From the first row $U_{m^*1}(\mathbf{x}^*) < 0$, $U_{m^*1}(\mathbf{x}) < 0$. Since there is change at x_1 due to sign, that implies $x_1 > 0$, $x_1^* > 0$, and $x_1^{**} > 0$. From row 2, $x_1^{**} < C_U$. Hence $C_L < x_1 < C_U$ and also $C_L < x_1^* < C_U$. In particular $x_1^* > 0$. This contradicts the sign change rule at \mathbf{x}^* .

Case 16: Note that a sign action took place at \mathbf{x} and at \mathbf{x}^{**} . Applying F1, we have $C_L < |x_1^*| < C_U$. Now if $x_1 < 0$ that implies $U_{m^*1}(\mathbf{x}) > 0$ which in turn implies $U_{m^*1}(\mathbf{x}^*) > 0$ and $U_{m^*1}(\mathbf{x}^{**}) > 0$. Since there is a sign change at \mathbf{x}^{**} then we have $x_1^{**} < 0$ which means $x_1^* < 0$ and a sign change should have occurred at \mathbf{x}^* . This is a contradiction. Now if $x_1 > 0$, similar reasoning leads to a contradiction at \mathbf{x}^* .

Case 17: From the first two rows we conclude $C_L < |x_1^*|$ and $|x_1^{**}| < C_L$ implying $x_1^* < 0$ and therefore $x_1 < 0$. The sign stage at \mathbf{x} implies $U_{m^*1}(\mathbf{x}) > 0$ and therefore $U_{m^*1}(\mathbf{x}^*) > 0$. But $|x_1| < C_U$ and $|x_1^{**}| < C_L$ implies $|x_1^*| < C_U$.

Hence we have $C_L < |x_1^*| < C_U$ and $x_1^* < 0$ and $U_{m^*1}(\mathbf{x}^*) > 0$. This contradicts the sign stage at \mathbf{x}^* .

Case 18: From the first two rows we have $|x_1| < C_L$ and $|x_1^*| > C_L$ so $x_1^* > 0$ implying $x_1^{**} > 0$. Since there is a sign action at \mathbf{x}^{**} , $|x_1^{**}| < C_U$ and $U_{m^*1}(\mathbf{x}^{**}) < 0$ implying $U_{m^*1}(\mathbf{x}^*) < 0$. Hence $x_1^* > 0$, $x_1^{**} > 0$, $C_L < x_1^* < x_1^{**} < C_U$, implies there should have been a sign action at \mathbf{x}^* . This is a contradiction. □

References

- AN, H. AND GU, L. (1985). On the selection of regression variables. *Acta Mathematicae Applicatae Sinica* 2, 27–36.
- BENJAMINI, Y. AND HOCHBERG, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Ser. B* 57, 289–289. [MR1325392](#)
- BUNEA, F., WEGKAMP, M.H., AND AUGUSTE, A. (2006). Consistent variable selection in high dimensional regression via multiple testing. *Journal of Statistical Planning and Inference* 136, 12, 4349–4364. [MR2323420](#)
- CHEN, C., COHEN, A., AND SACKROWITZ, H.B. (2009). Multiple testing in ordinal data models. Submitted.
- CHRISTENSEN, R. (1987). *Plane Answers to Complex Questions: The Theory of Linear Models*. Springer. [MR0897102](#)
- COHEN, A. AND SACKROWITZ, H.B. (2005a). Decision theory results for one-sided multiple comparison procedures. *The Annals of Statistics* 33, 1, 126–144. [MR2157798](#)
- COHEN, A. AND SACKROWITZ, H.B. (2005b). Characterization of Bayes procedures for multiple endpoint problems and inadmissibility of the step-up procedure. *The Annals of Statistics* 33, 1, 145–158. [MR2157799](#)
- COHEN, A. AND SACKROWITZ, H.B. (2007). More on the inadmissibility of step-up. *Journal of Multivariate Analysis* 98, 3, 481–492. [MR2293009](#)
- COHEN, A. AND SACKROWITZ, H.B. (2008). Multiple testing to two-sided alternatives with dependent data. *Statistica Sinica* 18, 1593–1602. [MR2469325](#)
- COHEN, A., KOLASSA, J., AND SACKROWITZ, H.B. (2007). A smooth version of the step-up procedure for multiple tests of hypotheses. *Journal of Statistical Planning and Inference* 137, 11, 3352–3360. [MR2363261](#)
- COHEN, A., SACKROWITZ, H.B., AND XU, M. (2009). A new multiple testing method in the dependent case. *The Annals of Statistics* 37.
- DUDOIT, S. AND VAN DER LAAN, M.J. (2008). *Multiple Testing Procedures with Applications to Genomics*. Springer Verlag. [MR2373771](#)
- JOHNSON, N.L. AND KOTZ, S. (1970). *Continuous univariate distributions-2*. Houghton Mifflin.
- LEHMANN, E.L. AND ROMANO, J.P. (2005). *Testing Statistical Hypotheses*, Third ed. Springer. [MR2135927](#)

- MATTHES, T.K. AND TRUAX, D.R. (1967). Tests of composite hypotheses for the multivariate exponential family. *The Annals of Mathematical Statistics* 38, 681–697. [MR0208745](#)
- MILLER, A. (2002). *Subset Selection in Regression*, Second ed. Chapman and Hall/CRC. [MR2001193](#)