APPLICATIONES MATHEMATICAE
43,2 (2016), pp. 157–172

Konrad Furmańczyk (Warszawa)

VARIABLE SELECTION USING STEPDOWN PROCEDURES IN HIGH-DIMENSIONAL LINEAR MODELS

Abstract. We study the variable selection problem in high-dimensional linear models with Gaussian and non-Gaussian errors. Based on Ridge estimation, as in Bühlmann (2013) we are considering the problem of variable selection as the problem of multiple hypotheses testing. Under some technical assumptions we prove that stepdown procedures are consistent for variable selection in a high-dimensional linear model.

1. Introduction. In many data problems in biology, medical and economical studies the number of explanatory variables p may greatly exceed the sample size n. Recently, a rich literature is devoted to variable selection for high-dimensional problems (see references to [2] and [13]). We focus on the variable selection problem in the high-dimensional linear model

$$\mathbb{Y} = \mathbb{X}\boldsymbol{\beta} + \varepsilon, \tag{1}$$

where $\mathbb{Y} = (Y_1, \ldots, Y_n)'$, $\mathbb{X}$ is a fixed $n \times p$ design matrix, $\boldsymbol{\beta}$ is a true $p \times 1$ parameter vector and ε is an $n \times 1$ stochastic error vector with $\varepsilon_1, \ldots, \varepsilon_n$ i.i.d. having $\mathbb{E}(\varepsilon_1) = 0$, $\mathrm{Var}(\varepsilon_1) = \sigma^2 < \infty$ and p is much larger than n ($p \gg n$); more precisely, we have $p = p(n)$, and $p(n)/n \to \infty$ as $n \to \infty$. We assume that $\beta_j \neq 0$ for $j \in I_0$ and $\beta_j = 0$ for $j \in I_1 := \{1, \ldots, p\} \setminus I_0$ ($|I_0| = p_0$) and p_0 is fixed and does not depend on n. For parameter estimation we use Ridge regression

$$\begin{aligned}\hat{\beta} &= \arg\min_{\boldsymbol{\beta}}(\|\mathbb{Y} - \mathbb{X}\boldsymbol{\beta}\|_2^2/n + \lambda\|\boldsymbol{\beta}\|_2^2) \\ &= (n^{-1}\mathbb{X}'\mathbb{X} + \lambda\mathbf{I})^{-1}n^{-1}\mathbb{X}'\mathbb{Y},\end{aligned} \tag{2}$$

2010 *Mathematics Subject Classification*: Primary 62J15, 62F03; Secondary 62J05.
Key words and phrases: multiple hypothesis testing, stepdown procedure, variable selection, high-dimensional linear model, ridge regression.
Received 28 December 2015; revised 6 June 2016.
Published online 25 August 2016.

DOI: 10.4064/am2286-6-2016

where $\lambda = \lambda_n$ is a regularization parameter, and $\mathbf{I}$ is the identity matrix. Let $\mathbb{X} = \mathbf{RSV}'$ be the SVD decomposition. Denote by $\mathcal{R}(\mathbb{X}) \subset \mathbb{R}^p$ the linear space generated by the n rows of $\mathbb{X}$. Then the projection of $\mathbb{R}^p$ onto $\mathcal{R}(\mathbb{X})$ has the form $P_{\mathbb{X}} = \mathbb{X}'(\mathbb{X}\mathbb{X}')^{-}\mathbb{X} = \mathbf{VV}'$ and we set $\boldsymbol{\theta} := P_{\mathbb{X}}\beta = \mathbf{VV}'\boldsymbol{\beta}$, where $(\mathbb{X}\mathbb{X}')^{-}$ denotes the pseudo-inverse of the matrix $\mathbb{X}\mathbb{X}'$. In [13] we can find a characterization of identifiability in a high-dimensional linear model with fixed design $\mathbb{X}$. In particular, if $p > n$ and $\beta \in \mathcal{R}(\mathbb{X})$, then $\boldsymbol{\beta}$ is identifiable.

It is well known that $\hat{\beta}$ is a biased estimator of $\boldsymbol{\beta}$ with bias under the null hypothesis $H_j : \beta_j = 0$ (see [2]) given by

$$\sum_{k \neq j} (P_{\mathbb{X}})_{j,k} \beta_k.$$

Taking an initial estimator $\hat{\beta}_{\text{int}}$ of $\boldsymbol{\beta}$ to be the Lasso estimator ([16]), we have an estimator of this bias

$$\sum_{k \neq j} (P_{\mathbb{X}})_{j,k} \hat{\beta}_{\text{int},k},$$

and a corrected Ridge estimator (see [2])

$$\hat{\beta}_{\text{corr},j} = \hat{\beta}_j - \sum_{k \neq j} (P_{\mathbb{X}})_{j,k} \hat{\beta}_{\text{int},k} \tag{3}$$

for $j = 1, \ldots, p$. Let $\hat{\mathbf{\Sigma}} = n^{-1}\mathbb{X}'\mathbb{X}$. Then $\operatorname{Cov}(\hat{\beta}) = n^{-1}\sigma^2\mathbf{\Omega}$, where

$$\mathbf{\Omega} = \mathbf{\Omega}(\lambda) = (\hat{\mathbf{\Sigma}} + \lambda\mathbf{I})^{-1}\hat{\mathbf{\Sigma}}(\hat{\mathbf{\Sigma}} + \lambda\mathbf{I})^{-1}.$$

Set

$$a_{n,j} := \sqrt{n}\,\sigma^{-1}\mathbf{\Omega}_{j,j}^{-1/2}.$$

We consider the following assumptions:

(A) There are constants $\Delta_{j,n} > 0$ such that

$$\mathbb{P}\Big(\bigcap_{j=1}^{p}\Big\{\Big|a_{n,j}\sum_{k \neq j}(P_{\mathbb{X}})_{j,k}(\hat{\beta}_{\text{int},k} - \beta_k)\Big| \le \Delta_{j,n}\Big\}\Big) \to 1 \quad \text{as } n \to \infty.$$

(B) The regularization parameter $\lambda = \lambda_n$ satisfies

$$\lambda_n(\mathbf{\Omega}_{\min}(\lambda_n))^{-1/2} = o(n^{-1/2}\|\boldsymbol{\theta}\|_2^{-1}\lambda_{\min\neq 0}(\hat{\mathbf{\Sigma}})) \quad \text{as } n \to \infty,$$

where $\mathbf{\Omega}_{\min}(\lambda) = \min_{j=1,\ldots,p}\mathbf{\Omega}_{j,j}(\lambda) > 0$ and $\lambda_{\min\neq 0}(\hat{\mathbf{\Sigma}})$ is the smallest non-zero eigenvalue of $\hat{\mathbf{\Sigma}}$.

A thorough discussion of conditions (A)–(B) is given in [2]. The assumption $\mathbf{\Omega}_{\min}(\lambda) > 0$ is very mild and (B) is fulfilled for λ_n sufficiently small. In Remarks 3–5 we also make some new comments on these conditions.

We test the multiple hypotheses

$$H_i : \beta_i = 0 \quad \text{versus} \quad H_i' : \beta_i \neq 0, \quad \text{for } i = 1, \ldots, p. \tag{h$_0$}$$

Similarly to [2], we assume that the p-value for single hypothesis testing H_i vs. H_i' has the form

$$\pi_i = 2\big(1 - \Phi((a_{n,i}|\hat{\beta}_{\mathrm{corr},i}| - \Delta_{i,n})_+)\big) \tag{4}$$

for $i = 1, \ldots, p$, where Φ is the c.d.f. of the standard normal random variable. As already mentioned, the problem of variable selection in linear regression can be viewed as multiple testing (h_0). In a linear Gaussian model, based on the p-values π_i, Bühlmann [2] constructed a method related to the Westfall–Young procedure [19], where the corrected p-values $P_{\mathrm{corr},i}$ have the form

$$P_{\mathrm{corr},i} = F_Z(\pi_i + \zeta),$$

where $\zeta > 0$ is an arbitrarily small number, and F_Z is the distribution function of $\min_{1\le i\le p} 2(1-\Phi(a_{n,i}|Z_i|))$, where $(Z_1, \ldots, Z_p) \sim N_p(0, \sigma^2 n^{-1}\mathbf{\Omega})$. Hypothesis H_i is rejected if $P_{\mathrm{corr},i} \le \alpha$ $(0 < \alpha < 1)$. In [2] it is shown that under assumptions (A)–(B) the procedure asymptotically controls the familywise error rate on level α.

In testing problem (h_0), we use stepdown procedures ([1], [8]–[11]), which we describe as follows. Let $\pi_1, \ldots, \pi_p$ be the p-values for individual tests, let $\pi_{(1)} \le \cdots \le \pi_{(p)}$ denote these p-values ordered, and let $H_{(1)}, \ldots, H_{(p)}$ stand for the corresponding null hypotheses. Let in addition $\alpha_1 \le \cdots \le \alpha_p$ be given thresholds that may depend on n. We proceed according to the following scheme. If $\pi_{(1)} > \alpha_1$, we reject no null hypotheses. Otherwise, if

$$\pi_{(1)} \le \alpha_1, \ldots, \pi_{(r)} \le \alpha_r, \tag{h_1}$$

we reject the hypotheses $H_{(1)}, \ldots, H_{(r)}$, where the largest r satisfying (h_1) is used. Stepdown procedures with data-dependent thresholds can be found in [12], [15], [18]. A selection procedure for the linear regression model may be described by the set $\hat{I}$ of all indices $i \in I_0 \cup I_1$ for which the null hypothesis H_i is rejected, and it is called *consistent* if

$$\mathbb{P}(\hat{I} = I_0) \to 1 \quad \text{as } n \to \infty.$$

We assume that this convergence holds for any parameter vector β since $\mathbb{P}$ belongs to a class of probability measures dependent on $\boldsymbol{\beta}$, i.e., $\mathbb{P} = \mathbb{P}_{\boldsymbol{\beta}}$. In our further considerations we will assume that all convergences with $\mathbb{P}$ hold for any parameter vector $\boldsymbol{\beta}$. Let R be the total number of rejections, and V the number of false rejections for the multitesting problem (h_0), (h_1). It is easy to check that a selection procedure is consistent (see [3]) if

$$\mathbb{P}(R = p_0, V = 0) \to 1 \quad \text{as } n \to \infty. \tag{5}$$

This condition holds if

$$\mathbb{P}(V \ge 1) \to 0 \quad \text{and} \quad \mathbb{P}(R \ne p_0) \to 0$$

as $n \to \infty$. Since

$$\mathbb{P}(V \geq 1) \leq \sum_{j \in I_1} \mathbb{P}(\pi_j \leq \alpha_p) \leq p \max_{j \in I_1} \mathbb{P}(\pi_j \leq \alpha_p), \tag{6}$$

and

$$\mathbb{P}(R \neq p_0) \leq \sum_{j=1}^{p_0} \mathbb{P}(\pi_{(j)} > \alpha_j) + \mathbb{P}(\pi_{(p_0+1)} \leq \alpha_{p_0+1}),$$

and if $\max_{j \in I_0}(1 - F_j(\alpha_p)) \to 0$ as $n \to \infty$, where F_j is the distribution function of the p-value π_j for $j \in I_0$, in the sparse model when $|I_0| = p_0$ is fixed and independent of n (for details see [6]–[7]), we have

$$\begin{aligned}
&\sum_{j=1}^{p_0} \mathbb{P}(\pi_{(j)} > \alpha_j) + \mathbb{P}(\pi_{(p_0+1)} > \alpha_{p_0+1}) \\
&\qquad = \mathcal{O}\Big(\max_{j \in I_0}(1 - F_j(\alpha_j)) + \sum_{j \in I_1} \mathbb{P}(\pi_j \leq \alpha_p)\Big) \\
&\qquad = \mathcal{O}\Big(\max_{j \in I_0}(1 - F_j(\alpha_1)) + p \max_{j \in I_1} \mathbb{P}(\pi_j \leq \alpha_p)\Big).
\end{aligned}$$

Of course $\max_{j \in I_0}(1 - F_j(\alpha_p)) \leq \max_{j \in I_0}(1 - F_j(\alpha_1))$. Hence, we obtain

Consistency of a stepdown selection procedure. *A stepdown selection procedure is consistent if*

(i) $p\mathbb{P}(\pi_i \leq \alpha_p) \to 0$ *as* $n \to \infty$ *for* $i \in I_1$*;*
(ii) $\max_{j \in I_0}(1 - F_j(\alpha_1)) \to 0$ *as* $n \to \infty$ *for* $j \in I_0$*.*

In Section 2 we establish asymptotic control of the familywise error rate and consistency of the stepdown procedure when the errors in (1) are Gaussian (Proposition 1, Theorem 2) and when they are non-Gaussian (Proposition 7, Theorem 8). In Remarks 3–5 and 9 we discuss conditions under which our main results are satisfied. All proofs are given in the Appendix. A simulation study supports the results obtained.

2. Main results

2.1. Gaussian model. We assume that the random errors in (1) have Gaussian distribution. We consider the following assumptions:

(C) $p\alpha_p \to 0$ as $n \to \infty$.
(D) $|(P_{\mathbb{X}})_{j,j}\beta_j| \geq n^{-c}$ for some $c \in (0, 1/2)$ and $\Delta_{j,n} = \mathcal{O}(1)$, $a_{n,j} > \sqrt{n}$ for all $j \in I_0$, and $n^{1/2-c} - \Phi^{-1}(1 - \alpha_1/2) \to \infty$ as $n \to \infty$.

Proposition 1. *For any stepdown procedure satisfying* (A)–(C), *we have*

$$\limsup_{n\to\infty} \mathbb{P}(V \geq 1) - p\alpha_p \leq 0. \tag{7}$$

THEOREM 2. *Any stepdown procedure satisfying* (A)–(D) *is consistent for the variable selection problem in the linear regression model* (1).

REMARK 3. Since

$$\Big|a_{n,j}\sum_{k\neq j}(P_{\mathbb{X}})_{j,k}(\hat{\beta}_{\text{int},k}-\beta_k)\Big| \le a_{n,j}\max_{k\neq j}|(P_{\mathbb{X}})_{j,k}|\,\|\hat{\beta}_{\text{int}}-\beta\|_1,$$

(A) holds for $\Delta_{j,n} = a_{n,j}\max_{k\neq j}|(P_{\mathbb{X}})_{j,k}|\,\|\hat{\beta}_{\text{int}}-\beta\|_1$. For the compatibility conditions with constant $\phi^2_{0,n}$ such that $\liminf_{n\to\infty}\phi^2_{0,n} > 0$ (see [17]) for sparse linear models (see [2, Lemma 2]) if we take the Lasso estimator for an initial estimator of $\boldsymbol{\beta}$ with $\lambda_{\text{Lasso}} = 2\sqrt{\log(p)/n}$ (in our case $|I_0| = p_0$ is constant and $\xi = 0$) we have

$$\|\hat{\beta}_{\text{Lasso}}-\beta\|_1 = \mathcal{O}_P\big(\sqrt{\log(p)/n}\big),$$

and

$$\Delta_{j,n} = \max_{k\neq j} a_{n,j}|(P_{\mathbb{X}})_{j,k}|\sqrt{\log(p)/n} \tag{8}$$

satisfies condition (A).

REMARK 4. If we assume the sparsity condition on $\boldsymbol{\theta}$ ([13, condition (C1)])

$$\|\boldsymbol{\theta}\|_2 = \mathcal{O}(n^a), \tag{9}$$

where $\|\boldsymbol{\theta}\|_2$ is the l_2 norm of the vector $\boldsymbol{\theta}$, and ([13, condition (C2)])

$$\lambda^{-1}_{\min\neq 0}(\hat{\boldsymbol{\Sigma}}) = \mathcal{O}(n^{1-b}) \tag{10}$$

as $n\to\infty$ for some $0<a<b<1$, then (B) holds if

$$\lambda_n = o(n^{-(3/2+a-b)}) \quad \text{as } n\to\infty. \tag{11}$$

REMARK 5. In (D), we have $\Delta_{j,n} = \mathcal{O}(1)$ if

$$\sqrt{\log(p)}\,\boldsymbol{\Omega}^{-1/2}_{\min}(\lambda_n)\max_{k\neq j}|(P_{\mathbb{X}})_{j,k}| = \mathcal{O}(1). \tag{12}$$

This follows from (8) and from the bound

$$a_{n,j} \le \sqrt{n}\sigma^{-1}\boldsymbol{\Omega}^{-1/2}_{\min}(\lambda_n). \tag{13}$$

We also have $a_{n,j} > \sqrt{n}$ if $\sigma^{-1}\boldsymbol{\Omega}^{-1/2}_{\max}(\lambda_n) > 1$, where

$$\boldsymbol{\Omega}_{\max}(\lambda_n) = \max_{j=1,\dots,p}\Omega_{j,j}(\lambda).$$

Of course working in real data we must estimate σ from the data by a consistent estimator. For large a, $1-\Phi(a)\le\varphi(a)$, where φ is the density function of the standard normal r.v. If we take a large $\tilde{a}$ such that $\varphi(\tilde{a})\le\alpha_1/2$, we have $1-\Phi(\tilde{a})\le\alpha_1/2$ and then $n^{1/2-c}-\Phi^{-1}(1-\alpha_1/2)\ge n^{1/2-c}-\tilde{a}$. Setting $\tilde{a}=\sqrt{\log(1/\alpha_1^2)}$ and $\alpha_p := \exp\big(-\frac{1}{2}\log(p)\big)/p$, $\alpha_1 = \alpha_p/c_p$ for some

$c_p \to \infty$, and $c_p \le p$, we obtain $p\alpha_p \to 0$ as $n \to \infty$, and for large n, we have $n^{1/2-c} - \Phi^{-1}(1-\alpha_1/2) \ge n^{1/2-c} - \sqrt{\log(p^5)}$. Assuming

$$n^{1/2-c} - \sqrt{\log(p^5)} \to \infty \quad \text{as } n \to \infty, \tag{14}$$

we have $n^{1/2-c} - \Phi^{-1}(1-\alpha_1/2) \to \infty$ as $n \to \infty$. In the special case when $p = n^\beta$ for some $\beta > 1$ or in ultra-high dimension when $p = \exp(n^\gamma)$ for some $\gamma \in (0, 1-2c)$, we obtain (14). Similarly, taking $\alpha_p := 1/(p\log(p))$, $\alpha_1 = \alpha_p/c_p$ for some $c_p \to \infty$, and $c_p \le p$, we find that $n^{1/2-c} - \Phi^{-1}(1-\alpha_1/2) \to \infty$ if

$$n^{1/2-c} - \sqrt{\log(p^4\log^2(p))} \to \infty \quad \text{as } n \to \infty, \tag{15}$$

which is a weaker assumption than (14).

Example 6. We consider the following stepdown procedures:

(a) the *Holm procedure* with

$$\alpha_j = \frac{q_n}{p+1-j},$$

(b) a *generalization of the Holm* (UHolm) *procedure* [11] with

$$\alpha_j = \frac{([\gamma j]+1)q_n}{p+[\gamma j]+1-j} \quad \text{for some } 0 < \gamma < 1,$$

for some $q_n \to 0$ as $n \to \infty$,

(c) the *Bonferroni procedure* with

$$\alpha_j = \exp\left(-\tfrac{1}{2}\log(p)\right)/p,$$

where $j = 1, \ldots, p$. It is easy to check that (14) holds if $q_n = \exp\left(-\frac{1}{2}\log(p)\right)/p$ and (15) holds if $q_n = 1/(p\log(p))$.

2.2. Non-Gaussian model. We assume that the errors $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. with $\mathbb{E}(\varepsilon_1) = 0$, $\mathbb{E}\,|\varepsilon_1|^3 < \infty$. Instead of (B) we consider the condition

(B_1) The regularization parameter $\lambda = \lambda_n$ satisfies

$$\lambda_n(\boldsymbol{\Omega}_{\min}(\lambda_n))^{-1/2} = o(p^{-1}n^{-1/2}\|\boldsymbol{\theta}\|_2^{-1}\lambda_{\min\neq 0}(\hat{\boldsymbol{\Sigma}})) \quad \text{as } n \to \infty.$$

Proposition 7. *For any stepdown procedure satisfying* (A), (B_1), (C), *we have* (7).

Theorem 8. *Any stepdown procedure satisfying* (A), (B_1), (C), (D) *is consistent for the variable selection problem in the linear regression model* (1).

Remark 9. Assuming (9)–(10) and

$$\lambda_n = o(p^{-1}n^{-(3/2+a-b)}) \tag{16}$$

as $n \to \infty$, we obtain (B_1). As in Remarks 3–5, (8) implies (A), and (12)–(14) for $\alpha_p = \exp\left(-\frac{1}{2}\log(p)\right)/p$ or (12)–(15) for $\alpha_p = 1/(p\log(p))$ imply (D).

3. Simulation study. First, we generated p independent vectors X_j from the standard normal distribution, $j = 1, \ldots, p$; second, we generated p vectors X_j from the normal distribution with covariance matrix $\mathbf{\Sigma}$, where $\sigma_{i,i} = 1$ and $\sigma_{i,j} = 0.3$ for $i \neq j$; and in the end we generated p vectors X_j from the normal distribution with covariance matrix $\mathbf{\Sigma}$, where $\sigma_{i,i} = 1$ and $\sigma_{i,j} = 0.6$ for $i \neq j$.

We consider two families of true models:

$$\text{(M1)} \qquad Y = \sum_{j=1}^{p_0} X_j + \varepsilon,$$

$$\text{(M2)} \qquad Y = \sum_{j=1}^{p_0} 1.5 X_j + \varepsilon,$$

for $p_0 \leq p$, where ε is a vector generated from the standard normal distribution ($\sigma = 1$) or Student's t distribution with $df = 5$ ($\sigma = \sqrt{5/3}$). In our simulations, in all models we considered two cases for $n = 100$: $p = 500$, $p_0 = 5$; $p = 2000$, $p_0 = 5$, and four cases for $n = 200$: $p = 500$, $p_0 = 5$; $p = 500$, $p_0 = 10$; $p = 1000$, $p_0 = 5$; $p = 1000$, $p_0 = 10$.

Table 1. Checking conditions (B) and (D) for $\mathbf{\Sigma} = \mathbf{I}$

n	100	100	200	200	200	200
	p=500,p_0=5	p=2000,p_0=5	p=500,p_0=5	p=500,p_0=10	p=1000,p_0=5	p=1000,p_0=10
$\mathbf{\Omega}^1_{\min}(\lambda_n)$	0.03	0.001	0.20	0.19	0.04	0.04
$\max_{j\in I_0} a^1_{n,j}$	44.97	192.57	28.47	28.91	69.65	69.02
$\max_{j\in I_0} \varDelta^1_{j,n}$	0.69	1.16	0.38	0.44	0.53	0.59
$\mathbf{\Omega}^2_{\min}(\lambda_n)$	0.03	0.001	0.18	0.18	0.03	0.04
$\max_{j\in I_0} a^2_{n,j}$	49.85	229.79	28.57	31.99	66.81	70.90
$\max_{j\in I_0} \varDelta^2_{j,n}$	0.73	1.03	0.37	0.33	0.18	0.17
$\lambda_{\min\neq 0}(\hat{\mathbf{\Sigma}})$	1.52	12.19	0.33	0.35	1.60	1.56
$\min_{j\in I_0} \lvert(P_{\mathbb{X}})_{j,j}\beta^{\mathrm{M1}}_j\rvert$	0.17	0.04	0.35	0.32	0.18	0.17
$\min_{j\in I_0} \lvert(P_{\mathbb{X}})_{j,j}\beta^{\mathrm{M2}}_j\rvert$	0.26	0.06	0.53	0.48	0.27	0.26
$\lVert\boldsymbol{\theta}^{M1}\rVert_2$	1.04	0.54	1.43	2.02	0.98	1.40
$\lVert\boldsymbol{\theta}^{M2}\rVert_2$	1.56	0.81	2.15	3.03	1.47	2.1

We simulated from the above linear models and we recorded the numbers of true models selected from each of $N = 1000$ MC replications with the use of the following stepdown procedures: Holm's, a generalization of Holm's (UHolm for $\gamma = 0.01, 0.1, 0.5, 0.9$) for $q_n = \exp\big(-\frac{1}{2}\log(p)\big)/p$ (superscript 1 in Tables 4–10) and for $q_n = 1/(p\log(p))$ (superscript 2 in Tables 4–10) and Bonferroni's (Bonf) for $\alpha_j = \exp\big(-\frac{1}{2}\log(p)\big)/p$ (see Example 6). In each replication we have fixed a design matrix $\mathbb{X}$ that corresponds to a

Table 2. Checking conditions (B) and (D) for $\sigma_{i,j} = 0.3$ for $i \neq j$ and $\sigma_{i,i} = 1$

n	100	100	200	200	200	200
	$p{=}500, p_0{=}5$	$p{=}2000, p_0{=}5$	$p{=}500, p_0{=}5$	$p{=}500, p_0{=}10$	$p{=}1000, p_0{=}5$	$p{=}1000, p_0{=}10$
$\boldsymbol{\Omega}^1_{\min}(\lambda_n)$	0.04	0.002	0.27	0.28	0.05	0.05
$\max_{j\in I_0} a^1_{n,j}$	39.97	181.80	24.46	25.74	55.45	58.22
$\max_{j\in I_0} \Delta^1_{j,n}$	0.63	0.99	0.35	0.32	0.49	0.47
$\boldsymbol{\Omega}^2_{\min}(\lambda_n)$	0.05	0.002	0.24	0.27	0.05	0.05
$\max_{j\in I_0} a^2_{n,j}$	41.73	165.30	25.00	24.29	60.50	57.08
$\max_{j\in I_0} \Delta^2_{j,n}$	0.63	0.91	0.32	0.40	0.46	0.45
$\lambda_{\min\neq 0}(\hat{\boldsymbol{\Sigma}})$	1.10	8.62	0.25	0.25	1.10	1.11
$\min_{j\in I_0} \lvert(P_{\mathbb{X}})_{j,j}\beta^{\mathrm{M1}}_j\rvert$	0.17	0.04	0.35	0.32	0.18	0.17
$\min_{j\in I_0} \lvert(P_{\mathbb{X}})_{j,j}\beta^{\mathrm{M2}}_j\rvert$	0.26	0.06	0.53	0.48	0.27	0.26
$\lVert\boldsymbol{\theta}^{\mathrm{M1}}\rVert_2$	1.01	0.53	1.39	1.99	1.00	1.40
$\lVert\boldsymbol{\theta}^{\mathrm{M2}}\rVert_2$	1.52	0.80	2.09	2.99	1.5	2.10

Table 3. Checking conditions (B) and (D) for $\sigma_{i,j} = 0.6$ for $i \neq j$ and $\sigma_{i,i} = 1$

n	100	100	200	200	200	200
	$p{=}500, p_0{=}5$	$p{=}2000, p_0{=}5$	$p{=}500, p_0{=}5$	$p{=}500, p_0{=}10$	$p{=}1000, p_0{=}5$	$p{=}1000, p_0{=}10$
$\boldsymbol{\Omega}^1_{\min}(\lambda_n)$	0.08	0.004	0.46	0.42	0.08	0.09
$\max_{j\in I_0} a^1_{n,j}$	30.00	131.00	17.94	19.36	41.87	44.18
$\max_{j\in I_0} \Delta^1_{j,n}$	0.45	0.60	0.29	0.27	0.33	0.37
$\boldsymbol{\Omega}^2_{\min}(\lambda_n)$	0.07	0.004	0.47	0.48	0.08	0.09
$\max_{j\in I_0} a^2_{n,j}$	34.81	133.19	18.15	18.39	44.09	43.29
$\max_{j\in I_0} \Delta^2_{j,n}$	0.43	0.69	0.25	0.28	0.34	0.39
$\lambda_{\min\neq 0}(\hat{\boldsymbol{\Sigma}})$	0.62	4.94	0.15	0.14	0.63	0.61
$\min_{j\in I_0} \lvert(P_{\mathbb{X}})_{j,j}\beta^{\mathrm{M1}}_j\rvert$	0.19	0.05	0.40	0.36	0.17	0.18
$\min_{j\in I_0} \lvert(P_{\mathbb{X}})_{j,j}\beta^{\mathrm{M2}}_j\rvert$	0.29	0.08	0.6	0.54	0.26	0.27
$\lVert\boldsymbol{\theta}^{\mathrm{M1}}\rVert_2$	1.02	0.52	1.50	1.94	1.05	1.52
$\lVert\boldsymbol{\theta}^{\mathrm{M2}}\rVert_2$	1.53	0.78	2.25	2.91	1.58	2.28

linear model with fixed design. As initial estimator of $\boldsymbol{\beta}$ in (3) we used the Lasso estimator with regularization parameter $\lambda_{\mathrm{Lasso}} = 2\sqrt{\log(p)/n}$ from the library glmnet in the R package [5]. This choice guarantees condition (A) (for more details see Remark 3). The parameter λ for Ridge regression (2) was chosen to be $\lambda = 1/n$, which satisfies (11) for Gaussian errors, and $\lambda = 1/(pn)$, which satisfies (16) for Student-t errors (in both cases $0 < a < b < 1$ and $b > a + 1/2$—see Remarks 4, 9). We compared our procedures with the SCAD algorithm with tuning parameter $\lambda = 1/n$ (see [4] for details of this algorithm). The results for the SCAD algorithm are given in Table 12.

Table 4. Frequencies of the true model that are selected by multiple procedures in 1000 simulations for M1 models when $\mathbf{\Sigma} = \mathbf{I}$ for Gaussian errors, and $\lambda = 1/n$

n	100	100	200	200	200	200
	p=500,p_0=5	p=2000,p_0=5	p=500,p_0=5	p=500,p_0=10	p=1000,p_0=5	p=1000,p_0=10
Bonf	767	499	947	728	867	404
^{1} Holm	897	941	1000	995	1000	977
^{1} UHolm_0.01	897	941	1000	995	1000	977
^{1} UHolm_0.1	897	941	1000	990	1000	960
^{1} UHolm_0.5	928	949	999	975	999	927
^{1} UHolm_0.9	939	949	999	965	998	919
^{2} Holm	991	979	998	996	998	987
^{2} UHolm_0.01	991	979	998	996	998	987
^{2} UHolm_0.1	991	979	998	994	998	974
^{2} UHolm_0.5	974	965	995	982	994	940
^{2} UHolm_0.9	964	960	991	974	993	917

From Tables 1–3, we can see that the assumptions on $P_{\mathbb{X}}, \mathbf{\Omega}_{\min}, \lambda_{\min\neq 0}(\hat{\mathbf{\Sigma}})$, $\max_{j\in I_0} a_{n,j}$, $\max_{j\in I_0} \mathit{\Delta}_{j,n}$ are reasonable for design matrices $\mathbb{X}$ (see conditions (B), (D) and Remarks 3–5). From M1 and M2 models, we note that $\beta_j^{\mathrm{M1}} = 1$ for $j \in I_0$, $\beta_j^{\mathrm{M1}} = 0$ for $j \in I_1$ and $\beta_j^{\mathrm{M2}} = 1.5$ for $j \in I_0$, $\beta_j^{\mathrm{M2}} = 0$ for $j \in I_1$. Similarly $\boldsymbol{\theta}^{\mathrm{M1}} = P_{\mathbb{X}}\boldsymbol{\beta}^{\mathrm{M1}}$ and $\boldsymbol{\theta}^{\mathrm{M2}} = P_{\mathbb{X}}\boldsymbol{\beta}^{\mathrm{M2}}$, where $\beta^{\mathrm{M1}} = (\beta_j^{\mathrm{M1}}, j \in I_0 \cup I_1)$, $\boldsymbol{\beta}^{\mathrm{M2}} = (\beta_j^{\mathrm{M2}}, j \in I_0 \cup I_1)$. In Tables 1–3 the superscript 1 on $\mathbf{\Omega}_{\min}$, $a_{n,j}$, $\mathit{\Delta}_{j,n}$ corresponds to $\lambda = 1/n$, and the superscript 2 corresponds to $\lambda = 1/(pn)$.

Table 5. Frequencies selected in 1000 simulations for M1 models when $\sigma_{i,j} = 0.3$ for $i \neq j$ and $\sigma_{i,i} = 1$ for Gaussian errors, and $\lambda = 1/n$

n	100	100	200	200	200	200
	p=500,p_0=5	p=2000,p_0=5	p=500,p_0=5	p=500,p_0=10	p=1000,p_0=5	p=1000,p_0=10
Bonf	960	856	980	981	993	988
^{1} Holm	670	365	1000	992	998	997
^{1} UHolm_0.01	670	365	1000	992	998	997
^{1} UHolm_0.1	670	365	1000	994	998	997
^{1} UHolm_0.5	769	432	999	997	1000	998
^{1} UHolm_0.9	802	472	999	996	1000	999
^{2} Holm	700	287	1000	999	1000	1000
^{2} UHolm_0.01	700	287	1000	999	1000	1000
^{2} UHolm_0.1	700	287	1000	998	1000	1000
^{2} UHolm_0.5	777	362	1000	998	1000	1000
^{2} UHolm_0.9	815	407	1000	998	1000	1000

Table 6. Frequencies selected in 1000 simulations for M1 models when $\sigma_{i,j} = 0.6$ for $i \neq j$ and $\sigma_{i,i} = 1$ for Gaussian errors, and $\lambda = 1/n$

n	100	100	200	200	200	200
	p=500,p_0=5	p=2000,p_0=5	p=500,p_0=5	p=500,p_0=10	p=1000,p_0=5	p=1000,p_0=10
Bonf	385	465	971	896	991	935
[1] Holm	23	9	668	370	922	513
[1] UHolm_0.01	23	9	668	370	922	513
[1] UHolm_0.1	23	9	668	440	922	580
[1] UHolm_0.5	36	24	770	556	952	683
[1] UHolm_0.9	52	37	808	608	960	723
[2] Holm	26	28	856	552	956	795
[2] UHolm_0.01	26	28	856	552	956	795
[2] UHolm_0.1	26	28	856	635	956	846
[2] UHolm_0.5	71	56	912	757	978	902
[2] UHolm_0.9	90	72	928	794	985	926

Table 7. Frequencies selected in 1000 simulations for M2 models when $\sigma_{i,j} = 0.6$ for $i \neq j$ and $\sigma_{i,i} = 1$ for Gaussian errors, and $\lambda = 1/n$

n	100	100	200	200	200	200
	p=500, p_0=5	p=2000, p_0=5	p=500, p_0=5	p=500, p_0=10	p=1000, p_0=5	p=1000, p_0=10
Bonf	986	976	982	986	991	989
[1] Holm	954	723	1000	1000	1000	998
[1] UHolm_0.01	954	726	1000	1000	1000	998
[1] UHolm_0.1	954	726	1000	1000	1000	999
[1] UHolm_0.5	977	799	1000	1000	1000	1000
[1] UHolm_0.9	980	819	1000	1000	1000	1000
[2] Holm	966	919	1000	1000	1000	1000
[2] UHolm_0.01	966	919	1000	1000	1000	1000
[2] UHolm_0.1	966	919	1000	1000	1000	1000
[2] UHolm_0.5	986	953	1000	999	999	1000
[2] UHolm_0.9	966	961	1000	999	999	1000

3.1. Conclusions based on simulation studies. We observe that in all the models considered, UHolm and Holm procedures worked better than the Bonferroni method when the design matrix was simulated from uncorrelated predictors ($\mathbf{\Sigma} = \mathbf{I}$, Tables 4, 9). When the correlation of the predictors is stronger (0.6 compared to 0.3), the Bonferroni method works

Table 8. Frequencies selected in 1000 simulations for M2 models when $\sigma_{i,j} = 0.6$ for $i \neq j$ and $\sigma_{i,i} = 1$ for Student-t errors, and $\lambda = 1/(pn)$

n	100	100	200	200	200	200
	p=500, p_0=5	p=2000, p_0=5	p=500, p_0=5	p=500, p_0=10	p=1000, p_0=5	p=1000, p_0=10
Bonf	683	742	580	594	653	695
[1] Holm	698	633	971	974	996	984
[1] UHolm_0.01	698	633	971	974	996	984
[1] UHolm_0.1	698	633	971	962	996	976
[1] UHolm_0.5	740	685	957	948	989	965
[1] UHolm_0.9	764	702	948	940	981	962
[2] Holm	890	896	953	950	965	971
[2] UHolm_0.01	890	896	953	950	965	971
[2] UHolm_0.1	890	896	953	948	965	958
[2] UHolm_0.5	894	900	917	914	937	934
[2] UHolm_0.9	893	899	911	892	922	914

Table 9. Frequencies selected in 1000 simulations for M1 models when $\mathbf{\Sigma} = \mathbf{I}$ for Student-t errors, and $\lambda = 1/(pn)$

n	100	100	200	200	200	200
	p=500, p_0=5	p=2000, p_0=5	p=500, p_0=5	p=500, p_0=10	p=1000, p_0=5	p=1000, p_0=10
Bonf	127	210	455	262	401	52
[1] Holm	423	896	968	922	961	821
[1] UHolm_0.01	423	896	968	922	961	821
[1] UHolm_0.1	423	896	968	892	961	758
[1] UHolm_0.5	431	851	934	852	934	658
[1] UHolm_0.9	453	831	925	812	921	604
[2] Holm	741	666	933	794	935	587
[2] UHolm_0.01	741	666	933	794	935	587
[2] UHolm_0.1	741	666	933	736	935	489
[2] UHolm_0.5	667	571	879	605	890	339
[2] UHolm_0.9	632	543	851	539	866	280

better and has greater power than the other stepdown procedures in M1 models for Gaussian errors (Table 6) and for $n = 100$, $p = 500$, $p_0 = 5$; $p = 2000$, $p_0 = 5$, and $n = 200$, $p = 1000$, $p_0 = 10$ for Student-t errors (Table 11); however, when increasing the sample size, the power of UHolm procedures approaches the power of the Bonferroni method when we use $q_n = 1/(p \log(p))$. In M2 models Holm and UHolm procedures work better than the Bonferroni method (Table 7) except for two cases: $n = 100$, $p = 500$, $p_0 = 5$; $p = 2000$, $p_0 = 5$. In M1 models with Student-t errors UHolm

Table 10. Frequencies selected in 1000 simulations for M1 models when $\sigma_{i,j} = 0.3$ for $i \neq j$ and $\sigma_{i,i} = 1$ for Student-t errors, and $\lambda = 1/(pn)$

n	100	100	200	200	200	200
	$p{=}500, p_0{=}5$	$p{=}2000, p_0{=}5$	$p{=}500, p_0{=}5$	$p{=}500, p_0{=}10$	$p{=}1000, p_0{=}5$	$p{=}1000, p_0{=}10$
Bonf	656	536	603	584	625	716
[1] Holm	514	187	946	884	973	929
[1] UHolm_0.01	514	187	946	884	973	929
[1] UHolm_0.1	514	187	946	906	973	936
[1] UHolm_0.5	596	239	935	915	970	935
[1] UHolm_0.9	620	187	929	904	964	927
[2] Holm	260	320	939	926	967	965
[2] UHolm_0.01	260	320	939	926	967	965
[2] UHolm_0.1	260	320	939	922	967	956
[2] UHolm_0.5	341	400	899	904	943	932
[2] UHolm_0.9	374	434	886	890	933	919

Table 11. Frequencies selected in 1000 simulations for M1 models when $\sigma_{i,j} = 0.6$ for $i \neq j$ and $\sigma_{i,i} = 1$ for Student-t errors, and $\lambda = 1/(pn)$

n	100	100	200	200	200	200
	$p{=}500, p_0{=}5$	$p{=}2000, p_0{=}5$	$p{=}500, p_0{=}5$	$p{=}500, p_0{=}10$	$p{=}1000, p_0{=}5$	$p{=}1000, p_0{=}10$
Bonf	194	56	526	413	644	628
[1] Holm	9	2	499	115	702	406
[1] UHolm_0.01	9	2	499	115	702	406
[1] UHolm_0.1	9	2	499	147	702	447
[1] UHolm_0.5	18	3	571	225	763	564
[1] UHolm_0.9	25	3	599	256	783	597
[2] Holm	39	19	551	224	673	273
[2] UHolm_0.01	39	19	551	224	673	273
[2] UHolm_0.1	39	19	551	274	673	312
[2] UHolm_0.5	60	28	619	389	736	410
[2] UHolm_0.9	70	39	633	427	761	445

procedures work better than the Bonferroni method in both cases for $q_n = 1/p^{3/2}$ and $q_n = 1/(p\log(p))$ except in two cases: $n = 100$, $p = 500$, $p_0 = 5$; $p = 2000$, $p_0 = 5$ (Tables 9–11). The power of all procedures is increasing with the sample size and decreasing with the number of predictors p. We may also observe the poor power of stepdown procedures for dependent predictors for M1 models and with small sample sizes (Table 6, Table 11). In comparison to the stepdown procedures the SCAD algorithm (Table 12) works poorly for M1 models with uncorrelated predictors. The power the

Table 12. Frequencies selected in 1000 simulations for SCAD algorithm for M1, M2 models with Gaussian or Student-t errors, when $\sigma_{i,j} = 0.3$ or 0.6 for $i \neq j$ or $\mathbf{\Sigma} = \mathbf{I}$

n	100	100	200	200	200	200
	$p{=}500, p_0{=}5$	$p{=}2000, p_0{=}5$	$p{=}500, p_0{=}5$	$p{=}500, p_0{=}10$	$p{=}1000, p_0{=}5$	$p{=}1000, p_0{=}10$
M1, Gauss, $\mathbf{\Sigma} = \mathbf{I}$	36	44	47	0	46	1
M2, Gauss, $\mathbf{\Sigma} = \mathbf{I}$	803	797	968	708	960	683
M1, Gauss, $\sigma_{i,j} = 0.3$	960	917	999	999	999	999
M1, Gauss, $\sigma_{i,j} = 0.6$	816	767	976	738	976	646
M2, Gauss, $\sigma_{i,j} = 0.6$	824	709	989	739	991	674
M1, $t(5)$, $\mathbf{\Sigma} = \mathbf{I}$	30	31	42	0	57	3
M1, $t(5)$, $\sigma_{i,j} = 0.3$	915	392	999	999	573	567
M1, $t(5)$, $\sigma_{i,j} = 0.6$	744	306	972	630	975	578

SCAD algorithm for M2 models is greater than for M1 models but still less than that of the stepdown procedures. In the case of Gaussian errors with correlated predictors ($\sigma_{i,j} = 0.6$) for $n = 100$ and $p = 500$, $p_0 = 5$ and $p = 2000$, $p_0 = 5$ the SCAD algorithm has greater power than the power of the stepdown procedures. For Student-t errors with correlated predictors ($\sigma_{i,j} = 0.6$) we can observe that the SCAD algorithm works better than the stepdown procedures.

4. Appendix

Proof of Proposition 1. By [2, Proposition 2], for any $j = 1, \dots, p$, we have the decomposition

$$\hat{\beta}_{\mathrm{corr},j} = Z_j + (P_{\mathbb{X}})_{j,j}\beta_j - \sum_{k \neq j} (P_{\mathbb{X}})_{j,k}(\hat{\beta}_{\mathrm{int},k} - \beta_k) + b_j, \tag{17}$$

where $Z_j = \hat{\beta}_j - \mathbb{E}(\hat{\beta}_j)$, $W_j := a_{n,j}Z_j \sim N(0,1)$, $b_j = \mathbb{E}(\hat{\beta}_j) - \theta_j$. For $j \in I_1$, we have

$$\begin{aligned}\mathbb{P}(\pi_j \leq \alpha_p) &= \mathbb{P}\big(a_{n,j}|\hat{\beta}_{\mathrm{corr},j}| \geq \Delta_{j,n} + \Phi^{-1}(1 - \alpha_p/2)\big)\\ &= \mathbb{P}\Big(\Big|W_j - a_{n,j}\sum_{k \neq j}(P_{\mathbb{X}})_{j,k}(\hat{\beta}_{\mathrm{int},k} - \beta_k) + a_{n,j}b_j\Big| \geq \Delta_{j,n} + \Phi^{-1}(1 - \alpha_p/2)\Big).\end{aligned}$$

By (A) for large n, with probability tending to 1,

$$\Big|W_j - a_{n,j}\sum_{k \neq j}(P_{\mathbb{X}})_{j,k}(\hat{\beta}_{\mathrm{int},k} - \beta_k) + a_{n,j}b_j\Big| \leq |W_j| + \Delta_{j,n} + \|a_{n,j}b_j\|_\infty.$$

Note that from (B), we have $\|a_{n,j}b_j\|_\infty \to 0$ (see [2]) and for large n, with probability tending to 1,

$$\Big|W_j - a_{n,j}\sum_{k \neq j}(P_{\mathbb{X}})_{j,k}(\hat{\beta}_{\mathrm{int},k} - \beta_k) + a_{n,j}b_j\Big| \leq |W_j| + \Delta_{j,n}.$$

Therefore for $j \in I_1$, for sufficiently large n,

$$\mathbb{P}(\pi_j \le \alpha_p) \le \mathbb{P}\big(|W_j| \ge \Phi^{-1}(1-\alpha_p/2)\big) = \alpha_p.$$

Thus from (6) and (C), we get (7). ∎

Proof of Theorem 2. By the Introduction, we must show conditions (i)–(ii). From (7), we obtain (i).

For all $j \in I_0$,

$$\begin{aligned} 1 - F_j(\alpha_1) = \mathbb{P}(\pi_j \ge \alpha_1) &= \mathbb{P}\big(2(1-\Phi(a_{n,j}|\hat{\beta}_{\mathrm{corr},j}| - \Delta_{j,n})) \ge \alpha_1\big) \\ &= \mathbb{P}\big(a_{n,j}|\hat{\beta}_{\mathrm{corr},j}| \le \Delta_{j,n} + \Phi^{-1}(1-\alpha_1/2)\big). \end{aligned}$$

By (17) and the triangle inequality,

$$a_{n,j}|\hat{\beta}_{\mathrm{corr},j}| \ge a_{n,j}|(P_{\mathbb{X}})_{j,j}\beta_j| - |W_j| - a_{n,j}\Big|\sum_{k\ne j}(P_{\mathbb{X}})_{j,k}(\hat{\beta}_{\mathrm{int},k} - \beta_k)\Big| - |a_{n,j}b_j|.$$

From (A) we have, with probability tending to 1,

$$a_{n,j}|\hat{\beta}_{\mathrm{corr},j}| \ge a_{n,j}|(P_{\mathbb{X}})_{j,j}\beta_j| - |W_j| - \Delta_{j,n} - |a_{n,j}b_j|.$$

Since $\|a_{n,j}b_j\|_\infty \to 0$, with probability tending to 1 for sufficiently large n we obtain

$$a_{n,j}|\hat{\beta}_{\mathrm{corr},j}| \ge a_{n,j}|(P_{\mathbb{X}})_{j,j}\beta_j| - |W_j| - \Delta_{j,n},$$

and

$$\begin{aligned} &\mathbb{P}\big(a_{n,j}|\hat{\beta}_{\mathrm{corr},j}| \le \Delta_{j,n} + \Phi^{-1}(1-\alpha_1/2)\big) \\ &\qquad \le \mathbb{P}\big(|W_j| \ge a_{n,j}|(P_{\mathbb{X}})_{j,j}\beta_j| - 2\Delta_{j,n} - \Phi^{-1}(1-\alpha_1/2)\big). \end{aligned}$$

Therefore, we have (ii) if $a_{n,j}|(P_{\mathbb{X}})_{j,j}\beta_j| - 2\Delta_{j,n} - \Phi^{-1}(1-\alpha_1/2) \to \infty$ as $n \to \infty$. This may be obtained from (D). ∎

Proof of Proposition 7. As in (17), we obtain

$$\hat{\beta}_{\mathrm{corr},j} = \tilde{Z}_j + (P_{\mathbb{X}})_{j,j}\beta_j - \sum_{k\ne j}(P_{\mathbb{X}})_{j,k}(\hat{\beta}_{\mathrm{int},k} - \beta_k) + b_j, \tag{18}$$

where $\tilde{W}_j := a_{n,j}\tilde{Z}_j$ and $\tilde{Z}_j := \hat{\beta}_j - \mathbb{E}(\hat{\beta}_j) = ((n^{-1}\mathbb{X}'\mathbb{X}+\lambda\mathbf{I})^{-1}n^{-1}\mathbb{X}'\varepsilon)_j$ is the j-component of the vector. Since (B_1) implies (B), we have $\|a_{n,j}b_j\|_\infty \to 0$, and using the same arguments as in the proof of Proposition 1, we obtain, for $j \in I_1$,

$$\mathbb{P}(\pi_j \le \alpha_p) \le \mathbb{P}\big(|\tilde{W}_j| \ge \Phi^{-1}(1-\alpha_p/2)\big).$$

If we show

$$p\mathbb{P}\big(|\tilde{W}_j| \ge \Phi^{-1}(1-\alpha_p/2)\big) \to 0 \tag{19}$$

as $n \to \infty$, then we obtain (i). It is obvious that we can get (19) from (C) and from the condition

$$p\sup_{x\in\mathbb{R}} \|\mathbb{P}(\tilde{W}_j \le x) - \Phi(x)\|_\infty \to 0 \tag{20}$$

as $n \to \infty$. Let $(w_{i,1}, \ldots, w_{i,n})$ be the ith row of $(n^{-1}\mathbb{X}'\mathbb{X} + \lambda\mathbf{I})^{-1}n^{-1}\mathbb{X}'$. Then $\tilde{W}_j = a_{n,j}\sum_{i=1}^{n} w_{j,i}\varepsilon_i$ and $\sigma^2 a_{n,j}^2 \sum_{i=1}^{n} w_{j,i}^2 = 1$, and from the Berry-Esséen bound (see [14]),

$$\sup_{x\in\mathbb{R}} \|\mathbb{P}(\tilde{W}_j \le x) - \Phi(x)\|_\infty = \mathcal{O}(a_{n,j}\max_i |w_{i,j}|).$$

Using the bound of the bias of $\hat{\beta}$ given in [13, proof of Theorem 1], we have

$$\max_i |w_{i,j}| \le \lambda\|\theta\|_2 \lambda_{\min\neq 0}^{-1}(\hat{\mathbf{\Sigma}}).$$

Now (13) and (B_1) imply (20) and we obtain (7). ■

Proof of Theorem 8. From (7), we obtain (i). Replacing W_j by $\tilde{W}_j$, we can obtain condition (ii) as in the proof of Theorem 2. ■

Acknowledgements. I would like to thank the referee for helpful comments which helped me to improve the paper.

References

[1] Y. Benjamini and W. Liu, *A step-down multiple hypotheses testing procedure that controls the false discovery rate under independence*, J. Statist. Plann. Infer. 82 (1999), 163–170.

[2] P. Bühlmann, *Statistical significance in high-dimensional linear model*, Bernoulli 19 (2013), 1212–1242.

[3] F. Bunea, M. H. Wegkamp and A. Auguste, *Consistent variable selection in high dimensional regression via multiple testing*, J. Statist. Plann. Infer. 136 (2006), 12, 4349–4364.

[4] J. Fan and R. Li, *Variable selection via nonconcave penalized likelihood and its oracle properties*, J. Amer. Statist. Assoc. 96 (2001), 1348–1360.

[5] J. Friedman, T. Hastie, N. Simon and R. Tibshirani, *glmnet: Lasso and elastic-net regularized generalized linear models*, R package version 2.0, 2015.

[6] K. Furmańczyk, *Selection in parametric models via some stepdown procedures*, Appl. Math. (Warsaw) 41 (2014), 81–92.

[7] K. Furmańczyk, *On some stepdown procedures with application to consistent variable selection in linear regression*, Statistics 49 (2015), 614–628.

[8] Y. Ge, S. C. Sealfon and T. P. Speed, *Some step-down procedures controlling the false discovery rate under dependence*, Statistica Sinica 18 (2008), 881–904.

[9] S. Holm, *A simple sequentially rejective multiple test procedure*, Scand. J. Statist. 6 (1979), 65–70.

[10] G. Hommel, *A stagewise rejective multiple test procedure based on a modified Bonferroni test*, Biometrika 75 (1988), 383–386.

[11] E. L. Lehmann and J. P. Romano, *Generalizations of the familywise error rate*, Ann. Statist. 28 (2005), 1–25.

[12] J. P. Romano and M. Wolf, *Exact and approximate stepdown methods for multiple hypothesis testing*, J. Amer. Statist. Assoc. 100 (469): (2005), 94–108.

[13] J. Shao and X. Deng, *Estimation in high-dimensional linear models with deterministic design matrices*, Ann. Statist. 40 (2012), 812–831.

[14] G. R. Shorack, *Probability for Statisticians*, Springer, New York, 2000.

[15] P. N. Somerville, *FDR step-down and step-up procedures for the correlated case*, in: Recent Developments in Multiple Comparison Procedures, IMS Lecture Notes Monogr. Ser. 47, Inst. Math. Statist., Beachwood, OH, 2004, 100–118.
[16] R. Tibshirani, *Regression shrinkage and selection via the lasso*, J. Roy. Statist. Soc. Ser. B 58 (1996), 267–288.
[17] S. van de Geer, *The deterministic Lasso*, in: JSM Proceedings, Amer. Statist. Assoc., paper no. 489, 2007.
[18] M. J. Van der Laan, S. Dudoit and K. S. Pollard, *Multiple testing. Part II. Step-down procedures for control of the family-wise error rate*, Statist. Appl. Genet. Molec. Biol. 3 (2004), http://www.bepress.com/sagmb/vol3/iss1/art14.
[19] P. Westfall and S. Young, *Resampling-based Multiple Testing: Example and Methods for p-value Adjustment*, Wiley, New York, 1993.

Konrad Furmańczyk
Department of Applied Mathematics
Warsaw University of Life Sciences (SGGW)
Nowoursynowska 159
02-776 Warszawa, Poland
E-mail: konfur@wp.pl